Quantitative studies of the core planar polarity protein complex stoichiometry in *Drosophila melanogaster*

Jessica Mary Allen

Submitted for the degree of Doctor of Philosophy
September 2013

MRC Centre for Developmental and Biomedical Genetics
Department of Biomedical Science
University of Sheffield

Acknowledgements

Professor Alfonso Martinez-Arias gave me my first job after graduation. He and his lab kindled a lifelong love of *Drosophila*, and subsequently firmly nudged me down the doctoral path. His encouragement was the catalyst that sparked this PhD into life.

Professor David Strutt kindly accepted my application, a decision that I suspect he may have regretted more than once. He has always been very generous with his time, advice, expertise and assistance in every aspect of this project. His rigorous, fair and insightful criticisms were supplied in abundance and gratefully received. A more thorough apprenticeship in How To Science would be extremely difficult to find.

Helen, Sam, Ros, John, Chloe, Amy and Liz are a diverse and brilliant group of scientists united by dry wit and excessive consumption of cake. Chloe's pragmatism, Amy's quiet determination, Liz's no-nonsense attitude, Sam's incomparable work ethic, John's good-humoured unflappability, Ros' openhearted friendliness and Helen's constant calming presence have all been individually inspirational in times of crisis and doubt.

Dr. Tanya Whitfield and Dr. Martin Zeidler were excellent project advisors, keeping everything on the straight and narrow and providing an inexhaustible stream of reassurance and new ideas. Special thanks to Martin for the loan of his printer!

Thanks are also due to the members of the other D floor fly labs for discussion and ideas, both at fly meetings and in passing, to Kath and Josh for looking after the flies, and to the BMS PhD students, who are a fantastic group of colleagues.

To Mum, Dad and Harriet, yes, it's finally finished, and no, the army of giant mutant fruit flies is not quite ready yet, sorry. Thank you for everything; you've been behind me all the way and kept me going through the tough times.

To my dearest, freshly-minted but long-suffering husband Matt, who has been keeping me fed and watered and relatively sane for the last few months, a thousands thanks would barely scratch the surface. Your constant love and support is entirely undeserved but very much appreciated. Normal service will be resumed shortly, you'll see.

Abstract

Planar polarity describes the ability of cells to orient themselves in the plane of a tissue. This process is partly mediated by the six core planar polarity proteins: Flamingo, Frizzled, Strabismus, Prickle, Dishevelled and Diego. These proteins localise in an asymmetric, punctate manner at apicolateral cell junctions. Strabismus and Prickle are found on the proximal side of the cell, Frizzled, Dishevelled and Diego on the distal side, and Flamingo on both sides. All six core proteins are required for planar polarity, but the molecular mechanism of the establishment of planar polarity has not been identified.

Several lines of evidence suggest that the core proteins form an asymmetric complex across cell membranes. The puncta observed are therefore hypothesised to represent clusters of core complexes with the same polarity. Elucidating the stoichiometry of the putative core complex will provide insights into the core complex and the mechanisms of planar polarity propagation. The simplest possible core complex model that fits the existing data is two Flamingo molecules, with one molecule each of Frizzled, Strabismus, Prickle, Dishevelled and Diego.

EGFP-tagged core protein constructs were generated for *in vivo* expression at endogenous levels, as changes in protein level could potentially affect the stoichiometry. Images of these fusion proteins in the live *Drosophila melanogaster* pupal wing were gathered by confocal microscopy. A custom method was then developed to detect the puncta in the images and quantify their brightness. The relative stoichiometries of the core proteins in puncta was determined at 28 hours after pupation, when core protein localisation is strongly asymmetric, and also at 20 hours after pupation, when no asymmetry is visible. The stoichiometry of the core proteins in puncta is more complex than the simple initial model. The data additionally reveal an interesting distinction in behaviour between the transmembrane and cytoplasmic core proteins.

Contents

Chapter 1: Introduction

Chapter 2: Generating a suite of endogenously-expressed EGFP-tagged core planar polarity proteins in *Drosophila*

Chapter 3: Measuring the stoichiometry

Chapter 4: Materials and Methods

Chapter 5: Discussion

Figures

Tables

Chapter 1

Introduction

Contents

Figures

1.0 Introduction

This project, along with countless others, was made possible by two of the most monumental bodies of work in the field of biology. The century of study of the fruit fly *Drosophila melanogaster*, pioneered by Thomas Hunt Morgan (eg. Morgan, 1910), has made it an ideal model organism for genetic experiments to address a wide range of questions. The extraordinary depth of genetic data and exquisitely powerful tools available to manipulate its genome are unmatched in any other model system.

The green fluorescent protein (GFP) was discovered by O. Shimomura in 1962. Its subsequent cloning and expression in bacteria by M. Chalfie, followed by R. Tsien's work to optimise its behaviour for experiments, completely revolutionised *in vivo* microscopy and allowed beautiful images of proteins in living cells to be captured and studied (Shimomura et al., 1962; Chalfie et al., 1994; Heim and Tsien, 1996).

One field that has benefited from coupling *Drosophila* work with the ability to visualise proteins *in vivo* is planar polarity. Planar polarity describes the ability of cells to orient themselves in the plane of a tissue. Despite being one of the most fundamental processes of metazoan life, the molecular mechanisms of this process are still relatively poorly understood. The best-characterised mediators of planar polarity are the six core planar polarity proteins, which colocalise in a punctate manner at the apicolateral membranes of neighbouring cells.

Several groups have exploited certain properties of GFP to measure protein levels in bacteria and yeasts, but so far this has only been reported once in *Drosophila*. This thesis describes an attempt to use GFP fluorescence in the live *Drosophila* pupal wing to measure the relative stoichiometry of the core planar polarity proteins in puncta.

1.1 A brief history of planar polarity

In 1940 Professor Sir Vincent Brian Wigglesworth, a prolific entomologist best known for his discovery of the insect juvenile hormone, was working on the blood sucking bug *Rhodnius prolixus*. The experiment involved burning away the bristle-bearing cuticular 'plaques' of the nymphs and observing their regeneration after successive moults. In passing, he writes, *"It is worth noting that, with few exceptions, the regenerated* *plaques and bristles show the normal orientation. This raises a new problem which has not been further investigated."* (Wigglesworth, 1940, p189).

This new problem was not further investigated until late in the next decade, when H. Piepho (1955) and M. Locke (1959) published their work on the cuticles of the wax moth *Galleria* and *Rhodnius*, respectively. Both excised small pieces of cuticle and rotated them before re-implanting them into the animals and looking at the resulting patterns of the healed cuticle after moulting. Piepho was the first to show that cells coordinate their polarity with their neighbours, while Locke proposed that polarity was specified by some kind of segmentally-repeating gradient in the cuticle. Both demonstrated communication of polarity between neighbouring cells.

P. Lawrence (1966) built on this idea, using the bug *Oncopeltus fasciatus* to model the concepts of gradients and 'flow'. He determined that cells must actively maintain the polarising gradient, rather than passively reading and responding to it. H. Stumpf (1967) modelled the putative polarity factor as a diffusible molecule that determined cell differentiation – a concept very close to that of the morphogen, which was being developed at the time by Lawrence, L. Wolpert and others. From this idea emerged the concept of a single gradient that controlled both cell fate and cell polarity.

The dawn of the genetic era, and the increasing popularity of *Drosophila* as a model organism, paved the way for Gubb and García-Bellido's 1982 analysis of the genetic loci affecting planar polarity that were known at the time. They examined the phenotypes of the mutants that only affected polarity and not any other kind of patterning. From this they made some important observations, including the fact that some mutants show non-autonomous effects – i.e. that null mutant clones of some planar polarity genes can also disturb the polarity of neighbouring wild-type cells. They additionally demonstrated that the altered polarity patterns of the wings of planar polarity mutant flies were consistent and specific to each particular gene, and that the pattern was not affected by physical factors such as wing and cell size and shape. They were the first authors to formally propose the existence of a dedicated planar polarity pathway.

It was not until K. Nübler-Jung's work in the late 1980s (e.g. Nübler-Jung, 1987; Nübler-Jung et al., 1987) which introduced the phrase 'planar polarity', that the presence of a global polarising cue was separated from the intrinsic ability of cells to align with their neighbours. This local communication of planar polarity is a theme that forms the basis of most planar polarity models.

Today there are more than a dozen known genes with a role in planar polarity in *Drosophila* alone, including both genes that are specifically required for planar polarity, and others that merely exert an indirect influence on polarity. Alleles of some of these genes were discovered well before Wigglesworth's time – the first mention of the upstream factor *dachsous* dates back to Bridges and Morgan in 1919 – but it is only relatively recently that they have all been identified explicitly as polarity genes and integrated into the current (somewhat confused) pathway models. Even so, there is still some disagreement about the exact definition of a planar polarity gene and which genes merit inclusion in this list.

1.2 The core planar polarity proteins

The best-characterised module of the planar polarity pathway is the six-member group of 'core' proteins, encoded by the core planar polarity genes. This group consists of three of the oldest known polarity genes, *frizzled*, *dishevelled* and *prickle*, and the more recent discoveries *strabismus/Van Gogh*, *flamingo/starry night* and *diego*. The core proteins are required for planar cell polarity, rely on each other for their subcellular localisation and give distinctive polarity phenotypes in multiple tissues (Gubb and García-Bellido, 1982; Strutt and Strutt, 2009; Axelrod, 2009; Goodrich and Strutt, 2011; Lawrence and Casal, 2013). The six genes encoding the core polarity proteins are the primary focus of this project, and are described individually in detail below.

Much of the work has been done in the *Drosophila* wing. The adult wing is shown in **Figure 1.1 A**. It consists largely of an array of hexagonal cells, each of which produces a single distal-pointing actin-based trichome (hair) from its distal vertex (**Figure 1.1 B**). In the wild-type wing all the hairs align with each other (**Figure 1.1 C**), but in a core planar polarity mutant such as *strabismus*^{6} the cells lose global coordination and the hairs adopt swirling patterns instead (**Figure 1.1 D**).

1.2.1 Frizzled

Prior to Gubb and García-Bellido (1982), *frizzled* (*fz*) mutants had only been used as markers for chromosome mapping. Nonautonomous effects of *fz* mutant clones were

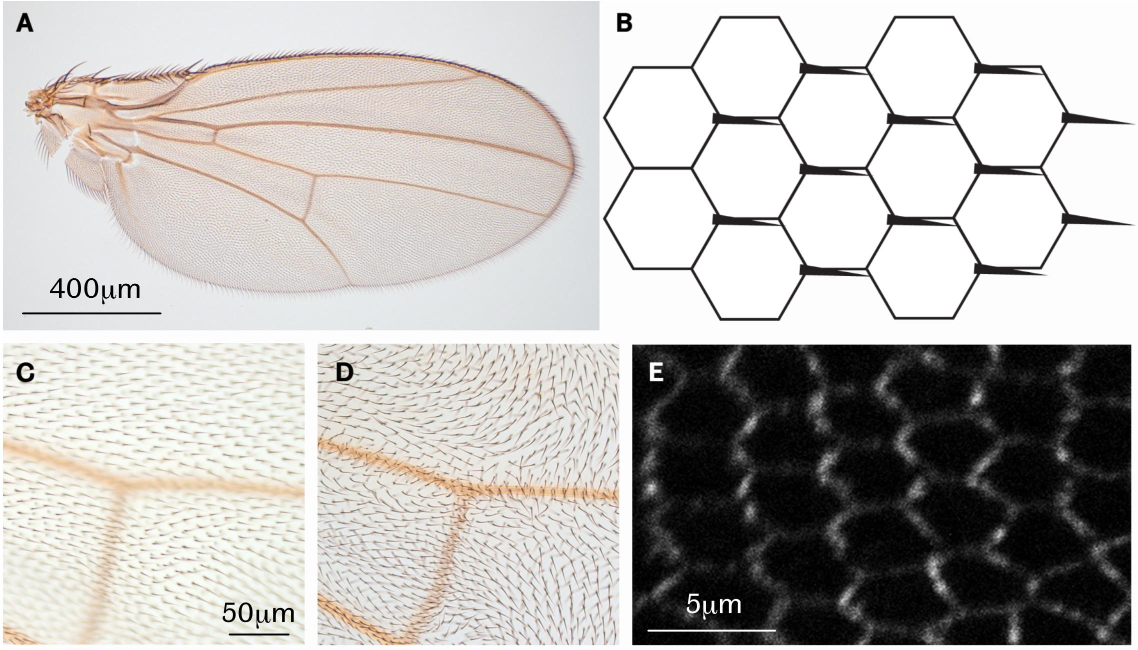


Figure 1.1: Planar polarity in the *Drosophila* wing

A: The wing of the adult fly consists largely of a bilayered epithelium, with several veins. **B:** The wing epithelium is a regular array of hexagonal cells, each of which produces a single distal-pointing trichome (hair) from its distal vertex. **C:** In the wild-type wing, the hairs all align with each other and point in the distal direction. **D:** In a planar polarity null mutant (*stbm^6*), the hairs lose their global coordination and adopt swirling patterns. **E**: In the pupal wing 28 hours after the onset of pupation, the core polarity proteins adopt a distinctive asymmetic localisation. Image shows EGFP-Stbm.

reported in Gubb and García-Bellido (1982), and in more detail in Vinson and Adler (1987). These studies showed that *fz*-induced polarity disturbances propagate only in wild-type cells distal to the clone, never proximally. Expressing high levels of *frizzled* at the distal tip of the wing causes the trichomes in this area to reorient to point proximally (Adler et al., 1997). Vinson and Adler (1987) additionally showed that some *fz* alleles are cell autonomous, suggesting that Fz functions in both the reception and transmission of the planar polarity signal.

Cloning and sequencing of the *fz* locus revealed a 62 kDa, 582 amino acid seven-pass transmembrane protein with an extracellular N-terminus (Vinson et al., 1989; Park et al., 1994). Most proteins of this class are G protein-coupled receptors (GPCRs), but the Frizzled receptors are generally considered to form a non-G protein-binding sub-group within the GPCR family (Bjarnadottír et al., 2006; Krishnan et al., 2012). However, this membrane-spanning structure suggested a mechanism by which the Fz protein could both receive and transmit the polarity signal. Fz is glycosylated. It also localises to the cell membrane and is transcribed evenly throughout the pupal wing (Park et al., 1994). An analysis of several autonomous and non-autonomous alleles of *fz* revealed that the autonomous alleles all encode mutations in the intracellular domains, whereas the nonautonomous lesions are distributed throughout the protein (Jones et al., 1996; Wu and Mlodzik; 2008; Strutt et al., 2012), consistent with the cytoplasmic domain being involved in intracellular signalling.

The members of the Frizzled (Fz) protein family are best known as receptors for Wingless (Wg) signalling (Bhanot et al., 1996), which bind Wnt ligands on the extracellular side of the cell membrane, and Dishevelled (Dsh) on the intracellular side. Wg signalling is involved in patterning and cell fate decisions during development, and severe *wg* mutations cause anterior-posterior patterning defects and lethality (eg. Klingensmith et al., 1994). Ironically, Frizzled itself is not greatly involved in Wg signalling in *Drosophila* – this role is largely filled by its homologue, Dfz2. Dfz2 is responsible for the majority of the Wg signalling (Bhanot et al., 1996; Bhat, 1998) and is incapable of signalling in the planar polarity pathway (Rulifson et al., 2000). On the other hand, Fz is specialised for planar polarity signalling, but is capable of low levels of Wg signalling, although its affinity for Wg is approximately 10-fold lower than that of Dfz2 (Bhat, 1998; Kennerdell and Carthew, 1998; Rulifson et al., 2000).

1.2.2 Dishevelled

The *dishevelled*^{1} (*dsh*^{1}) allele was first mentioned in a new mutants report by Fahmy and Fahmy in 1959. Perrimon and Mahowald (1987) used the embryonic segment polarity phenotype of lethal *dsh* alleles to place it into the *wingless* signalling pathway, whereas Adler (1992) identified *dsh* as a member of the planar polarity pathway.

The *dsh* gene was cloned independently by two groups in the same year (Klingensmith et al., 1994; Theisen et al., 1994). The protein consists of 623 amino acids and is approximately 70 kDa (Klingensmith et al., 1994). Klingensmith et al. used the lethal *dsh*^{V26} null allele to identify Dsh as an intracellular mediator of Wingless (Wnt1) signalling. *wg* and *dsh* null clones have similar phenotypes and the proteins are required together in the same tissues at the same stages of development; flies null for *wg* or *dsh* are larval lethal. A potential kinase-binding site, glycosylation and myristoylation sites, and a possible nuclear localisation signal were also identified. Clones in the dorsal thorax showed abnormal trichome polarity, revealing a function for Dsh in the planar polarity pathway as well as the Wg tissue patterning system (Adler et al., 1997; Klingensmith et al., 1994). Dsh also contains a DEP domain, a DIX domain and a PDZ domain (Axelrod et al., 1998).

Theisen et al. (1994) further investigated the *dsh* polarity phenotype, using the viable hypomorph *dsh*^{1}. Examination of *dsh*^{1} adults showed planar polarity phenotypes in the abdomen, thorax and eye. They also noted that the proteins most similar to Dsh are involved in junctional complexes, which suggested that Dsh might have a similar role. Both Theisen et al. (1994) and Klingensmith et al. (1994) noted the autonomous phenotype of *dsh* clones, suggesting that Dsh is involved in responding to signals rather than signal propagation.

Yanagawa et al. (1995) studied the embryonic cellular localisation of Dsh by immunostaining. They found that Dsh was largely cytoplasmic, enriched at apicolateral membranes, and absent from the nucleus. Additionally, cell culture experiments showed that Dsh is hyperphosphorylated in response to Wg and that the hyperphosphorylated form is more tightly localised to membranes. Krasnow et al. (1995) showed that Fz interacted with Dsh independently of Wg signalling.

Finally, Axelrod et al. (1998) and Boutros et al. (1998) demonstrated that the Wg signalling and planar polarity functions of Dsh were separable. Fz recruits Dsh from the cytoplasm to the membrane. Mutation of the DEP domain of Dsh abolishes the membrane localisation and disrupts only planar polarity signalling. In contrast, the *Drosophila* Wingless receptor Dfz2 does not cause relocalisation of Dsh, and the DEP domain is not required for Wg signalling. This difference in localisation could be a mechanism by which the two pathways can use the same cytoplasmic signalling molecule. The same papers showed that the DIX domain is solely required as an anchoring module, assisting with protein localisation. The PDZ domain was not required for Dsh to localise at the cell membrane, but deletion of this domain gave a dominant-negative planar polarity phenotype, indicating that this domain is involved in polarity signalling. The PDZ domain additionally binds Naked Cuticle, a component of the Wg signalling pathway (Rousset et al., 2001).

1.2.3 Prickle

prickle (*pk*) was another of the early planar polarity loci to be identified (new mutants report, Goldschmidt, 1944), and, along with *spiny-legs* (*sple*) and *frizzled*, was included in Gubb and Garíca-Bellido's 1982 analysis of the genetic factors controlling planar polarity. The *pk* locus was cloned by Gubb et al. (1999), revealing that *pk* and *sple* are in fact isoforms of the same gene. Pk and Sple differ only at their N-termini. *pk* uses the most 5' exon of the gene, *sple* starts at the second exon, and the rest of the exons are common to both isoforms (see **Figure 2.6 A**). The Pk protein is 93 kDa and Sple is 129 kDa. A third isoform, PkM, does not have any phenotype in the adult and is not considered further in this thesis. *pk-sple*^{13}, the null allele used throughout this project, is null for both *pk* and *sple*.

The initial identification of *pk* and *sple* as two separate genes was due to their different phenotypes. Loss of *pk* affects the wing and notum, while *sple* manifests in the legs, abdomen and eye (Gubb et al., 1999). The cause of these different phenotypes is a subject of ongoing research. A *pk/pk-sple* or *sple/pk-sple* mutation gives a less severe phenotype than *pk* or *sple* alone, suggesting that the balance between the two is important (Gubb et al., 1999; Strutt et al., 2013). However, *pk* and *sple* are expressed at similar levels throughout the fly (Gubb et al., 1999, Lin and Gubb, 2009).

Subsequent work on Pk and Sple showed that the short N-terminus of Pk is not required for planar polarity (Lin and Gubb, 2009). Pk and Sple both contain a PET domain followed by three LIM domains (Gubb et al., 1999). The LIM domain contains two zinc-finger motifs, common domains that are usually involved in protein binding. Pk additionally has a C-terminal prenylation motif, which is farnesylated *in vivo*, and is required for its normal localisation and function in the planar polarity pathway (Veeman et al., 2003; Maurer-Stroh et al., 2007; Lin and Gubb, 2009; Strutt et al., 2013). Prenylation is often associated with membrane localisation of cytoplasmic proteins (Zhang and Casey, 1996). Pk is known to bind to Strabismus (Stbm) (Jenny et al., 2003; Bastock et al., 2003), and prenylation of Pk has recently been shown to promote its interaction with Stbm (Strutt et al., 2013).

This project focussed on the *Drosophila* wing, so Pk was the appropriate isoform to study and the role of Sple was not investigated.

1.2.4 Strabismus

Strabismus (Stbm) (Wolff and Rubin, 1998), also called Van Gogh (Vang) (Taylor et al., 1998), is another transmembrane core protein. It has 584 amino acids and is 65 kDa (Wolff and Rubin, 1998; Lindqvist et al., 2010), with four transmembrane domains and a C-terminal PDZ-binding motif (Wolff and Rubin, 1998). PDZ-binding motifs are short sequences that interact with PDZ domains (Songyang et al., 1997). Wolff and Rubin identified a recessive, autonomous allele of *stbm* based on its rough eye phenotype, whereas Taylor et al. found a dominant, nonautonomous allele in their screen of the adult wing. Since the Wolff and Rubin paper was published before the Taylor et al. results, this thesis will refer to this gene as *strabismus*.

The nonautonomy of *stbm* clones propagates in the proximal direction, the opposite direction to *frizzled*. Likewise, the reoriented hairs point away from the clone rather than towards the clone as in *fz*. Generating *fz* clones in a *stbm* mutant background, or vice versa, suppressed the nonautonomy, showing that Fz and Stbm function in the same pathway. (Taylor et al, 1998). Taylor et al. (1998) and Wolff and Rubin (1998) both found duplication of tarsal segments in *stbm* legs and the usual general disturbances of planar polarity throughout the adult fly.

A recent paper identified multiple phosphorylation sites in Vangl2, the mouse homologue of Stbm (Gao et al., 2011). mRNAs encoding mouse Vangl2 with mutated phosphorylation sites were injected into zebrafish embryos. In wild-type or *tri/+* fish (*trilobite* is the fish Stbm homologue) the injected mRNAs caused severe planar polarity phenotypes. In *tri/tri* fish the mutated mRNA failed to rescue the PCP phenotype. This indicates that the phosphorylation state of Stbm is important for its function.

1.2.5 Starry Night/Flamingo

Like *pk*, *flamingo* (*fmi*) also has two isoforms with two different names. One isoform, *flamingo*, was discovered by Usui et al. (1999), just one month before the *starry night* (*stan*) isoform was published by Chae et al. (1999). The isoforms differ from each other in the inclusion of a 7bp microexon before the final exon. *stan* omits this microexon and splices straight from exon 3 into exon 5, whereas *fmi* includes the microexon between exons 3 and 5, but consequently splices into exon 5 in a different reading frame to *stan*.

The net result is that the more abundant Stan isoform terminates with a PDZ-binding motif that is absent in the less common Fmi (Wasserscheid et al., 2007), see also **Figure 2.6E**. In this thesis, this gene is generally referred to as *fmi*. However, the *fmi* construct used in the experiments produces only the Stan isoform, so *stan* and *fmi* are used to refer to the respective gene and protein isoforms when the distinction is relevant to the discussion. Note that every *fmi* mutation used in this project affects both the Fmi and Stan proteins.

Fmi is a large protocadherin, around 300 kDa, with almost 3,600 amino acids (Usui et al., 1999; Chae et al., 1999). The classical cadherins are cell adhesion molecules, whose large extracellular domains bind to each other in *trans* across cell membranes in a calcium-dependent manner (reviewed in Shapiro and Weis, 2009). Fmi has nine extracellular cadherin domains, two laminin G motifs and four EGF motifs, followed by seven transmembrane domains and a much smaller intracellular tail (Usui et al., 1999; Chae et al., 1999). The intracellular domains of classical cadherins link to the cytoskeleton via catenins, but Fmi lacks a catenin binding site, suggesting that it is not involved in cell adhesion processes (Usui et al., 1999; Chae et al., 1999). Like Fz, its similarity to G-protein coupled receptors suggested that it might be involved in signal reception (Usui et al., 1999; Chae et al., 1999), but Usui et al. additionally showed that Fmi had some homophilic binding activity when expressed in culture cells. This suggested that Fmi might have a role in both intercellular adhesion and signal transduction.

The mutant phenotype of *fmi* placed it into the core planar polarity family. While null alleles cause lethal defects in neurogenesis (Usui et al., 1999), hypomorphs or clones of null mutant tissue in adults show the familiar phenotype of swirling wing hairs, misrotated ommatidia, and defects in bristle alignment (Usui et al., 1999; Chae et al., 1999). Null clones of *fmi* are autonomous; having no effect on the polarity of neighbouring wild-type cells (Usui et al., 1999; Chae et al., 1999).

1.2.6 Diego

Diego (Dgo), the final core planar polarity protein, was identified by Feiguin et al. (2001) in a notum-based overexpression screen for new polarity mutants. The null mutant shows the classic wing and eye phenotypes, but does not affect the polarity of bristles on the notum (Feiguin et al., 2001). Its polarity phenotypes are less severe than those of the other core proteins in all tissues examined (eg. Das et al., 2004). Dgo is a 106 kDa protein with 7 N-terminal ankyrin repeats (Feiguin et al., 2001). Ankyrin repeats tend to be involved in membrane binding (Bennett and Chen, 2001), so these domains may facilitate localisation of Dgo at the cell membrane. With no other obvious domains or homology to any other known proteins, it is difficult to predict its molecular function.

1.3 Putting the pieces together: genetic and physical interactions

The similar systemic phenotypes of the six core proteins strongly suggested that they were all members of the same pathway. Wong and Adler (1993) were one of the first to put the known components of planar polarity signalling into an epistatic order and sketch out a potential structure of the pathway, placing Fz, Dsh and Pk firmly into the same group.

Further genetic work established firm links between the core proteins. The interactions between Fz and Dsh were cemented by Krasnow et al. (1995). Since the phenotypes of Fz, Dsh and Pk were too similar to be distinguished by a traditional epistasis experiment, they instead overexpressed *fz* at a late stage of development. This caused formation of multiple trichomes per wing cell, an effect that was blocked by removal of *dsh*. This showed that Dsh functions downstream of Fz and is required for its function in planar polarity. These authors additionally showed that removing one copy of *dsh* ameliorated their *fz* overexpression phenotype, further strengthening the Fz-Dsh link. Pk did not interact with Fz in this manner, showing that although it was clearly a key member of the core polarity system it did not function directly with Fz. Removal of *fmi* or *stbm* prevents *fz* overexpression from reorienting polarity towards the zone of higher *fz* expression (Taylor et al., 1998; Chae et al., 1999), adding Fmi and Stbm to the emerging group. Gubb et al. (1999) also showed no clear epistasy between *pk* and *fz*, although they were able to confirm from a *dsh*; *pk* double mutant that Dsh and Pk both acted in the same pathway.

In addition to these extensive genetic and phenotypic interactions between the core proteins, some physical interactions have also been reported. Several of the core proteins contain domains that are often involved in protein interactions, for example, the Fmi cadherin repeats (Usui et al., 1999; Chae et al., 1999) and the Pk LIM domains (Gubb et al., 2009).

Tree et al. (2002) used transfected human osteosarcoma cells to show that *Drosophila* Pk binds to the Dsh DEP domain via its PET and LIM domains. Dgo interacted with a region close to the C-terminus of Pk and with the intracellular domain of Stbm in a yeast two-hybrid assay, and the ankyrin repeats of Dgo interacted with Pk and Stbm in a pull-down assay (Das et al., 2004). The regions of Stbm and Pk that interacted with the Dgo anykrin repeats had previously been shown to bind to each other, and Stbm to itself, in a separate pull-down assay by the same group (Jenny et al., 2003).

The same authors also showed that Dsh and Dgo co-immunoprecipitated (co-IP) when expressed in HEK293 cells (human embryonic kidney), and that Dsh could pull down Dgo, Pk and Stbm (Jenny et al., 2005). Park and Moon (2002) were able to co-IP endogenous Stbm and Dsh from *Xenopus* embryos and showed that the Dsh PDZ domain binds to Stbm, but not via the PBM of Stbm. Bastock et al. (2003) were also able to co-IP Stbm and Dsh, this time in COS7 cells (African green monkey kidney), and additionally showed that Stbm could recruit both Pk and Dsh to the cell membrane. However, these results should perhaps be interpreted with caution, due to the unphysiological nature of the assays used. These interactions are summarised in **Figure 1.3**.

1.4 Core protein localisation

1.4.1 The core proteins are localised asymmetrically

The planar polarity proteins Fz and Stbm show directional nonautonomy in mutant clones, in that the polarity of adjacent wild-type cells are affected by the clone but only on the proximal or distal side of the clone, never both. Clones of cells overexpressing any core protein produce the same effect (Feiguin et al., 2001; Tree et al., 2002; Strutt and Strutt, 2008). This unusual directional nonautonomy suggested that the distribution of the nonautonomous polarity proteins might also be polarised.

The subcellular localisation of the core planar polarity proteins has been examined by immunostaining and by expression of GFP-tagged fusion proteins *in vivo*. These studies revealed that prior to trichome formation all six of the core proteins localise to the apicolateral membrane in a strikingly asymmetric fashion, being strongly enriched on the proximal-distal axis (Usui et al., 1999 (Fmi); Axelrod, 2001 (Dsh); Feiguin et al., 2001 (Dgo); Strutt, 2001 (Fz); Tree et al., 2002 (Pk); Bastock et al., 2003 (Stbm)). This localisation in the hexagonal cells of the pupal wing results in a distinctive zig-zag pattern (see **Figure 1.1 E**). The core proteins localise in the same apico-basal plane as the adherens junctions, but do not colocalise with adherens junction components in the plane of the tissue.

Furthermore, analysis of clones showed that Pk and Stbm localise only to the proximal edges of the cell, while Fz, Dgo and Dsh are found on the distal side (Strutt, 2001; Axelrod, 2001; Tree et el., 2002; Bastock et al., 2003; Das et al., 2004). Fmi is the only core protein that is found on both the proximal and the distal membranes (Strutt and Strutt, 2008). This localisation tallies with the putative ability of Fmi to form homodimers (Usui et al., 1999), as it would allow Fmi homodimers to form across cell boundaries, which could potentially be a molecular mechanism for communication of the polarity signal from cell to cell.

When any one of the core planar polarity proteins is removed, the asymmetric localisation of the remaining core proteins is lost (eg. Shimada et al., 2001; Tree et al., 2002; Bastock et al., 2003), revealing a mutual dependence among the core proteins for their junctional apicolateral localisation. In some cases the protein still reaches the apicolateral membrane but is evenly distributed on all sides, whereas for others the membrane localisation is almost entirely abolished (Shimada et al., 2001). Feiguin et al., (2001) demonstrated that Fmi was required for the normal localisation of Dgo, and Shimada et al. (2001) showed that Fmi and Dsh were mutually reliant on each other for their localisation. Pk requires Stbm for its localisation, and Fmi requires Fz (Usui et al., 1999; Bastock et al., 2003; Rawls and Wolff, 2003).

Loss of asymmetric localisation of the core proteins is always accompanied by a corresponding loss of planar polarisation of trichomes, and the localisation of the core proteins always corresponds with the direction of polarity, even in abnormally-polarised cells (Strutt, 2001; Axelrod, 2001). The position of trichome initiation also correlates with the polarised localisation of the core proteins (eg. Strutt and Strutt, 2007). This asymmetric localisation and clear link to cell polarity suggested a cellular mechanism for the propagation of planar polarity.

In the wing, this asymmetric localisation changes throughout development. At 25°C, the core proteins are asymmetrically localised from around 22 hours after pupation until after the formation of the trichomes. The strongest asymmetry is seen around 28 to 32h after pupation (Shimada, 2001; Axelrod, 2001). At earlier stages, including around 20h when the wing is undergoing much cell division and growth, the core proteins are distributed all around the edges of the cells, and at even earlier stages they are once again asymmetric (Usui et al., 1999; Shimada et al., 2001; Axelrod, 2001; Classen et al, 2005; Aigouy et al., 2010; Strutt et al., 2011; Brittle et al., 2012, Sagner et al., 2012). The dynamic nature of the core protein localisation suggests that it is an active process, which is either developmentally controlled or constantly functional but perturbed by morphogenetic processes.

It has been demonstrated that overexpression of Fmi preferentially recruits the distally-localised proteins Fz and Dsh to the membrane, reducing membrane-associated levels of the proximally-localised proteins Stbm and Pk (Strutt and Strutt, 2008). Fmi appears to have an inherent preference for binding to other Fmi molecules that are themselves able to interact with Fz. To form stable intercellular Fmi homodimers, Fz is required at the distal edge of the proximal cell and Stbm:Pk are required at the proximal edge of the distal cell. In the absence of Stbm or Fz, the symmetric localisation of Fmi is lost and more Fmi is seen in endocytic vesicles. These results suggest that the asymmetric complex is more stable than any alternative symmetric configuration, and both Fz and Stbm are required to stabilise homophilic Fmi interactions (Strutt and Strutt, 2008).

1.4.2 The core protein complex

These various genetic and physical interactions, coupled with the codependence of the core proteins for their subcellular localisation, show that the functions of the six core proteins are intimately linked. The similar phenotypes of the six core proteins, along with the numerous genetic and physical interactions described above, strongly suggests that all six act together as members of a protein complex (Tree et al., 2002; Strutt and Strutt, 2007).

This putative core protein complex would therefore form across cell membranes. It is based on a Fmi homodimer, with Fz on the distal edge of the more proximal cell mirrored by Stbm on the proximal edge of the more distal cell. The cytoplasmic factors Dsh and Dgo would then localise to the Fz intracellular domain, and Pk would localise with Stbm (**Figure 1.2**). The asymmetric localisation of the core proteins thus forms the basis of an asymmetric complex, which could be integral to the propagation of the planar polarity signal.

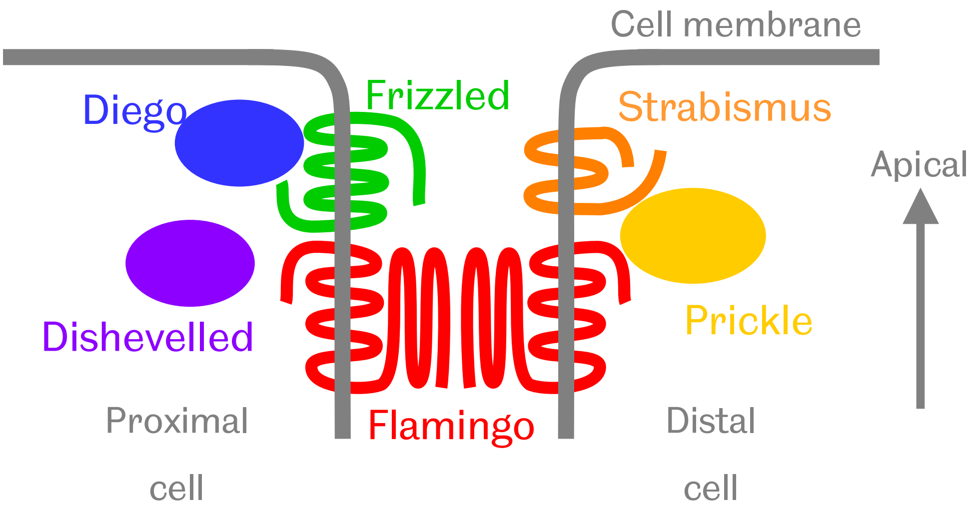


Figure 1.2: The putative core protein complex

The complex is asymmetric and forms across cell boundaries. A Fmi homodimer is accompanied by Fz, Dgo and Dsh on the distal edge of the proximal cell, and by Stbm and Pk on the proximal edge of the more distal cell.

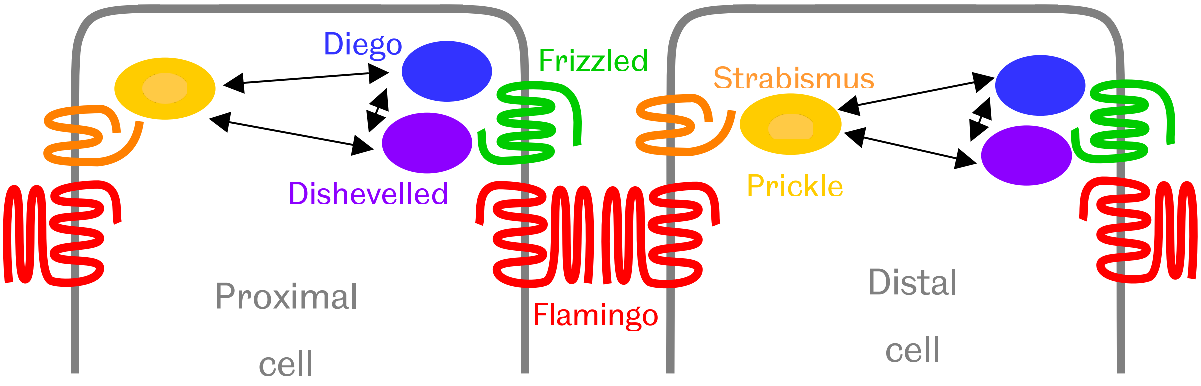


Figure 1.3: Core protein interactions

Fz recruits Dsh from the cytoplasm to the membrane. Pk binds to Stbm, but does not interact directly with Fz. Pk, Dgo and Dsh may interact with each other in the cytoplasm. Dgo and Dsh may also interact with Stbm. Fmi forms homodimers.

1.4.3 The core proteins are localised in a punctate manner

Another interesting aspect of core protein localisation is its punctate nature. The core proteins do not form a smooth line along the membranes where they localise, instead forming several bright spots, or 'puncta'. These puncta are where the core proteins all colocalise together and are the points of maximum core protein asymmetry. The puncta are thus very likely to be the sites where the core proteins are actively involved in planar polarity signalling. Presence of puncta correlates with stable planar polarity, and loss of puncta with loss of polarity. A recent study comparing the dynamics of puncta with regions of membrane that do not contain bright puncta showed that the turnover of core proteins in puncta is much less than that of core proteins not in puncta (Strutt et al., 2011). This difference in dynamics between the puncta and the rest of the membrane strongly suggests that the puncta are actively established and maintained (or conversely, are not actively removed against a background of polarity protein turnover).

With a mean area in the region of $0.3\mu m^2$ (see **Figure 3.3**), puncta are much too large to represent individual core protein complexes. The puncta could therefore be explained as clusters of individual core complexes (Strutt and Strutt, 2008). If the core complex is the primary transducer of the planar polarity signal then the polarity of the complexes must be coordinated across cell boundaries. Each cell expresses all six core proteins, so is capable of generating both Fz:Fmi and Stbm:Fmi sub-assemblies. For correct polarised localisation of the core proteins, Fz:Fmi must be restricted to the distal edge of the cell, and Stbm:Fmi interactions to the proximal edge. One way to achieve this would be for complexes of the same polarity to agglomerate into puncta, while simultaneously destabilising or otherwise removing complexes of the opposite polarity.

1.5 The wider polarity network

As described above, Pk, Stbm, Dsh, Dgo, Fz and Fmi form the core module of the planar polarity pathway and mutations in any one of these affects multiple parts of the fly body. However, several other planar polarity genes have been identified, which are involved either in interpretation of the planar polarity signal, or in regulatory roles that function upstream of, or parallel to, the core proteins.

1.5.1 The effectors

Some genes only affect planar polarity in a specific tissue. These genes are designated the ‘effectors’, and their role is to couple the planar polarity signal generated by the main planar polarity machinery into the relevant output for each particular tissue. From the early beginnings of the planar polarity field, several genes had been identified that specifically affected planar polarity in just one or a few tissues. *multiple wing hairs*, *inturned* and *fuzzy* are three of the oldest known polarity genes. These three all affect the trichomes of the wing, with *inturned* and *fuzzy* mutants additionally manifesting in bristles. *inturned* and *fuzzy*, together with *fritz*, respond to the planar polarity signal to orient the trichomes and bristles of the wing and abdomen (Gubb and García-Bellido, 1982; Wong and Adler, 1993; Collier and Gubb, 1997; Collier et al., 2005). All four genes restrict each cell to produce a single hair, again in response to the planar polarity signal (Gubb and García-Bellido, 1982; Wong and Adler, 1993; Collier et al, 2005).

In the eye, planar polarity is required for correct rotation of the developing ommatidia. The serine/threonine kinase *nemo*, the GTPase *RhoA,* and *Drok* (*Drosophila* Rho-associated kinase) mediate ommatidial rotation in response to planar polarity signalling (Choi and Benzer, 1994; Strutt et al., 1997; Winter et al., 2001). Mutations in these genes cause ommatidial misrotation phenotypes.

The effectors are often identified by their planar polarity phenotypes, but their effects are due to a failure of the cells to respond to the polarity signal rather than a failure to generate it. The effectors function strictly downstream of the core and are responsible for translating the core proteins’ polarity signal into the appropriate output for their particular tissue. This covers diverse processes including cilia alignment, directed cell growth, migration or division, as well as trichome orientation and ommatidial rotation as discussed above. In many cases, the exact molecular links between the core and the effectors are not known, but the localisation and behaviour of the core proteins is always independent of the effectors (Strutt and Strutt, 2009; Axelrod, 2009; Goodrich and Strutt, 2011). Due to their tissue-specific nature, and their position downstream of the core proteins, the effectors are not considered further in this project.

1.5.2 The ‘upstream’ module

While it is clear that the core proteins are instrumental in mediating planar polarity, they must receive an upstream input to dictate the axis of planar polarity in the tissue, as nothing that is currently known about them suggests they could achieve this themselves.

The protocadherins Fat (Ft) and Dachsous (Ds), together with the Golgi protein Four jointed and the myosin Dachs (Cho and Irvine, 2004), form a module that functions upstream of the core planar polarity proteins, hereafter termed the upstream module. Observations in some fly tissues imply that the Ft/Ds system functions strictly upstream of, and feeds into, the core proteins (Adler et al., 1998; Amonlirdviman, 2005), whereas other work suggests that the relationship between the two modules is more complicated (Strutt and Strutt, 2002). Tree et al. (2002) suggest that this upstream module could fill the role of the global polarity-coordinating cue alongside another upstream factor such as a Wnt. However, Casal et al. (2006) showed that the Ft/Ds system is able to generate planar polarity in the absence of the core proteins, indicating that it can determine planar polarity independently of the core complex. Lawrence et al. (2007) highlight some residual questions about this model, reframing the debate to ask how the core and upstream modules generate polarity independently and how they relate to each other.

Ft and Ds are large proteins: Ds has 27 extracellular cadherin repeats and Ft has 34, in contrast to the usual 4 repeats found in classical cadherins (Mahoney et al., 1991; Clark et al., 1995). Ft and Ds interact with each other via their cadherin repeats (Casal et al., 2006), and Ds and Fj are expressed in gradients (Clark et al., 1995; Zeidler et al., 1999). Graded expression of Fj is expected to polarise Ft (reviewed in Lawrence et al., 2007). Like the core proteins, Ft and Ds were also recently shown to be asymmetrically localised in the cells of the wing and eye imaginal discs (Ambegeonkar et al., 2012, Brittle et al., 2012). Dachs localises to the membrane in a polarised manner in response to Ft and Ds signalling (Mao et al., 2006). As for the core complex, this asymmetric localisation suggests a mechanism by which the upstream module could provide a planar polarity signal.

The upstream module differs from the core complex in that its members affect both planar polarity and tissue growth. *ft* mutants develop large tumourous overgrowth of imaginal tissues, which often have reversed planar polarity (Mahoney et al., 1991). *ds* mutations do not cause overgrowth but result in generalised defects in morphogenesis, including shortened legs, rough patches in the eye and wing hinge defects (Clark et al., 1995).

There is evidence that Ft and Ds might polarise the core proteins only in the proximal regions of the pupal wing (Sagner et al., 2012). The model is that the cells compare levels of active Ft and Ds on both sides of their membranes and use this signal to polarise the core proteins to amplify said signal. Brittle et al. (2012) showed Dachs asymmetry was not lost in a *fz* mutant wing disc, again showing that Ft/Ds can polarise in the absence of a functional core complex. However, *dachs*, *ft dachs* and *dachs ds* mutant wings show subtle planar polarity phenotypes in the proximal wing (Brittle et al., 2012). While the core complex is not absolutely dependent on Ft/Ds for asymmetric localisation of the core proteins, loss of *ft* or *ds* causes an associated loss of core protein polarisation in the proximal wing at early stages of development (Brittle et al., 2012).

The core pathway defects seen in Ft/Ds module mutations may be partly due to tissue overgrowth, as polarity defects are reduced if cell overgrowth is suppressed in a *ft* or *ds* background (Brittle et al., 2012). Cell polarity is lost during mitosis and must be re-established in the daughter cells. In a tissue experiencing excessive cell proliferation, the constant loss of asymmetry may be great enough to disrupt the polarity of the entire tissue. A related idea is suggested by Lawrence et al. (2007) to explain some of the contradictory results seen in different tissues; these authors suggest that the severe polarity disruptions seen in *fz* eyes and wings disturb these tissues to the extent that they are unable to respond appropriately to the Ft/Ds system.

If cells respond both to the gradient of Ft/Ds and to the polarity signals from their neighbours' core complexes this could explain some of the interplay between the two systems. Sagner et al. (2012) propose that multiple factors are responsible for coordination of the direction of planar polarity across a tissue, including the process of tissue growth and morphogenesis itself. The relationships between the different components of the planar polarity machinery are thus not completely understood. However, the core proteins and the Ft/Ds module are active in almost every tissue that has been studied. Fortunately for this project, the self-contained nature of the core complex renders it amenable to focussed study without the need to elucidate the workings of the wider polarity network.

1.6 Planar polarity in vertebrates

While much of the work on planar polarity has been done in *Drosophila*, there are also unique insights to be gained from vertebrates. The components of the planar polarity machinery are conserved in vertebrates, and the vertebrate homologues of the core planar polarity complex appear to have very similar roles to their *Drosophila* counterparts. The vertebrate homologue of *flamingo* is the somewhat less imaginatively named *CELSR* (cadherin EGF LAG seven-pass G-type receptor). Celsr1, Vang-like 1 and Vang-like 2 (Vangl, the vertebrate Stbm/Vang), Pk2, Dvl2 and Dvl3 (Dsh) and the multiple Frizzled receptors all have polarised localisation in vertebrates (reviewed in Goodrich and Strutt, 2011).

While it appears to follow the same general principles as *Drosophila*, the vertebrate planar polarity system is much more complicated than that of flies. The ancestral vertebrate linage experienced two full genome duplications after splitting from the arthropods (Sidow, 1996), which means it is common to find multiple homologues of *Drosophila* genes in the vertebrate model organisms. The planar polarity genes are no exception, so, although the planar polarity machinery is largely conserved through evolution, research on vertebrate polarity is often complicated by redundancy and difficulties with identifying the most homologous homologue to the usually better-known *Drosophila* gene. For example, the mouse has three *dsh* homologues and 10 *frizzled*s (Simons and Mlodzik, 2008). Additionally, the body plan of vertebrates is more complicated than that of flies. Morphogenesis is consequently a much more complex procedure, and polarity mutations produce more complex phenotypes which can be more difficult to interpret.

However, vertebrate polarity research can answer questions that cannot be investigated in *Drosophila*. The process of neural tube closure has no direct analogue in invertebrates. Planar polarity is involved in the morphogenetic processes that shape the neural tube and severe defects in neural tube closure are a hallmark of vertebrate planar polarity mutants. Open neural tubes and exencephaly are commonly seen when core planar polarity complex components are deactivated in mouse (Kibar et al., 2001; Curtin et al., 2003) and *Xenopus* (Darken et al, 2002). Failure of neural tube closure is fatal in vertebrates, but very mild defects in posterior tube formation result in spina bifida in humans. *Vangl1* mutations have been linked to neural tube defects in humans (Kibar et al., 2007), and *Looptail* (*stbm*) mice (Kibar et al., 2001).

Neural tube formation and body axis elongation are both processes that require convergent extension, a process in which the cells of a tissue converge along one axis, intercalate, and extend in the perpendicular direction. Mutations in the planar polarity system result in a shortened body axis due to failure of convergent extension (Park and Moon, 2002; Darken et al., 2002; reviewed in Keller, 2002), which is another classic vertebrate planar polarity phenotype.

Additionally, many human planar polarity disorders manifest in organs that rely on ciliated epithelia for their function. Flies largely lack cilia, so this work has to be done in vertebrates. At early stages of development, the cells of a structure termed the ventral node each produce multiple cilia. When the planar polarity system is functioning, these cilia align and their coordinated beating results in directed fluid flow, which determines the left-right asymmetry of the body (Antic et al., 2010; Hashimoto et al., 2010; Song et al., 2010). Brain development is affected by the flow of cerebrospinal fluid, which is controlled by coordinated cilia in the ventricular epithelium (Hirota et al., 2010).

As in *Drosophila*, simple examples of alignment of arrays of cells can also be found in vertebrates. The node, which was mentioned above, is one example of this. Another classic vertebrate manifestation of planar polarity is seen in the sensory hair cells of the inner ear. These cells produce stereocilia in a chevron formation behind a single kinocilium. Under normal circumstances the chevrons all point in the same direction, but planar polarity mutants have incorrectly oriented hair cells (e.g. in mice, Curtin et al., 2003). Ciliated cells in the *Xenopus* epithelium form in more basal layers and migrate apically, only polarising on reaching the surface layer. Wild-type cells arriving at a polarity mutant surface fail to polarise (Mitchell et al., 2009). The orientation of the multicellular follicles that produce the mouse fur is also affected in a *Fz6*, *Vangl2* or *Celsr1* mutant (Guo et al., 2004; Devenport and Fuchs, 2008).

Planar polarity has an essential role in the control of the major morphogenetic movements that occur at the very beginning of life. Pk1 has been implicated in gastrulation movements and neuronal migration in zebrafish (Carreira-Barbosa et al., 2003; Veeman et al, 2003). Fly embryos can tolerate such mutations. Perhaps this is because they are smaller and their cells have to cover less distance, their tissues have fewer cells to coordinate and they do not have so many complicated internal organs as vertebrates.

Failure of oriented cell division, another process under the control of planar polarity, can also have severe consequences. Polycystic kidney disease (PKD) is one of the leading causes of kidney failure in humans, affecting 1 in 800 live births (Wilson, 2004). In PKD, the renal tubules fail to elongate properly, resulting in widened ducts that develop into large cysts (Fischer et al., 2006; Saburi et al., 2008).

Diseases related to planar polarity often have systemic effects, reflecting the ubiquity of planar polarity throughout the animal body. For example, polycystic kidney disease is often associated with cysts in the intestine, pancreas and liver, in addition to heart valve defects and an increased risk of brain aneurysm (Wilson, 2004). While many failures of the planar polarity system cause severe and fatal congenital defects in humans, the milder defects such as spina bifida and polycystic kidney disease have a severe impact on life expectancy and quality of life.

1.7 Planar polarity in plants

Plants must also coordinate cells in the planes of their tissues. One of the main differences between plants and animals from the polarity perspective is the malleable body plan. Plant morphogenesis also proceeds largely by controlling the rate and direction of tissue growth, rather than by cell migration and coordinated tissue remodelling. This means that planar polarity is also likely to be quite different in these organisms.

The main determinant of plant polarity seems to be auxin. Auxin is a growth-promoting hormone, which is transported through cells to set up local concentrations and gradients in the tissues. Auxin production and distribution is tightly regulated by an array of different long- and short-distance transport mechanisms. At the cellular level, the PIN proteins mediate auxin influx, while the AUX proteins mediate efflux, and the polarised localisation of these proteins is an important contributor to polarised auxin transport (reviewed in Petrášek and Friml, 2009).

The PIN and AUX proteins are localised to opposite sides of the cells, and auxin is transported both proximally and distally throughout most tissues (reviewed in Grebe, 2004; Nakamura et al., 2012). Particularly in the root, polarity is thought to be cued by a gradient set up by auxin synthesis in the root tip, which is then transported distally back up the root. Each root cell produces a small protrusion at its distal end, and these are thought to be a manifestation of planar polarity. However, the localisation of the PIN proteins is complex and regulated on both the planar and apicobasal axes. However, PIN1 is localised along the proximal-distal axis at early stages of leaf development and appears to be intimately linked to both auxin regulation and the anisotropic growth of leaf tissue (Nakamura et al., 2012).

1.8 Modelling planar polarity

Despite the long history of work on planar polarity, several unresolved questions remain. Given the wealth of available genetic interaction data it should be a simple matter to understand the architecture of the planar polarity pathway, but due to the nonlinearity of the planar polarity network this remains one of the main questions in the field. Many different screens have been done, so it is likely that all, or almost all, of the components of the pathway have been identified. Any additional components must be hiding in plain sight, masked by redundancy or pleiotropy, or have very subtle effects on polarity. More molecular data are required to understand how the planar polarity system determines and coordinates the polarity signal but, in the absence of such data, mathematical modelling of planar polarity is providing some insights to the possible mechanisms of asymmetric localisation.

Any model must be consistent with the existing data. The local coordination of cells in core complex mutants demonstrates the presence of an additional polarising influence, whether this is the Ft/Ds module or some other mechanism. The nonautonomous behaviour of *fz* and *stbm* clones gives some clues to core protein behaviour – the fact that wild-type cells reorient themselves in these situations shows that the core proteins would rather make complexes on the 'wrong' cell edges than make no complexes at all. This also shows that the core proteins are capable of independently reorganising polarity, so they must play an active part in planar polarisation rather than passively respond to some underlying cue. At the cellular level, the drive to form complexes is prioritised over the need to coordinate with the global tissue polarity.

The residual local coordination of cells that persists in core protein mutants suggests that such cells are still able to communicate a polarity signal across short distances, but are unable to respond to the global polarity cue. There is some evidence that small amounts of Fz are transported in a polarised manner along microtubules (Shimada et al., 2006), but this could only work as part of a feedback mechanism, reinforcing an existing asymmetry within the cell.

A likely mechanism is that the upstream cue induces a small initial asymmetry in the core proteins, perhaps via Fz, which is then amplified by a feedback loop (Strutt, 2001; Axelrod, 2001). Localisation of a protein at a cell edge could locally inhibit or destabilise the localisation of another core protein that belongs at the opposite end of the cell. Such a mechanism could result in sorting the correct proteins to the correct ends of the cell. Polarisation could therefore be achieved by removal of proteins from the incorrect cell edges, with or without an associated increase in delivery to the correct cell edge.

Based on experiments with Pk, Tree et al. (2002) proposed a negative feedback model whereby Pk at the proximal edge of the cell inhibits localisation of Dsh, which in turn inhibits proximal localisation of Fz in the same cell, thus permitting distal localisation of Fz and Dsh. Some intercellular interaction between Fz and Pk would then either decrease distal Pk in the proximally adjacent cell, or further increase Pk on the proximal edge of the cell. In this model, Fmi homodimers are relegated to a complex-stabilising role and the cytoplasmic factors Pk and Dsh are promoted as the key mediators of feedback. The authors updated their negative feedback model in 2005 (Amonlirdviman et al.) to include Stbm and mathematical modelling of the interactions between the various core proteins as binding and dissociating from complexes.

Le Garrec et el. (2006) present a similar model, but with the negative feedback loop operating in the reverse direction. They centre their model on cell-cell communication via Fmi homodimers, where Fmi on one side of the membrane is bound to Fz and on the other side to Stbm. This model has two layers of intracellular feedback – Fz inhibits Stbm binding to Fmi, and Dsh inhibits Pk binding to Stbm. The authors emphasise that polarity is established simultaneously in all cells, rather than propagating from one end of the tissue to the other. Additionally, they find that in their model the upstream polarising cue should be weak (approximately 4% difference across a cell), transient, and should not significantly overlap the period during which polarity is established.

Burak and Shraiman (2009) approach the problem from the opposite direction to derive a phenomenon-based model, instead of trying to fit a model to the existing data. They have Fz and Stbm interacting across adjacent cell membranes, with Fz complexes inhibiting formation of Stbm complexes and vice versa. While the Tree-Amonlirdviman and le Garrec models are robust in the face of noise, the Burak and Shraiman model actively requires stochastic fluctuations for the cells to polarise in a realistic timescale. Their model has two stable steady states, one with polarisation centring on a single cell edge, and one centred on a vertex, incorporating the two adjacent edges. However, in this model the core proteins polarise before tissue-wide coordination of polarity emerges, whereas in the previous models polarity and coordination occurred simultaneously. In the absence of a global orienting cue, this results in a locally-coordinated swirling phenotype similar to that seen in the wings of core protein mutants.

Returning to the molecular perspective, Strutt et al. (2011) show that apicolaterally-localised Fmi is stabilised by binding to Fz or Stbm, and any Fmi that does not incorporate into puncta is removed from the membrane. Dsh, Pk and (weakly) Dgo are required for the clustering of the core proteins into puncta. The stability of Fz and Fmi in puncta was assessed by their decreased fluorescence recovery after photobleaching (FRAP) compared to membrane regions that did not contain bright core protein puncta, where these proteins are turned over more rapidly. The clustering of core proteins into puncta, and the increased stability of core proteins in puncta, suggests that this could be the cellular subdomain where any feedback mechanisms take place. This clustering mechanism also suggests a possible positive reinforcement of asymmetry, contrasting with the negative feedback loops in the mathematical models.

Cell packing has also been suggested to have a role in planar polarity (Gubb et al., 1999; Classen et al., 2005). At early stages of wing development the cells are more irregularly shaped than at later stages, when the familiar array of hexagonal cells emerges (Classen et al., 2005). If cell packing affects polarity, this could explain the distinctive and reproducible swirling wing hair phenotypes of core protein mutants. Under normal circumstances in the wing, each cell has a distal-most vertex and this is the point of prehair initiation (Wong and Adler, 1993). When cell packing is less regular, cells may end up with more or fewer vertices than normal, and present a straight edge towards the distal end of the wing. This could cause problems if polarity is normally actively oriented towards vertices as in the Burak and Shraiman model. Aigouy et al. (2010) note that the contraction of the wing hinge region during wing morphogenesis exerts a mechanical force on the tissue, which corresponds with the resulting axis of polarity. Mechanical forces could be another way of specifying a long-range polarity signal, which could be coupled to asymmetric distribution of the core proteins via the cytoskeleton.

Several studies show that the levels of the core proteins at the cell membrane must be regulated. The overexpression phenotypes of the core proteins are dose-dependent (Axelrod et al., 1998; Gubb et al., 1999; Strapps and Tomlinson, 2001; Lin and Gubb, 2009), showing that further imbalance in core protein levels produces correspondingly more severe phenotypes. Strutt et al. (2013) showed that ubiquitination of Dsh is an important mechanism for controlling core protein levels at the membrane, and in the absence of this regulation the levels of all six core proteins at the membrane increase, with an associated loss of asymmetry and polarity. Taken together, these results and models suggest that the stoichiometry of the core complex might be important for planar polarity.

1.9 The stoichiometry of the core complex

It has been shown that the molecular mechanism of the establishment of planar cell polarity has so far proved an intractable problem. This has now been distilled down to a case of distinguishing between the different feedback models that have been proposed. Understanding the stoichiometry of the core complex should thus inform a number of different approaches to this problem.

Firstly, it has never been formally demonstrated that the core proteins do actually form a complex, and indeed, it would be very difficult to do so. Pulldown experiments have demonstrated physical interactions between various pairs of core proteins, but under artificial conditions, which might induce artificial interactions. However, the core complex hypothesis is supported by the colocalisation of the core proteins and the relative stability of the core protein puncta. Additionally, several pairs of core proteins have been shown to bind to each other *in vitro* (although one should perhaps still be mindful that the proteins may behave abnormally in un-physiological assays). If the core proteins were present in the puncta in stoichiometric ratios, this would be another piece of evidence consistent with the core complex hypothesis. If no sensible stoichiometry emerged this would perhaps suggest that this idea is incorrect, or that there is not a simple relationship between the core complex and the puncta.

The stoichiometry can also help shed light on how the individual proteins interact within the complex. For example, if there are 4 molecules of X to every one molecule of its binding partner Y, there must either be 4 binding sites for X on every molecule of Y, or there must be some X present which is not bound directly to Y – perhaps some molecules of X must multimerise in the complex. In this way, stoichiometric information can guide much further experimentation and restrict the possible configurations of the core complex.

Measuring the stoichiometry of the complex through development can also provide clues to molecular interactions over time. If the proteins arrive in the puncta at different times then this might indicate some temporal control mechanism, or show that one particular protein forms the foundation of the complex. It would be surprising if the cytoplasmic proteins Dsh and Dgo reached the membrane before the transmembrane ones, as they would have nothing to tether them at the membrane unless they were bound to some other membrane-anchored protein. Pk, with its prenylation site, could be able to localise at the membrane independently.

1.9.1 Methods for measuring stoichiometry

A method must be employed that allows measurement only of the core proteins contained in core complexes or puncta, as the stoichiometry of the complex may differ from the ratios of core proteins in the cell as a whole. Dsh particularly is known to have a significant cytoplasmic fraction, due to its role in Wnt signalling.

The ideal way to measure the stoichiometry of the putative core complex would be to purify intact complexes and quantitatively detect each protein using antibodies or mass spectrometry or some related technique. However, the purification would be technically challenging, because the complex forms across cell membranes and would presumably be disrupted by cell lysis. The proteins could be covalently cross-linked before purification by fixing the tissue, but this could affect the structure of the proteins or their relative stoichiometries. Even the simplest model of the core complex, with one of each core protein plus one additional Fmi molecule (to form the dimer), is a large structure of around 1,000 kDa that would be difficult to handle *in vitro*.

Quantitative Western blots would reveal the relative amounts of the core proteins in the tissue as a whole, but for the core complex one would only want to detect the proteins in complexes, not the total cellular pools, so it would be more complicated than just making extracts of specific tissues. Calibration of the antibodies would again be an issue, and both blots and immunofluorescence rely on the availability of suitable antibodies, which is not guaranteed.

One could express the proteins individually in bacteria, make solutions of equal molarity and mix them, then either sediment and analyse any resulting complexes, or measure the concentration of any protein left in solution. Laser light scattering has been successfully used to determine the stoichiometry of a complex in solution (Mogridge, 2004). However, this would require all six core proteins to be soluble, and the membrane-bound ones may not adopt the required shapes for complex formation when purified *in vitro*. This method would also rely on the core proteins spontaneously forming complexes without needing any energy input, or modifications such as phosphorylation or acetylation or any post-translational processing. Such modifications may not reliably occur when expressed in prokaryotic cells, and it has already been shown that several of the core proteins have important post-translational modifications applied.

Quantitative mass spectrometry has been used to measure the stoichiometry of the U1 small nuclear ribonucleoprotein (snRNP), a key component of the spliceosome. Hochleitner et al. (2004) purified and digested the U1 snSNP complex and synthesised heavier matching peptides using heavier isotopes. By adding specific known quantities of the heavier peptides to the digested complex they were able to use the ratio of the heavy and light peptides to determine the original amounts of the lighter, endogenous peptides in the complex. These authors also added a fluorescent tag to the digested peptides and used this to quantify the protein amounts. They note that complete digestion and denaturation of the protein constituents are required and that it is difficult to resolve large numbers of different peptides. A similar method, using GFP-tagged extracts from HeLa cells, has also recently been published (Smits et al., 2013). The largest component of the U1 snRNP is 70kDa, a quarter the size of Fmi. Complete digestion of Fmi would therefore be likely to yield too many different peptide components to be resolved by mass spectrometry.

The stoichiometry could also be determined by immunofluorescence. However, any immunofluorescence-based method would be complicated by the probable non-linear relationship between the number of core proteins present and the fluorescence observed. Waters (2009) discusses several limitations of immunofluorescence for quantification. Antibodies have been shown to bind more readily to more concentrated regions of epitopes (Mason and Williams, 1980), and any such effect would be amplified by the secondary antibody. The process of fixation could also affect the numbers and distributions of the core proteins (Melan and Sluder, 1992). Each antibody would also have a differing affinity for its target peptide, and would require extensive calibrations and controls. This would probably be possible, by making calibration curves for each antibody using known concentrations of each protein, but expression and purification of each core protein would present another technical challenge.

The only option that remains is therefore *in vivo* methods. Attaching a fluorophore to each core protein would allow visualisation of the tagged protein by confocal microscopy. There would be a simple 1:1 relationship between the number of core protein molecules and the number of fluorescent molecules. Studying the complexes in live tissue should ensure that they are behaving normally. The tissue does not need to be interfered with and can be observed *in situ*, which is preferable to fixation and immunostaining. The main concern from the experimental point of view would then be whether the addition of the fluorophore to each core protein was affecting its behaviour in any way, which can be easily tested by looking at the phenotypes of flies expressing the tagged proteins.

1.9.2 Previous *in vivo* stoichiometry work in yeast and *Drosophila*

The proposed method of GFP-based protein measurement has been successfully used by several groups using *S. cerevisiae* and *S. pombe*. Chiu et al. (2001) used His-tagged GFP coupled to beads to demonstrate that the relationship between the number of GFP molecules present and the measured fluorescence was a straight line over a wide range of concentrations. The GFP levels could be measured with an accuracy of approximately 20%, which might reflect differences in sample preparation between the GFP beads and the neural cells. The same group used the GFP beads to calibrate measurement of the numbers of GFP-tagged GAT1 molecules in various mouse neural subcellular compartments, detecting a range from 640 to 7.8 million GAT1 molecules (Chiu et al., 2002).

Dundr et al. (2002) highlighted the difficulty of calibrating fluorescence measurements *in vivo* for quantitation of protein levels. Any standard used must be imaged under the same conditions as the experimental sample, and this can be difficult for living specimens. To solve this problem, these authors tagged one of the coat proteins of a modified rotavirus and introduced the assembled GFP viruses to their samples. Due to their self-assembling nature, virus particles usually contain the same number of each particular molecule in every particle. This meant that the virus particles acted as an internal fluorescence standard for a known number of GFP molecules.

Although the viruses in that study were not taken up into the experimental cells, it should be possible to induce uptake so that the fluorescence of the viruses can be compared to the experimental protein of interest in exactly the same conditions. While this is a very satisfying method, it would be difficult to apply to the *Drosophila* wing because the developing wing sits within a fluid-filled membrane. It would not be possible to deliver the virus particles to the wing without damaging the membrane, and if the viral genes were added to the fly genome there is no guarantee that they would express and assemble into viral particles properly.

A notable successful measurement of protein levels in *S. pombe* by GFP fluorescence used quantitative immunoblotting to calibrate the results (Wu and Pollard, 2005). YFP was fused to proteins involved in cytokinesis, and the measured fluorescence was directly proportional to the protein levels detected by blotting. For low-abundance proteins the error bars were proportionately quite large, but the method was robust and the procedure straightforward. Joglekar et al. (2006) applied a very similar method to the kinetochore proteins of *S. cerevisiae*, and Geng et al. (2006) applied Wu and Pollard's methods to the autophagy proteins of *S. pombe*, demonstrating that quantitation of fluorescence was applicable to many different complexes.

McGill et al. (2009) are the only group so far to apply this method to Drosophila. They used GFP-tagged E-cadherin, Armadillo (β-catenin) and Bazooka (α-catenin) to measure the levels of these proteins at spot adherens junctions in the cellularising fly embryo. Intensity measurements from confocal images were combined with whole-embryo protein counts by ELISA to work out the absolute and relative levels of these three proteins.

1.10 Drosophila as a model organism

The fundamentals of our understanding of planar polarity resulted from work in the fruit fly (or vinegar fly) *Drosophila melanogaster*. *Drosophila* has a plethora of advantages as a model organism, many of which make it particularly suitable for this project. While vertebrates are (quite rightly) protected by legal restrictions on the kinds of experiments and procedures that can be done, no such legislation is extended to insects. In terms of time, cost and space its efficiency is unmatched by any metazoan apart from the nemaotode worm *Caenorhabditis elegans*. It is easy to culture, requiring just a small vial with some fly media in the bottom to complete its whole life cycle. Development from embryo to imago takes 10 days at 25°C, and this short generation time can be manipulated simply by changing the incubation temperature of the culture. Development from embryo to larva takes just 24 hours, prompting a wise professor to name *Drosophila* 'the Ferrari of development'.

In terms of genetics, *Drosophila* is extremely easy to work with. Males and females are sexually dimorphic. Freshly-eclosed females are both phenotypically different from mature females and reluctant to mate, while males take a day or two to reach full fertility. This means that virgin females can be easily collected, which greatly facilitates crossing schemes. Once a female has mated she is fertile until old age, and each female can produce around 200 offspring. This means that very large populations can be produced in a very short time, which makes it possible to screen for rare genetic events such as recombination between closely-linked loci, or a specific combination of several marker genes. The flies have numerous physical traits that are tolerant of mutations, so there is a wide range of dominant and recessive markers for every chromosome. This means that individual chromosomes, or even individual loci, can be followed through the generations, allowing stocks with complex combinations of mutations and transgenes to be easily made. Rearranged chromosomes, called balancer chromosomes, can be used to prevent recombination and preserve such stable stocks.

The *Drosophila* genome is also very compact, with only four chromosomes. The fourth chromosome is very small and often ignored (unless it contains one's gene of interest), so there are only two autosomes and one pair of sex chromosomes to track. Whereas vertebrates often have multiple copies of genes flies usually have just the one, so redundancy between related genes rarely affects experiments. The sequencing of the

Drosophila genome allowed unprecedented control of genes and gene regulation, and with modern techniques the genome can be manipulated with exquisite precision.

Transgenesis in *Drosophila* is also a simple matter of injecting DNA directly into embryos with an active transposon system and screening the survivors for transformants. There are several companies that specialise in *Drosophila* microinjection (BestGene Inc. and GenetiVision were both used for this project), so getting transgenic flies is often as easy as posting some DNA to America and waiting a few months to receive the resulting fly lines in the post. The *Drosophila* community is generally very generous with fly stocks and related reagents, and almost any mutant imaginable can usually be obtained from one of the many stock libraries around the world. Addition of genes and DNA sequences from bacteriophage, bacteria and yeasts allows fine control of many aspects of gene expression. The site at which an inserted transgene integrates can be dictated by using the ΦC31 machinery from phage, mitotic clones can be generated at will using the *Saccharomyces cerevisiae* Flp/FRT system (Golic and Linquist, 1989) and gene expression can be induced at specific times in specific tissues using the yeast UAS-Gal4 system (Brand and Perrimon, 1993).

In terms of planar polarity research, the fruit fly is an ideal model. Compared to vertebrates it has very few epithelial-derived internal organs. Additionally, cilia are restricted to specific non-essential sensory organs. These properties make the fly very tolerant of disruptions to the planar polarity pathway. Mutations that would kill a vertebrate before gastrulation can exist as healthy and fertile adult flies. The outer surface of the fly is also covered almost entirely in small structures that reveal the underlying polarity of the tissue. The notum, abdomen, legs and wings are all festooned with planar-polarised bristles and trichomes, while the rotation and packing of the ommatidia of the eye is also controlled by planar polarity. As described above, the genetic tractability of *Drosophila* has also contributed greatly to our current understanding of planar polarity.

This project looks almost exclusively at the *Drosophila* wing at various stages of development. The wing begins in the larva as an imaginal disc. During metamorphosis the disc everts and elongates to form the adult structure. The wing is physically accessible throughout development, and after pupation it can be studied *in vivo*, without destroying the animal. While much of the work on planar polarity has been done in the eye and abdomen, these are complex structures that might not give as clear a picture of polarity regulation as a simpler tissue. The adult wing is about as uncomplicated as it gets, consisting of a flat bilayer of regular hexagonal cells interrupted only by the five longitudinal veins and the two perpendicular crossveins. To study the *Drosophila* wing is therefore to look at the simplest tissue in the most streamlined model organism. If the mechanisms of planar polarity cannot be elucidated here then there may not be much hope for other tissues and organisms.

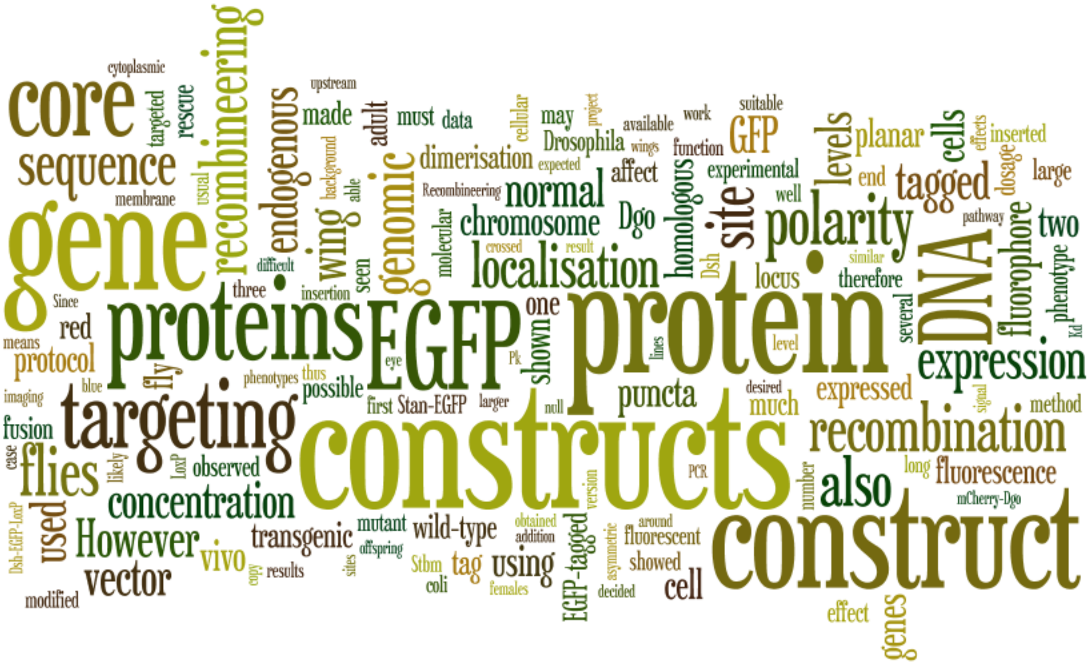
1.13 Scope of project

While the *in vivo* GFP-measurement method described above should also work in flies, applying it to *Drosophila* may not be entirely straightforward. Yeasts are single-celled organisms, which means they are easy to apply to a slide for imaging and there is nothing but suspension media and cell membrane between the slide and the fluorophore. McGill et al. (2009) used fly embryos, which are very small and have little intervening material between the experimental tissue and the coverslip. Fly pupae are much larger, and the wing tissue is enclosed in a membrane, which is itself inside a pupa, so the experimental material is not quite as accessible to the laser. For this reason, it was decided not to attempt to quantify the actual numbers of molecules in puncta, but to establish the relative stoichiometry of the core proteins.

The main aim of this project was therefore to generate fluorescently-tagged core proteins, measure the fluorescence, and work out the relative stoichiometry of the core proteins in puncta. Great care will have to be taken to design fluorescently-tagged core proteins so that they are expressed at the normal level and are able to function normally, otherwise the stoichiometry observed may not reflect the true ratios of the endogenous core proteins. This is explored in more detail in Chapter 2. A method of detecting and measuring the brightness of the tagged proteins had to be developed; this is discussed in Chapter 3 along with the results of the stoichiometry experiment.

Chapter 2

Generating a suite of endogenously-expressed EGFP-tagged core planar polarity proteins in *Drosophila*



Contents

Figures

Tables

2.1 Introduction

The main aim of this project was to determine the relative stoichiometry of the core planar polarity proteins in the puncta *in vivo*. It was proposed to achieve this objective by measuring the brightness of a fluorophore that was genetically fused to each core protein, and then using the measured brightness as a proxy for the amount of protein present. If this strategy was to work then two main requirements had to be met. Firstly, exactly the same fluorophore must be used for every construct; otherwise the fluorescence measured would not be directly comparable between constructs. Secondly, the modified gene must be expressed at endogenous levels. The tagged protein must then behave normally, incorporating into complexes and puncta in the normal proportion. Few of the existing transgenic constructs in the lab fulfilled both these criteria, necessitating the creation of a matched suite of fluorescently-tagged core protein constructs. This chapter describes the generation and testing of these constructs.

2.2 Choosing a fluorophore

When making the constructs it was essential to use the brightest and most reliable fluorophore available. All images would have to be taken at the same microscope settings so that data from different images and genotypes could be directly compared. It was anticipated that images would be taken of the constructs in both the homozygous and heterozygous state, so the fluorophore had to be bright enough to easily visualise one copy of the gene for the least abundant core protein at settings which would not over-saturate images of two copies of the most abundant.

The original GFP was discovered in, and isolated from, the jellyfish *Aequoria victoria*, and it is no exaggeration to say that its discovery and application to genetics completely revolutionised biological research. Osamu Shimomura, Martin Chalfie and Roger Tsien were awarded the 2008 Nobel Prize for Chemistry for their work on GFP. Cormack et al. (1996) generated three GFP mutants with improved solubility, faster folding and up to 35 times brighter fluorescence than the wild-type molecule. Zhang et al. (1996) took the brightest of Cormack et al.'s mutants and re-coded the DNA sequence; replacing the original *A. victoria* codons with preferred codons from selected highly-expressed human proteins, resulting in the EGFP gene. The ease of expression and improved brightness, solubility and folding kinetics of this mutated protein made it ideal for this project.

Constructs using red fluorescent proteins had not previously worked in our hands. Although fluorophores that emit in the orange, yellow or blue regions of the spectrum are also available (both GFP-derived and independent), none of these except for EYFP (enhanced yellow fluorescent protein) had been tested in our lab. The observed brightness of the fluorophore is also dependent on the efficiency of excitation, which relies upon having the appropriate laser lines available. EGFP is strongly excited by the 488nm line of an Argon laser, is highly photostable and behaves well when fused to other proteins (Shaner et al., 2007). Taking all of the above into consideration, EGFP was chosen as the most appropriate fluorophore for this project. The requirement of using the same fluorophore for every construct was thus easily met by using enhanced green fluorescent protein (EGFP).

2.3 Transgenesis and protein expression in *Drosophila*

The second requirement, that of endogenous expression levels, is much more difficult to achieve in practice than standardisation of the fluorophore. Most of the existing tagged proteins in our lab, and in the wider *Drosophila* community, use the gene's cDNA sequence expressed under the control of a heterologous promoter, as this is by far the most convenient way to do transgenesis in flies. Commonly used promoters are those of the *Actin5C*, *armadillo* (β-catenin) and several ubiquitin genes, as they are generally ubiquitously expressed both physically and temporally. This usually results in reliable expression levels, which is ideal for visualising a relatively dim fluorophore fused to a relatively low-abundance protein when the actual levels of the fluorophore-tagged protein are not important. The significant drawback is that this can put the cell into an abnormal physiological state and sometimes cause abnormal phenotypes, and then It can be unclear whether a result is due to the tagged protein itself, or just a by-product of its abnormal expression profile.

These constructs have traditionally been inserted into the fly genome at a random locus via P element-mediated transposition (Rubin and Spradling, 1982, reviewed in Castro and Carareto, 2004 and Venken and Bellen, 2007), which can result in unpredictable expression effects caused by upstream enhancer elements or chromatin structure. As well as the usual promoter and enhancer elements, expression levels in flies seem to be strongly affected by chromatin packaging and organisation (reviewed in Farkas et al., 2000). The insertion site of a transgene can therefore have a significant effect on its expression profile, so it is desirable to insert a construct either into the endogenous locus or to have a set of constructs all in the same genomic locus. Using the same insertion site for every construct would control for any position effects, but care should also be taken to choose a site that does not suffer from any obvious position effects in the first place. Gene dosage will additionally have a strong effect on protein levels, so it is important to be able to have each tagged gene present at the endogenous dosage.

Other means of controlling cellular protein levels are certainly active in *Drosophila*, such as the endocytic degradation pathway and mRNA processing machinery. However, it is both difficult to predict and unfeasible to control these aspects of protein expression. The best that can be done from this perspective is to try and keep the gene structures and the resultant proteins as close as possible to their endogenous counterparts, and hope that the transgenic mRNA and fusion proteins are not singled out for special treatment by the relevant cellular processes. Since 5' and 3' UTRs (untranslated regions) can affect transcript stability and processing, these were preserved in their endogenous states as far as possible. Protein levels can also be regulated at the post-translational level, but this is not a concern as long as the constructs express normally and the fusion proteins are not affected by the addition of EGFP.

It may be the case that expression levels do not affect the levels of the core planar polarity proteins in the puncta at all. The amount of protein in the membrane or puncta may be controlled separately to the overall cellular pool, by either an active or a passive mechanism. In this scenario, as long as the expression levels fall between an upper and lower threshold, the actual amount of protein would be unimportant.

As mentioned in the introduction, it is likely that core protein levels are interconnected and co-dependent. This emphasises how desirable it would be to avoid altered expression levels in the EGFP-tagged fusion proteins. In any case, if the level of one of the EGFP-tagged core proteins is increased relative to wild-type, it will give an inaccurate value when compared to a tagged core protein expressed at the normal level. This will distort the measured stoichiometry of the core complex.

To generate clear, meaningful data from this experiment it was therefore essential that every tagged protein was expressed at its endogenous level, to avoid any possible disturbance to the core PCP complex due to misexpression. This meant preserving the genomic context of each gene as much as possible, paying particular attention to the region immediately upstream of the coding sequence, as this is where the promoter should be. Two relatively new transgenesis techniques were employed: 'recombineering' and *in vivo* gene targeting. Both methods utilise homologous recombination in place of traditional 'cut and paste' molecular cloning, and allow the upstream regulatory sequences to be preserved in the final construct.

diego (*dgo*), *dishevelled* (*dsh*) and *strabismus* (*stbm*) are smaller genes, which have transcription regions spanning 3.8, 2.8 and 3.6 kilobases respectively (FlyBase, version FB2013_04), were suitable candidates for recombineering. For each gene, its sequence, plus a good chunk of the upstream genomic DNA, would easily fit into the CHORI-322 P[acman] library, whose members have insertions of ~20kb (Venken et al., 2009).

The other core protein genes *frizzled* and *prickle* have very long introns and both span more than 70kb, making them too long even for the larger ~80kb inserts of the CHORI-321 P[acman] library – even if the gene itself fitted into the construct, there would not be room for the upstream regulatory elements, and regardless, DNA constructs of this size can be very difficult to work with. The constructs for *frizzled* and *prickle* were therefore made by *in vivo* gene targeting, described below. The other core protein gene, *starry night*, is about 50kb in length. The same construct for this gene was made using both recombineering and *in vivo* gene targeting; both methods are described below in section 2.6.

2.4 Recombineering of *diego*, *dishevelled* and *strabismus*

2.4.1 Overview of the recombineering protocol

Recombineering is a technique that uses homologous recombination to modify DNA in *Escherischia coli* hosts, instead of using restriction enzymes *in vitro*. Recombination-based methods are often preferable to cloning-based methods for a number of reasons, not least the absence of numerous digestion, purification and ligation steps. Since restriction enzymes are not used there is no need to find or add compatible sites to the DNA. As long as the sequence data for designing the homology arms (which facilitate the recombination) is available, there is no requirement for the gene of interest to be cloned.

However, homologous recombination was initially only widely used in yeast, due to its efficiency of transformation and recombination with linear DNA fragments (Szostak et al., 1983; Copeland et al., 2001). *E. coli* cells are inherently less amenable to uptake of linear DNA, as their *RecBCD* exonuclease tends to rapidly degrade it. Recombination in *E. coli* was thus largely restricted to sickly exonuclease-deficient strains, and was mainly used by groups who worked on bacterial biology. Most of the common *E. coli* strains used for generating transformation constructs have their endogenous recombinases disabled to protect the plasmid, which is the desired end product, so cannot be used for recombinogenic engineering anyway. Murphy (1998) solved this problem by using the recombination machinery of bacteriophage λ to achieve homologous recombination-mediated gene replacement in various strains of *E. coli*. The λ recombination genes *exo*, *bet* and *gam* were introduced into the *E. coli* chromosome, and used to recombine a PCR product containing an antibiotic resistance gene into the *LacZ* gene.

The Court lab expanded on this work to develop the recombineering protocol (thoroughly reviewed in Copeland et al., 2001). They added a temperature-sensitive repressor to the λ-carrying *E. coli* to suppress the phage λ genes, which allows induction of the recombination genes only when needed (Yu et al., 2000). The *gam* gene inhibits the bacterial exonuclease, protecting the transformed linear DNA while it recombines with its homologous target sequence. The same paper also tested how long the homology arms, which target the DNA fragment to the correct genomic locus, needed to be, and found that maximal recombination occurred using arms of 30-50bp. Yu et al. additionally showed that this method could be used to modify plasmids as well as the bacterial chromosome, and thus recombineering was born.

Recombineering is particularly suited to this project, as it facilitates modification of large (>20kb) DNA constructs. Handling such constructs *in vitro* can be difficult, as they are easily damaged and degraded by the usual molecular biology protocols, and the larger they get the lower the transformation efficiency becomes, into both bacteria and

Drosophila. Doing all the modification of the vector in the bacterial host means that the vector only needs to be transformed into the cells once, and thereafter the cells' own DNA replication and repair machinery maintains the vector (with the help of the usual antibiotic resistance gene in the vector, and the addition of the corresponding antibiotic to the culture media, to prevent loss of the vector). This means that large constructs of 20kb and above can be quickly and easily manipulated with exceptional precision, and without the need to find or add restriction sites to the DNA sequence. The recombineering protocol is summarised graphically in **Figure 2.1**, and in the text below.

The general procedure is to obtain a large genomic construct that contains the gene of interest, and to identify the DNA sequence around the locus to be modified (**Figure 2.1 A**). The base vector used for these constructs is P[acman], which contains P element ends, an *attP* site for site-specific integration into the fly genome, and an inducible high copy number replication origin (Venken et al., 2006; Venken and Bellen, 2007). The development of the P[acman] vector was precipitated by the restricted capacity of the P element, which can only accommodate inserts of up to 40kb, with an associated decrease in transgenesis efficiency. High copy P element-based vectors become unstable in bacteria when larger than 25kb (Venken and Bellen, 2007). The low copy number of P[acman] allows it to stably maintain a larger insert than a normal P element vector. Venken et al. added an inducible replication origin that allows controlled induction to high copy number for ease of plasmid recovery. P[acman] also contains an *attP* site, which allows ϕC31-mediated site-specific integration into the fly genome (see below).

The required P[acman] construct can be made by retrieving the desired genomic DNA from a BAC into an empty P[acman] vector, or ready-made P[acman] clones can be ordered from a library (e.g. Hoskins et al., 2000). Long DNA primers (around 50bp, primers A and B in **Figure 2.1 A**) are designed, which have homology to both the identified genomic locus and to the DNA which is to be added to the genomic construct, thus bridging the junction between the genomic and the exogenous DNA. For this project, the EGFP source vector was PL452-N-EGFP (**Figure 2.1 B**), for tagging the N-terminus of a protein, and PL452-C-EGFP, for tagging the C-terminus. The only difference between these two vectors is the location of the kanamycin selection cassette relative to the EGFP gene.

Primers A and B are used to generate, via PCR, the ‘targeting DNA’. This consists of the sequence to be added to the genomic construct, plus a floxed (‘LoxP-flanked’) antibiotic resistance gene, all flanked by the homologous sequence from the genomic construct that was included in the long primers (**Figure 2.1 C**). Particular care is needed when designing these primers, as the reading frame must be preserved across the junction between the gene and the tag sequence after removal of the kanamycin selection cassette. It is therefore sometimes necessary to add one or two base pairs in the junction to maintain the reading frame. The final sequence must also be checked for accidental introduction of stop codons. The single remaining LoxP site in the final construct serves as a short additional linker sequence between the core protein gene and EGFP.

In most commercial *E. coli* strains the endogenous recombinases are inactivated to protect the transformed plasmid, but the strains developed for use in recombineering protocols carry the genes for the bacteriophage λ recombinase machinery under heat-shock control, allowing recombination to be induced at the correct time. The P[acman] construct is transformed into one of these recombineering strains, the transformed cells are cultured and heat shocked, then the targeting DNA is electroporated into the cells (**Figure 2.1 D**). This results in the recombination of the targeting DNA with its homologous sequence in the P[acman] construct (**Figure 2.1 E**). The transformed bacteria are grown on agar plates, and colonies carrying the correctly modified P[acman] construct are identified by their acquisition of kanamycin resistance from the targeting DNA. These colonies are isolated and the antibiotic resistance gene is removed by inducing *Cre* expression in the recombineering cells (**Figure 2.1 F**). The modified construct is then recovered by miniprep.

Under ideal conditions, with all reagents already in hand, the protocol to this point takes just one to two weeks. The modified section is amplified from the purified P[acman] and sequenced to check for any mutations or rearrangements before being microinjected into *Drosophila* eggs. The construct can also be restriction digested before injection to check for any major structural rearrangements.

As mentioned above, the ϕC31-mediated transformation system allows insertion of transgenic constructs into specific genomic loci (Groth et al., 2004). This is another bacteriophage mechanism that has been exploited for transgenesis applications. The ϕC31 integrase catalyses recombination of an *attB* site in the vector with an *attP* site in the target locus. This process destroys the *attP* and *attB* sites, converting them into an

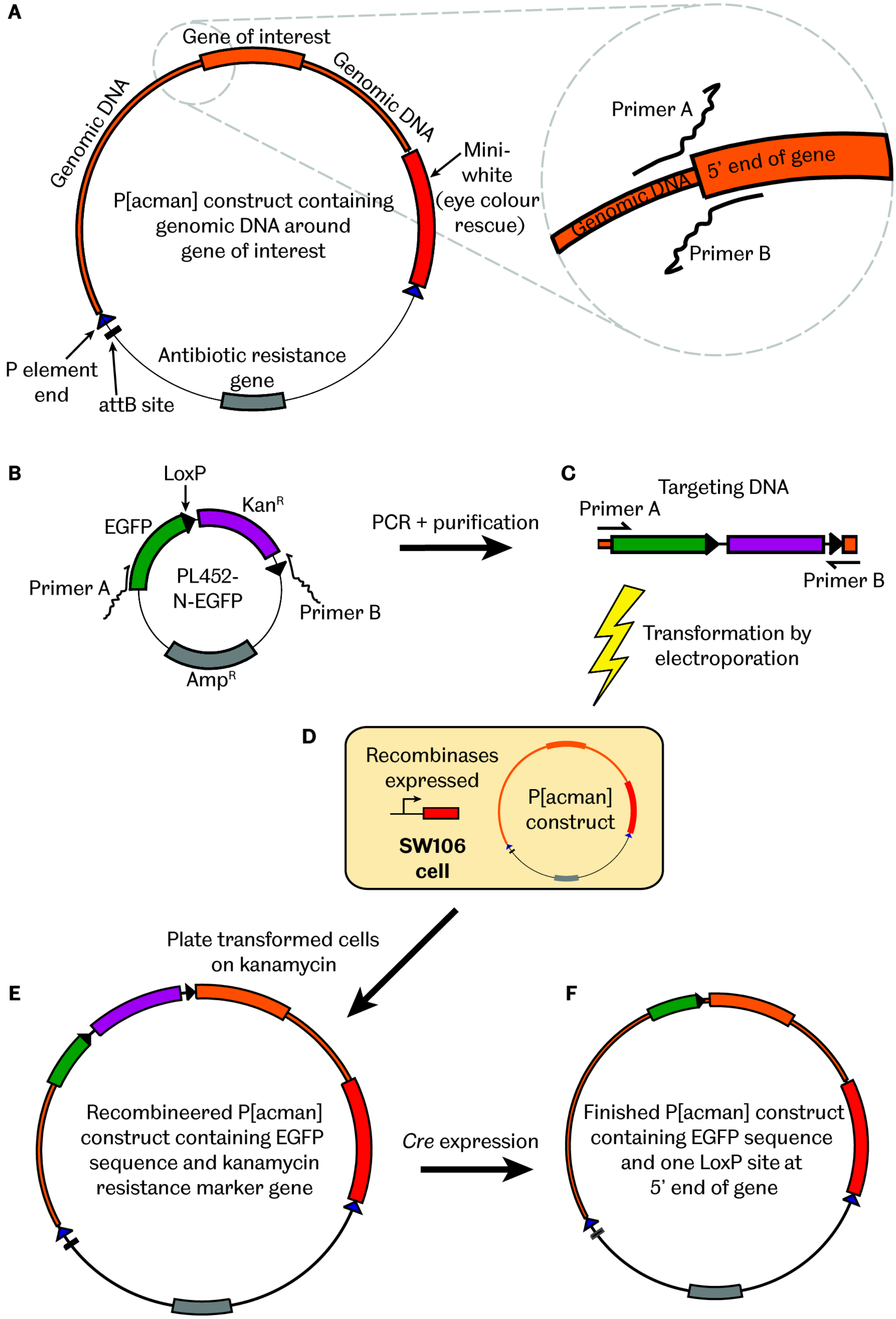


Figure 2.1: Overview of the recombineering protocol.
A: The P[acman] vector contains the gene of interest and the surrounding genomic DNA. The 5' ends of primers A and B are homologous to the sequence flanking the intended transgene insertion site. **B:** The 3' ends of primers A and B anneal to the EGFP source vector pL452-EGFP, which contains a floxed kanamycin selection cassette. **C:** PCR from pL452 using primers A and B adds the homology arms to the EGFP cassette. **D:** The targeting DNA is transformed into SW106 cells and recombines with the P[acman] vector. **E:** The recombineered P[acman] is selected for using kanamycin media. **F:** The kanamycin cassette is removed by inducible expression of *Cre*.

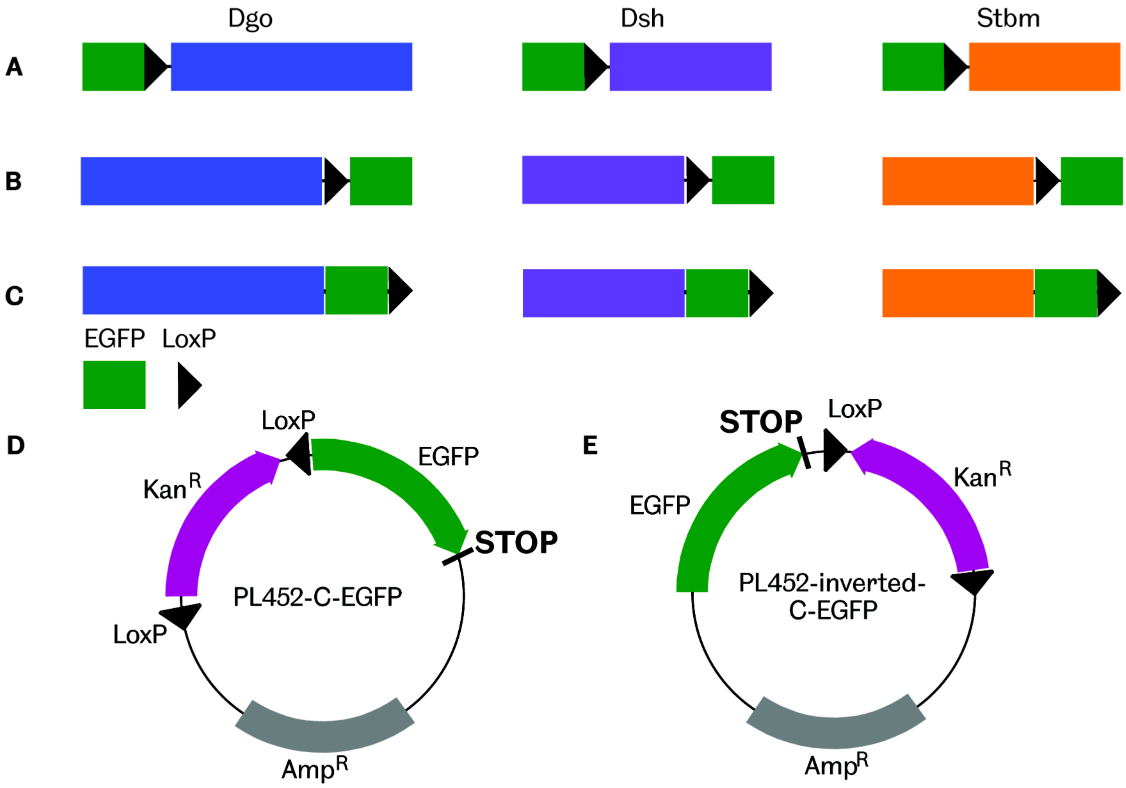


Figure 2.2: The recombineered constructs
Genes and EGFP (coloured bars) in A-C are shown to scale, LoxP (black triangle) is not. **A:** The N-tagged constructs. **B:** The C-tagged constructs derived from PL452-C-EGFP. **C:** The C-tagged constructs with EGFP and LoxP reversed, derived from PL452-inverted C-EGFP. **D:** The original PL452-C-EGFP vector, with EGFP downstream of the kanamycin cassette. **E:** The plasmid in D with EGFP upstream of the kanamycin cassette. EGFP Stop codon is present in C-tagged constructs.

attL and an *attR* site, thus preventing excision of the insert by the same mechanism (Thorpe and Smith, 1998). The majority of the recombineered constructs used in this project were inserted into the *attP40* site on 2L, as this site was found to be one of the most reliable (Markstein et al., 2008). 2L is also a convenient chromosome arm, because *pk*, *stbm*, *fmi* and *dgo* are all on 2R. Inserting all constructs into the same genomic locus controls for any effect of local chromatin structure on the expression of the transgene.

A full list of the constructs and their insertion sites can be found in the Materials and Methods, section 4.3.1 on page 132. The constructs are also listed in **Table 2.5** on page 75 with descriptions of their expression and phenotype. Adding the modified DNA construct to the fly introduces an extra copy of the gene of interest, so when transgenic flies are obtained the functional dosage of the modified gene must be restored to normal. This is achieved by crossing the transgenic chromosome into the appropriate mutant background.

2.4.2 Construct design: *diego*, *dishevelled* and *strabismus*

Schematics of the constructs are shown in **Figure 2.2 A**-**C**. Since the molecular biology part of the recombineering protocol is a very quick process it was decided to make two constructs each for *dgo*, *dsh* and *stbm*, one with the EGFP tag at the N-terminus of the target protein (**Figure 2.2 A**), and another with the tag at the C-terminus (**Figure 2.2 B**). The PL452-C-EGFP vector, which was the template for the targeting DNA PCR, is shown in **Figure 2.2 D**. However, this vector gave poor results (see below, section 2.7). The PL452-C-EGFP vector was re-engineered to place the kanamycin cassette, and hence the final LoxP site, downstream of the EGFP sequence. This vector was designated PL452-inverted C-EGFP (**Figure 2.2 E**). This vector resulted in the C-tagged constructs for Stbm, Dgo and Dsh shown in **Figure 2.2 C**, which did show GFP fluorescence. Note that tagging the C-terminus of Stbm will block the PBM. Tagging the N-terminus of Stbm should not affect its membrane localisation – In most multipass proteins the first transmembrane domain acts as a membrane localisation signal.

The tagged constructs were crossed or recombined into the appropriate null mutant background to restore normal gene dosage. All P[acman] constructs were inserted into *attP40* on the second chromosome. EGFP-Stbm was initially inserted into *VK23* on the third chromosome, but was subsequently re-injected to be in the same site as the others. Stbm constructs were recombined or double-balanced with $stbm^{6}$ on II, Dgo constructs with dgo^{380} on II, and Dsh constructs were double-balanced with dsh^{V26} on the X chromosome.

2.5 *In vivo* gene targeting of *frizzled* and *prickle*

2.5.1 Overview of the gene targeting protocol

Like recombineering, gene targeting also exploits homologous recombination, but with this technique the recombination takes place in the fly germline instead of in *E. coli*. Genetic engineering by homologous recombination was developed in vertebrates long before it became possible in *Drosophila*. This was largely due to uncertainties about the mechanisms of DNA repair in flies and difficulty in establishing parameters for efficient targeting constructs (discussed in Rong and Golic, 2001) coupled with unsuccessful attempts to achieve targeting in males (Bellaiche et al., 1999). The ‘ends in’ method, which duplicates the targeted locus, was first successfully reported in *Drosophila* by

Rong and Golic (2000). In 2003 Gong and Golic demonstrated successful ‘ends out’ targeting, and innovations by Huang et al. (2008) greatly increased the power and efficiency of screening for recombinant flies. Our targeting protocol is largely based on that published in Huang et al. (2008), and uses their fly lines and reagents.

In terms of the end product, ends-out *in vivo* targeting is the ideal transgenesis method for endogenous expression, although the protocol is slower and less reliable than recombineering. As with recombineering, this technique enables very precise manipulation of the genome, but when targeting it is the endogenous locus that is modified. This means that the surrounding genomic DNA is not duplicated, and there is no need to restore normal dosage of the tagged gene. Crucially, the modified gene exists in its normal genomic context, so identification and control of promoters and enhancers becomes irrelevant and concerns about expression levels are erased. The only remaining worry is any effect that the tag might have on the behaviour or dynamics of the resulting fusion protein. The only disadvantage is that it takes much more time and effort to obtain the finished transgenic fly stocks, as the process involves several generations of crosses and a screening step. If several different versions of a construct are desired then recombineering is a more suitable method.

The targeting itself is achieved by first generating the targeting cassette *in vitro*. This means adding the DNA sequence of the desired modification, in this case EGFP, to the pRK2 vector, as shown in **Figure 2.3 A**. Large homology arms (around 3kb each) must also be ligated into this vector using traditional molecular cloning techniques (**Figure 2.3 B**). The targeting cassette for *in vivo* recombination is much larger than that used for recombineering and much longer homology arms are needed to facilitate the recombination event. This is probably because there is a much larger amount of non-homologous DNA in the fly genome than in a bacterial cell, so the targeting cassette may be less likely to find its homologous locus, and also there is no specific induction of recombinases. As with recombineering, the appropriate genomic DNA can be obtained from existing *Drosophila* genomic libraries (e.g. Hoskins et al., 2000).

Once the targeting vector has been made, it is microinjected into *Drosophila* eggs to create a ‘donor’ fly strain. The DNA inserts via P-element mediated recombination at a random genomic locus – everything shown between the P element ends in **Figure 2.3 B** is inserted. The chromosome carrying the inserted P element, shown in **Figure 2.5 A**, is hereafter termed the ‘donor’ chromosome, and flies carrying this chromosome are donor flies.

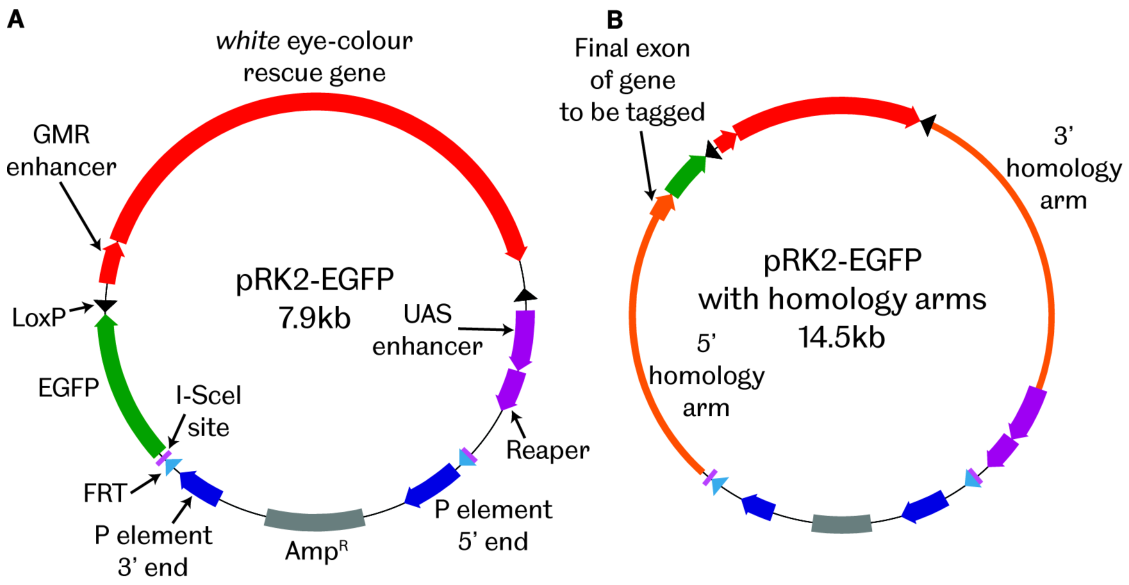


Figure 2.3: Using pRK2-EGFP to generate the donor cassette for in vivo gene targeting
A: The pRK2-EGFP vector was made by simply ligating the EGFP sequence into pRK2. When injected into flies, the DNA between the P element ends (dark blue) integrates at a random genomic locus. The *FRT* (light blue) and *I-SceI* sites (pink) are for excision of the targeting cassette. *GMR* (red) is a strong enhancer active in the eye, which drives the *white* gene (red) to give a strong red eye colour in flies carrying this DNA. The *LoxP* sites (black) are used for subsequent removal of the *GMR>white* cassette. *UAS>Reaper* (purple) facilitates the screening step by reducing the number of false positives. **B:** Genomic homology arms (orange) must be ligated into pRK2-EGFP to target the EGFP cassette to the corresponding genomic locus. The example shown is tagging the C-terminus. The reading frame must be carefully preserved across the gene:EGFP junction.

Virgin donor females are used to start the simplified crossing scheme shown graphically in **Figure 2.4 A**-**C** and **D**. The following protocol is described in more detail in the Materials and Methods and Huang et al. (2008). The corresponding molecular events, which take place within the flies, are shown in **Figure 2.5 A**-**F**.

The first cross (**Figure 2.4 A**) generates offspring in which *hs-FLP* and *hs-I-SceI* are expressed. The action of FLP (**Figure 2.5 B**) and I-SceI (**Figure 2.5 C**) excises and linearises the targeting DNA, resulting in offspring with mosaic eyes (**Figure 2.4 B**). Linearisation is essential, because the ends of the fragment are recognised by the fly cells as double-strand breaks and this is what stimulates recombination with the homologous genomic sequence (Rong and Golic, 2000). The *hs-hid* gene on the Y and CyO chromosomes is also activated by the heat shock. Expression of this gene causes the death of all male offspring and all offspring that did not inherit the *FLP, I-SceI* chromosome, greatly facilitating the collection of large numbers of mosaic virgin females. The desired homologous recombination event happens very infrequently, so

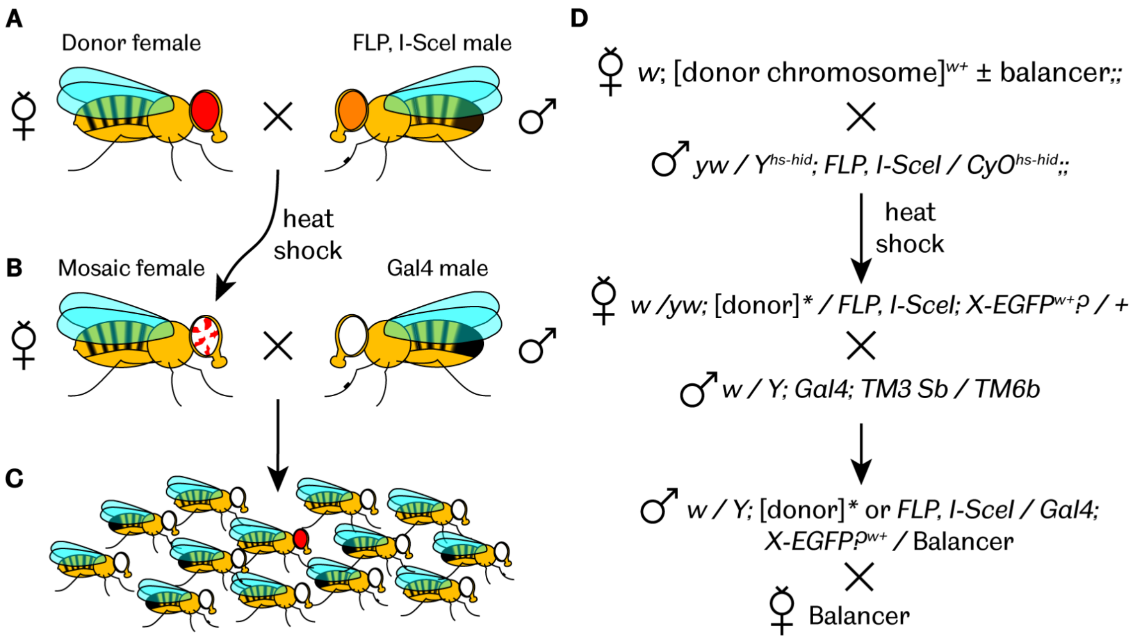


Figure 2.4: The in vivo gene targeting crossing scheme
A: Virgin females carrying the donor chromosome are crossed to males carying FLP and I-SceI under heat shock control. The resulting larvae are heat shocked at 38°C for two hours every day until pupation. **B:** The heat-shocked daughters of the cross express FLP and I-SceI, enzymes which excise and linearise the targeting DNA, resulting in a mosaic eye phenotype. Potential targeting events are generated in their germlines. These virgins are crossed to males which express Gal4. **C:** The offspring of this cross are screened for hits, which will have bright red eyes. Any flies which retain UAS-Rpr from the donor chromosome will be killed by expression of Gal4. Hits are crossed to balancer virgin females, to check that the red eye gene maps to the expected chromosome.

the second cross is needed to generate many thousands of offspring, which can then be screened for the transgenic individuals (**Figure 2.4 B**).

The targeting DNA **(Figure 2.5 D**) contains an eye colour rescue gene under a strong promoter, and the whole protocol is conducted in a white-eyed background, allowing potential targeting events to be identified by their bright red eyes (**Figure 2.4 C**). The vast majority of the offspring will not be hits, and have white eyes – their targeting DNA has excised but not reintegrated, so they have lost their eye colour. Any flies that have retained the targeting DNA, without homologous recombination taking place, will consequently also retain *UAS-Reaper*. Such individuals are killed by expression of Gal4 (**Figure 2.5 E**). Flies that integrate the targeting DNA by homologous recombination will lose UAS-Reaper, and survive expression of Gal4.

Any red-eyed hits are thus either the desired targeting event or false positives. These flies are collected and crossed individually to balancer flies to establish stable stocks. During this process it becomes clear whether the red eye colour segregates on the targeted chromosome; only these lines are kept. The presence of the EGFP DNA

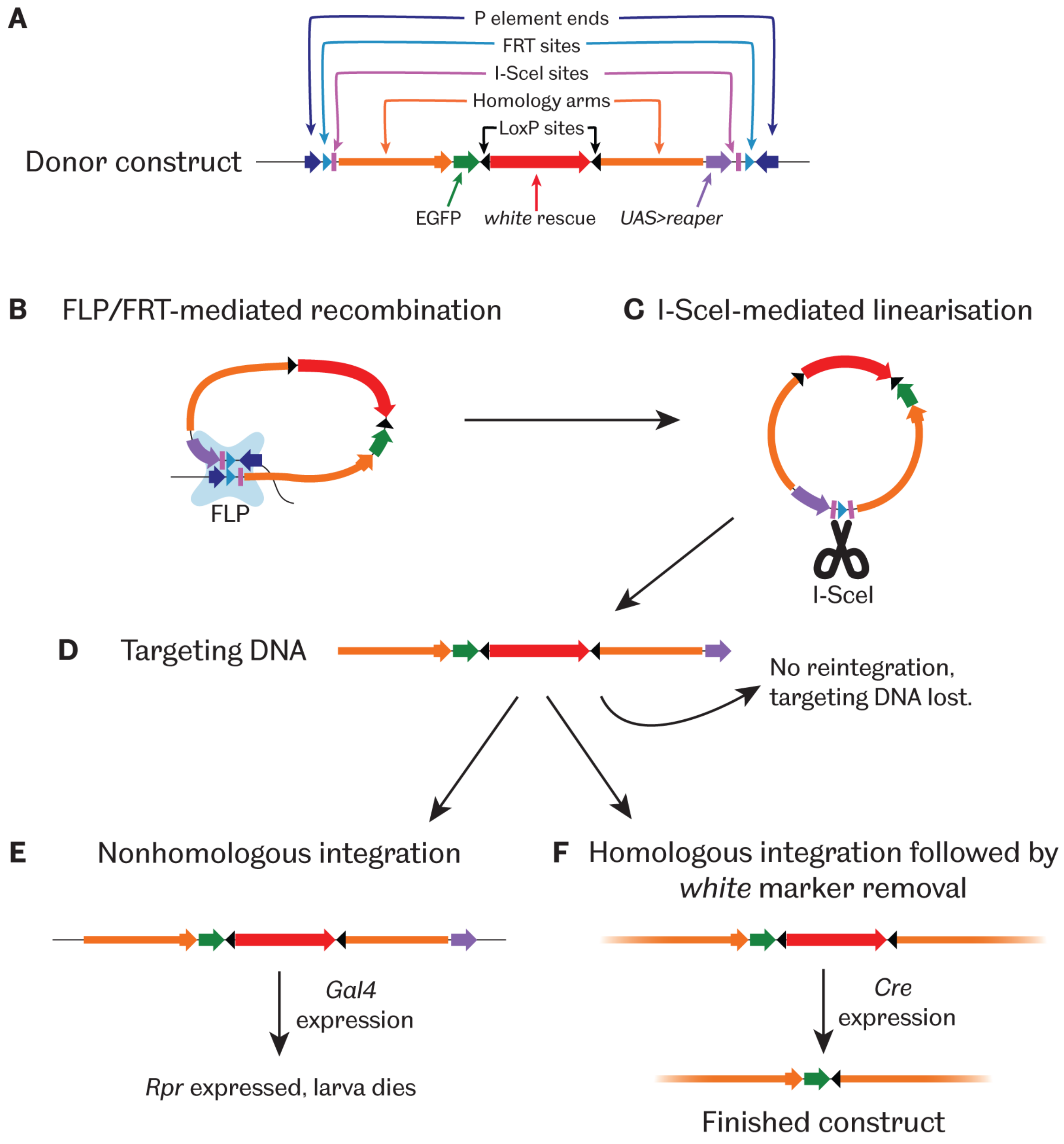


Figure 2.5: Gene targeting from the molecular perspective
A: The targeting donor construct is based on a P element. The FRT and I-SceI sites are needed to excise and linearise the targeting DNA. The genomic homology arms are added to the vector *in vitro*; these are what target the EGFP and floxed *white* rescue cassette to their homologous locus in the fly genome. *UAS>reaper* is present to facilitate screening of the potential hits. **B:** In the next generation, FLP-catalysed recombination between the FRT sites excises the targeting DNA as a closed circle. **C:** I-SceI cuts at the I-SceI site to linearise the targeting DNA. **D:** The linearised targeting DNA follows one of three fates. In most cases it does not reintegrate into the genome and is lost. **E:** The targeting DNA can reinsert in a random location, or be otherwise retained in the genome. If it has not integrated by homologous recombination it will retain *UAS-Reaper*, and the unfortunate fly carrying this insertion will express Gal4, and hence Reaper, and die. **F:** In the case of homologous integration the fly survives Gal4 expression and is crossed to a *Cre*-carrying fly to remove the floxed white rescue gene, leaving only the EGFP sequence and a single LoxP site at the targeted genomic locus.

sequence is confirmed by PCR, and the red eye colour rescue gene is removed from the fly stocks before they are used for experiments. The end result is an unmarked insertion of EGFP, with a short linker sequence and a single LoxP site, into the endogenous genomic locus (**Figure 2.5 F**).

2.5.2 Construct design: *prickle* and *frizzled*

The *in vivo* targeting protocol becomes quite labour-intensive once the screening for hits starts, so only one construct was designed for each gene. It is difficult to predict what effect, if any, adding the fluorophore to the protein will have, especially when so little is known about the molecular function of the target protein. However, due to the large number of transgenic constructs in our lab's collection and published in the literature, in some cases it was possible to infer suitable tag positions from existing transgenic constructs for each protein.

Prickle (Pk) has a prenylation site at its C-terminus and prenylation is required for its function in the planar polarity system (Veeman et al., 2003; Maurer-Stroh et al., 2007; Lin and Gubb, 2009; Strutt et al., 2013b), so the EGFP sequence had to be placed at the N-terminus. As discussed in the introduction Pk has two isoforms, named Pk ('Prickle-Prickle') and Sple ('Prickle-Spiny Legs'), which differ only at the N terminus, as shown in **Figure 2.6 A**. These isoforms have different phenotypes in the wing and the leg, but Pk has a much stronger wing phenotype than Sple. Sple is expressed at very low levels in the wing and is dispensable for planar polarity in this tissue (Gubb et al., 1999; Lin and Gubb, 2009; Strutt et al., 2013b). The tag was consequently placed immediately upstream of the first exon of the Pk isoform, as shown in **Figure 2.6 B**. These flies should still be able to produce endogenous Sple.

The *frizzled* gene is comparatively uncomplicated, with only one isoform (**Figure 2.6 C**). Three Frizzled (Fz) constructs tagged at the N-terminus were unable to fully rescue the mutant phenotype (D. Strutt, personal communication). The intracellular transmembrane loops might be an attractive target for tagging, but these are required for Fz's function in both planar polarity and Wg signalling (Jones et. al, 1996; Povelones et al., 2005; Wu et al., 2008), and placing a tag here could potentially disrupt the configuration of the transmembrane domains. The tag was consequently placed at the C-terminus of the protein (**Figure 2.6 D**). However, Fz has a C-terminal PDZ-binding

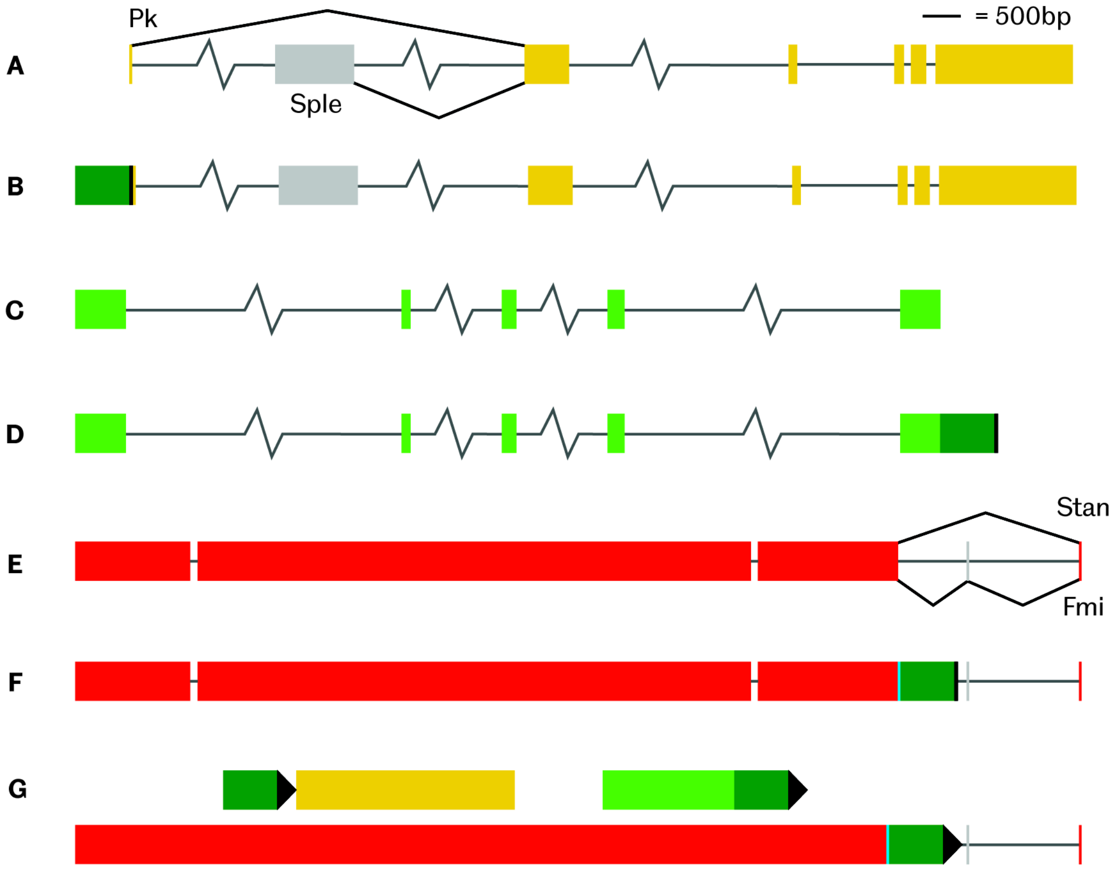


Figure 2.6: EGFP-Pk, Fz-EGFP and Stan-EGFP gene structures
A: Gene structure of *pk*^{pk} and *pk*^{sple}. The *sple*-specific exon is shown in grey. **B:** Location of EGFP (dark green) and LoxP site (black) in the *EGFP-Pk* construct. **C:** Gene structure of *fz*. **D:** Location of EGFP and LoxP in the *Fz-EGFP* construct. **E:** Gene structure of *stan* and *fmi*. The *fmi* isoform includes the 7bp exon shown in grey, and splices into the final exon in a different reading frame to *stan*. **F:** For *Stan-EGFP*, the final exon of *stan* was copied and placed immediately adjacent to the penultimate exon (blue), followed by EGFP, the LoxP sequence and a PolyA sequence (not shown). The *fmi* exon is still present, but the splice donor site of the 3rd exon has been destroyed by the addition of the 5th exon and following sequences. **G:** The relative lengths of EGFP-Pk, Fz-EGFP and Stan-EGFP. LoxP sites (black triangles) are not to scale.

All exons (wide coloured rectangles) and short introns (grey lines) are shown to the scale indicated. Long introns have been truncated, this is indicated by zigzags.

motif (PBM) (Itoh et al., 1998; Strutt et al., 2012). PBMs are 9bp motifs that must be at the extreme C-terminus of the protein in order for them to interact with PDZ domains (Kornau et al., 1995; Niethammer et al., 1996), so the PBM would presumably be blocked by the addition of a C-terminal EGFP. On the other hand, an existing Fz-GFP construct expressed under the *actin5C* or *Armadillo* promoter localised asymmetrically and was largely able to rescue the *fz*^{P21} mutant phenotype in the wing (Strutt, 2001). It was therefore decided to proceed with tagging the C-terminus.

2.6 Recombineering and *in vivo* gene targeting of *starry night*

While *starry night* (*stan*) is too large for the CHORI-322 P[acman] library mentioned above, there were two suitable clones in the CHORI-321 P[acman] library. Both CH321-66D09 and CH321-13B09 adequately covered the *stan* gene and the upstream and downstream sequences at least until the start of the next gene. Unfortunately, these constructs are 94kb and 105.5kb long respectively, which were considered too large for comfortable *in vitro* work and for efficient integration into the fly genome. Attempts to recover a chunk of the genomic DNA around *stan* into an empty P[acman] vector to make a reasonably-sized P[acman] construct failed. However, since the recombineering protocol was working very well, it was decided to try to tag Stan by recombineering, using both the 66D09 and 13B09 clones. As a back-up, the same construct was prepared in the pRK2-EGFP vector for *in vivo* targeting, in case we were unable to obtain transgenic flies using the large recombineered P[acman] constructs.

2.6.1 Construct design: *starry night*

As discussed in the introduction, the *flamingo* gene has two isoforms with two different names. Starry night (Stan) terminates with a PDZ-binding motif, whereas Fmi does not (Wasserscheid et al., 2007), see also **Figure 2.6E**. Since this construct was made, four additional isoforms have been predicted in FlyBase. However, these are only *in silico* predictions that have not been observed experimentally. In the unlikely event that these transcripts exist *in vivo* it is equally unlikely that they are expressed at levels that would affect or be affected by the Fmi-EGFP construct.

Wasserscheid et al. (2007) showed by reverse-transcriptase PCR that the *stan* transcript is much more abundant than *fmi* in third-instar wing discs, so it was decided to tag Stan in such a way as to prevent the flies from producing any Fmi at all. This was achieved by copying exon 5, changing the codon usage, and fusing it to the end of exon 3, with the EGFP tag and a PolyA sequence following immediately afterwards (**Figure 2.6 F**). While the information to produce the Fmi isoform will still be present in the flies, this modification should destroy the splice site needed to make the *fmi* mRNA. The PolyA sequence should additionally discourage polymerases from reading through and facilitate correct mRNA processing to eliminate Fmi. A potential disadvantage of this construct was the possibility that the 3' UTR contained elements affecting mRNA stability. Modifying the 3' end of the gene in this way might therefore have a knock-on effect on the resulting protein level. However, this was considered a minor risk, with no obvious means of avoiding it. Additionally, in this construct EGFP will block the PBM.

2.6.2 Recombineering and *in vivo* gene targeting of *starry night*

The main difficulty with recombineering the Stan-EGFP construct was the low transformation and recombination efficiency, which is to be expected when handling such large pieces of DNA *in vitro*. In the end, this was challenging but not impossible. No special modifications were made to the recombineering protocol, but extra care was taken to treat minipreps gently and always transform fresh DNA, not frozen-thawed – this appeared to be the main factor affecting the success of transformation. The recombination step was repeated several times, and very few colonies were recovered, but in the end a single correct recombinant clone was obtained for both the 13B09 and the 66D09 vectors.

Roughly 200 injected embryos per construct were received and allowed to eclose as adults before being crossed to white-eyed but otherwise wild-type flies for screening. Injected males were crossed individually to four female virgins, whereas injected females were crossed in pairs to three males. The 400 crosses were flipped into fresh vials every 2 days until the parents died or stopped producing larvae. The offspring were screened on the alternate days, looking for any rescue of the white eye phenotype. After several weeks of intensive screening, one transgenic fly was obtained for each construct. Unfortunately, the 66D09-derived insertion was on the wrong chromosome and was assumed to be either a false positive or a non-specific integration event. Screening of an estimated three hundred thousand flies thus resulted in just one correct transgenic line. However, the eye colour rescue was extremely weak, giving a very pale, almost creamy, colour rather than the more usual yellow or pale orange of a P[acman] transgenic, so it is likely that several transformants were in fact missed and discarded. Based on this experience with injecting large constructs, it was decided not to pursue this strategy in future.

The *in vivo* targeting method was very straightforward with no complications, and several verified hits were obtained.

2.7 Testing the EGFP-tagged constructs

2.7.1 Initial observations

Having generated EGFP-tagged transgenic fly lines for all six of the core proteins, it was necessary to check that the EGFP-tagged proteins were functionally equivalent to their endogenous counterparts. The first test was a cursory examination of the pupal wings on the confocal microscope to see whether any EGFP fluorescence could be detected. All constructs passed this initial test, except for P[acman] Stbm-LoxP-EGFP and P[acman] Dsh-LoxP-EGFP. No fluorescence at all was detected for these lines apart from the weak autofluorescence of the cell nuclei, which is also observed in wild-type flies. Neither construct was analysed further, as they were replaced by the functional C-tagged EGFP-LoxP versions. The presence of the EGFP gene was confirmed for all constructs by PCR on single-fly genomic DNA preps. The insertion site of the P[acman] constructs was also confirmed by PCR.

Another simple assay for the function of the P[acman] Dsh constructs was to check whether they could rescue the *dsh*^{V26} null mutation to viability, and this was the case for both the EGFP-Dsh and Dsh-EGFP-LoxP construct. The failed Dsh-LoxP-EGFP construct, which was made first, was not crossed into the mutant background. A similar test was conducted for the P[acman] Stan-EGFP construct with the lethal null mutant *fmi*^{E59}, but unfortunately it did not rescue to viability. Another null mutation, *fmi*^{192}, and *Df(2R)ED2076* (a deletion covering *fmi* and *dgo*), did not rescue either. P[acman] Stan-EGFP is viable in the wild-type background. Fortunately, the completion of the targeted version of the Stan-EGFP construct meant that the P[acman] version was no longer needed.

2.7.2 Adult wing phenotypes

The next simplest assay for function in the PCP pathway is to look at the adult wing phenotype. As mentioned in the introduction, the adult wing is a simple bilayered epithelium consisting of an array of hexagonal cells. Each cell produces a single hair, or trichome, from its distal edge and in the wild-type fly the hairs all point distally (**Figure 2.7 A**). Mutations in any of the six core complex genes cause characteristic swirling

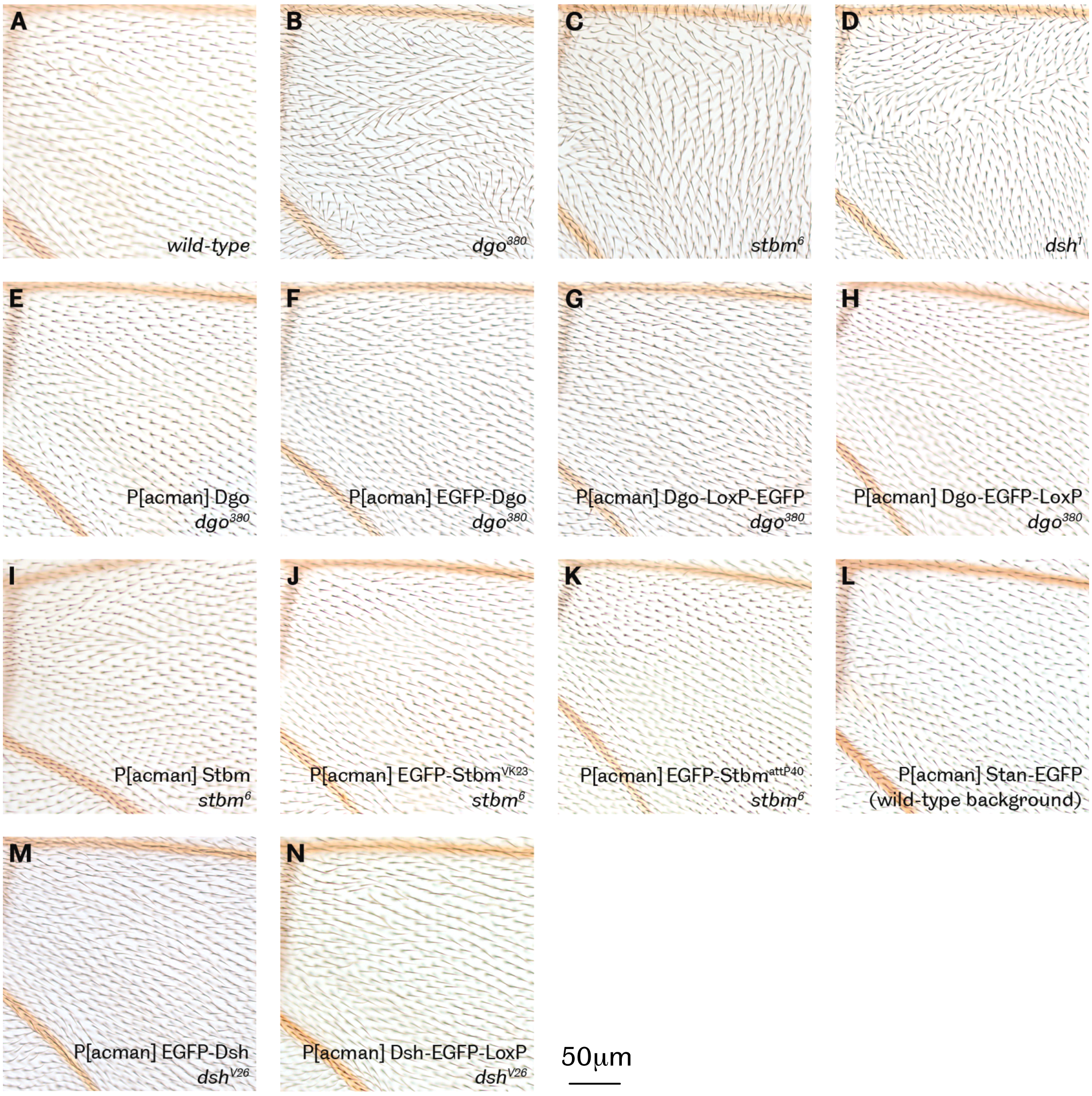


Figure 2.7: Endogenously-expressed untagged and EGFP-tagged P[acman] constructs rescue their respective mutant phenotypes

A: In the wild-type wing the hairs are all aligned and pointing in the distal direction. **B, C, D:** When components of the core planar polarity complex are mutated, the hairs adopt a swirling pattern. **E, I:** Untagged Dgo (E) or Stbm (I) expressed from the P[acman] vector rescues the relevant mutant phenotype. **F, G, H:** All three versions of tagged Dgo rescue the mutant phenotype. **J, K:** The same P[acman] EGFP-Stbm construct rescues the mutant phenotype in two different insertion sites, VK23 on III (J) and attP40 on II (K). **L:** The P[acman] Stan-EGFP construct does not cause overt polarity defects when in the wild-type background. **M, N:** EGFP-tagged Dsh rescues the *dsh* null polarity phenotype. dsh^{1} is a viable hypomorph, whereas dsh^{V26} is lethal.

Images are taken immediately posterior to the posterior crossvein. Distal is to the right and anterior is up. All flies are homozygous for the P[acman] insertion and/or the mutant, as applicable. All wings are from males, except for those in panels M and N which are from females. All P[acman] constructs were inserted into the *attP40* site on 2L, except for P[acman] EGFP-Stbm in panel J, which is in VK23 on III.

patterns in the wing trichomes (**Figure 2.7 B, C, D** and **Figure 2.8 B, F, K**), and this can also be caused by over-expression (Lu et al., 1999; Feiguin et al., 2001; Tree et al., 2002; Bastock et al., 2003) (**Figure 2.8 C** and **G**). Additionally, defects in the core PCP network can cause production of multiple trichomes per wing cell instead of the usual single trichome, as exemplified by the downstream PCP effector *multiple wing hairs* (Gubb and Garcia-Bellido, 1982). Mutations in *sple*, the other isoform of *pk*, cause defects in the distal leg joints (compare **Figure 2.8 E** and **E'**), and all core proteins can affect the packing of ommatidia in the eye, which is visible as varying degrees of roughness under the dissecting microscope.

As can be seen from **Figure 2.7**, all of the recombineered EGFP-tagged constructs that showed EGFP fluorescence were able to rescue their respective mutant phenotypes to a level comparable to wild-type. These flies were homozygously null for the endogenous gene and homozygous for the P[acman] construct, showing that the tagged protein has the same functionality as the endogenous version. The exception to this was P[acman] Stan-EGFP which, as described above, did not rescue the lethality of its mutant.

No differences were observed between untagged P[acman] Dgo (**Figure 2.7E**) and its three EGFP-tagged derivatives (**Figure 2.7 F, G, H**), or between untagged P[acman] Stbm (**Figure 2.7 I**) and its two functional EGFP-tagged versions (**Figure 2.7 J** and **K**) (Stbm-LoxP-EGFP did not show any fluorescence and was not tested further). This again shows that addition of EGFP does not affect the planar polarity of the adult wing for these constructs. The P[acman] Stan-EGFP construct did not have a polarity phenotype in the wild-type background (**Figure 2.7 L**). P[acman] EGFP-Dsh (**Figure 2.7 M**) and Dsh-EGFP-LoxP (**Figure 2.7 N**) rescued the lethality of the null mutation dsh^{V26}, with no sign of the polarity disturbances seen in the viable hypomorph dsh^{1} (**Figure 2.7 D**). Unfortunately, no untagged P[acman] Dsh insertion was obtained for comparison. No multiple wing hairs were observed in the adult wings of any P[acman] line.

For the targeted constructs, it was important that they did not perturb the normal polarity of the wing, as this would again indicate that the tagged protein was not able to function in the same way as the endogenous. No significant deviations from normal polarity were observed (**Figure 2.8 A, D, H, L**). No multiple hairs were observed in the adult wings of any targeted construct.

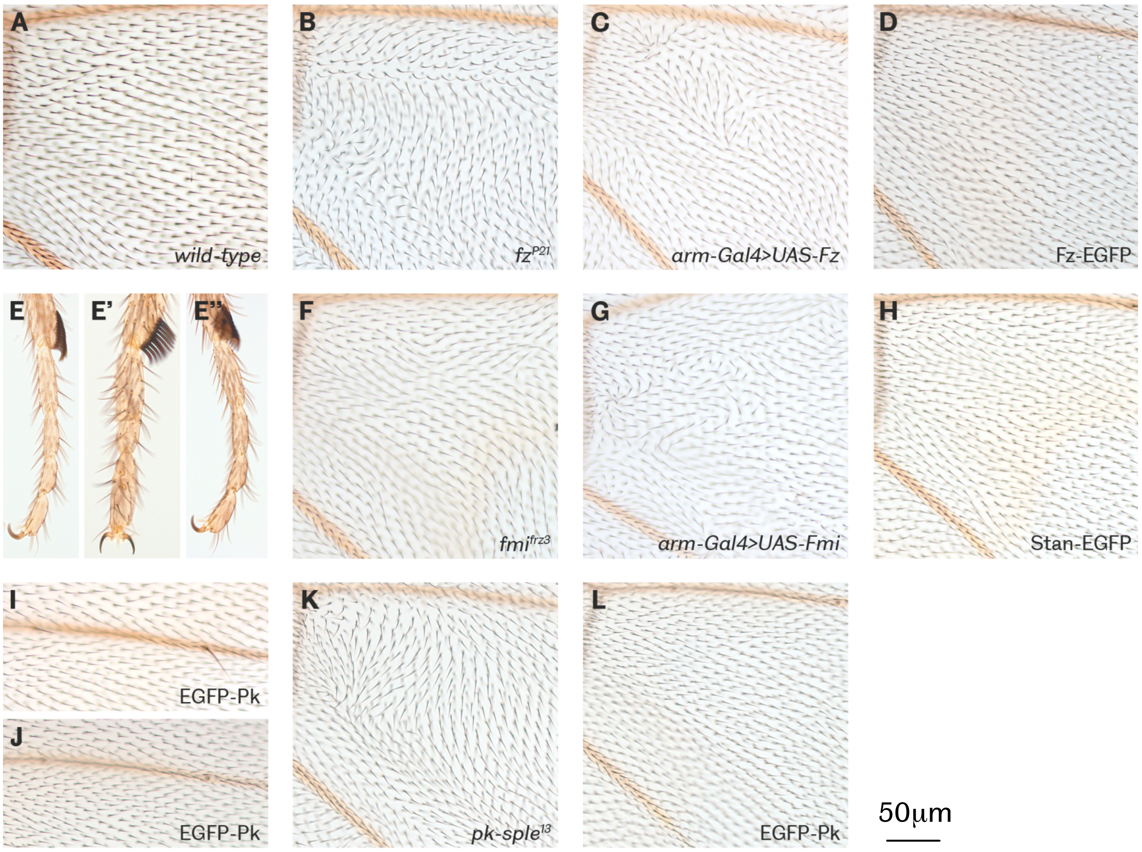


Figure 2.8: Core proteins tagged by gene targeting do not perturb normal trichome polarity
A: In the wild-type wing the trichomes are all aligned. **B, F, K:** Mutations in *fz* (B), *fmi/stan* (F) or *pk* (J) cause distinctive swirling hair phenotypes. **C, G:** Overexpression of *fz* (C) or *fmi/stan* (G) causes similar swirling phenotypes. **D, H, L:** The EGFP-tagged constructs generated for Fz (D), Stan (H) and Pk (K) do not perturb wing hair polarity. **E:** The wild-type fly leg. **E':** The *pk-sple*^{13} leg shows joint and bristle polarity defects. **E'':** The EGFP-Pk leg is identical to wild-type. **I:** Occasional conversion of campaniform sensillae to sensory bristles was observed in some individuals for all-constructs. Part of vein 4 distal to the posterior crossevein in an EGFP-Pk wing is shown. **J:** The campaniform sensillae of an unaffected wing in the same region and genotype as shown in I.

Except for I and J, wing images are taken immediately distal to the posterior crossvein; distal is to the right and anterior is up. The UAS flies are heterozygous for *arm-Gal4* and the *UAS* construct. All other flies are homozygous for the indicated mutations and/or insertions. All wings and legs are from males.

Occasionally, some campaniform sensillae in the wing veins were converted to sensory bristles (compare **Figure 2.8 I** and **J**). This affected some, but not all, individuals of most of the constructs. Differentiation of these sensory organs is controlled by the *wingless* signalling pathway, and overexpression of Dfz2 causes a similar ectopic bristle phenotype in the wing blade (Rulifson et al., 2000). Since this effect was only seen in some flies and did not affect the polarity of the wings this result was not investigated further. Other aspects of adult morphology, including notum hair polarity and ommatidia packing, appeared normal for all constructs but were not formally recorded.

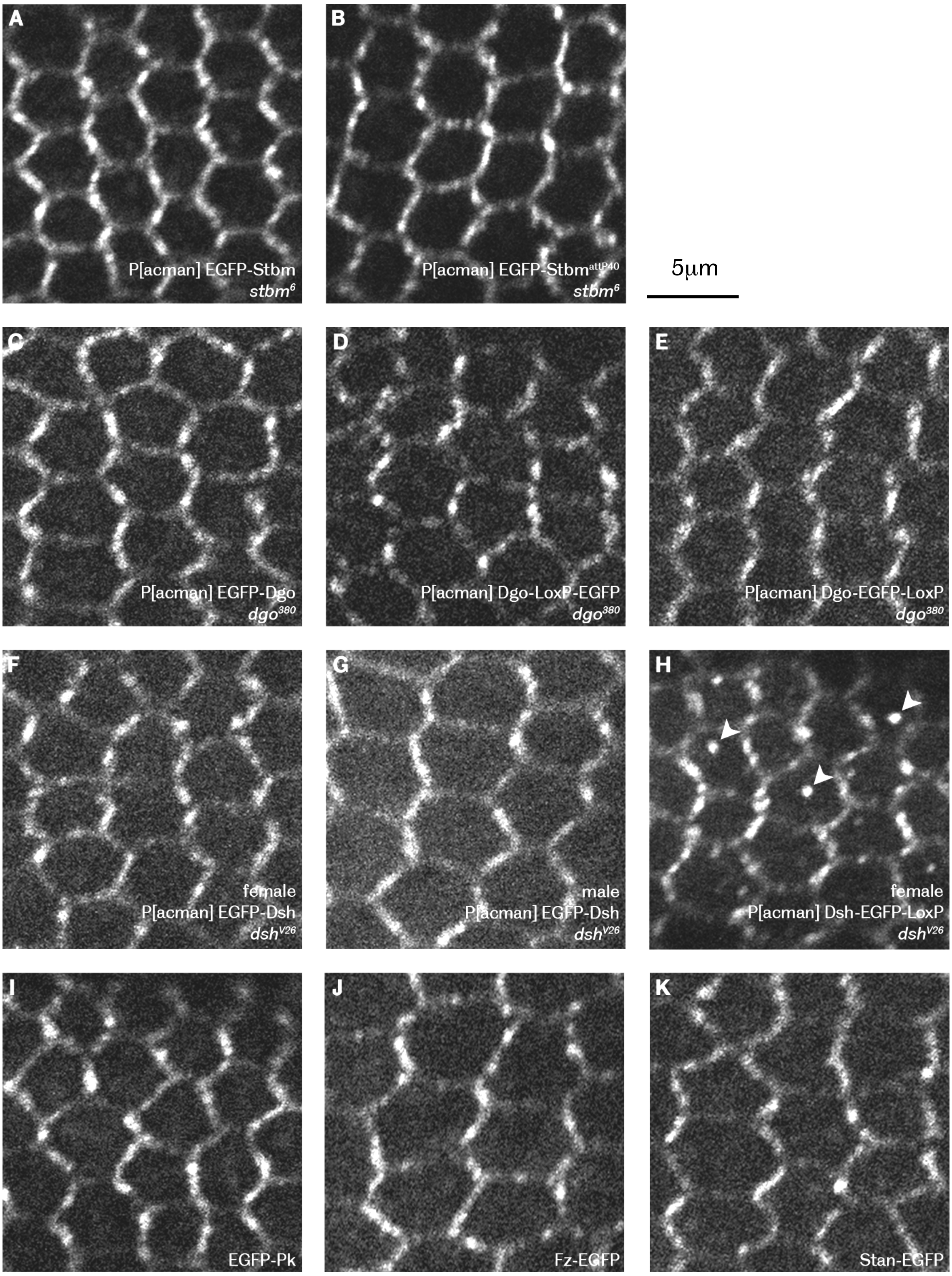


Figure 2.9: EGFP-tagged core protein localisation in the 28h APF pupal wing
A - **K:** All of the EGFP-tagged core proteins that showed fluorescence adopted the normal asymmetric localisation. **D:** Dgo-LoxP-EGFP puncta were more rounded and discrete than usual (arrowheads). **H:** In addition to the normal asymmetric localisation, Dsh-EGFP-LoxP showed very bright cytoplasmic puncta (arrowheads).

Distal is to the right and anterior is up. All flies are homozygous for the indicated transgenes and mutations.

2.7.3 EGFP-tagged core protein localisation in the pupal wing

Since the adult flies appeared to have normal polarity in all external tissues examined it was expected that the subcellular localisation and behaviour of the tagged protein at the molecular level would mimic that of the endogenous untagged proteins. To confirm this, the wings of live pupae at 28h APF (hours after puparium formation) were imaged on a Nikon A1 confocal microscope as described in the Materials and Methods. As shown in **Figure 2.9**, all constructs that displayed EGFP fluorescence showed the familiar zig-zag localisation pattern.

P[acman] EGFP-Stbm inserted into the *attP40* site on 2L showed the same asymmetric subcellular distribution as the same construct in the *VK23* site on the third chromosome (**Figure 2.9 A** and **B**). Likewise for the three Dgo constructs, the tagged proteins adopted asymmetric localisation. However, there is a subtle difference in the puncta in cells expressing Dgo-LoxP-EGFP. While EGFP-Dgo puncta look normal (**Figure 2.9 C**), those in Dgo-LoxP-EGFP animals look more ‘spotty’, discrete and rounded than usual. (**Figure 2.9 D**). This effect is not seen in Dgo-EGFP-LoxP puncta (**Figure 2.9 E**), which look normal and identical to EGFP-Dgo.

dsh is on the X chromosome, so females usually carry two copies, whereas males have just the one. However, P[acman] EGFP-Dsh is inserted on the second chromosome, so both the males and females in the experimental stock carry two copies. Due to dosage compensation, the males might be expected to express twice as much EGFP-Dsh as the females. No obvious difference in subcellular localisation was seen between male and female pupae expressing P[acman] EGFP-Dsh (**Figure 2.9 F** and **G**). While the Dsh-LoxP-EGFP flies did not show any fluorescence, the Dsh-EGFP-LoxP line did. Unfortunately, the vast majority of Dsh-EGFP-LoxP animals showed abnormalities at the cellular level. In addition to the usual asymmetric punctate localisation, large bright cytoplasmic spots were also visible (**Figure 2.9 H**), which are not seen in wild-type flies (Axelrod, 2001; Shimada et al., 2001). This could be indicative of any number of problems with expression, trafficking or recycling of the fusion protein, so this line was not used for gathering experimental data. The targeted constructs EGFP-Pk, Fz-EGFP and Stan-EGFP (**Figures 2.9 I**, **J** and **K** respectively) showed normal asymmetric localisation.

In wild-type animals, Stbm and Pk are found on the proximal side of the cell, Dgo, Dsh and Fz on the distal side and Stan on both sides. This localisation was confirmed for Stan by generating clones of cells expressing the fluorescent construct adjacent to wild-type cells using the *Flp*-FRT system. While it is theoretically possible that the other fusion proteins are localising to the wrong side of the cell, such localisation is invariably associated with total loss of asymmetry and swirling wing trichomes, neither of which were observed. **Figures 2.10 A**, **B** and **C** show clones from three different EGFP-Stan wings. The mean membrane fluorescence was measured on proximal and distal membrane segments around clone boundaries, using a 10 pixel wide line selection in ImageJ. The mean ratio of proximal to distal mean membrane brightness was 0.933, indicating a slightly greater fluorescence on distal membranes, but this was not a significant result (two-tailed ratio paired t test, P = 0.1011). 13 images from 9 different pupae were analysed.

2.8 Other fluorophores

All of the tagged polarity proteins for measuring the stoichiometry of the core complex had to use the same fluorophore, otherwise the brightness measurements would not be comparable. However, as the recombineering protocol was working efficiently it was decided to try replacing EGFP with several other fluorophores. Two-colour imaging *in vivo* has many potential applications, as it would allow simultaneous observation of two or even three core proteins. Previous attempts using mCherry and tdTomato (both red) (Shaner et al. 2004) or the photoconvertible tetrameric green-and-red Eos (Wiedenmann et al., 2004) protein under heterologous promoters did not result in visible fluorescence in the flies. Most of these constructs had been made using the transmembrane core proteins, so it was possible that something to do with the membrane localisation was preventing the maturation or fluorescence of the red fluorophores. It was therefore decided to work on the cytoplasmic protein Dgo in the first instance.

mCherry, a monomeric red fluorescent protein, and tdTomato, a tandem dimer red fluorescent protein, were added to P[acman] Dgo. Both mCherry and tdTomato are derived from mRFP1 (Shaner et al. 2004). mRFP1 is an extensively mutated version of dsRed, from the sea anemone *Discosoma striata* (Matz et al., 1999). No fluorescence was seen in the P[acman] tdTomato-Dgo lines, so this construct was not crossed into

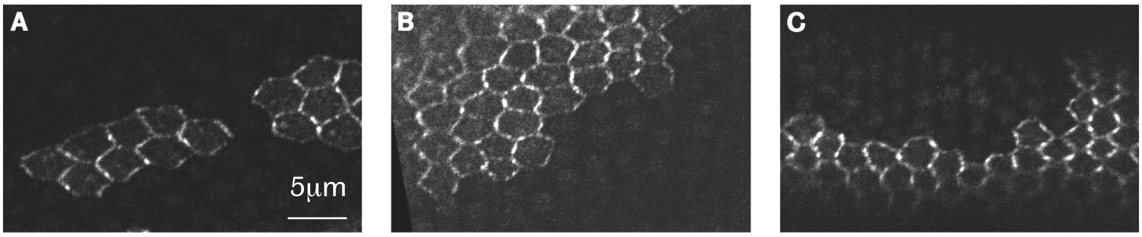


Figure 2.10: Stan-EGFP mitotic clones
A, B, C: Clone of Stan-EGFP in a wild-type background. Stan-EGFP localises to both the proximal and distal edges of the cells. Overall gene dosage is preserved throughout the tissue. Distal is to the right and anterior is up.

the mutant background. P[acman] mCherry-Dgo was able to rescue the mutant adult wing phenotype and fluoresced weakly but with normal asymmetric localisaton (**Figure 2.11 A** and **I**). Although the mCherry-Dgo construct did not give a strong enough signal for quantification, the qualitatively similar localisation to EGFP-Dgo suggested that neither the EGFP nor the mCherry tag affected the behaviour of Dgo.

TagBFP and Cerulean constructs were then made for Dgo, Dsh and Stbm, as well as a P[acman] mCherry-Dsh construct. As discussed above, the other cytoplasmic core protein gene *pk* was not suitable for recombineering, so was not a candidate for experimentation with alternative fluorophores. Stbm is a transmembrane protein, but was easy to recombineer due to its small size.

Cerulean is an improved version of the GFP-derived ECFP (enhanced cyan fluorescent protein), which shifts the excitation and emission peak wavelengths toward the blue end of the spectrum (Heim and Tsien, 1996; Rizzo et al., 2004). No signal was seen in the cerulean constructs beyond autofluorescence (data not shown). However, our Nikon A1 confocal microscope does not have a suitable laser line for optimal excitation of Cerulean, so these constructs may be weakly fluorescent.

TagBFP is a mutation of TagRFP (Subach et al. 2008), a red fluorescent protein from the sea anemone *Entacmaea quadricolor* (Merzlyak et al., 2007). The TagBFP-Dsh and TagBFP-Stbm constructs gave a comparatively good signal (**Figure 2.11 B** and **C** respectively), but the TagBFP-Dgo construct fluoresced very weakly with abnormal localisation (not shown). TagBFP-Stbm pupae additionally showed abnormal localisation, sometimes with bright cytoplasmic spots, reminiscent of the localisation of Dsh-EGFP-LoxP.

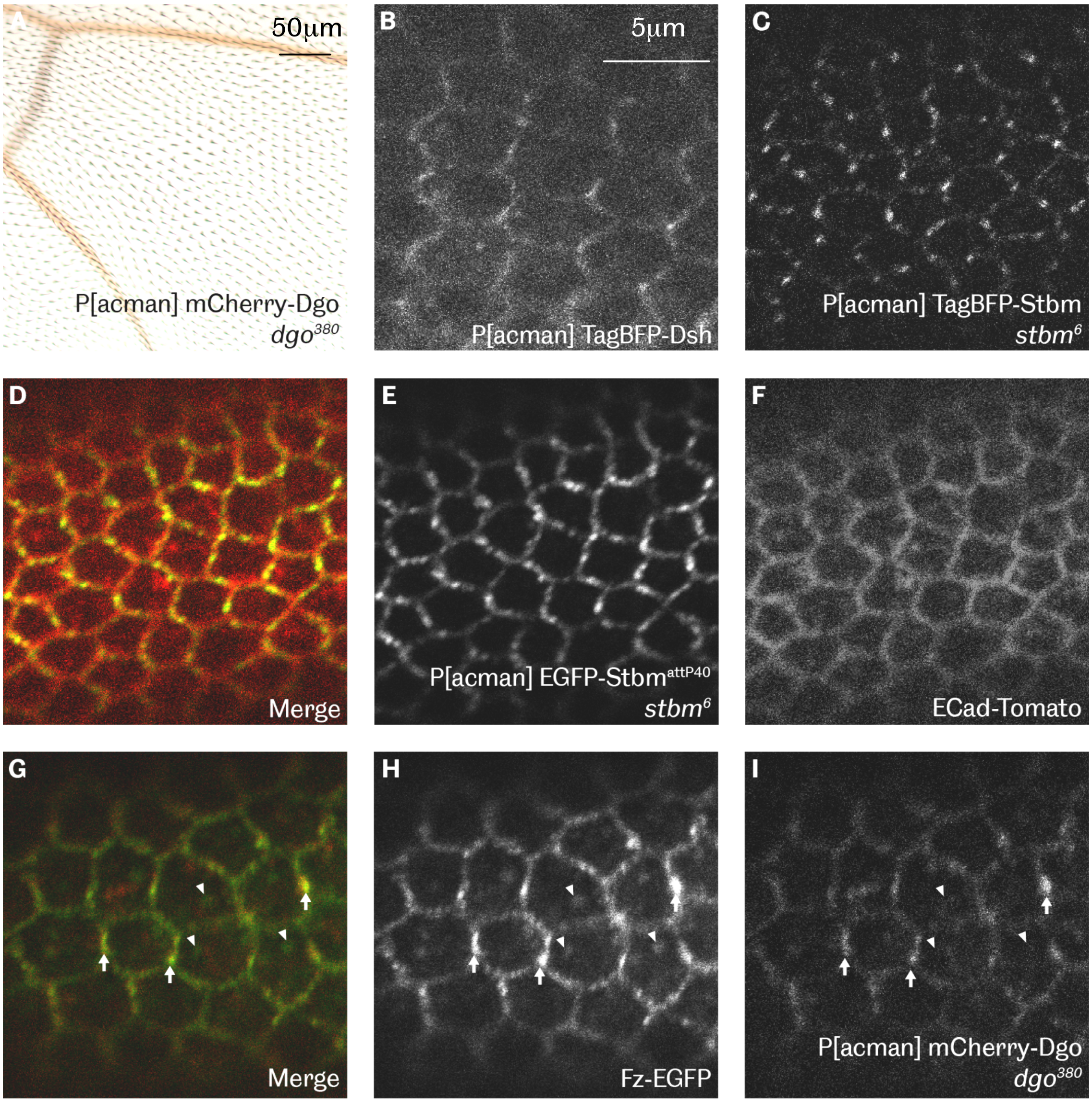


Figure 2.11: mCherry- and TagBFP-tagged core protein localisation in the 28h APF pupal wing

A: P[acman] mCherry-Dgo rescues the mutant adult wing phenotype. **B:** P[acman] TagBFP-Dsh expresses and localises normally in the wild-type background with a reasonably strong signal. **C:** P[acman] TagBFP-Stbm fluoresces relatively brightly, but shows abnormal localisation. **D, E, F:** Two-colour imaging of EGFP-Stbm (E) and a targeted DE-Cad:mTomato construct (F) met with limited success. **G, H, I:** A pupal wing homozygous for Fz-EGFP (H) and P[acman] mCherry-Dgo *dgo*^{380} (I). The two fusion proteins colcalise in membrane puncta (arrows) but only Fz-EGFP is seen in cytoplasmic spots (arrowheads).

The brightness and contrast of the pupal wing images have been enhanced. Distal is to the right and anterior is up. All flies are homozygous for the indicated transgenes and mutations.

Two-colour imaging met with limited success. A DE-Cadherin-mTomato targeted construct from Huang et al. (2009) expressed, localised and fluoresced relatively well in combination with P[acman] EGFP-Stbm$^{\text{attP40}}$ *stbm*^{6} (**Figure 2.11 D**, **E** and **F**). When Fz-EGFP and P[acman] mCherry-Dgo *dgo*^{380} were expressed together (**Figure 2.11 G**, **H** and **I**), colocalisation of Fz-EGFP and mCherry-Dgo could be seen in puncta.

Cytoplasmic spots of Fz-EGFP did not contain mCherry-Dgo. This shows that two-colour imaging could be useful, but would require red fluorophores that expressed better and optimisation of imaging settings.

A full list of all of the red and blue core protein constructs made is shown in the Materials and Methods.

2.9 Discussion

2.9.1 Recombineering *versus* gene targeting

Both recombineering and *in vivo* gene targeting by homologous recombination are reliable and satisfactory methods for making endogenously-expressed fusion proteins. Recombineering is the more versatile method and thus more suited to making several versions of a construct. This is useful when there is little data available to assist with tag placement, but requires a reliable null mutation to restore normal gene dosage. The *attP* landing site must be carefully chosen both to avoid position effects (Markstein et al., 2008) and to facilitate making experimental stocks that combine the tagged protein with other mutants.

Gene targeting is particularly suitable for larger genes or cases where only one version of the tagged construct is desired. However, it can be a risky option if there is no molecular data available to suggest a suitable location for the tag. Since establishing this protocol in the lab, four further constructs have been successfully generated using this method with satisfactory results, and it is now our preferred method for experiments that would benefit from endogenous expression of a tagged protein. A summary of the gene targeting results is located in the Materials and Methods chapter.

2.9.2 Function of the EGFP-tagged core proteins

The EGFP-tagged fusion proteins are able to function normally in the planar polarity pathway, as demonstrated by the lack of adult planar polarity phenotypes. For P[acman] EGFP-Stbm, the one construct that was injected into two different genomic landing sites, no differences in adult phenotype were observed. This is probably because these two sites were selected specifically because they are known not to hinder the expression of inserted transgenes, and this result should not be extrapolated to every genomic ΦC31 landing site. However, this result does suggest that all the other constructs, which were inserted into *attP40*, should be able to express normally, and that the failed constructs were not suffering from position effects.

Something else must have caused the failure of Stbm-LoxP-EGFP and Dsh-LoxP-EGFP. For both of these constructs moving the LoxP site to the C-terminal side of EGFP, leaving just a short linker (which is present in all constructs) between the polarity protein and the fluorophore, resulted in functional fluorescent fusion proteins. The presence of the LoxP site is unavoidable with this recombineering method, because the floxed Kanamycin marker is required for selection of recombinant colonies. Its subsequent removal by Cre-mediated recombination will always leave a single LoxP site behind.

It is possible that the LoxP site contains some sequence that adopts a secondary structure that hinders the folding of EGFP or the function of the polarity protein, or some RNA sequence that interferes with mRNA processing. The construct sequences were thoroughly checked before starting the molecular biology, and were also checked by sequencing before injection, but it is possible that there is some design flaw in the original versions which was not carried through to the second versions. It is interesting that the adult phenotype of C-tagged Dgo was unaffected by the position of the LoxP site, but Dgo is somewhat notorious for having the weakest phenotypes of the core proteins so perhaps this is unsurprising.

Interestingly, for Stbm, Dsh and Dgo in the adult wing it does not make any difference whether the N or C terminal is tagged. This suggests that the termini of these three core proteins do not mediate any essential molecular functions, which would presumably be blocked by the addition of the fluorophore. This contrasts with what is known for Pk, which relies on C-terminal prenylation for stable localisation at the membrane. This also implies that the PBM of Stbm is not necessary for its function.

In most cases the fusion protein adopts the normal asymmetric subcellular localisation. The tagged core proteins must therefore be expressed at, and functioning at, levels comparable to their endogenous counterparts. Additionally, where PBM domains have been blocked (Stan-EGFP and Fz-EGFP) this has had no overt effect on protein function. (C-tagged Stbm was not used for experiments). The functional recombineered constructs, and all four targeted constructs, fluoresce at levels that are easily visible using confocal microscopy. All of the EGFP-tagged core protein constructs are homozygous viable and fertile, so the experiments can be done in the absence of the wild-type proteins.

Ideally, the levels of the EGFP-tagged core proteins could be compared to their endogenous counterparts by Western blotting. While antibodies exist that work well for some of the core proteins, unfortunately there was not time to complete these experiments.

2.9.3 Examining the other genes included in the P[acman] constructs

In the P[acman] constructs, gene dosage is only restored for the gene of interest. As these constructs contain several kilobases of genomic DNA around the gene of interest they do also contain other genes, which will be present at twice the usual gene dosage in the experimental stocks. Full lists of these genes are shown at the end of this chapter in **Tables 2.1 – 2.4**. Around half of these genes are predicted genes, with no experimental data to support them, so they may not actually exist, or they may be unidentified exons of nearby genes. The majority of them are unlikely to be a cause for concern, with the exception of the following.

Rab3, in the P[acman] Stan-EGFP construct, is involved in synaptic vesicle trafficking (DiAntonio et al., 1993; Graf et al., 2009), and the experimental tissues are all epithelial in nature. Even if the P[acman] Stan-EGFP flies demonstrated a Rab3 overexpression phenotype it would be unlikely to manifest in the planar polarity pathway. Anyway, the targeted version of Stan-EGFP was used for all the experiments in this project so this is not a concern for the experimental data. The *G protein-coupled receptor kinase interacting ArfGAP* and *CG7220*, a predicted E2 ubiquitin ligase, in the same construct may have an effect on muscle development and protein degradation respectively, but can also be discounted for this reason.

tsunagi, in the P[acman] Stbm construct, could potentially affect oocyte patterning when overexpressed, but its effects appear to be limited to the female reproductive tract (Mohr et al., 2001). There were no overt fertility problems in females carrying this construct. The P[acman] Dgo construct contains *CG12344*, a predicted chloride channel, but again, no overt ill effects were observed. The fact that this gene only has a CG number rather than a real name shows that this gene has not yet been formally identified and may not display the predicted activity *in vivo*.

The most concerning protein is *hopscotch* (*hop*), in the P[acman] Dsh construct, which is the sole *Drosophila* homologue of the vertebrate JAK. The JAK/STAT signalling pathway controls cell proliferation and differentiation, particularly in the haemolymph/blood and immune systems. Flies overexpressing *hop* develop melanotic tumours (Hanratty and Ryerse, 1981; Harrison et al., 1995). However, the 5' end of the first exon of *hop* is not included in P[acman] Dsh, so very low expression, if any, would be expected from this copy of the gene. No tumours were seen in the P[acman] EGFP-Dsh and Dsh-EGFP-LoxP flies. Genes included in the P[acman] Dsh constructs would also be expected to experience dosage compensation. In *Drosophila,* dosage compensation is achieved by up-regulating genes on the male X chromosome to match the expression levels from the female's two copies of the X This effect generally persists even when a gene from the X is translocated to an autosome, so pupae carrying any Dsh construct were always sexed before imaging.

Overall, these additional genes are unlikely to cause problems. When overexpression is deliberately induced to give experimental phenotypes it is driven to express at much higher levels than would be caused by a doubling of gene dosage, so any overexpression effects here are likely to be mild or non-existent. Additionally, the experimental data is gathered in a specific tissue at a specific developmental stage, and these genes may not even be active at these times in the pupal wing.

2.9.4 Properties and dimerisation of EGFP

As mentioned above, the codon usage of the EGFP gene is optimised for human cells. In 2010, Pfeiffer et al. published a *Drosophila*-optimised GFP containing the same brightness-enhancing mutations as the EGFP sequence. Unfortunately at this time work on the EGFP constructs had already begun, so, even if it were deemed necessary, it would have been too late to change fluorophore. However, EGFP has been widely and successfully used in *Drosophila*, both in cell culture and *in vivo*, since at least 2000 (Schotta and Reuter, 2000; Morrow et al., 2000).

Many sources report that GFP has a somewhat distressing tendency to dimerise (Phillips Jr., 1997; Jain et al., 2001; Zacharias et al., 2002; Chalfie and Kain, 2005), although it is difficult to find a through examination of this behaviour. Dimerisation is reported both in crystals (Phillips Jr., 1997) and when expressed in cultured cells (Jain et al., 2001). The EGFP molecule used in this study has not had this ability ablated, so would be expected to retain the dimerisation kinetics of the parent GFP molecule. A monomeric form of GFP is available (Zacharias et al., 2002), but we did not use it. Many studies have been conducted using dimerisation-competent GFP and EGFP, and where this is mentioned, few authors have reported a significant effect of dimerisation on their results. The exception to this is where EGFP is attached to a secretory signal and expressed in endocrine cells in culture, in which case it very obviously oligomerises (Jain et al., 2001), and in Zacharias et al. (2002), which targeted GFP on its own to lipid rafts.

Dimerisation of EGFP could affect the stoichiometry of the core proteins by interfering with complex formation or turnover. Is the dimerisation likely to be sufficiently strong to affect the results? Phillips Jr. (1997) estimated the dissociation constant (K_d) of GFP by analytical ultracentrifugation, and found it to be 100μM, similar to the value of 0.11mM reported for YFP in Zacharias et al. (2002). He describes this as a 'weak association', but how weak is 'weak'?

The dissociation constant of a complex is defined as the concentrations of the free components multiplied together and divided by the concentration of the dimer. In the case of GFP dimerisation, where A and B are the same molecule, K_d is simply equal to the concentration of monomeric GFP, squared, and divided by the concentration of the dimer:

$$K_d = [A]^x[B]^y / [A_xB_y] \quad \rightarrow \quad K_d = [\mathrm{mGFP}]^2 / [\mathrm{GFP{:}GFP}]$$

Thus, when the concentration of monomeric GFP (mGFP) is equal to the dissociation constant, so is the concentration of the GFP dimer:

At $[\mathrm{mGFP}] = K_d$, $\quad [\mathrm{mGFP}] / [\mathrm{GFP{:}GFP}] = 1$

With a K_d of 100μM, this means that the concentration of free GFP would have to reach 100μM before there was an equal concentration of dimeric GFP. Each dimer consists of two molecules of GFP, so this equates to a cellular concentration of GFP monomers of

300μM. Is this a realistic physiological concentration? The concentration of EGFP in the cells will be equal to the concentrations of the individual core planar polarity proteins.

The concentrations of the core planar polarity proteins have not been measured, but figures are available for many other proteins in various organisms and cell culture lines. Ubiquitin, so named for its ubiquitous expression profile, was measured at approximately 85μM in HEK293 cells (Kaiser et al., 2011). The concentration of actin in the cytoplasm and organelles in *S. pombe* was estimated at 8.7μM (Takaine and Mabuchi, 2007). Ghaemagammi et al. (2003) calculated the cellular abundance of over 3,000 *S. cerevisiae* proteins, from which an average cellular concentration of 0.4-1.4μM was calculated. By contrast, the concentration of β-catenin, a component of the adherens junction, in the *Xenopus* embryo is only 35nM (Lee et al., 2003).

Actin and Ubiquitin are very highly-expressed proteins, which are involved in basic cellular processes. The planar polarity proteins have a much more specific and specialised role, so their concentrations can reasonably be assumed to be much lower than that of Actin or Ubiquitin. The adherens junctions are in a similar subcellular domain to the core protein puncta and appear to be of a similar size and abundance (Bastock et al., 2003), therefore it is likely that the concentrations of the core planar polarity proteins are of the same order of magnitude of that of the adherens junction component β-catenin. Taken together, these inferences suggest that dimerisation will not be a significant problem at the physiological concentrations of the core proteins.

This is all assuming that the polarity proteins are floating freely within the cell, but of course this is not the case. The core proteins coalesce into bright puncta at the apicolateral cell membrane. This means that the local concentration of core proteins in the puncta will be much higher than the estimated cellular concentration. The rough dimensions of a pupal wing cell are 5μm in diameter and 17.5μm in height, giving an approximate volume of 350μm^3. Puncta in my images tend to have areas approaching 0.2μm^2; if this two-dimensional area is modelled instead as a sphere of the same radius (which is probably a bit of an over-estimate) then the rough volume of an average punctum is 0.065μm^3. With roughly 6 puncta per cell, the total puncta volume of a cell is 0.39μm^3.

Assuming all molecules of the core protein are localised in puncta (which they are not), the local concentration of a planar polarity protein in a punctum would therefore be around 30μM, which is below the measured GFP dissociation constant of 100μM, but on the same order of magnitude. Rearrangement of the dissociation constant equation gives a dimer concentration equal to monomer concentration squared, divided by K_d. So, at this concentration of EGFP, the concentration of EGFP dimers in the puncta would be predicted to be 9μM, or almost 30% of the total. This is quite a generous estimate, as all assumptions made were chosen to model the greatest possible dimerisation. This also does not take into account the fact that the EGFP is physically tethered to the polarity protein. This physical constraint may favour dimerisation by bringing EGFP molecules into close proximity, or it may hinder dimerisation by preventing the EGFP monomers from accessing each other's dimerisation sites.

In conclusion, from these very rough estimates of EGFP dimerisation, it is likely that a fraction of the EGFP molecules will dimerise in puncta (steric help or hindrance notwithstanding). However, any effect of dimerisation is difficult to predict and impossible to prevent when using a dimerisation-competent molecule. In the end, any inaccuracies introduced by EGFP dimerisation are likely to be absorbed into the other sources of error inherent in the attempt to measure the brightness of the puncta in the pupal wing.

Finally, the possibility remains that attaching EGFP to the core proteins might affect their behaviour, rather than that of EGFP. For the recombineered constructs, where N- and C-tagged versions of the same protein can be directly compared, some differences in protein localisation were observed. Dsh-EGFP-LoxP was seen in bright cytoplasmic spots as well as its usual punctate membrane distribution and diffuse cytoplasmic fluorescence. There is also a subtle difference between EGFP-Dgo and Dgo-LoxP-EGFP, but the localisation of Dgo-EGFP-LoxP is identical to that of EGFP-Dgo.

Comparisons between untagged and tagged P[acman] Stbm and P[acman] Dgo adult wings show no differences, indicating that, for these constructs, addition of EGFP to the gene of interest does not affect their function sufficiently to perturb trichome orientation. However, tests for differences in trafficking or turnover dynamics were not conducted. At 28h APF the core proteins are at their most stable localisation, so any subtle time-sensitive differences in trafficking or turnover would be expected to have resolved by this time. At the end of the day, the flies did not show any polarity defects, which shows that the tagged core proteins and the planar polarity pathway as a whole are all functioning within normal parameters.

2.9.5 Red and blue fluorophores

While the red and blue constructs did not give as strong a signal as the EGFP ones, some conclusions can be drawn from these preliminary two-colour imaging results. The similar localisation of EGFP-Dgo and mCherry-Dgo suggests that for this construct these two tags do no affect the localisation of the fusion protein. However, the TagBFP-Stbm construct, which was identical to EGFP-Stbm apart from the fluorophore, showed abnormal localisation that was obviously different to that of EGFP-Stbm. Conversely, TagBFP-Dsh showed normal localisation in the wild-type background. It seems that the type of fluorophore can sometimes affect the behaviour of the fusion protein in an unpredictable manner.

Due to their low signal, and correspondingly low-contrast images, the red and blue constructs are unlikely to be useful for image quantitation applications. However, they could still be used in qualitative studies. For example, the preliminary imaging of Fz-EGFP and mCherry-Dgo shows that they are not trafficked together – unsurprising, as Fz is a transmembrane protein and Dgo is cytoplasmic. Applying this method to other protein pairs would allow trafficking and protein dynamics to be observed *in vivo*, especially if brighter blue or red fusions could be made.

Chapter 3 explains some of the difficulties in detecting cells based on a punctate and asymmetric marker. The ECad-Tomato knock-in construct could be very useful as a membrane marker to help solve this problem, as it localises evenly around the cell membrane in the same Z plane as the core proteins. Unfortunately the stoichiometry imaging was well under way when this construct was obtained, but it has been combined with the best EGFP construct for each core protein to facilitate further experiments of this nature.

Table 2.1: Genes contained in the P[acman] Dgo constructs

P[acman] CH322-99N23 (Dgo)

Gene	Evidence	Functions
mms4/CG12936	Predicted	Nuclease activity, DNA binding. DNA metabolic process
CG12341	Predicted	Unknown
CG7222	Predicted	Unknown
dgo	Experimental	Planar polarity
Prosbeta5	Experimental, inferred	Proteasome b5 subunit, endopeptidase activity
Lsm10	Experimental	snRNP, histone mRNA processing
Ribosomal protein S15Ab	Experimental, inferred	Ribosomal subunit
CG12343	Experimental, inferred	mRNA splicing
CG12325	Experimental, predicted	Neurogenesis
CG12344	Experimental, inferred	Chloride channel

Table 2.2: Genes contained in the P[acman] Dsh constructs

P[acman] CH322-187A06 (dsh)

Gene	Evidence	Functions
CG1597	Predicted	Oligosaccharide metabolism
CG11756	Predicted	Unknown
CG1738	Predicted	Unknown
CG11752	Predicted	Unknown
CG1737	Predicted	DNA binding
dsh	Experimental	Planar polarity
PTIP associated 1	Experimental	Histone methylaton
*hopscotch**	Experimental	*Drosophila* JAK homologue, protein tyrosine kinase

* *hop* runs in the reverse direction and CH322-187A06 truncates the first exon with no 5' UTR present. Would thus expect very low expression levels, if any. Overexpression causes tumours, lethality and patterning defects (Harrison et al, 1995), which are not seen in the P[acman] Dsh stocks, so any overexpression must be mild enough to be ignored.

Table 2.3: Genes contained in the P[acman] Stbm constructs

P[acman] Stbm

Gene	Evidence	Functions
SKI3/CG8777	Experimental	Nuclear-transcribed mRNA catabolism
stbm/Vang	Experimental	Planar polarity
Phosphomannomutase 45A	Predicted	Phosphohexomutase
tsunagi/Y14	Experimental	Oocyte patterning
Mystery 45a	Experimental, predicted	Neurogenesis, ribosome biogenesis

Table 2.4: Genes contained in the P[acman] Fmi constructs

P[acman] CH321-31B09 and P[acman] CH321-66D09 (Fmi/Stan)

Gene	Evidence	Functions
CG11883	Predicted	Nucleotide catabolism
G protein-coupled receptor kinase interacting ArfGAP	Experimental	Somatic muscle development, ARF GTPase activator
Elongator complex protein 2/Stip	Predicted	Proteasome assembly, transcription elongation
CG12934	Predicted	Unknown
stan	Experimental	Planar polarity
Rab3	Experimental	GTP-ase, synaptic vesicle trafficking
CG12309	Predicted	Carbonate dehydratase
CG12338	Predicted	D-aspartate oxidase
*CG7220**	Predicted	E2 ubiquitin ligase

*CG7220 is present in CH321-66D09 only.

Table 2.5: List of fluorescent constructs made during this project.

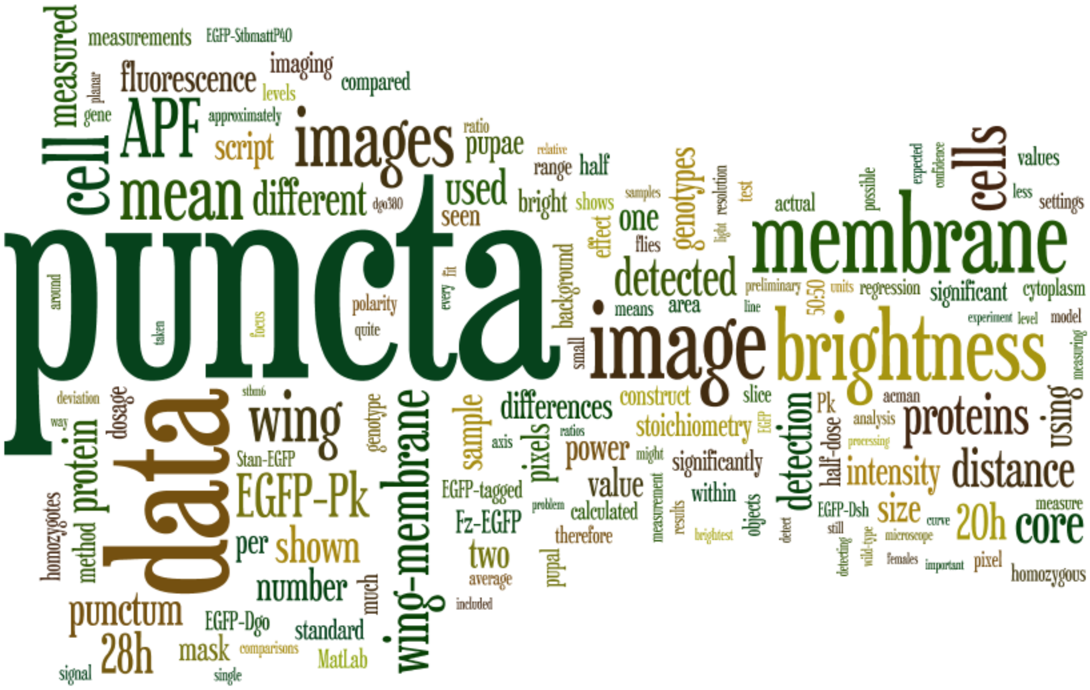
Constructs highlighted in **bold** were used for the stoichiometry experiments.

Construct	Expression	Phenotype /localisation	Notes
P[acman] EGFP-Stbm	**Yes**	**Normal**	
P[acman] Stbm-LoxP-EGFP	No	-	PBM blocked
P[acman] Stbm-EGFP-LoxP	Yes	Normal	PBM blocked
P[acman] EGFP-Dgo	**Yes**	**Normal**	
P[acman] Dgo-LoxP-EGFP	Yes	Subtly abnormal	Puncta slightly 'spotty' and discrete
P[acman] Dgo-EGFP-LoxP	Yes	Normal	
P[acman] EGFP-Dsh	**Yes**	**Normal**	**'Fuzzy' cytoplasmic signal is normal for Dsh**
P[acman] Dsh-LoxP-EGFP	No	-	
P[acman] Dsh-EGFP-LoxP	Yes	Abnormal	Bright cytoplasmic puncta
EGFP-Pk	**Yes**	**Normal**	
Fz-EGFP	**Yes**	**Normal**	**PBM blocked**
Stan-EGFP	**Yes**	**Normal**	**PBM blocked**
P[acman] Stan-EGFP	Yes	Normal	PBM blocked
P[acman] Cerulean-Stbm	No*	-	
P[acman] TagBFP-Stbm	Yes	Abnormal	Bright cytoplasmic puncta, poor asymmetry
P[acman] mCherry-Dgo	Yes	Normal	Weak fluorescence
P[acman] tdTomato-Dgo	No	-	
P[acman] Cerulean-Dgo	No*	-	
P[acman] TagBFP-Dgo	Yes	Abnormal	Very weak, misclocalised in some pupae
P[acman] mCherry-Dsh	Yes	Normal	Weak, best red construct
P[acman] Cerulean-Dsh	No*	-	
P[acman] TagBFP-Dsh	Yes	Normal	'Fuzzy' cytoplasmic signal is normal for Dsh

* The Nikon A1 confocal microscope used does not have an optimal laser line for excitation of Cerulean, so potentially these constructs do express and fluoresce, only not with our system.

Chapter 3

Measuring the stoichiometry



Contents

Figures

Tables

3.1 Introduction

This chapter describes the development of a robust method for measuring the relative stoichiometries of the core planar polarity proteins. This was accomplished *in vivo* in the *Drosophila melanogaster* pupal wing by measuring the brightness of images of fluorescently-tagged core proteins acquired with a confocal microscope. Imaging methodology and controls and validation of the image measurement technique are discussed in detail. Stoichiometry was measured at 28 hours after puparium formation (hAPF), when the core proteins are maximally asymmetric, and also at 20h APF, when the planar polarity system is in a state of flux.

The stoichiometry of the putative core complex will further understanding of the molecular mechanisms of planar polarity. While individual core complexes cannot be directly observed, the core proteins all colocalise in distinctive punctate structures in the apicolateral cell membrane. These puncta are the functional sites of planar polarity. It was therefore expected that the stoichiometry of the core proteins in puncta would be equal to that of the core protein complex. Little is known about interactions between the core proteins. Fmi is expected to form homodimers across cell boundaries, and is found on both sides of each cell, whereas the other core proteins localise to either the proximal or the distal side. The simplest core complex structure that fits the available data is therefore a 1:1 ratio of every core protein except for Fmi, which would be present at twice the level of the other five.

As discussed in more detail in the introduction and Chapter 2, the *Drosophila* pupal wing is an ideal tissue for studying planar polarity. At 28 hours after pupation at 25°C the core proteins have reached more or less maximal asymmetry prior to the formation of the wing trichomes. This should therefore be the ideal time to measure the stoichiometry of the core complex when it is both active and stable. By contrast, in the 20h APF wing the core planar polarity proteins are not asymmetrically localised, although some weak asymmetry may remain (Aigouy et al., 2010). At this stage there is much cell division, and the wing surface is buckled and blistered. It was decided to also look at the stoichiometry in the 20h APF wing, to see whether there were any differences compared to 28h APF.

3.2 Preparing pupae for pictures

It is necessary to collect prepupae and age them to the right stage, because pupae at 28h APF cannot be picked out from the general population. This time point is in the middle of pupal stage six, which spans approximately 19 to 32h APF (*Drosophila* Developmental Ontology[1]). Although no morphological alterations occur between these timepoints, the planar polarity system undergoes dramatic changes, so it is important to age the pupae accurately.

The beauty of gathering data from fly pupae is that it is very simple to collect many individuals at the same developmental stage. When third instar larvae are ready to pupate they enter the 'wandering' stage, where they climb out of the food and up the walls of the culture container. At the onset of pupation the anterior spiracles are everted and movement ceases. Pupae at this stage of development are termed 'prepupae'. Within 30 to 60 minutes the white larval cuticle begins to 'tan', and turns an orangey-brown colour. The larval cuticle is now a puparium, hence why pupal development times are given as 'hours after puparium formation' (hAPF). White prepupae can therefore be collected from the walls of the bottles or vials, transferred to a fresh vial, and incubated for the desired amount of time before imaging, to give a cohort of pupae all aged within 1 hour of each other.

To prepare the aged pupae for imaging, a small window is cut in the cuticle over the developing wing, taking care not to damage the wing. To reduce the chances of nicking the wing, the cut was made along the line between the two legs closest to the wing. The pupal cuticle is translucent, so it is easy to position the window correctly. A drop of viscous halocarbon oil is applied to the exposed wing. This has the dual function of preventing desiccation of the developing pupae and aiding the positioning of the pupae on a glass dish for imaging. The microscope used is of the inverted style, so the wing is placed against the base of the dish, which is in turn placed above the objective lens (**Figure 3.1 A**).

The light path is thus somewhat less than ideal, as the laser comes up through the lens, then the lens oil, passes through the glass dish and the halocarbon oil, then through the wing membrane and some thickness of aqueous wing fluid before hitting the actual wing epithelium (**Figure 3.1 B**). The emitted fluorescence is of course transmitted back

---

[1] http://bioportal.bioontology.org/ontologies/45239/?p=terms&conceptid=FBdv%3A00005353

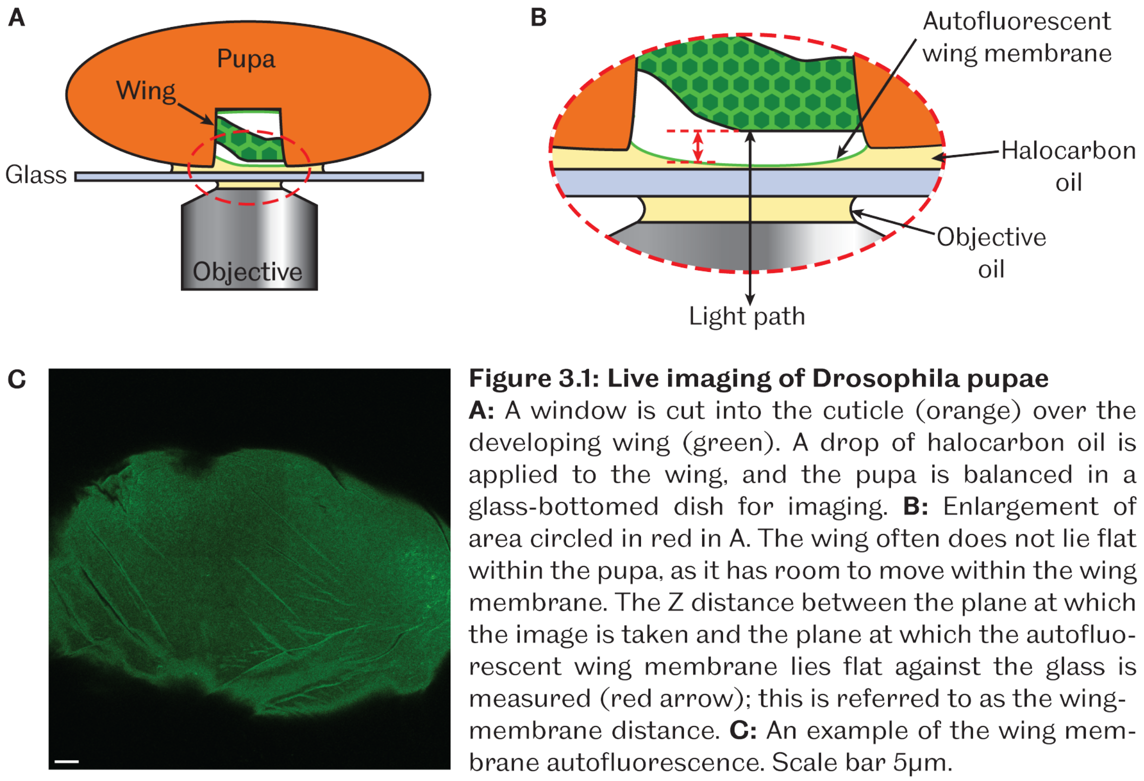


Figure 3.1: Live imaging of Drosophila pupae
A: A window is cut into the cuticle (orange) over the developing wing (green). A drop of halocarbon oil is applied to the wing, and the pupa is balanced in a glass-bottomed dish for imaging. **B:** Enlargement of area circled in red in A. The wing often does not lie flat within the pupa, as it has room to move within the wing membrane. The Z distance between the plane at which the image is taken and the plane at which the autofluorescent wing membrane lies flat against the glass is measured (red arrow); this is referred to as the wing-membrane distance. **C:** An example of the wing membrane autofluorescence. Scale bar 5μm.

down the same path. The optical densities of the various fluids in the light path are not optimal, and that of the wing fluid is not known, so spherical aberration will be introduced. The objective lens used was not fitted with a correction collar. This means that the optical resolution in the resulting images will not be quite as good as the capability of the instrument. However, it is extremely difficult to remove the developing wing from the pupa without destroying it, so this setup is unavoidable.

Note that the wing membrane is strongly autofluorescent, even in wild-type pupae. It can be easily seen at the settings used for gathering the experimental images. An example of this is shown in **Figure 3.1 C**.

3.3 Imaging settings and optimisation

All of the images needed for measuring the stoichiometry must be taken using exactly the same microscope settings, to ensure that measurements from different images can be directly compared. This means that the image capture parameters must be set so that the dimmest genotype can be still be seen at settings that do not over-saturate images of the brightest genotype, at settings which also avoid photobleaching the samples. This also means that the imaging settings will not be optimal for any one genotype. Several structures within the *Drosophila* pupa are strongly autofluorescent, so a confocal scanning laser microscope was a more appropriate instrument to use than a wide-field system.

It was also important to get the highest resolution images possible. Puncta are small objects, which often occur close to each other, so to get the most accurate data possible it is important to take images at maximum resolution. The theoretical limit of optical resolution is calculated as 0.61 multiplied by fluorescence emission wavelength, divided by the numerical aperture (NA) of the objective lens (reviewed in Waters, 2009). The emission wavelength of GFPmut1, which is the same molecule as EGFP, is 507nm (Cormack et al., 1996; Zhang et al., 1996). When imaging EGFP using a 1.4NA objective the theoretical limit of resolution is thus 221nm. This means that objects less than 200nm apart cannot physically be resolved using this combination of fluorophore, lens and imaging system. This value is valid for a system where every variable it at its optimal level. Due to the physical structure of the sample, discrepancies in optical densities between different substances in the light path and the noise inherent in the imaging equipment the maximum resolution will rarely be achieved in practice, and certainly not in this project.

Measuring the point spread function (PSF) of the microscope would have given an indication of the actual resolution achieved with this setup. The PSF can be measured by taking images of fluorescent beads on a slide and inspecting the resulting brightness profile of the beads in the image. The beads must be smaller than the limit of resolution – 100nm beads are typically used for confocal microscopes. The bead acts as a point source of light, so the profile of its fluorescence shows how the microscope records sub-resolution objects. The actual resolution achieved is given by taking the width, in pixels or nm, of the peak at half maximum brightness. This distance tells you how close together two objects can be in your image before they cannot be resolved.

Due to the complex nature of the light path in the experimental samples it was decided not to attempt to measure the PSF. It would be extremely difficult to get the beads into the same imaging conditions as the pupal wing, which is in a fluid of unknown composition and variable depth, inside a membrane, which is bathed in oil. The differing optical properties of these substances would all add up to a considerable difference in image quality compared to simple beads on a slide, so the measured PSF would be unlikely to be representative of the actual imaging conditions.

The sampling rate describes the size of the pixels relative to the structure being studied. Under-sampling results in sub-optimal resolution, whereas over-sampling gives 'empty' data – the number of pixels is increased with no associated increase in the actual data present in the image. The camera should be set so that the pixel size is approximately half that of the resolution limit. This means that the smallest detectable object in the image will always be sampled by at least four pixels (Waters, 2009). Using a 60x objective at 3x zoom on the Nikon A1 microscope gave good images of the pupal wing cells. At 1024 x 1024 pixels per image, the pixel size was 70nm. This represents approximately 3x over-sampling, which is acceptable. The resolution of the images would not be limited by pixel size.

Images of live pupal wings were collected using a Nikon A1 confocal microscope according to the description in the Materials and Methods. The laser power and gain were adjusted so that the dimmest genotypes were just visible, and the brightest genotypes were not over-saturating the detector. The settings selected gave an acceptable balance between signal-to-noise ratio, absolute image brightness, photobleaching and resolution. It was possible to do 2x averaging without significant photobleaching, which helped to reduce noise in the images.

3.4 Background measurements

The background signal of the images had to be measured so that it could be subtracted from the data. There are two kinds of background to consider. Firstly, there is the noise introduced by the microscopy machinery. This is affected by the microscope imaging settings, including the detector gain and the offset, as well as random fluctuations in the electronics of the PMT. Images to measure this noise can be acquired by capturing an image while no laser light is reaching the sample. This was achieved by reducing the laser transmission to zero after taking an image of a pupa, and was repeated approximately every 15 to 20 minutes throughout each imaging session. The focus was not changed from the plane at which the pupal wing image was taken, so the background images are taken at a range of focal depths. Since no light should have been actually reaching the detector, this should not affect the results.

The mean intensity of each background image was measured (**Figure 3.2 A**). The machine background would always increase by one or two intensity units per hour

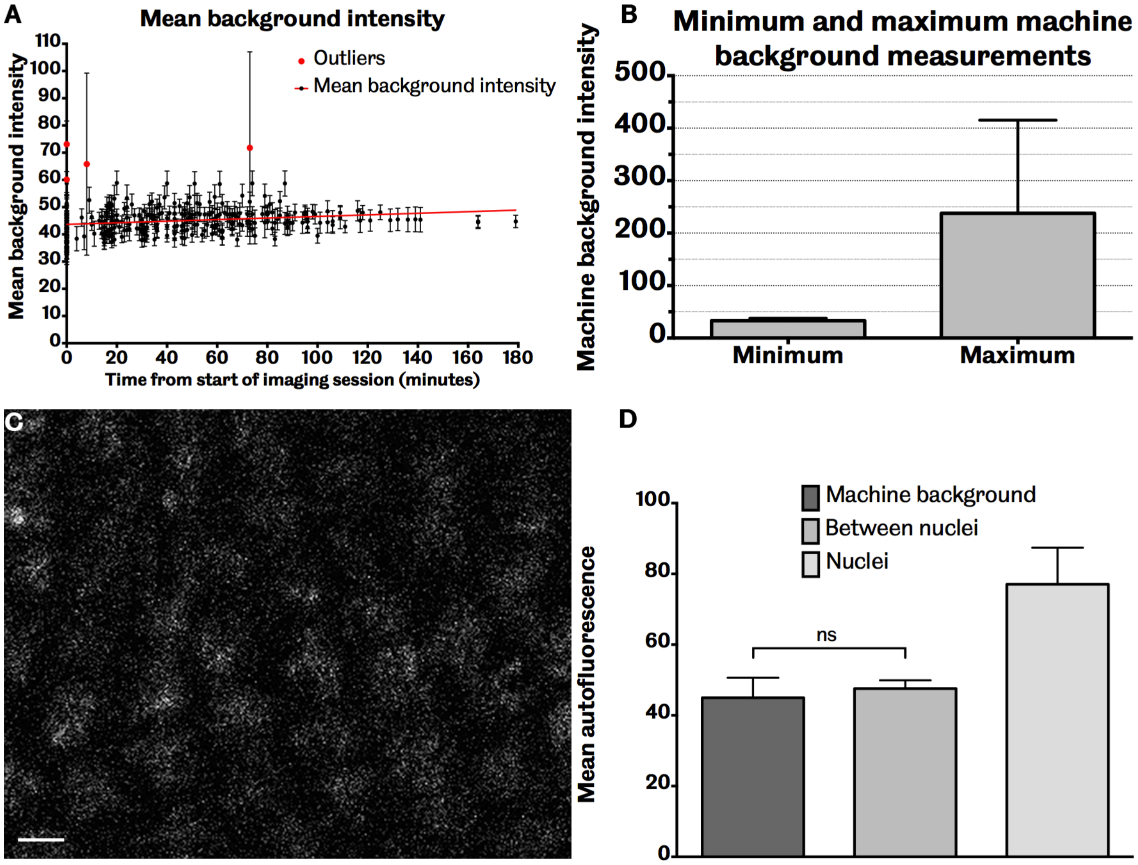


Figure 3.2: Measurements of machine background and autofluorescence.
A: The mean brightness of the machine noise. Bet fit line generated by linear regression. 285 images were measured. **B:** Comparison of the mean minimum intensity and the mean maximum intensity of the machine noise. **C:** An example of nuclear autofluorescence in a wild-type 28h APF pupa, taken at the same imaging settings used for the stoichiometry experiments. Brightness and contrast have been greatly enhanced for visibility. Scale bar 5µm. **D:** Comparison of mean machine noise, mean autofluorescence of areas between nuclei and mean autofluorescence of nuclei. Measurements between nuclei are the average brightness of 7 pixel wide lines drawn between the brighter nuclei. Measurements of nuclei are the average brightness of circular selections drawn in the centre of the nuclear patches.

All error bars are standard deviations.

throughout the imaging session if the microscope was turned on at the beginning of the session. If the instrument had already been running for some time before the start of the session, then the background was usually stable. However, this increase is of negligible importance, as the detector is capable of registering 4096 intensity levels and this full range is used when imaging the pupae. The four outliers indicated by large red symbols represent occasions when the brightfield lamp was accidentally left on while taking the background image. The mean background signal was 45 intensity units, with a standard deviation of 5 units. The best fit line was calculated by linear regression.

Figure 3.2 B shows the average minimum and maximum pixel intensity recorded in each background image. The brightest pixel in a background image averaged at 200 units, with occasional occurrences of up to 2,100 units. All of these values are very large when compared to the standard deviation of the mean brightness, which is only 5 units. This indicates that very few pixels in each image fall into this range. When coupled with the fact that puncta occupy a very small proportion of the area of each cell, it is very unlikely that these rare bright background pixels would coincide with a punctum and thus have an effect on the measured data. The dimmest background pixel value measured was 18 brightness units. The offset setting on the microscope was deliberately set so that pixels of zero brightness would never occur; this was to ensure that no data was being missed by falling outside of the lower limit of the detector range.

Another source of background signal could be autofluorescence from the wing tissue itself. It has already been shown that the wing membrane is strongly autofluorescent when excited by a 488nm laser (**Figure 3.1 C**). When testing new fly lines it was possible to locate the wing tissue in non-EGFP-expressing strains by faint autofluorescence from the cell nuclei. An example of this, taken at the same imaging settings used for gathering the experimental image, but shown with enhanced brightness and contrast, is shown in **Figure 3.2 C**. The mean intensities of the nuclear fluorescence and the darker areas between the nuclei are shown in **Figure 3.2 D**. The nuclear signal was significantly greater than both the machine background and the inter-nuclear fluorescence (P <0.0001). The inter-nuclear signal was not significantly different from the machine background signal (P = 0.472) (Ordinary one-way ANOVA, F(2, 296) = 107.6, P <0.0001, with Tukey’s test for multiple comparisons) Please see below for a description of ANOVA reporting.

Interpretation of ANOVA results: The ANalysis Of VAriables method determines whether there is a statistically significant difference between the means and associated variances of grouped data. ANOVA results are typically reported as F(*a*, *b*) = *x*, where *a* is the degrees of freedom between the groups and *b* is the degrees of freedom within the groups. *x*, the F ratio, is calculated from these degrees of freedom and the associated sums-of-squares from the between- and within-group analysis. A larger F value indicates greater variance between groups than would be expected by chance. The statistical software used takes the F value along with *a* and *b* to look up the P value associated with these values. Reporting F, *a* and *b* along with the P value gives a fuller description of the results of the ANOVA.

While the nuclei are a few microns basal to the level of core protein localisation, this should be a fair representation of membrane autofluorescence. It was therefore concluded that autofluorescence of the cell membranes would not make a significant contribution to the EGFP intensities measured using these microscope settings. 45 intensity units were subtracted from all EGFP intensity measurements to compensate for the machine background signal.

3.5 Determining puncta detection parameters

Several properties of puncta had to be determined before the puncta measurements could be automated. Knowledge of their numbers, size, shape and fluorescence profile was required.

3.5.1 Puncta numbers and sizes

The preliminary images acquired during imaging optimization were used for this purpose, by manually selecting puncta in ImageJ and measuring them. Three non-adjacent cells from each of six images were measured, along with the membrane and the cytoplasm. The mean brightness of the cells varied from image to image, but averaging across six images should have minimised this effect. While manual selection of puncta will introduce bias, at this point human judgement was the most accurate way of detecting them. This ensured that every object that could possibly be a real punctum was detected, which is important when establishing the detection parameters.

Five genotypes were measured: P[acman] EGFP-Dgo *dgo*^{380}, P[acman] Dgo-LoxP-EGFP *dgo*^{380} and EGFP-Pk, EGFP-Pk / + and EGFP-Pk / *pk-sple*^{13}. The Pk dosages were included to check whether the size and number of puncta changed with gene dosage. In EGFP-Pk / + (the ‘50:50 dosage’) there should be the usual amount of Pk expressed, but only half of it is EGFP-tagged, so the other half is invisible. In EGFP-Pk / *pk-sple*^{13} (the ‘half-dose’) there is only one copy of the Pk gene, so there might be half as much protein in the cells.

There were an average of 4±1.18 puncta per cell, with some small significant differences seen between genotypes (ordinary one-way ANOVA with Tukey’s multiple comparison test, F(4, 85) = 4.471, P = 0.0025) (**Figure 3.3 A**). However, this is probably more due to the small number of cells measured (18 per genotype) and the narrow, discrete range of the Y axis than to any real significant differences. At the settings used for the stoichiometry experiment a single image can contain anywhere from 20 to 200 cells, so this sample size is not representative of the final dataset. No significant differences were observed between the different EGFP-Pk gene dosages. Either there is no reduction in protein level in the half-dose, or any reduction is not sufficient to affect the number of puncta (perhaps the cellular Pk pool is redistributed to compensate), or the cells are making the same number of puncta but they are smaller than usual.

However, no significant differences in puncta area were observed in any of the genotypes measured (Ordinary one-way ANOVA with Tukey's multiple comparison test, F(4, 355) = 2.344, P = 0.0545) (**Figure 3.3 B**). In the case of the EGFP-Pk half-dose, this could mean that the same amount of protein gets into the membrane as in EGFP-Pk homozygotes, implying a redistribution of the available cellular Pk pool, or that puncta are the same size in the half-dose but somehow less 'dense', which would be seen as a reduction in brightness. Under normal circumstances Pk is continually sent to the proteosome for degradation (Strutt et al., 2013), so this could also represent a reduction in protein turnover. The average punctum area was 0.33±0.15 μm^2. The distribution for most of the genotypes was quite skewed, with most puncta clustering around the mean but a few much larger ones too. The larger ones may represent two puncta in close proximity that were not resolved in the image. The puncta detection parameters in the final method would have to allow for this.

3.5.2 Puncta brightness

The most appropriate measurement of puncta brightness also had to be determined. The edges of cells were traced around with a 10-pixel wide line and the resulting fluorescence profiles plotted (three examples shown in **Figure 3.3 C', C"** and **C"'**). Each graph shows data from a single cell's membrane and the membranes shared with adjacent cells. The puncta are visible as sharp spikes in the traces. This suggested that the peak intensity might be a suitable measure. However, the sharp spikes imply that the peak value is not very representative of the general brightness of the punctum. Also, some peaks have flatter tops than others, which could complicate detection.

The data shown in **Figure 3.3 C** are averaged from 10 pixels for each point, so the plots look smoother than they actually are. **Figure 3.3 D** shows the raw image of the cell measured in **Figure 3.3 C'**, false-coloured for clarity. It can be seen that the raw image is quite noisy, even using 2x averaging when capturing the image. This is probably

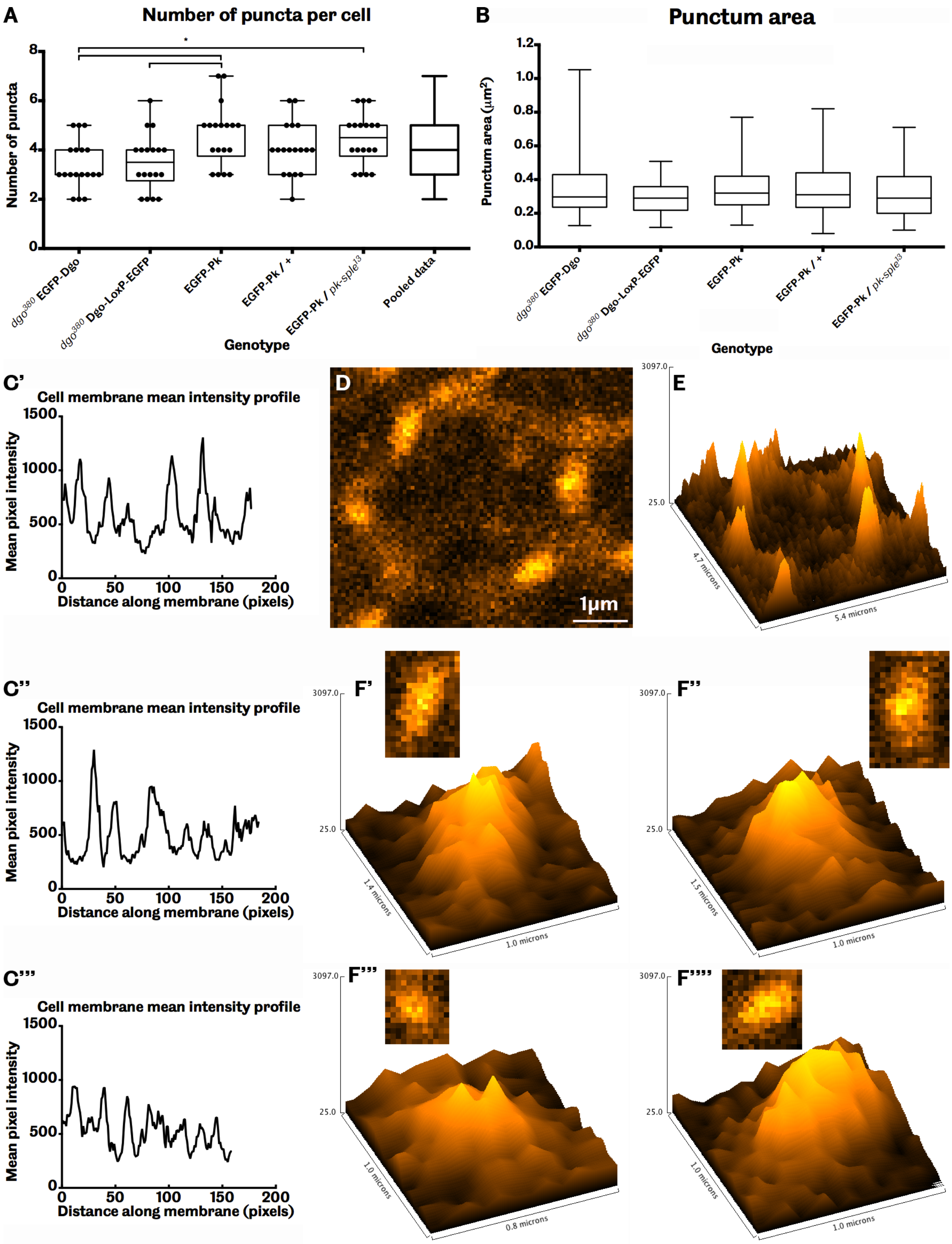


Figure 3.3: Puncta properties and profiles

A: The number of puncta per cell and the mean area were measured by manually selecting and drawing around the features in ImageJ. The genotypes measured have an average of 4±1.18 puncta per cell. Some small significant differences were seen, but the sample size is small (18 cells per genotype) (1-way ANOVA with Tukey's test for multiple comparisons, F(4, 85) = 4.471, P = 0.0025). **B:** The average size of a punctum is 0.33±0.15 µm2. No significant differences were observed between the genotypes tested (1-way ANOVA with Tukey's test for multiple comparisons, F(4, 355) = 2.344, P = 0.0545). **C’, C”, C”’:** Traces were generated by tracing around cell membranes using a 10-pixel wide line. The cell in C’ is the same as in D. Puncta are visible as spikes. Local background around each spike is variable. **D:** Appearance of a Fz-EGFP cell, false-coloured to highlight differences in pixel intensity. **E:** Surface intensity plot of the cell in D. **F’-F””:** Enlargements and surface plots of the four bright puncta in the cell shown in D.

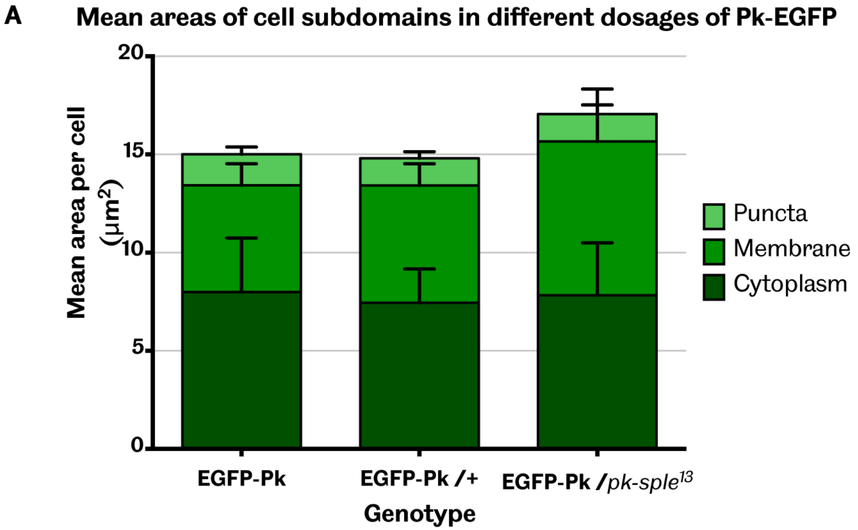
because the puncta are quite small, so even at a pixel size of 70nm and using an appropriate oversampling rate there are still not many pixels per punctum. Coupled with sub-optimal resolution and all the other sources of noise, this produces a rather jagged surface profile (**Figure 3.3 E**). Additionally, some of the puncta measured had two peaks at their centres (**Figure 3.3 F'–F''''**). This further indicated that taking a simple maximum intensity from each detected punctum would not be very representative of the actual punctum fluorescence. This was unfortunate, because detecting a single brightest pixel is much easier to implement than more complicated measurements such as mean or median brightness.

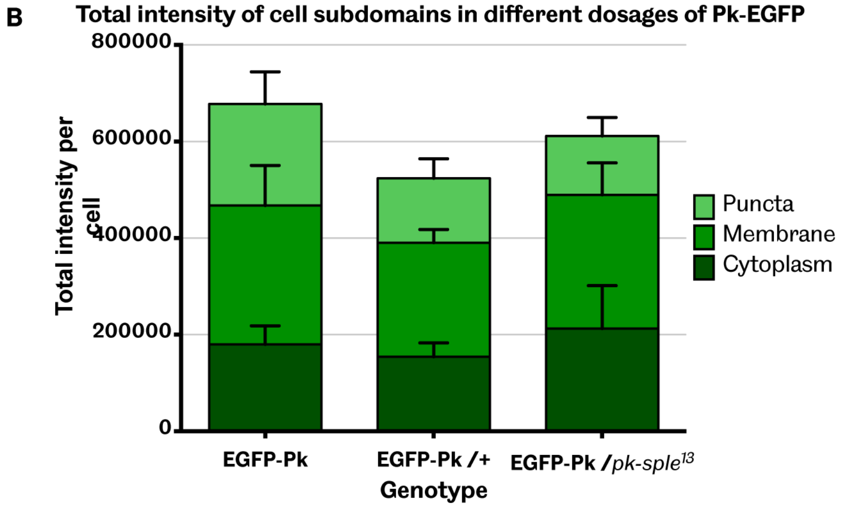
From the surface plots in **Figure 3.3 F** it was decided that the mean brightness of a punctum would give the most accurate representation of punctum fluorescence. The surface plots and the line profiles in **Figure 3.3 C** showed that the brightest pixels tended to lie at the centre of puncta. This meant that finding the bright centres of puncta would be good starting point for automated puncta detection.

3.5.3 Preliminary data results

The preliminary data were also used to try to see whether the Pk protein was redistributing in the EGFP-Pk / *pk-sple*^{13} half dose. The areas and brightness of the puncta, membrane and cytoplasm (the interior area of the cell excluding the cell membrane) were measured and compared for EGFP-Pk, EGFP-Pk / + and EGFP-Pk / *pk-sple*^{13}. As expected, the areas of the puncta, membrane and cytoplasm were not significantly affected by EGFP-Pk dosage (2-way ANOVA, F(2, 153) = 3.016, P = 0.0519) (**Figure 3.4 A**). This shows that if there are any changes in protein levels in membrane and puncta in different gene dosages it does not increase the observed areas of these cell subdomains.

As expected, significant differences were seen in the brightness of the different genotypes (2-way ANOVA, F(2, 153) = 10.83, P <0.0001) (**Figure 3.4 B**). This effect is largely restricted to differences in puncta brightness (Tukey's test for multiple comparisons: EGFP-Pk vs EGFP-Pk / +, P = 0.0003 and EGFP-Pk vs EGFP-Pk / *pk-sple*^{13}, P <0.0001), with a slight effect on the membrane and cytoplasm (Tukey's test for multiple comparisons: membrane EFGP-Pk vs EGFP-PK / +, P = 0.0216; cytoplasm EGFP-Pk / + vs EFGP-Pk / *pk-sple*^{13}, P = 0.0075). The puncta brightness was approximately halved in the





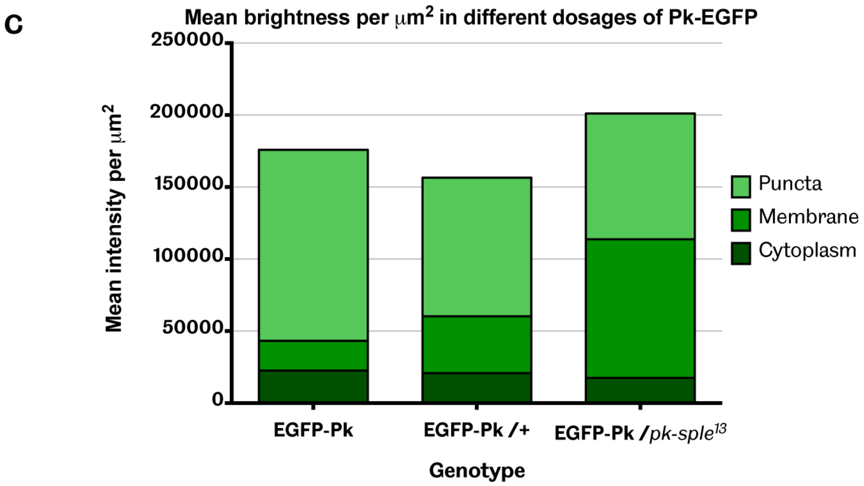


Figure 3.4: The relative contributions of puncta, membrane and cytoplasm to the area and fluorescence of the cell in different dosages of EGFP-Pk and endogenous Pk
Images of 28h AFP pupal wings were manually measured by drawing around the features of interest in ImageJ. Puncta data represents the mean of all puncta in the cell. 3 cells were measured in each of 6 images, for a total of 18 cells per genotype. EGFP-Pk flies carry two copies of EGFP-Pk, EGFP-Pk /+ have one copy of EGFP-Pk and one copy of the wild-type gene, and EGFP-Pk /*pk-sple*^{13} flies have only one copy of EGFP-Pk with no endogenous protein. **A:** The mean areas of the puncta, membrane and cytoplasm are not significantly affected by Pk dosage (2-way ANOVA, F(2, 153) = 3.016, P = 0.0519). **B:** As expected, genotype has a significant effect on the brightness of the cells (2-way ANOVA, F(2, 153) = 10.83, P <0.0001) but this effect is largely restricted to the differences in puncta brightness (Tukey's multiple comparisons test; EGFP-Pk vs EGFP-Pk /+, P = 0.0003 and EGFP-Pk vs EGFP-Pk /*pk-sple*^{13}, P < 0.0001) with a slight effect on the membrane and cytoplasm (Tukey's multiple comparisons test; membrane EGFP-Pk vs EGFP-Pk /+, P = 0.0216: cytoplasm EGFP-Pk /+ vs EGFP-Pk /*pk-sple*^{13}, P = 0.0075). All other comparisons in this test were not significant. **C:** Normalsing brightness to unit area shows that the majority of the fluorescence is concentrated in the puncta, the smallest cellular subdomain by area.

EGFP-Pk / + cells compared to the EGFP-Pk homozygotes – there should be the same amount of protein, but in the 50:50 dosage only half of it is visible. Similarly, in the half-dose puncta where one would expect half the amount of protein to be present, the brightness of the puncta was also approximately halved compared to the EGFP-Pk homozygotes. Although the size of the half-dose puncta appears to be the same as the homozygotes' puncta, it should be noted that many puncta are close to the limit of resolution. This means that puncta area measurements are necessarily upper estimates – puncta smaller than the limit of resolution will appear larger than they actually are. Additionally, due to the nature of the samples and the sub-optimal light path the microscope would not have achieved maximum resolution either.

When the brightness is normalised by area, the distribution of fluorescence in the cells becomes clear (**Figure 3.4 C**). The vast majority of the EGFP-tagged Pk is found in the puncta, the smallest cellular subdomain, with hardly any in the cytoplasm. There is a much greater proportion of fluorescence in the membrane (compared to the cytoplasm and puncta) of the half-dose images compared to both EGFP-Pk homozygotes and EGFP-Pk / +, suggesting that additional protein was drafted in from the cytoplasm to compensate for the reduced Pk levels.

3.6 The wing-membrane distance problem

While analysing the preliminary images it was noticed that puncta in some images were more than twice as bright as puncta in other images of the same genotype. In fact, there was often more difference in brightness between images of the same genotype than between genotypes (**Figure 3.5 A**). While allowing for some individual differences in protein expression between pupae, it was expected that all puncta of a particular genotype should have more or less the same brightness. This was a problem, because such noisy data would make it more difficult to accurately measure any differences between the different core proteins.

On closer examination, there was found to be much variation in the Z-plane at which each image was taken. This meant that the distance the emitted light travelled through the sample from the wing surface to the lens was very variable, with an accompanying variable loss of signal. The culprit was eventually identified as variations in the distance

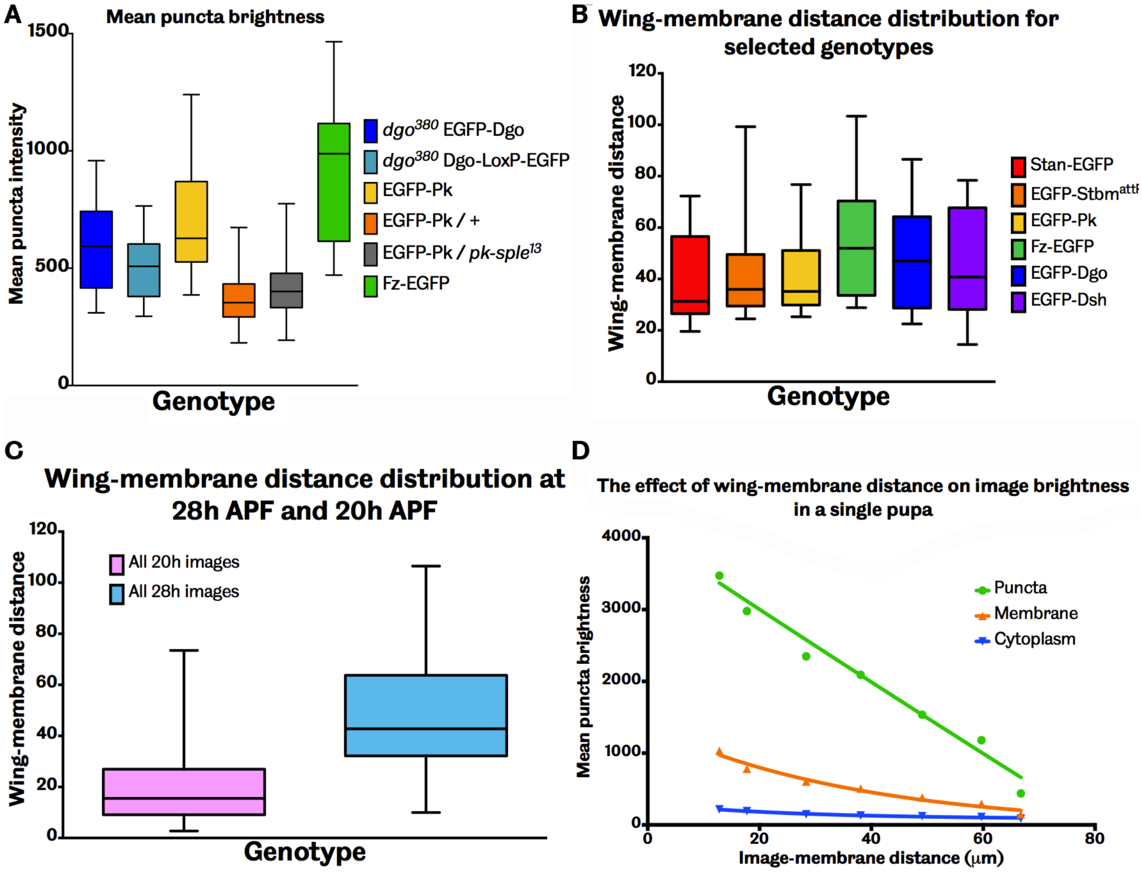


Figure 3.5: Asessment of wing-membrane distances
A: The distribution of the mean puncta fluorescence in the genotypes used for preliminary imaging. This data is taken from the final dataset. **B:** The distribution of wing-membrane distances for the homozygous stocks used to calculate the stoichiometry. While the means are not significantly different from each other (ordinary one-way ANOVA, F(5, 115) = 1.709, P = 0.138), the ranges and interquartile ranges are large. **C:** Comparison of the wing-membrane distances for every image measured at 28h APF (453 images) and 20h APF (194 images) shows a large significant difference (two-tailed t test, P<0.0001). **D:** Measurements of mean puncta, membrane and cytoplasm brightness from a single pupa with a wing spanning a wide range of wing-membrane distances. Curves were fitted by nonlinear regression using a one-phase decay model with no constraints.

between the wing membrane and the wing epithelium. The wing membrane fits quite snugly around the wing in the proximal-distal and anterior-posterior directions. However, the wing has much scope for movement within the membrane in the medial-lateral direction, which is roughly perpendicular to the plane of the wing epithelium. Even when the wing membrane is flat against the glass dish, the actual wing tissue (and hence the plane at which the image is taken) can be anywhere from 10 to 100μm away from the membrane. Unfortunately when imaging the pupae the light path is roughly parallel to medial-lateral direction, so the brightness of the images suffers from the full effect of this variation in wing-membrane distance.

The range of wing-membrane distances for each homozygous line used for measuring the stoichiometry is shown in **Figure 3.5 B**. While the differences are not significant (ordinary one-way ANOVA, F(5, 115 = 1.709, P = 0.138), the ranges and interquartile ranges are large. This means that there will be much variation in the measured puncta brightness, which will decrease the accuracy of the final measurements unless this source of variation can be compensated for. This will make it more difficult to see any significant differences in mean puncta brightness. Additionally, the mean wing-membrane distance for all images taken at 20h APF is significantly different to that at 28h APF (**Figure 3.5 C**) (two-tailed t test, P < 0.0001). If these two datasets are to be compared then the differences in wing-membrane distance must be accounted for.

Numerous attempts to change dissection and mounting techniques to avoid this problem were unsuccessful, meaning that the wing-membrane distance would have to be measured for each image and compensated for in the final analysis. Fortunately the wing membrane is both strongly autofluorescent (**Figure 3.1 C**) and very thin, which greatly facilitated accurate measurement of its Z-position. The effect of the wing-membrane distance on image brightness was confirmed by taking multiple non-overlapping images of a single pupa whose wing fortuitously spanned a wide range of Z-depths (**Figure 3.1D**). Curves were fitted to the points by unconstrained nonlinear regression using a one-phase decay model.

3.7 *A priori* power calculations

An interesting aspect of the statistics of this experiment was the expectation that in most cases the levels of the core proteins would be the same. The working hypothesis was that the core proteins adopted the simplest ratios that fit the available data, that is, two molecules of Fmi to form a homodimer, accompanied by one molecule each of the other five proteins. Usually, statistical tests are geared to assess *differences* between samples. For this project, the relevant value is the probability of *no* significant differences between samples, and it is important to have some understanding of how confident one can be when making these kinds of conclusions.

The familiar P value gives the probability of getting the observed results if the null hypothesis were true (“How likely is it that these results are false?”) – with a low P value indicating that the results are probably correct. However, the important question for this experiment is “If there was a difference between these two samples, would I have seen it?” This probability is described by the ‘power‘ of the experiment, which is formally defined as the probability of not making a type 2 error. A type 2 error occurs when a false null hypothesis is not rejected, concluding that there isn't an effect when there really is. The power value gives therefore gives the probability of correctly rejecting a false null hypothesis. A high-powered experiment would thus have a high chance of rejecting the null hypothesis if there was indeed a real difference between the two samples. This means that if the null hypothesis is *not* rejected in such an experiment one can be quite confident (depending on the alpha error probability chosen) in concluding that the two samples are not significantly different from each other.

Like the usual P value, the power of an experiment is related to sample size. For any given pair of experimental populations, a larger sample size will give a smaller p value, and a greater power. It was decided to use the concept of experimental power with the preliminary data to try to find an appropriate sample size for the stoichiometry measurements. This was done by using G*Power, a free power calculation application (Faul et al., 2007), to estimate answers to these questions: 1) how many pupae should be imaged to confidently detect a particular stoichiometric ratio? Conversely, 2) what is the highest stoichiometric ratio that could be detected at a particular sample size?

Effect size is a measurement used to calculate the experimental power; it is a way of quantifying the actual magnitude of the difference between two values calculated from two populations without considering sample size. It is equivalent to a Z-score from a pair of normal distributions, ie. a random value taken from sample 1 is, on average, [effect size] standard deviations greater than a random value from sample 2. It can be thought of as measure of the overlap between two distributions. The estimate of effect size used was Cohen's *d*, which is the mean of sample 1 minus the mean of sample 2, divided by the standard deviation of the two samples. This makes the assumption that the standard deviations of both samples is the same, which may not be the case in the real data.

The preliminary data passed normality tests, which means that the calculated figures should be valid for the data. The standard deviations of the average maximum puncta brightness per genotype ranged from 100 to 300 units, so a standard deviation of 200 was used for the *a priori* power calculations. A desired power of 0.95 and an α error probability of 0.05 (the same as a significance threshold for P) were set in G*Power, and a range of combinations of hypothetical results were tested. From this input G*Power calculated the achieved experimental power and the required sample size. The results of these calculations are shown in **Table 3.1** (at end of chapter). It can be seen that stoichiometric ratios of up to 3:4 can be reliably detected using a sample size of 20 pupae. 20 pupae per genotype is an achievable goal, and it was unclear whether stoichiometries of greater than 3:4 would represent a meaningful stoichiometry, or just noise in the measurements. It was consequently decided to gather data from 20 pupae from every genotype.

3.8 Using MatLab to process and measure images

Existing software was found to be unsatisfactory for detecting puncta and measuring the pupal wing images, as none of the programs tested (NIH's ImageJ, PerkinElmer's Volocity and Imaris' BitPlane) permitted sufficient control of detection and measurement parameters. In the case of ImageJ, extensive knowledge of Java would have been required to write a custom puncta detection and measurement plugin. It was consequently necessary to develop a new method to extract data from the images.

MatLab is a widely-used matrix-based programming platform, which allows users to manipulate large data sets with a wide range of functions. By combining the functions in different ways one can achieve almost any complex computational task in a controllable and reproducible manner. The software includes an image processing toolbox, which adds functions that are particularly useful for image manipulation applications. The other programs that I considered using did not permit such flexibility and control of the processing operations, so it was decided to write a custom-made MatLab script to detect and measure puncta. During the course of the preliminary imaging it became clear that it is very difficult to make the wings lie flat within the pupae, which meant that most images contained both in-focus and out-of-focus cells. The need to easily exclude out-of-focus cells from the analysis was another important reason for developing a bespoke image measurement method.

In writing the MatLab script I relied heavily on the image processing principles outlined in Russ (2011). The raw confocal images are noisy and have poor contrast, particularly in dimmer genotypes. The images would thus require some manipulation before puncta could be detected, although all the actual measurements would be taken from the raw data. These processing steps are shown in **Figure 3.6**. The raw image is loaded into MatLab using the LOCI BioFormats plugin (Linkert et al., 2010), which allows most of the commonly used image processing programs to import confocal images in their native file format.

3.8.1 Pre-processing images prior to detecting cells

The raw image was often dim, with low contrast and much noise (**Figure 3.6 A**). This is an unfortunate consequence of having to use the same imaging settings for a set of constructs of varying brightness. First, the contrast was enhanced to saturate the brightest and dimmest 0.1% of pixels (**Figure 3.6 B**). This effectively normalised the overall image brightness, shifting the puncta into the same region of the image histogram regardless of the original image brightness. This allows detection of puncta over a wide range of image intensities. Noise was then reduced using two filters. The first was a Gaussian averaging filter (**Figure 3.6 C**), which replaces the brightness of each pixel with the mean brightness of the pixels in its surrounding 5x5 pixel neighbourhood, weighted using the Gaussian profile shown in **Figure 3.6 D**. The weights of the Gaussian filter total 1, so overall image brightness is preserved. Gaussian averaging filters give a good trade-off between noise removal and minimal blurring of the image. This was followed by a median filter, which replaces the value of each pixel with the median brightness of the surrounding pixels, again using a 5x5 pixel neighbourhood (**Figure 3.6 E**). The median filter further reduces noise while preserving sharp changes in brightness, such as the transition from cytoplasm to membrane. This resulted in the processed image from which the cells and puncta would be detected.

3.8.2 Detecting the cells

Once the images had been processed, the strategy was to define the pixels comprising the cell membranes and the cytoplasm, and then detect only those puncta that lie within the membrane mask. Attempts to detect cell outlines using my processed images and MatLab were unsatisfactory. This was due in part to the patchy, asymmetric and discontinuous nature of the membrane fluorescence. Fortunately I was able to use a custom-written program, with the kind permission of the author (Aigouy et al., 2010). B. Aigouy developed PackingAnalyzer to analyse cell packing and polarity. While it struggled with raw images, PackingAnalyzer was able to segment the cells very well from the pre-processed images (**Figure 3.6 F**), and most incorrect lines can be

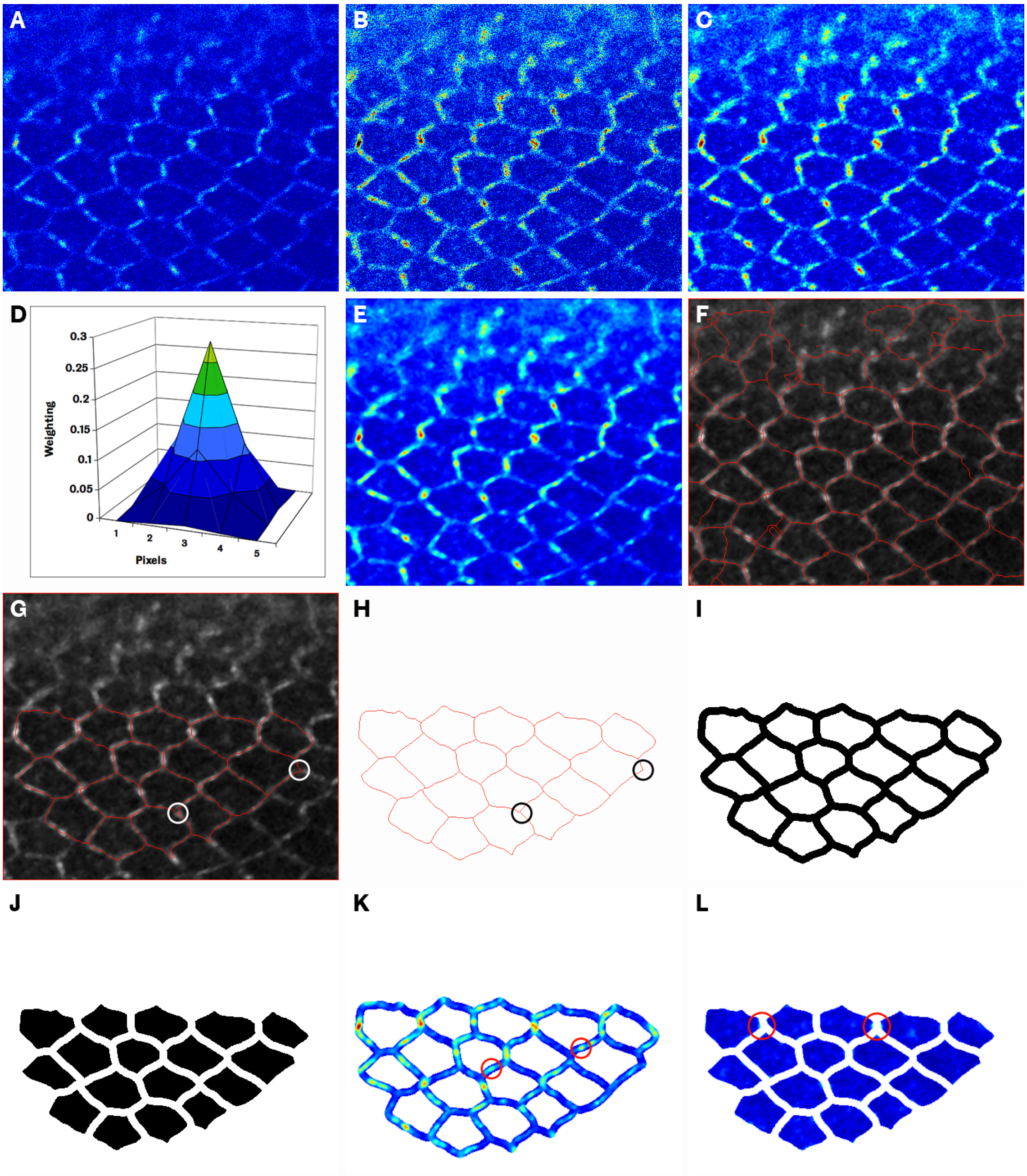


Figure 3.6: Use of MatLab to process images prior to puncta detection
A: The raw image is often dim with much noise and poor contrast. **B:** The contrast is adjusted to saturate the dimmest and brightest 0.1% of pixels. **C:** A Gaussian averaging filter is applied to reduce noise without blurring the details of the image. **D:** The profile of the Gaussian averaging filter used. Weights total 1, so overall image brightness is preserved. **E:** A median filter is applied to further reduce noise which preserving edges. **F:** The initial cell outlines generated by PackingAnalyzer. **G:** The manually-corrected cell outlines produced by PackingAnalyzer still contain some inaccuracies (circled). **H:** The refined cell skeleton after dilation and thinning is a better representation of the image (compare circles with those in G). **I:** The cell outline is dilated to a width of 10 pixels to form the membrane mask for puncta detection. **J:** The spaces inside the membrane mask are used as the cytoplasm mask. **K:** Applying the membrane mask to image E shows that the mask is a little over-generous. Dark blue can be seen around the edges of some puncta (red circles). **L:** Applying the cytoplasm mask to image E shows that bright fluorescence from the membrane is rarely included in the detected cytoplasm (red circles).

In colour images, dark blue represents zero brightness and dark red represents maximum brightness.

automatically removed by a series of adjustments within the program. Nevertheless, significant manual correction was still required to exclude out-of-focus cells from the analysis (**Figure 3.6 G**). The increased contrast of the pre-processed images simplified this step, as it was generally obvious which cells were in and out of focus. For further processing it was also important to remove any cells touching the edge of the image.

After generating the cell outline in PackingAnalyzer it was imported into MatLab for further processing. The outline was fattened to a width of 10 pixels and then thinned back down to a single pixel width; this had the effect of straightening the lines and pulling any distorted vertices closer to their real location (**Figure 3.6 H**).

The cell outline was then fattened again to a width of 10 pixels to form the membrane mask (**Figure 3.6 I**). The spaces inside the membrane mask were used as the cytoplasm mask (**Figure 3.6 J**). As can be seen from the overlay in **Figure 3.6 K**, the membrane mask is a little over-generous – dark blue can be seen around the edges of the puncta (red circles). This is intentional, to allow for imprecise centring of the membrane mask on the actual membrane signal and also for detection of unusually large or off-centre puncta. The overlay in **Figure 3.6 L** shows that bright membrane fluorescence is rarely included in the cytoplasm mask (red circles). Now the puncta can be detected.

3.8.3 Puncta detection method

From the preliminary data, it was decided to use mean puncta intensity as the primary measure of puncta brightness. The trouble with measuring the mean intensity is that it is highly dependent on the puncta detection method. If only the very centre of a punctum is detected, the dimmer pixels around the edge, which would be considered part of the punctum by a human observer, would be excluded. This would result in an artificially high average brightness. Conversely, if the puncta detection method is too generous there is a risk of including membrane or cytoplasm pixels in the punctum, which would lower the measured mean brightness. However, the final result from the puncta measurements will be the *relative* stoichiometries of the core proteins. This means that it does not matter so much whether the measured puncta brightness values are as accurate as possible, but it *is* crucial that they are consistent from image to image and genotype to genotype. On the other hand, the measured values must still be accurate enough not to distort the calculated stoichiometric ratios from their actual values.

When looking at images of fluorescently-tagged or immunostained core proteins, there is no sharp delineation between the puncta and the baseline membrane signal. This raises the problem of how to define the edge of a punctum. The best way would be to use the point at which the punctum intensity meets the average non-punctum membrane intensity. This could be achieved by measuring the average fluorescence of the membrane, excluding the puncta, and using this value as the cut-off intensity threshold for the edges of puncta. This is possible when selecting puncta by eye, as one can easily identify the puncta and assess the local membrane brightness. Unfortunately, this method only works computationally if the puncta have already been accurately defined and detected. The membrane fluorescence also fluctuates locally, and not all puncta have the same brightness, which further complicates matters. This can be seen in **Figure 3.3 C** – the fluorescence around the base of the puncta spikes varies both within and between cells.

Another way to detect puncta would be to look for local maxima in the membrane. This should reliably detect the bright centres of puncta, but does not solve the problem of how to define the edges of the detected objects, ie. when to stop detecting. A method based on detection of local maxima was implemented in the MatLab script.

Several parameters and considerations for the development of the automated measurement method emerged from the preliminary data in section 3.5 above. The puncta size range should be from 0.18 to around 0.9 μm^2, to be sure of including the smallest puncta and most of the largest ones. This should give an accurate enough detection of puncta area to be able to use the mean brightness as the measure of punctum intensity. It would also be desirable to have a way to measure the shape of the detected objects, so that merged puncta could be identified and discarded.

The processed image (**Figure 3.7 A**) was sliced into 16 brightness levels (**Figure 3.7 B**). This is similar to multi-level thresholding, or converting an 8-bit image (with 256 brightness levels) into a 2-bit image (which has 16). Pixels in the processed image with values between 256 and 240 are assigned to the brightest slice, values between 239 and 223 to the next slice, and so on. Only the brightest nine levels are used for puncta detection. The adjustment of the image contrast should have ensured that all the puncta were in the brighter end of the image, so this cut off was chosen to prevent

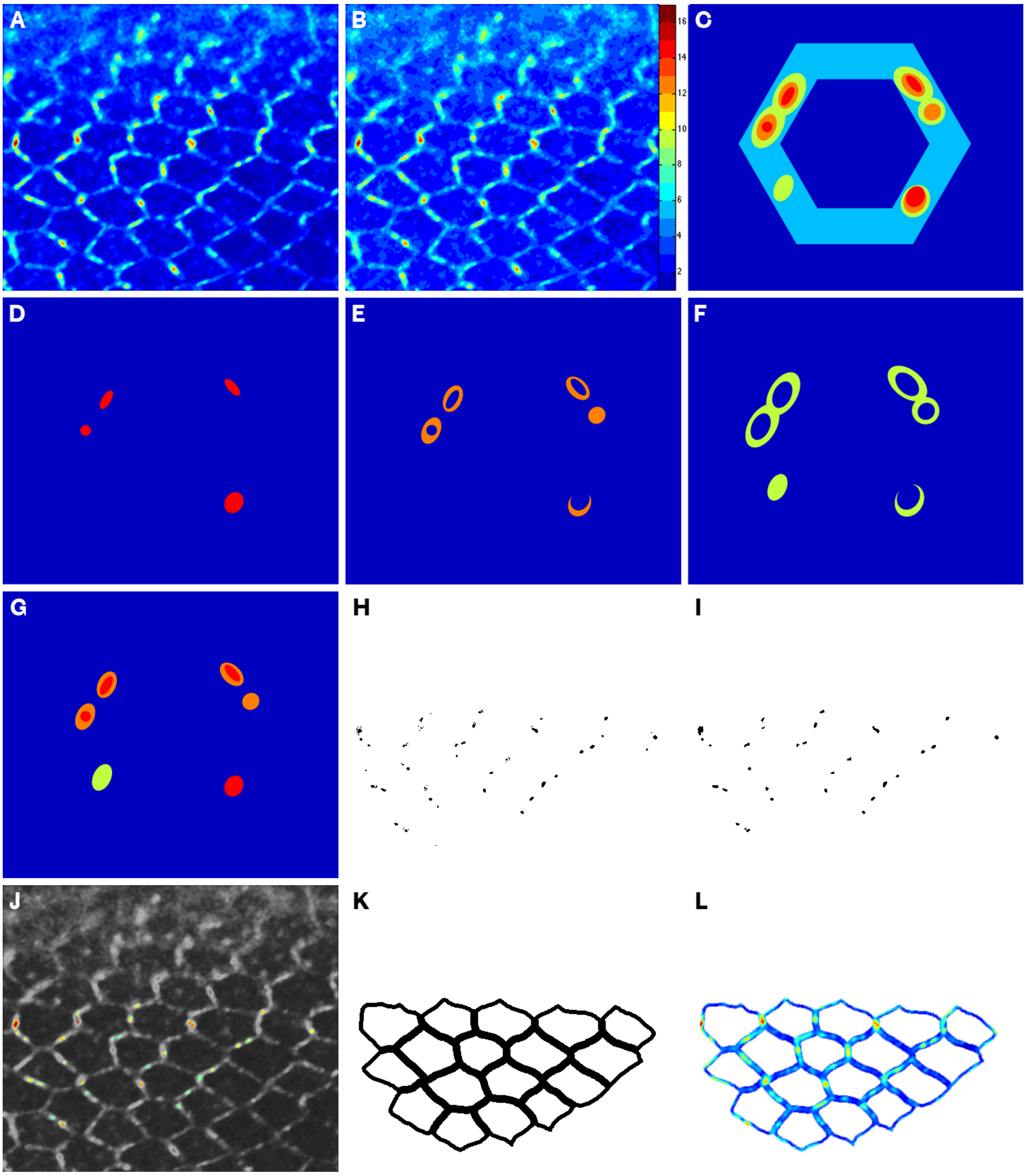


Figure 3.7: Use of MatLab to detect puncta
A: The preprocessed image. **B:** The preprocessed image is sliced into 16 levels. **C:** A simplified cell is shown to illustrate the puncta detection method. **D:** In the brightest slice, every group of pixels meets the criteria for inclusion in the final puncta mask. **E:** In the second brightest slice, a new punctum peak is detected, and one punctum appears as a crescent. **F:** In the third brightest slice two pairs of puncta are merged. **G:** The final puncta mask. **H:** Applying the same procedure to the image shown in A results in this puncta mask. **I:** The refined puncta mask. **J:** The puncta mask applied to the image in A. Detected puncta are shown in colour. **K:** For measuring the mean membrane fluorescence, the cell skeleton shown in Figure 3.6 H was overlaid on the membrane mask from Figure 3.6 I, and the exterior strip was removed to give this modified mask. **L:** The membrane measurement mask overlaid on the preprocessed image in A.

In colour images, dark blue represents zero brightness and dark red represents maximum brightness. Please see the accompanying description in Section 3.8.3 for a more thorough explaination of this figure.

detection of slightly brighter but non-punctate patches of membrane. This should also help to ensure that puncta are not detected in any dim, our-of-focus cells that were erroneously included in the membrane mask.

The detection strategy was to find the bright centres of puncta and then expand each detected spot by adding the neighbouring regions in the consecutive dimmer slices until one of two criteria were met. Either the lowest slice, and hence the dimmest brightness threshold, would be reached, or the detected punctum would merge with a neighbouring punctum.

To achieve this, the script looked at each slice in turn. The first slice contains every pixel that has a value of 15 in the sliced image (the levels are numbered from zero to fifteen). Every connected group of pixels is detected as an object in the image, and several of their properties are measured. The groups of pixels are then retained or discarded based on these measurements.

The first measurement is the circularity of the object, calculated using the formula $4\pi[\text{Area}] / [\text{Perimeter}]^2$. A perfect circle would have a circularity of 1, whereas a line would have a circularity of zero. Any objects with a circularity greater than 0.7 are kept, and those with a circularity of 0.7 or less are discarded. This threshold was determined empirically by trial and error. Puncta generally have a rounded shape.

The second measure is the Euler number, which is equal to the number of objects minus the number of holes in those objects. In this case, where the objects are taken one by one, the number of objects in each instance is one. A group of pixels with one hole would therefore have an Euler number of zero. Any objects with an Euler number less than zero are discarded.

This process of detecting objects and assessing their circularity and Euler number is repeated for each slice, and the detected objects in each slice are combined to generate the puncta mask. For clarity, this process is illustrated using an isolated model cell with only five intensity slices. The model cell contains six puncta, with differing peak intensities and brightness profiles (**Figure 3.7 C**).

Four of the puncta are represented in the brightest slice (**Figure 3.7 D**). These are detected by the MatLab script and added to the developing puncta mask. In the next slice, three of the puncta appear as rings (**Figure 3.7 E**). The pixels in the centre of these rings were present in the previous slice, and have already been detected. The three rings pass the Euler number test and are added to the puncta mask. The punctum at the bottom of the group appears in this slice as a crescent. This is because it has a very steep intensity profile, so few pixels are represented in slices between its peak and the level of the membrane fluorescence. The crescent does not meet the circularity criterion, so it is discarded. An additional punctum peak, which was not bright enough to be included in the previous slice, has also appeared in this slice.

In the third slice the peak of the sixth punctum appears (**Figure 3.7 F**). The upper two pairs of puncta have now merged – each is represented as a single object with two holes. These objects fail to meet the Euler number criterion, so they are rejected. The steep punctum at bottom right is again represented as a crescent, which is again discarded based on circularity. The next slice is mostly membrane, and is below the lower threshold for puncta detection. These detection steps result in the puncta mask shown in **Figure 3.7 G**. Puncta of different peak intensities have been accurately detected, including ones which spread across more than one brightness slice. This method should also ensure that a similar proportion of the area of each punctum is detected every time, maximising the accuracy of the mean intensity value of each punctum.

This method was applied to the preprocessed sample image shown in **Figure 3.7 A**, with the additional restriction that puncta must fall within the membrane mask shown in **Figure 3.6 I**. This resulted in the rough puncta mask shown in **Figure 3.7 H**. Some of the detected puncta are fragmented or very small. To fix this, unconnected pixels in close proximity are joined together, any holes are filled in, and any objects smaller than nine pixels ($0.0441\mu m^2$) are removed. This gave the final puncta mask shown in **Figure 3.7 I**. **Figure 3.7 J** shows the detected puncta in colour on the preprocessed image in grey. It can be seen that a reasonably large proportion of the puncta have been detected reasonably well. The script did not work quite as well in practice as it did in theory, but it was decided that this level of accuracy would suffice. This method appears to reliably detect the bright centres of puncta, but does not always extend the detection all the way to the edge of the punctum as judged by a human eye. However, the script should select the same proportion of puncta in every image, and it is capable of detecting a wide range of peak intensities, which is the most important thing.

The membrane mask shown in **Figure 3.6 I** was generated by fattening the cell skeleton. This mask was used in the puncta detection method to ensure that all detected puncta lay within the membrane. However, this mask is unsuitable for measuring the mean membrane intensity because cells on the edge of the mask have proportionately more membrane assigned to them than those on the inside. This is only generally a problem if one needs the membrane measurement per cell. This is not essential for this project, but to leave options for future work using this method the following fix was applied. The cell skeleton was overlaid on the original membrane mask and the outer strip of membrane was removed (**Figure 3.7 K**). If the skeleton defines the boundaries between cells, each cell now has the appropriate membrane area assigned to it. The membrane was detected and measured as a single object (**Figure 3.7 L**).

The finished cytoplasm, membrane and puncta masks were applied to the original raw image. The mean areas (including total puncta area per image), number of objects, and their mean, minimum and maximum intensities were exported with the corresponding standard deviations. The MatLab script used for all image analysis is shown in the Materials and Methods.

3.9 Puncta measurement validation

Before drawing any conclusions from the data, the validity of the script's output was assessed. The puncta masks showed obvious asymmetry along the proximal-distal axes of the images, accurately reflecting the distribution of the actual puncta in the images (examples shown in **Figure 3.8 A**, **B** and **C**). This strongly implied that most of the objects detected were genuine puncta rather than random slightly brighter patches of membrane.

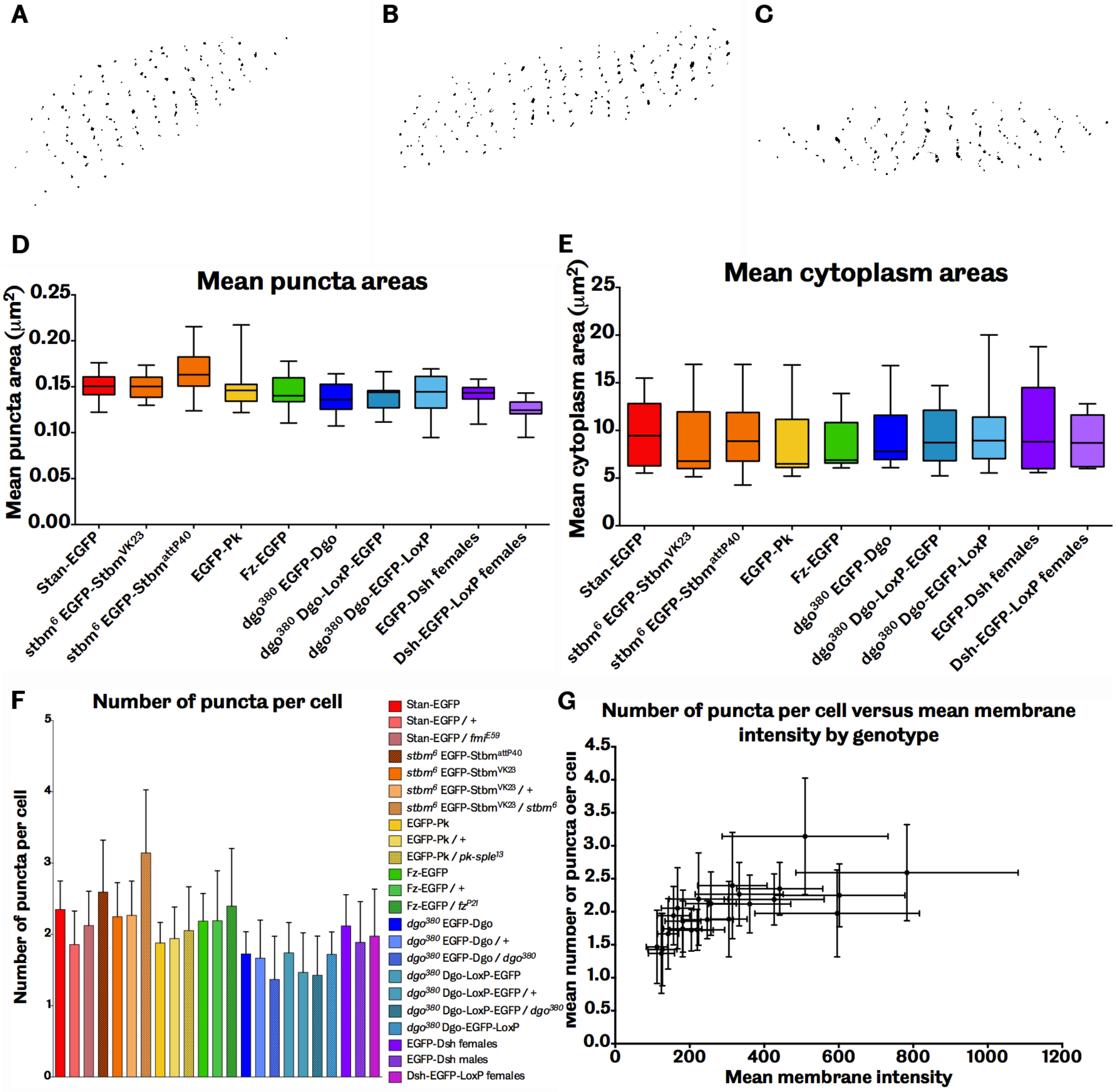
The mean punctum area for all the 28h APF images measured (446 images) was 0.135±0.022 μm^2. This is less than half of the mean punctum area calculated from the preliminary measurements, which was 0.33±0.15 μm^2. However, it was known from the puncta detection development that the MatLab script tended to detect only the bright central cores of puncta. Additionally, the manual puncta measurements may have been over-generous, as they were drawn on the image by hand. The manual measurements were also subject to bias, because puncta were selected by eye, and the sample was small. Due to the small puncta areas detected, the mean puncta intensities should be expected to err on the high side of reality.

The mean punctum area varied significantly between the homozygous flies (**Figure 3.8 D**) (ordinary one-way ANOVA, F(9, 182) = 5.619, P < 0.0001), but the multiple comparisons test showed that this was restricted solely to the data for EGFP-Stbm^{attP40} and Dsh-EGFP-LoxP females. The Dsh-EGFP-LoxP females mean punctum area was significantly different to EGFP-Pk (P = 0.040), EGFP-Stbm^{VK23} (P = 0.017), EGFP-Stbm^{attP40} (P < 0.0001) and Stan-EGFP (P = 0.005) (Tukey's multiple comparisons test). Due to its abnormal localisation in the 28h APF pupal wing, this construct was not used for the experimental data. The mean punctum area of EGFP-Stbm^{attP40} was significantly different to every other homozygous genotype, except for EGFP-Pk, Stan-EGFP and EGFP-Stbm^{VK23} (Tukey's multiple comparisons test). From **Figure 3.8 D** it can be seen that the mean punctum area of EGFP-Stbm^{attP40} is larger than average, which could decrease the measured mean puncta intensity relative to the other genotypes. For brevity, the punctum areas of each dosage test are not shown.

No significant differences in the mean cytoplasm area per cell were seen for any of the homozygous genotypes (**Figure 3.8 E**) (ordinary one-way ANOVA, F(9, 182) = 0.733, P = 0.678). This suggests that the cell skeletons were generated accurately and that the EGFP-tagged core proteins did not affect the size of the pupal wing cells.

The preliminary data were gathered by drawing around a whole cell and measuring all the puncta on its membrane. However, apart from Fmi, the core proteins each localise to only one side of the cell, so when puncta are counted in an individual cell all the shared puncta are included. When all the puncta in a field of cells are measured and divided by the number of cells, as in the MatLab script, the number of puncta per cell should thus be half the number measured in an individual cell. The preliminary measurements gave an average of around 4 puncta per cell. As expected, the number of puncta per cell for all 28h APF images measured by the script was exactly half of this value, at 2.03±0.67 puncta per cell.

The membrane mask used for puncta detection allows detection of puncta all around the edge of the detected patch of cells. This means that the number of puncta per cell will be slightly inflated, due to inclusion of puncta on the edge of the selected patch of cells that are contributed by cells immediately adjacent to but outside of the membrane mask. However, the cells at the edges of the membrane mask were often significantly dimmer than those towards the centre. Although these cells were still visible in the image, they were not quite in the same plane of focus as the brightest cells. This was sometimes difficult to see when correcting the membrane mask. Puncta



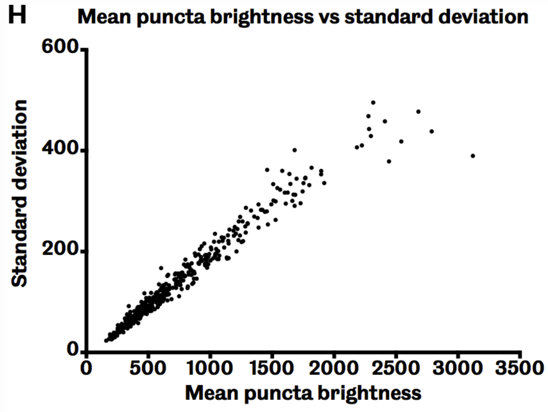


Figure 3.8: Evaluation of MatLab image measurements

A, B, C: The majority of puncta masks showed a clear asymmetric distribution of puncta. A, *dgo*^{380} EGFP-Dgo; B, Fz-EGFP; C, Stan-EGFP. **D:** The mean punctum area for all homozygous genotypes measured at 28h APF. Significant differences between genotypes were detected (ordinary one-way ANOVA, F(9, 182 = 5.619, P < 0.0001, see main text for details). **E:** The mean cytoplasm area for all homozygous genotypes measured at 28h APF. There were no significant differences between genotypes (ordinary one-way ANOVA, F(9, 182 = 0.733, P = 0.678). **F:** The mean number of puncta per cell for every 28h APF genotype. Differences between the genotypes are significant (ordinary one-way ANOVA, F(22, 249) = 11.65, P < 0.0001). Colour codes are the same as in D and E. **G:** Part of the variation in puncta numbers per cell is caused by the detection of more puncta in brighter images (Pearson's correlation coefficient = 0.71). Each point represents the mean data for a single genotype. **H:** Brighter images had a greater standard deviation of the mean puncta brightness (Pearson's correlation coefficient = 0.980).

from these cells should not be included in the measurements, as they are not as bright in the image as would be were they properly in focus, and indeed puncta were usually not detected in such cells. This would have the effect of decreasing the number of puncta measured per cell, which appears to have compensated for the expected increase described above.

The standard deviation for the overall mean number of puncta per cell is quite large, and significant differences were seen between the genotypes measured (**Figure 3.8 F**) (ordinary one-way ANOVA, $F(22, 249) = 11.65$, $P < 0.0001$). This effect can be largely explained by the effect of mean membrane brightness on the number of puncta detected. **Figure 3.8 G** shows a strong correlation between mean membrane intensity and the mean number of puncta per cell (Pearson's correlation coefficient = 0.71). Mean membrane brightness was not expected to affect puncta detection, as the preprocessing steps normalised image brightness. It seems that the decreased signal-to-noise ratio in dimmer images made it more difficult to detect puncta. However, the mean number of puncta per cell was within one standard deviation of the overall mean for every homozygous genotype used to measure the stoichiometry at 28h APF. Overall, the MatLab script seems to be detecting roughly the right numbers of puncta.

As has already been shown, the variable wing-membrane distance meant that the average image brightness also varied within each genotype. The question therefore became whether the wing-membrane distance could account for all of the variation in the data. This was assessed by plotting mean puncta brightness per image against wing-membrane distance and using nonlinear regression to fit the resulting data points to a one-phase exponential decay model. The one-phase decay model describes many physical phenomena, including absorption of light passing through a medium (Beer's Law), so a good fit to this model would suggest that the wing-membrane distance accounts for most of the variation in puncta brightness.

The regression was weighted by $1/Y^2$, because the standard deviations of the mean puncta intensities were greater in brighter images (**Figure 3.8 H**) (Pearson's correlation coefficient = 0.980). The plateau was constrained to the measured average background fluorescence of 45 units, because at an infinite wing-membrane distance the observed brightness would decay to zero with only background signal remaining. None of the genotypes deviated significantly from the one-phase decay model. As can be seen from the following figures (**Figures 3.9, 3.10** and **3.11**), the data fit the model well and none of the genotypes failed to fit the model. This shows that the intensity variations within genotypes due to variations in wing-membrane distance follow Beer's Law and can be compensated for in the data anaylsis.

The R^2 values of the nonlinear regressions are shown below in **Table 3.2**. An R^2 value is a way of describing the fraction of the variation in the data that is accounted for by fitting the curve, ie. the strength of the relationship between the two variables. A higher R^2 value indicates a better fit to the model.

Table 3.2: R^2 values for the 28h APF nonlinear regressions

Genotype	**R^2**	**Genotype**	**R^2**
Fz-EGFP	0.82	*dsh*^{V26} EGFP-Dsh females	0.70
Fz-EGFP / +	0.76	*dsh*^{V26} EGFP-Dsh males	0.30
Fz-EGFP / *fz*^{P21}	0.55	*dsh*^{V26} Dsh-EGFP-LoxP females	0.36
Stan-EGFP	0.68	*dgo*^{380} EGFP-Dgo	0.74
Stan-EGFP / +	0.63	*dgo*^{380} EGFP-Dgo / +	0.68
Stan-EGFP /*fmi*^{E59}	0.61	*dgo*^{380} EGFP-Dgo / *dgo*^{380}	0.67
stbm^{6} EGFP-Stbm^{VK23}	0.63	*dgo*^{380} Dgo-LoxP-EGFP	0.69
stbm^{6} EGFP-Stbm^{VK23} / +	0.58	*dgo*^{380} Dgo-LoxP-EGFP / +	0.39
stbm^{6} EGFP-Stbm^{VK23} / *stbm*^{6}	0.74	*dgo*^{380} Dgo-LoxP-EGFP / *dgo*^{380}	0.63
stbm^{6} EGFP-Stbm^{attP40}	0.75	*dgo*^{380} Dgo-EGFP-LoxP	0.67
EGFP-Pk	0.51		
EGFP-Pk / +	0.40		
EGFP-Pk / *pk-sple*^{13}	0.62		

3.10 Dosage testing

A final test of the EGFP-tagged core proteins' behaviour was conducted before measuring the stoichiometry of the core proteins. Despite the lack of adult polarity phenotypes and the normal apicolateral localisation of the tagged proteins, it was still possible that the EGFP was affecting the behaviour of the protein in a subtle way. Such effects could potentially alter the levels of core proteins in puncta without disturbing polarity. The behaviour of the EGFP-tagged core proteins was therefore assessed in two different ways. These tests served as a dual assessment of the behaviour of the constructs and the accuracy of the puncta detection and measurement.

Firstly, flies heterozygous for the EGFP-tagged gene and its endogenous counterpart were used. These animals were generated by crossing tagged, dosage-compensated homozygous virgin females to wild-type males. These animals thus contained one copy of the endogenous gene and one copy of its endogenous untagged counterpart. No dosage compensation is needed for the targeted lines, as EGFP is inserted directly into the endogenous locus. In the case of the P[acman] constructs, the wild-type chromosome contained only the endogenous gene, while the P[acman] chromosome carried the EGFP-tagged gene in the P[acman] construct recombined with a null mutation at the corresponding endogenous locus. The P[acman] flies therefore had three copies of the gene of interest, but only two were functional. Flies carrying one EGFP-tagged and one functional untagged copy of a gene are termed '50:50' flies.

Comparing 50:50 flies to the EGFP-tagged homozygotes would reveal any differences in EGFP-tagged protein behaviour when in the presence of the wild-type protein. If the EGFP fusion has no effect on the protein then the puncta of 50:50 flies should be exactly half as bright as those of EGFP homozygotes. If the puncta were more than half as bright, this might suggest that the EGFP-tagged protein out-competes the wild-type one for inclusion into puncta. Conversely, if the puncta were less than half as bright as those of the homozygotes this might indicate the opposite problem.

Flies carrying only the EGFP-tagged gene with no other functional copies are designated 'half-dose' flies. These flies were produced by crossing the homozygous, dosage-compensated EGFP-tagged lines to the appropriate null mutant. In the case of *fmi*, where null mutations are lethal when homozygous, *stan-EGFP/fmi*^{E59} heterozygotes were viable. While *dsh* null mutations are also lethal, due to extensive complications with making the EGFP-tagged Dsh stocks it was not possible to dosage-test these constructs within the time available.

In half-dose flies, the protein levels may be half that of wild-type flies, or the same as that of wild-type flies, or somewhere in between, depending on how any regulatory mechanisms respond to the change in gene dosage. This may or may not affect the intensity of the puncta. If the protein in question is the limiting factor for puncta formation then the puncta may be smaller, fewer, or dimmer than usual. If there is usually a surplus of the protein then a greater proportion of the available protein pool (which might be reduced compared to wild-type) might be incorporated into puncta compared to the wild-type situation. Such effects are difficult to predict in advance. The preliminary data for EGFP-Pk shown above in section 3.5 and **Figure 3.4** do show some differences in fluorescence between the different dosages. However, this data has a very small sample size, was acquired manually rather than using the MatLab script, and is not adjusted for wing-membrane distance.

3.10.1 Description of dosage testing graphs

The results of the dosage testing for each construct and the wing-membrane distance complications are discussed in detail below. Up to three genotypes are shown per graph, always derived from the same EGFP construct, but in different gene dosages. Additionally, two graphs are shown for each construct.

In each case, the graph on the left of the page show the mean puncta intensity and standard deviation for each pupa of the indicated genotype. Each point represents the mean punctum brightness and the standard deviation of this mean from a single pupa. The curve is the result of the nonlinear regression. These graphs should be used to assess the spread of data along the X axis (ie. the range of the wing-membrane distances) and the variation in puncta brightness within each pupa. The right-hand graph shows the nonlinear regression curve again (solid line), accompanied by dashed lines representing the 95% confidence intervals of the regression curve. This gives a visual representation of the fit of the data to the calculated curve.

When comparing curves on both kinds of graph it is tempting to look first at the Y-intercepts. While this is a common way of interpreting nonlinear decay graphs, it should be noted that there is generally very little data at this end of the X axis. This is reflected in the wider confidence intervals close to the Y axis. Data was extended to the Y axis mainly out of interest, to assess how the regression behaved when extrapolated beyond the range of the actual data points, and also for later comparison with the 20h APF data (see section 3.13).

3.10.2 Dosage testing of the EGFP-tagged transmembrane proteins: Fz, Stan and Stbm

For Fz-EGFP, the brightness of the 50:50 flies is approximately half that of the homozygous flies all along the X axis (**Figure 3.9 A**, light green and dark green data). This shows that the EGFP is not having any effect on the ability of the tagged protein to participate in puncta formation. The confidence intervals are relatively tight (**Figure 3.9 B**), indicating that the data is a good fit to the calculated curve. This pair of graphs also strongly suggests that the MatLab script is detecting and measuring the puncta with acceptable accuracy, and that the resulting data is a fair representation of the actual mean puncta brightness. It is very unlikely that any flaw in the puncta measurement method would be exactly compensated for by some quirk of the EGFP construct to produce such a close two-to-one dosage ratio.

Interestingly, in the half-dose flies the puncta are approximately 1.5 times as bright as the 50:50 puncta (**Figures 3.9 A** and **B**, black data). This suggests that there is a reduced protein level in these animals and that some mechanism has acted to increase the amount of Fz-EGFP that reaches the puncta. If no such mechanism existed then the half-dose puncta should be the same as the 50:50 ones, i.e. half as bright as the homozygotes.

The data for Stan-EGFP is not quite as satisfactory. The 50:50 dosage puncta brightness is generally half that of the homozygotes, but not as exact as the Fz-EGFP data (**Figures 3.9 C** and **D**, red and maroon data). This could be an effect of the larger standard deviations and tighter clustering of the data points around 20-40μm wing-membrane distance when compared to the Fz-EGFP data. As with Fz-EGFP, the half-dose puncta are more than half as bright as the homozygotes (**Figures 3.9 C** and **D**, black data). This again suggests that some mechanism is increasing the proportion of the available Fmi molecules in puncta when the cellular pool is reduced.

P[acman] EGFP-Stbm was injected into two genomic landing sites: *VK23* on III and *attP40* on II. This was the brightest construct, and some images of homozygotes had to be discarded from this dataset for containing too many saturated pixels. This was considered preferable to the alternative, which was an unacceptably low signal-to-noise ratio in the dimmest genotypes. Dosage testing was conducted for P[acman] EGFP-Stbm in *VK23* (**Figures 3.9 E** and **F**). While this data follows the same general trend as that of Fz-EGFP and Stan-EGFP, the half-dose data (black) is clearly aberrant, often reaching or exceeding the brightness of the homozygotes at the same wing-membrane distance.

The endogenous *stbm* gene is on the second chromosome, and the *stbm*^{6} allele used is recessive, unmarked, and fully rescued by the transgene. It is therefore possible that the homozygous *stbm*^{6}; P[acman] EGFP-Stbm stock was contaminated with a wild-type second chromosome, leaving the third homozygous for *EGFP-Stbm*. This would explain the results seen here, as the presence of additional wild-type Stbm molecules would be likely to decrease the brightness of the homozygous stock (**Figure 3.9 E** and **F**, orange data). This hypothesis is supported by the increased brightness of the EGFP-Stbm^{attP40} puncta (**Figure 3.9 G** and **H**, orange) compared to the EGFP-Stbm^{VK23} data.

However, the *VK23* 50:50 data is a perfect match for the homozygous P[acman] EGFP-Stbm^{attP40} data (**Figure 3.9 G** and **H**). For example, the 50:50 *VK23* data at 30μm wing-membrane distance is 1,000 units, and the corresponding value for the *attP40* homozygotes is 2,000 units. This suggests that the problems with the construct in *VK23* are restricted to the homozygous and half-dose data. Since both the *VK23* and *attP40* EGFP-Stbm lines can rescue *stbm*^{6} perfectly well, data from either line is likely to be comparable. Taken together, this data implies that the EGFP-Stbm protein itself functions normally and the *VK23* stock should not be trusted. The P[acman] EGFP-Stbm^{attP40} insertion was recombined with and sequenced for the presence of *stbm*^{6}, so, unlike the *VK23* stock, the transgene and the mutant were physically on the same DNA molecule. Any contamination of the P[acman] EGFP-Stbm^{attP40} *stbm*^{6} stock that could affect the relative *stbm* gene dosage would have been visible by a lightening of the P[acman]-derived orange eye colour, which was never observed.

3.10.3 Dosage testing of the EGFP-tagged cytoplasmic proteins: Pk, Dsh and Dgo

Comparison of the homozygous EGFP-Pk puncta to the 50:50 puncta shows a good two-to-one ratio all along the X axis (**Figure 3.10 A** and **B**, yellow and orange data), qualitatively similar to the results for Fz-EGFP and EGFP-Stbm. In contrast to the transmembrane proteins, the half-dose data is the same as the 50:50 (**Figure 3.10 A** and **B**, black data). This suggests that whatever causes the increase in brightness of half-dose puncta relative to 50:50 for the transmembrane proteins does not act on EGFP-Pk.

The data for the P[acman] EGFP-Dsh females in the *dshV26* null background fit the one-phase decay model well (**Figure 3.10 C** and **D**, purple data). The data from the males (**Figure 3.10 C** and **D**, pink data) was similar to the females, perhaps slightly brighter, but not enough male pupae were measured to confirm this. To avoid any potential

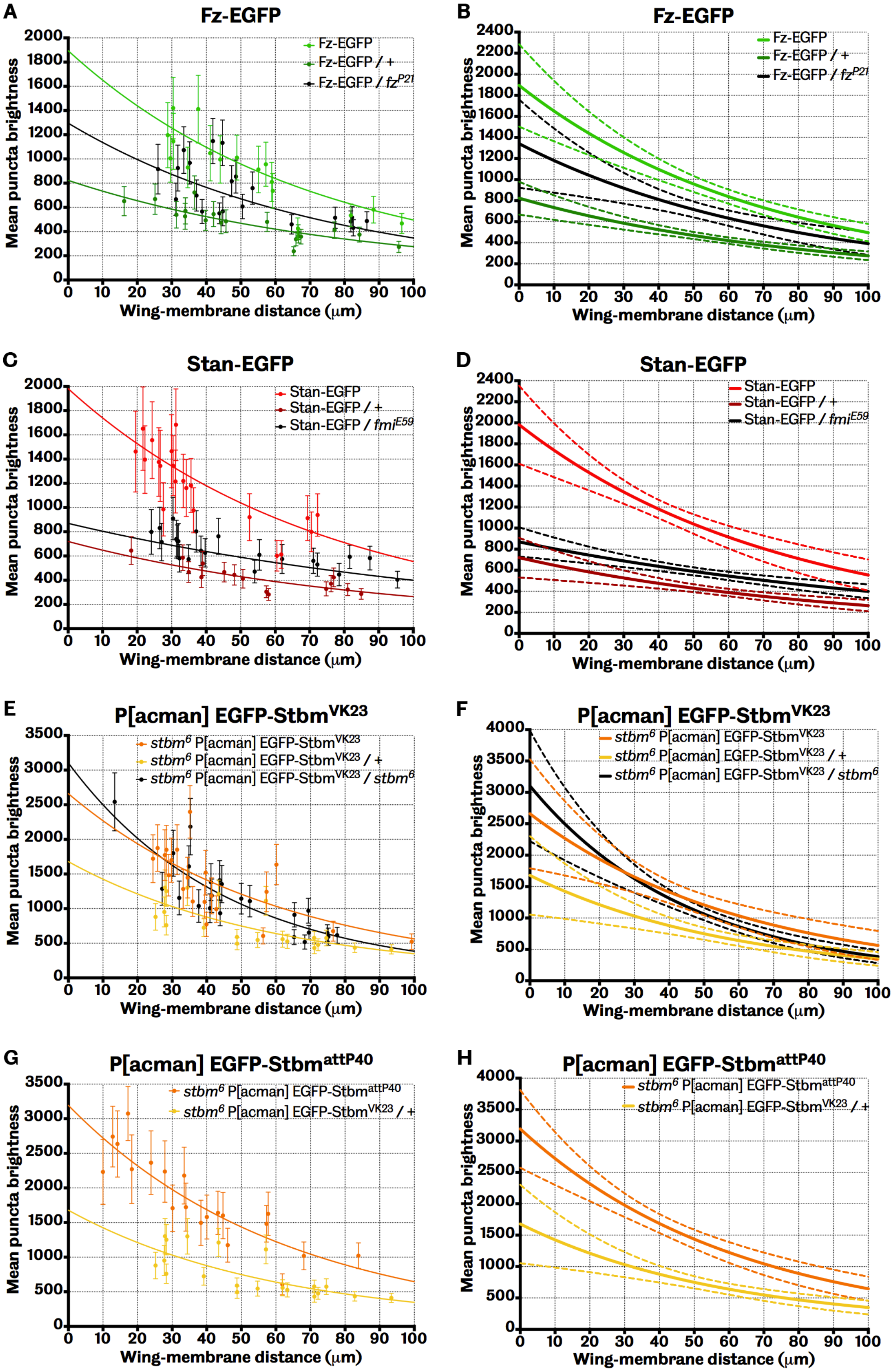


Figure 3.9: Dosage testing of the transmembrane core proteins

Lines are nonlinear regression curves. Error bars in A, C, E and G are standard deviations. Each point represents the mean puncta intensity of a single image. Dashed lines in B, D, F and H are the 95% confidence intervals of the nonlinear regression. Please see Section 3.10.2 in the main text for a thorough explanation of this figure.

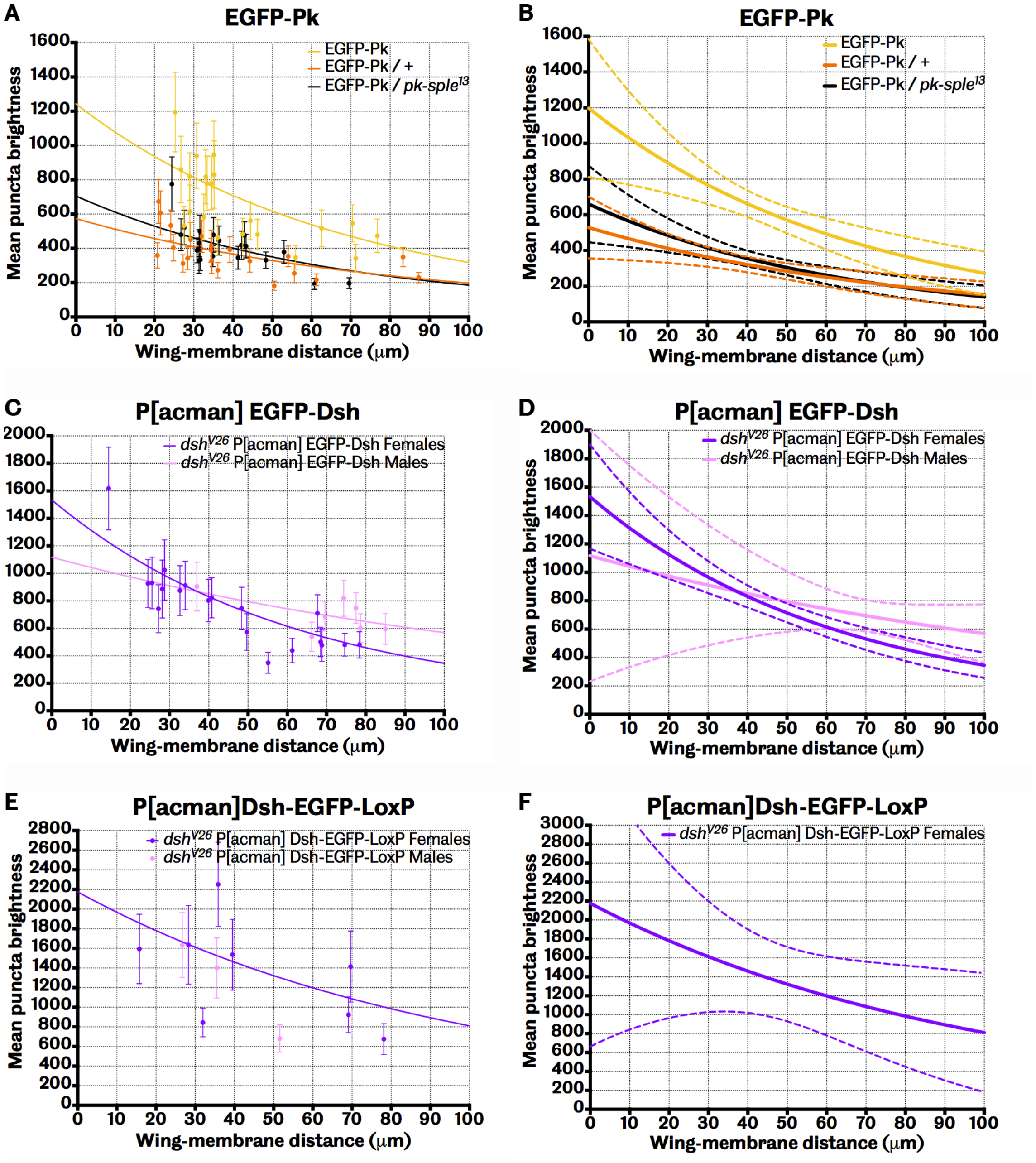


Figure 3.10: Dosage testing of EGFP-Pk and the P[acman] Dsh constructs
A, B: The puncta of homozygous EGFP-Pk pupae (yellow) are twice as bright as those of the 50:50 pupae (orange) or the half-dose (black) at almost any wing-membrane distance. **C, D:** The puncta of male EGFP-Dsh pupae (pink) are very similar to those of the females (purple), although not enough male pupae were measured to be certain that there is no difference. **E, F:** The puncta of Dsh-EGFP-LoxP males (pink) and females (purple) are generally much brighter than those of EGFP-Dsh. Due to the abnormal localisation of this construct, few pupae were measured.

Lines are nonlinear regression curves. Error bars in A, C and E are standard deviations. Each point represents the mean puncta intensity of a single image. Dashed lines in B, D and F are the 95% confidence intervals of the nonlinear regression. Please see main text for details.

dosage compensation effects in the males, only female pupae were used for the stoichiometry experiment. The puncta of the Dsh-EGFP-LoxP females were approximately twice as bright as those of the EGFP-Dsh females (**Figure 3.10 E** and **F**, purple data). As mentioned above, the C-tagged Dsh construct had very bright cytoplasmic puncta as well as the usual membrane puncta. These puncta made it very difficult to accurately segment the cells in PackingAnalyzer, and as it became clear that the localisation of this construct was very abnormal, fewer pupae were measured. This is reflected in the large confidence bands in **Figure 3.10 F**.

Dgo was the only core protein for which three functional fluorescent constructs were made. While EGFP-Dgo and Dgo-EGFP-LoxP localised normally, the puncta of Dgo-LoxP-EGFP were more discrete than usual (see **Figure 2.9 E**). The ratio of homozygous EGFP-Dgo to 50:50 puncta was approximately two-to-one (**Figure 3.11 A** and **B**, blue and light blue). The half-dose puncta line is remarkably similar to the 50:50 data (**Figure 3.11 A** and **B**, black data), suggesting that, like EGFP-Pk, either there is no decrease in protein level in the half-dose, or any change in protein level is not compensated for.

The Dgo-LoxP-EGFP data (**Figure 3.11 C** and **D**) was not dramatically different to the EGFP-Dgo data, with the homozygotes (**Figure 3.11 C** and **D**, dark blue) being a little dimmer than the EGFP-Dgo homozygotes (**Figure 3.11 A** and **B**, dark blue). The ratio of the homozygotes to the 50:50 puncta was closer to 2:1 in the EGFP-Dgo than the Dgo-LoxP-EGFP data. However, the data from the Dgo-EGFP-LoxP homozygotes (**Figure 3.11 E** and **F**), with the LoxP site moved downstream of EGFP, was closer to EGFP-Dgo than to its sister construct Dgo-LoxP-EGFP. This suggests that it was the position of the LoxP sequence between the protein and the fluorophore that caused the subtle differences in the Dgo-LoxP-EGFP protein distribution. Based on this data, the EGFP-Dgo construct was used for the stoichiometry measurements.

3.11 The stoichiometry of the core planar polarity proteins in the 28h APF pupal wing

For Dgo, Stbm and Dsh, where there was a choice of experimental stocks or constructs, the EGFP-Dgo, EGFP-Stbm^{attP40} and EGFP-Dsh females were selected to determine the

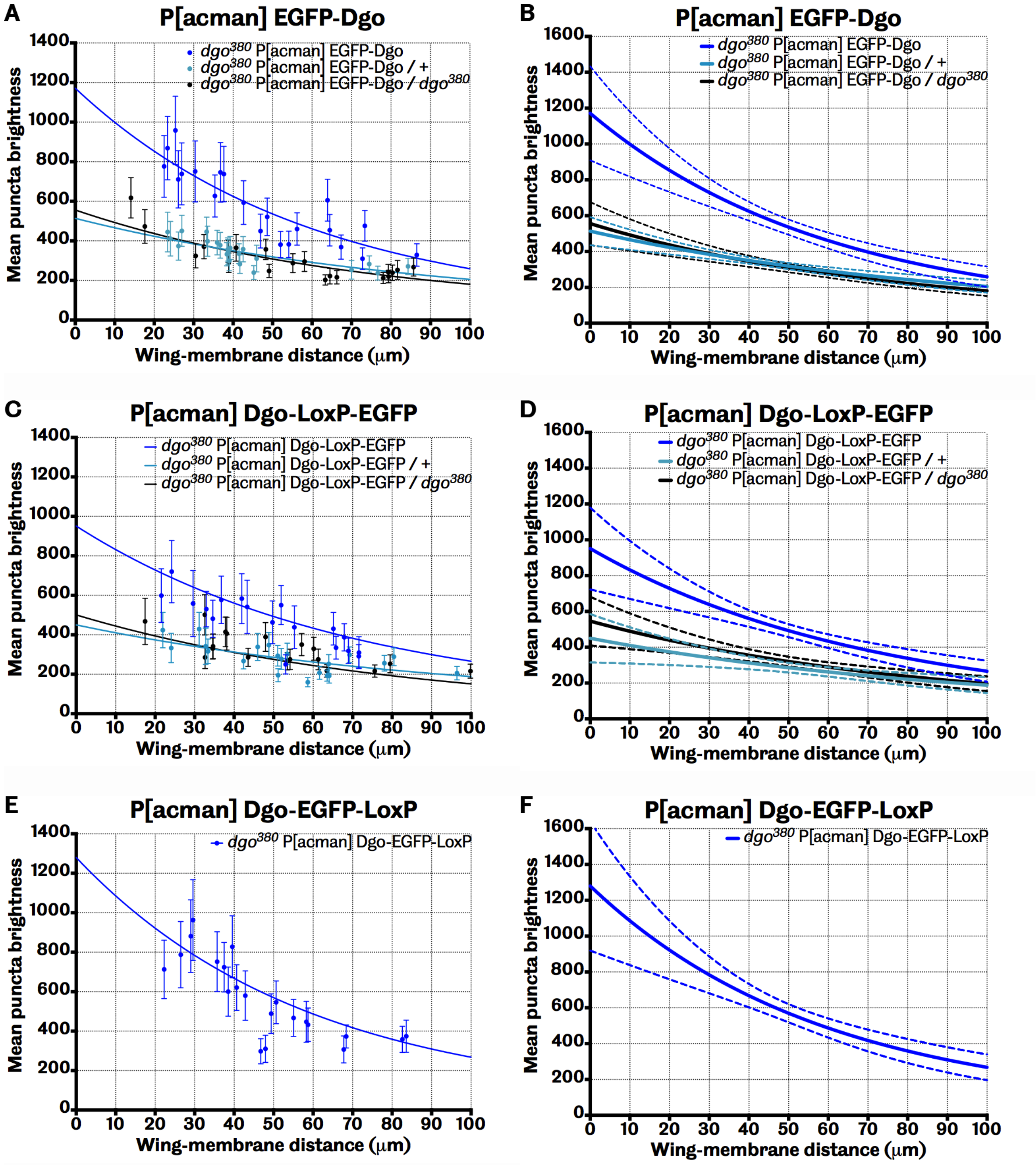


Figure 3.11: Dosage testing of the P[acman] Dgo constructs
A, B: The puncta of the 50:50 EGFP-Dgo pupae (light blue) and half-dose pupae (black) are approximately half as bright as those of the homozygotes (dark blue) at all wing-membrane distances. **C, D:** The equivalent data for the Dgo-LoxP-EGFP pupae (dark blue, light blue and black) does not follow this pattern as neatly. **E, F:** The puncta of the Dgo-EGFP-LoxP pupae are more simlar to EGFP-Dgo than Dgo-LoxP-EGFP, suggesting that it the Dgo-LoxP-EGFP construct does not function as well as the other two.

Lines are nonlinear regression curves. Error bars in A, C and E are standard deviations. Each point represents the mean puncta intensity of a single image. Dashed lines in B, D and F are the 95% confidence intervals of the nonlinear regression. Please see main text for details.

stoichiometry. For Stan-EGFP, EGFP-Pk and Fz-EGFP there was only one construct, but the dosage data showed that they behaved as expected. A single value had to be chosen to represent the mean puncta brightness for each construct. It was decided to interpolate the value at 50μm wing-membrane distance for each construct (**Figure 3.12 A**). This was generally the point at which the confidence intervals were tightest in the dosage testing data. The value of the regression curves at each point along the X axis are provided by the analysis software used.

F tests were used to check for statistically significant differences between the regression curves for the stoichiometry genotypes. For each pair of genotypes, the sum of squares and degrees of freedom from the original regression curves were summed. The data was then pooled into one sample and the curve was fitted to the pooled data using the same method as for the two separate samples. The F ratio was then calculated using the following formula, from the GraphPad Prism documentation:

$$F = \frac{(SS_{pooled} - SS_{summed})/(DF_{pooled} - DF_{summed})}{SS_{summed}/DF_{summed}},$$

where SS = sum of squares and DF = degrees of freedom. The value of the F ratio can be used along with the degrees of freedom of the numerator and denominator to look up the corresponding P value; this step was done using the Excel function FDIST.

The data for EGFP-Stbm^{VK23} was not significantly different to EGFP-Stbm^{attP40} (P = 0.078). EGFP-Pk was not significantly different to EGFP-Dgo (P = 0.064) or EGFP-Dsh (P = 0.115). Fz-EGFP was not significantly different to Stan-EGFP (P = 0.282). All other pairwise comparisons showed significant differences. A table of the P values is shown below (**Table 3.3**). The stoichiometric ratios at 28h APF and 50μm wing-membrane

	EGFP-Dgo	**EGFP-Dsh**	**EGFP-Pk**	**EGFP-Stbm^{attP40}**	**Fz-EGFP**	**Stan-EGFP**
EGFP-Dsh	0.4×10^{-6}					
EGFP-Pk	**0.0644**	**0.1151**				
EGFP-Stbm^{attP40}	0.1×10^{-19}	0.5×10^{-9}	0.2×10^{-15}			
Fz-EGFP	0.8×10^{-13}	0.0001	0.2×10^{-7}	0.5×10^{-9}		
Stan-EGFP	0.6×10^{-15}	0.1×10^{-7}	0.8×10^{-10}	0.6×10^{-9}	**0.2827**	

Table 3.3: The results of the F tests for the stoichiometry genotypes mean puncta data. Non-significant P values are shown in bold.

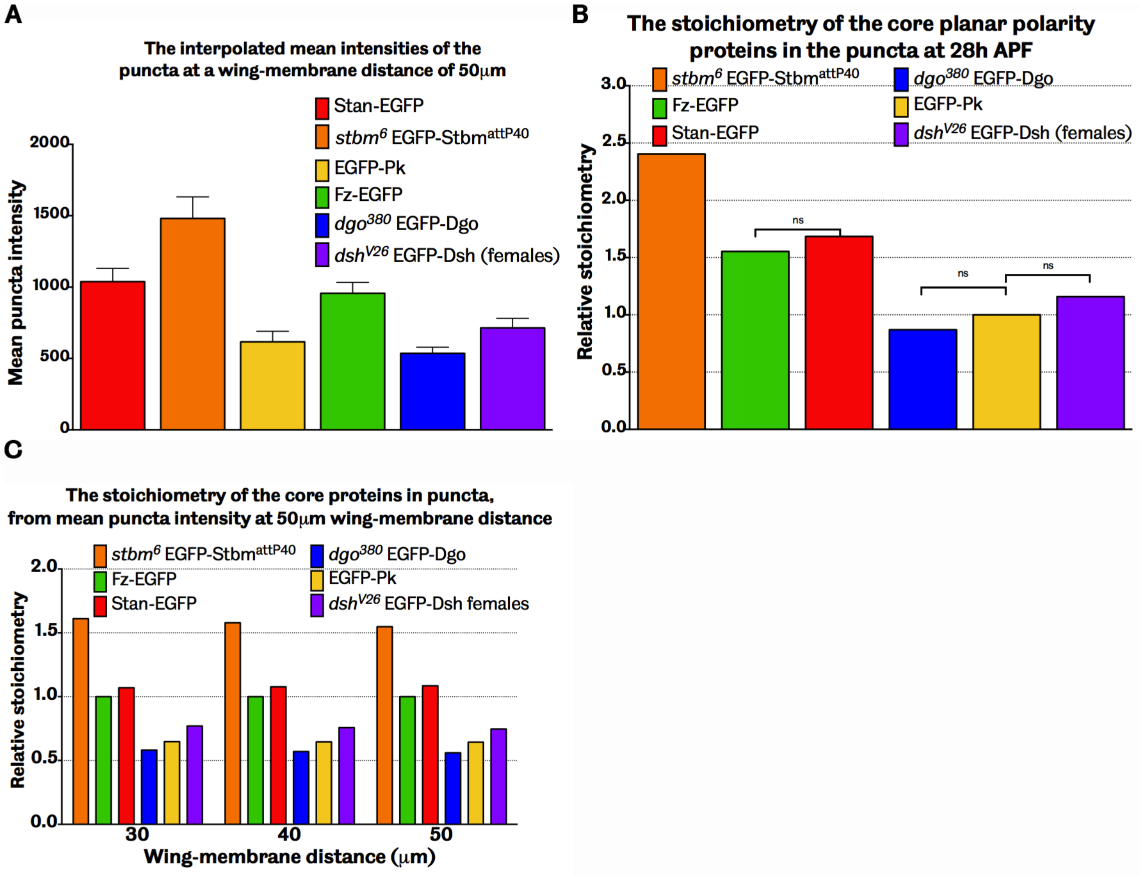


Figure 3.12: The stoichiometry of the core proteins in puncta at 28h APF
A: The value of the regression curve for each genotype was interpolated at a wing-membrane distance of 50μm. Error bars are 95% confidence intervals. EGFP-Pk is not significantly different to EGFP-Dgo ($P = 0.0644$) or EGFP-Dsh ($P = 0.1151$). Fz-EGFP is not significantly different to Stan-EGFP ($P = 0.282$). All other comparisons were significantly different to each other. Please see accompanying text in section 3.11 for an explanation of the statistical test used. **B:** The relative stoichiometry was calculated by normalising to EGFP-Pk. **C:** The stoichiometric ratios are the same when the values are interpolated at three different wing-membrane distances.

distance are shown in **Figure 3.12 B**, with the interpolated values normalised to Pk. The stoichiometric ratios were thus approximately equal levels of Pk, Dgo and Dsh, with 1.5 times as much Fz and Stan, and 2.5 times as much Stbm. The ratios remain the same when the values are interpolated at 30μm or 40μm wing-membrane distance (**Figure 3.12 C**).

3.12 The pupal wing at 20h APF

At 20h APF the developing wing is undergoing cell division and tissue growth. Cell size and shape are irregular, and the core planar polarity proteins are not asymmetrically localised (**Figure 3.13 A-F**). This loss of asymmetry is reflected in the puncta masks, examples of which are shown in **Figure 3.13 G, H** and **I**. Imaging of the 20h pupal wing

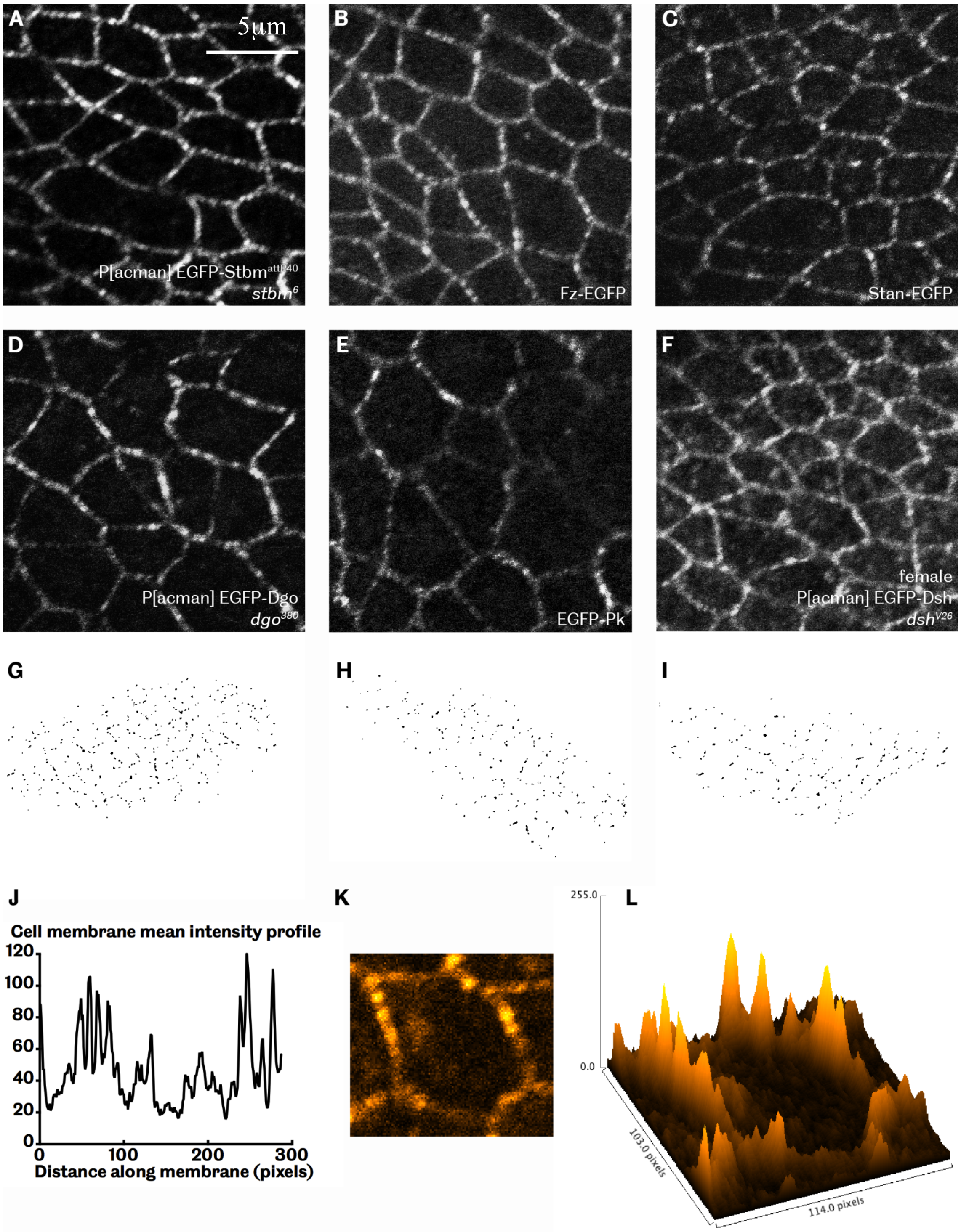


Figure 3.13: Puncta at 20h APF
A - F: At 20h APF, the core proteins are symmetrically distributed around the apicolateral membrane. Cell size and shape are very variable due to tissue growth and cell division. Images in these panels are presented as the same magnification as the 28h APF pupal wing images in Figure 2.9. **G, H, I:** The loss of asymmetry is reflected in the masks of the detected puncta. G, female *dsh*^{V26} EGFP-Dsh; H, Stan-EGFP; I, *dgo*^{380} EGFP-Dgo. **J:** Mean membrane intensity profile for the Fz-EGFP cell shown in K, generated by tracing along the cell membrane using a 10-pixel wide line. **K:** An example of a 20h Fz-EGFP cell, comparable to that shown in Figure 3.3 D. **L:** Surface plot of cell shown in K.

In pupal wing images and puncta masks distal is to the right and anterior is up. All flies are homozygous for the indicated transgenes and mutations.

was complicated by the uneven, blistered surface of the tissue, but it was still possible to acquire images of sufficient quality for quantitation. As shown in **Figure 3.5 C**, the wing-membrane distances tend to be much smaller at 20h APF than at 28h APF. This meant that many images had to be discarded due to over-saturation of the detector – the images had to be taken at the settings chosen for 28h APF so that the data could be compared between the two timepoints.

Figure 3.13 J shows the mean membrane profile of the Fz-EGFP cell shown in **Figure 3.13 K**, which has been false coloured to highlight differences in pixel intensity. It can be seen that puncta are more numerous and appear smaller than those shown in the comparable cell in **Figure 3.3 D**. The surface plot of this cell (**Figure 3.13 L**) shows that the basal membrane fluorescence is generally high, and the puncta peaks are very sharp. The mean number of puncta per cell at 20h APF is slightly but significantly greater than at 28h APF. Only the genotypes used at both 20h APF and 28h APF, excluding the dosage testing genotypes, were compared. The mean number of puncta per cell of these genotypes at 28h APF was 2.1±0.05 and at 20h APF it was 2.4±0.06 (unpaired t test, P <0.0001). The puncta areas were essentially the same, averaging $0.148\mu m^2$ at 28h APF and $0.143\mu m^2$ at 20h APF, although this difference was slightly significant (unpaired t test, P = 0.0382).

3.13 The stoichiometry of the core planar polarity proteins in the 20h APF pupal wing

The mean puncta intensity data for the 20h pupae was analysed in the same way as the 28h APF data. The 20h data for each construct is shown overlaid on the corresponding data at 28h APF (**Figure 3.14 A** - **F**). As with the 28h APF dosage data, a distinction was seen between the transmembrane core proteins Fz, Stan and Stbm and the cytoplasmic proteins Pk, Dsh and Dgo. For Fz-EGFP and Stan-EGFP, the 20h APF points tend to fall outside the 28h APF confidence bands (**Figure 3.14 A** and **B**). For EGFP-Stbm, the 20h APF data is very similar to the 28h APF data (**Figure 3.14 C**). However, the 20h APF data points for EGFP-Pk, EGFP-Dsh and EGFP-Dgo all tend to fall within the 28h APF confidence bands (**Figures 3.14 D**, **E** and **F** respectively).

The stoichiometry for the 20h APF mean puncta intensity was interpolated from the regression curves as for the 28h data (**Figure 3.15 A**). A wing-membrane distance of

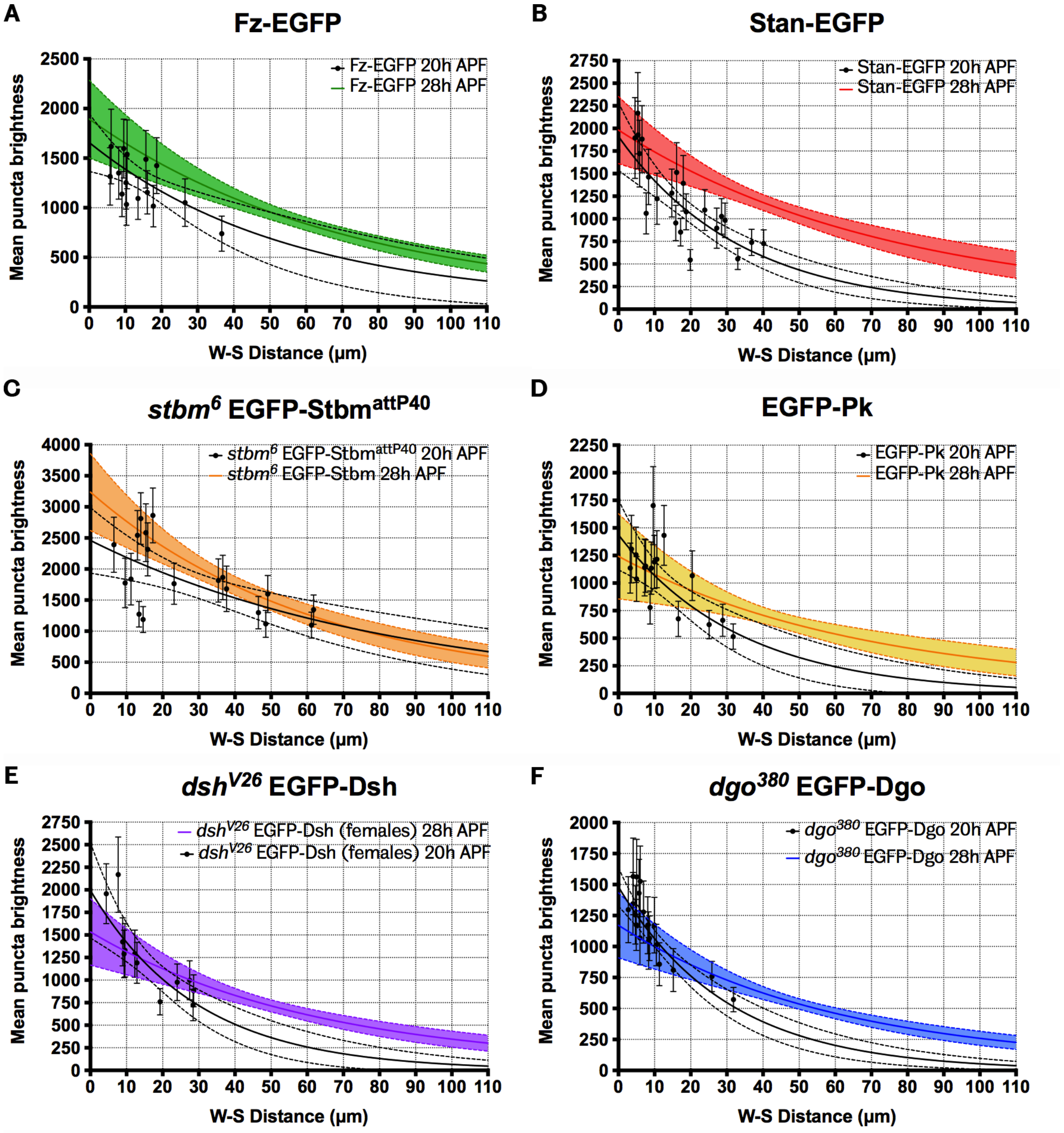


Figure 3.14: The stoichiometry of the core planar polarity proteins in puncta at 20h APF

In **A** - **F** the 20h APF data and regression curves have been overlaid in black on the regression curve and 95% confidence bands of the corresponding 28h APF data in colour. **A:** The 20h APF Fz-EGFP points tend to fall lower than the 28h data. **B:** The same trend is seen in Stan-EGFP as in Fz-EGFP. **C:** The 20h EGFP-Stbm data is very similar to the 28h data. **D:** The EGFP-Pk 20h data is similar to the 20h data, falling mainly within the confidence bands of the 28h data, as do EGFP-Dsh (**E**) and EGFP-Dgo (**F**).

15µm was chosen, as this was where the confidence intervals tended to be narrowest. F tests showed that Fz-EGFP was not significantly different to Stan-EGFP or EGFP-Dsh (P = 0.355 and P = 0.150 respectively) and additionally Stan-EGFP was not significantly different to EGFP-Dsh (P = 0.714). EGFP-Pk and EGFP-Dgo were not significantly different from each other either (P = 0.851), but all other pairwise comparisons were significantly different from each other. The relative stoichiometry was again calculated by normalising the values to Pk (**Figure 3.15 B**).

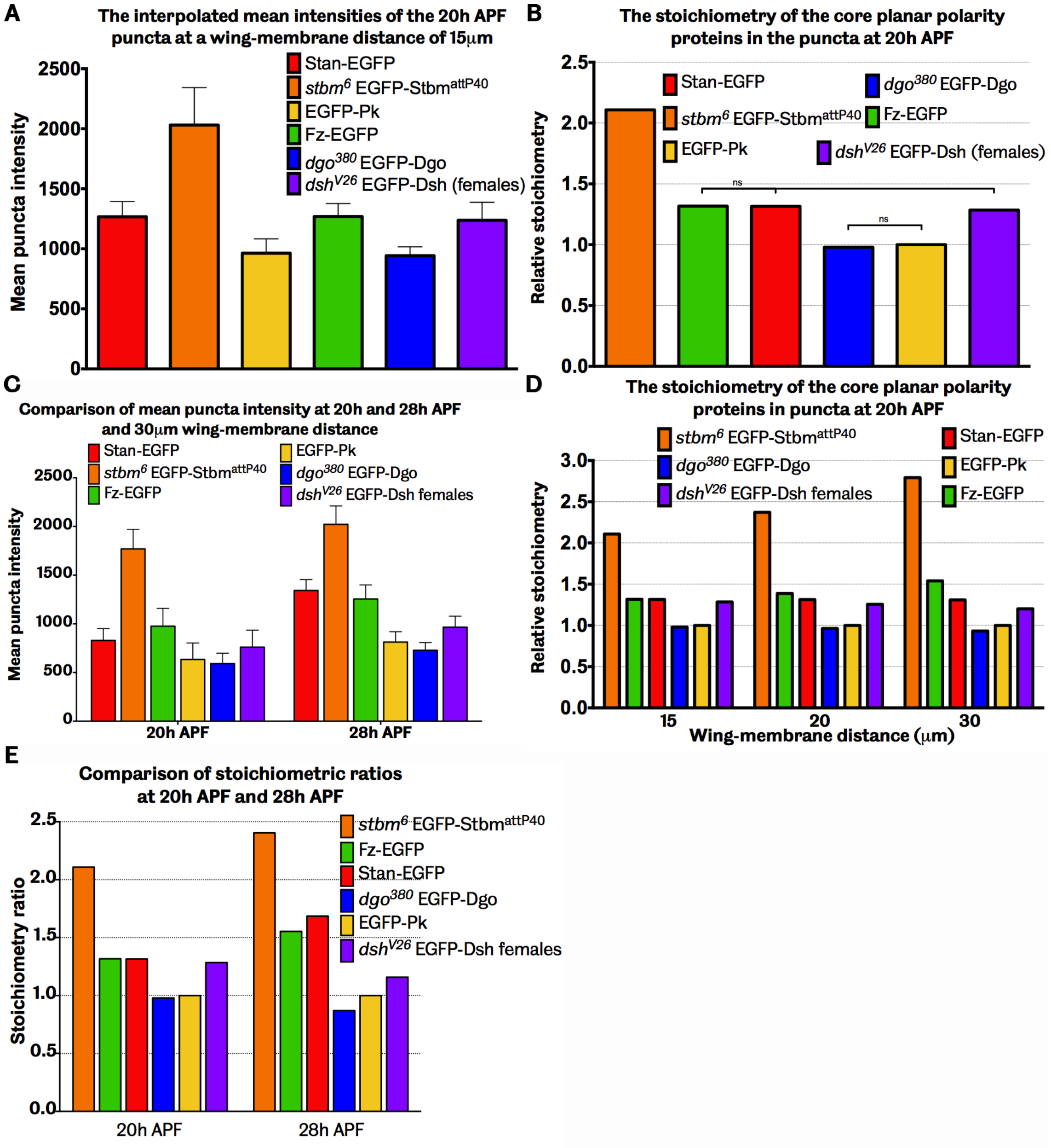


Figure 3.15: Comparison of stoichometry at 28h and 20h APF
A: The values of the nonlinear regression curves for the 20h data were interpolated at a wing-membrane distance of 15µm. EGFP-Pk and EGFP-Dgo are not significantly different to each other (P =0.851), nor are EGFP-Dsh and Fz-EGFP (P = 0.150), EGFP-Dsh and Stan-EGFP (P = 0.714) and Fz-EGFP and Stan-EGFP (P = 0.356) (F tests). **B:** The stoichiometry was calculated by normalising the figures in A to Pk. **C:** The values of the regression curves at 30µm wing-membrane distance for 20h and 28h APF mean puncta data. **D:** The stoichiometry of the core proteins at 20h APF changes depending on the point of interpolation. **E:** The stoichiometric ratios of 20h APF data at 15µm wing-membrane distance and 28h APF data at 50µm wing-membrane distance.

Error bars are 95% confidence intervals.

The 20h data points tended towards smaller wing-membrane distances than the 28h data. To give directly comparably data, the curves for both 20h and 28h were interpolated at a wing-membrane distance of 30µm (**Figure 3.15 C**). 30µm is not the optimal wing-membrane distance for either dataset, but it is a reasonable compromise.

At an equivalent wing-membrane distance, the puncta of 20h APF pupae are generally brighter than at 28h APF, implying that the absolute levels of the core proteins in puncta decrease from 20h to 28h APF. The 20h mean puncta data were compared to the equivalent 28h data using F tests. All 20h datasets were significantly different to their 28h counterparts, apart from EGFP-Pk (P = 0.2321). However, this could be largely an effect of the relative distribution of points along the X axis. There is little overlap between the wing-membrane distances of the 20h and 28h APF data. This means that regions where one curve has narrow confidence bands tend to coincide with the most tenuous regions of the corresponding curve at the other timepoint.

Figure 3.15 D shows that, unlike the 28h APF data in **Figure 3.12 C**, the stoichiometric ratios for the 20h APF data are affected by the wing-membrane distance at which the data is interpolated, even in quite a narrow range (15-30μm). The stoichiometric ratios for the 28h APF data at 50μm wing-membrane distance and for the 20h APF at 15μm wing-membrane were therefore compared (**Figure 3.15 E**). These ratio values, when calculated from the wing-membrane distance with the tightest confidence intervals, should also be directly comparable, although they cannot give information about absolute puncta intensity.

The differences between the levels of the core proteins in the puncta are not as marked at 20h APF as at 28h APF, although there is still an excess of Stbm. Perhaps as the puncta 'mature' between these two time points, becoming fewer and larger, the stoichiometry is also being fine-tuned as core proteins arrive and are removed from the puncta. The overall pattern of stoichiometric ratios between 20h and 28h APF is similar, with no dramatic changes. The stoichiometric ratios and their implications are discussed in more detail and a wider context in Chapter 6.

3.14 Post-hoc power calculations

The *a priori* power calculations described in Section 3.7 were used to choose a suitable sample size based on model data and a desired experimental power. Post-hoc power calculations determine the achieved power given the actual data and sample sizes. Post-hoc power calculations are complicated by the presence of the wing-membrane distance problem, which was not accounted for in the *a priori* power analysis.

Assessment of the achieved power of the nonlinear regression analysis would require a complicated method based on F tests and the noncentrality parameter, which was not attempted. It was instead decided to apply post-hoc power analysis to the simple mean values of the data. These were calculated by pooling the mean puncta data for each image within a genotype to produce an overall unweighted mean. While this was not the best method to use, it would give a lower bound for the power achieved. The actual power of the F tests, which include the contribution of the wing-membrane distance, would have been greater than the power values calculated here.

The procedure is the same as for the *a priori* power analysis, apart from a small modification to the estimate of the effect size. Cohen's *d* was used to calculate the effect sizes for the *a priori* calculations, which assumes that the standard deviations of the two sample means are the same. In post-hoc calculations, where the actual standard deviations of the sample means are available, the pooled standard deviation can be used instead. The pooled standard deviation is calculated as follows:

$$\sigma = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

where σ is the pooled standard deviation, s^2 is the sample variance and n is the sample size.

The results of the post-hoc testing are shown in **Table 3.4**. Powers less than the target value of 0.95 are highlighted in bold. For nine of the 15 comparisons, the target power has been achieved. For the three comparisons with powers between 0.8 and 0.9 (EGFP-Dgo and EGFP-Dsh, EGFP-Pk and Fz-EGFP and Fz-EGFP and Stan-EGFP), the actual power of the F tests on the regression curves is likely to be closer to the target value. For EGFP-Dgo and EGFP-Pk, EGFP-Dsh and EGFP-Pk, and EGFP-Dsh and Fz-EGFP, the actual powers are likely to still fall short of the target. This means that for these data sets, it is unlikely that any significant differences between them would have been detected. There is too much overlap between the distributions of these samples to detect any differences at the sample sizes achieved in this experiment. It is therefore possible that there are small significant differences between the levels of Dgo, Pk and Dsh that were not detected in the data.

3.15 Discussion

3.15.1 Solving the wing-membrane problem

The issue of the pupal wings not lying flat within their sacs caused several challenges for data collection and image quantitation. Many variations of dissection technique were tested to see whether it could be avoided. The window in the cuticle was extended posteriorly to allow the tip of the wing pouch to pop out of the pupa, the window was made very small, or extended up anteriorly past the hinge. Dissection was attempted with forceps, a razor blade and a surgical needle. Even when a large area of the wing sac was as flat as possible against the glass, the wing inside would still be at an angle. Every modification tested produced the same variable wing-membrane distance.

However, the fact that the data for every construct could be fit to a one-phase nonlinear regression model shows that the wing-membrane distance was the major source of variation in puncta brightness between the pupae. A whole curve is not convenient for comparisons between different genotypes, so a single measure of puncta brightness had to be chosen. For the 28h APF data, the common procedure of extrapolation to the Y axis to find the Y intercept was considered less reliable than interpolation of the curve at an arbitrary wing-membrane distance. This decision was guided both by the statistics (the confidence intervals were much wider at the Y axis than at larger values of X) and by the physical nature of the samples from which the data was obtained. A wing-membrane distance of zero *in vivo* would not be observed unless the wing sac was damaged during dissection, and then the strong autofluorescence of the membrane would occlude the signal from the tagged core protein. The extrapolated value would therefore never be actually observed *in vivo*, which makes it a less reliable measure than an interpolation.

However, for the 20h APF data, quite small wing-membrane distances did occur naturally. These pupae were dissected in the same way as the 28h pupae, suggesting that the wing-membrane distance varies throughout development. This is likely to be due to the morphogenetic processes that shape the developing wing, and should therefore always be monitored when imaging the wing *in vivo* for comparisons at different developmental stages.

3.15.2 Evaluation of puncta detection method

While the puncta detection method did not work as well as was hoped, sometimes fragmenting the puncta, not detecting all of a punctum or missing some genuine puncta, the puncta masks always showed the expected distribution (asymmetric at 28h APF and symmetric at 20h APF). The puncta masks could even be used to identify the occasional 20h APF pupa that had the beginnings of asymmetric localisation, which were not included in the final data set. The script detected roughly the expected numbers of puncta per cell, and the dosage data showed that the same proportion of puncta was being measured across different genotypes.

The only residual questions are whether the puncta detected by the script are representative of the actual mean puncta brightness of the image. If the script was consistently over- or under-estimating the puncta brightness this would not necessarily be reflected in the dosage data, but could affect the relative stoichiometries of different genotypes.

The most technically challenging part of the image processing procedure was generating the cell skeletons. The fluorescence from the core proteins is both asymmetric and punctate. This means that every cell has some edges that have very low fluorescence, and the edges with bright fluorescence have dark patches. This greatly complicates detection of cell edges, particularly at cell vertices. Consequently, the most time-consuming and error-prone step of the image analysis was hand-editing the cell outlines.

To facilitate segmentation of the cells it would have been much easier to include a membrane marker in another channel, and use this as a template for the membrane. As discussed in Chapter 2, *in vivo* imaging using fluorophores in channels other than green has been problematic in the past, and the red and blue constructs generated in the course of this project were underwhelming at best. A DE-Cadherin:mTomato knock-in construct from Huang et al. (2009) was recently published. This construct agglomerates in bright cytoplasmic spots, similar to the Dsh-EGFP-LoxP line, but it is also present in the usual uniform apicolateral distribution. This transgene has now been added to all of the EGFP-tagged core protein fly lines to aid in future experiments.

3.15.3 Potential improvements to the MatLab script

The script was largely developed using the Fz-EGFP test image shown in **Figures 3.6** and **3.7**, because it happened to have medium brightness with a good range of in-focus and fuzzy cells. Perhaps it is not a coincidence that the Fz-EGFP flies gave the cleanest dosage data. The script was tested on images of all the different constructs before analysing all the data, but it might be beneficial to refine the image processing and puncta detection parameters using more images of all different genotypes. Decreasing the number of slices into which the preprocessed image is divided might improve puncta detection, as each one would be spread across fewer slices. This would make the detected puncta less fragmented and facilitate detection of a larger area of each punctum.

The slanted wings meant that at 28h APF it was difficult to get an entire field of view full of wing cells. It was also common to have both in focus and out of focus cells in each image. It was therefore important to exclude out of focus cells from the analysis, as they would have dimmer puncta than the in-focus cells. Fortunately, it was reasonably easy to exclude out of focus cells in PackingAnalyzer when making the cell skeletons. Severely out of focus cells could be identified by their higher cytoplasmic signal – the nucleus is only a few microns below the level of the puncta, and its weak autofluorescence is enhanced by the preprocessing operations in the MatLab script.

However, cells that were only slightly out of focus were often included in the membrane masks. In most cases, these cells were too dim to have any puncta detected by the MatLab script, but on other occasions, puncta that were only a little dimmer than they should have been were detected and included in the analysis. This raises the problem of how to define a cut-off threshold when the signal is continuous. It might be possible to use the data produced in this project to improve exclusion of out of focus cells, perhaps by detecting each cell individually and comparing its mean membrane fluorescence to an empirically determined threshold. Alternatively, such a threshold could be calculated individually for each image. The mean membrane fluorescence of each detected cell could be used to calculate an average cell membrane fluorescence for the whole image. Any cell which falls some predetermined fraction of a standard deviation above or below this average could then be discounted from further analysis, prior to the puncta detection stage. This cell filtering step might reduce inaccuracy in the data produced by the script, resulting in better quality data.

Several modifications of convenience should also be added to the script. Batch processing of images would be useful. A couple of measurements that are not currently included in the raw data exported by the script, such as the number of puncta per cell for each image, could also be included. The ability to easily experiment with different detection parameters and thresholds within the script would also be preferable to having these hard-wired into the code.

3.15.4 Puncta profile plots

The puncta profile plots revealed that the majority of the fluorescence is concentrated in a bright core, usually with a sharp drop back down to the level of basal membrane fluorescence. However, puncta with two peaks, or broader, flatter peaks, were also seen. These two-peaked puncta could represent two separate puncta that were too close together to be resolved in the image. Puncta were also sometimes seen clearly off-centre relative to the membrane. These are likely to be deliveries or collections of core proteins to or from the membrane. The membrane could also be bent or kinked at these locations instead of straight, which could produce the same effect.

The resolution achieved in the experimental images was variable, due to the wing-membrane distances. Dimmer images were noisier, with a poor signal-to-noise ratio. Many interesting questions about the size, shape and behaviour of puncta cannot be answered using the confocal microscope. Super-resolution microscopy might give better insights into this aspect of planar polarity. Electron microscopy could potentially be informative, but the technical hurdles of sample preparation and acquisition of suitable antibodies would be a significant barrier to this line of inquiry. Shimada et al. (2006) show good TEM images of 30h APF pupal wings, but one gets the impression that these were not easy to obtain. Additionally, the immuno-EM images are quite different to what is seen in immunostained tissue imaged on a confocal or super-resolution instrument, suggesting that artefacts may be present.

3.15.5 Imaging methodology

When Geng et al. (2008) measured autophagy proteins in yeast, they took Z-stacks, 12 slices 0.5μm apart, to cover the whole depth of the yeast cell. The images used in this project are all single planes. It is unclear how much benefit there would be to taking stacks in the fly pupal wing. The cells are very deep in the apicobasal axis, around 30μm compared to the approximately 6μm yeast cells, so, depending on wing-membrane distance and transgenic construct, there might not be enough fluorescence to detect low levels of diffuse protein in the basal cytoplasm. The puncta are restricted to a narrow band around the apicolateral edge of the cell, and when imaging them it only took a very small movement of the focus control on its finest setting to move them out of focus.

It is likely that the spherical aberration introduced by the thickness of the tissue, distance from the coverslip and the different optical densities of the various fluids in the light path reduced the resolution in the Z axis to a point where taking stacks would not have yielded higher quality data. Taking Z stacks could also introduce photobleaching. The imaging settings were optimised to take single images of a wide range of mean brightnesses, and it is doubtful that it would be possible to find settings that allowed the capture of Z stacks without photobleaching for all genotypes.

3.15.7 The stoichiometry of the core planar polarity proteins in puncta

The stoichiometry of the core proteins at 28h APF is more complicated than the simplest configuration that would fit the current data, which is two molecules of Fmi to form a homodimer, accompanied by one molecule each of Stbm, Fz, Dsh, Dgo and Pk. Instead there are roughly equal amounts of the cytoplasmic proteins Dgo, Pk and Dsh, with approximately 1.5 times as much Fz and Fmi and up to 2.5 times as much Stbm. The ratios at 20h APF are broadly similar. Fz and Fmi are relatively slightly reduced, although still more abundant than Pk and Dgo. Dsh is increased to match Fz and Fmi, and the excess of Stbm is reduced to only twice that of Pk and Dgo. The implications of these results are discussed in detail in Chapter 6.

Hypothetical mean puncta brightness:

Stoichiometry	Brightness ratio	Mean 1	Mean 2	Effect size	Power achieved (3SF)	Required sample size to reliably detect this difference (pupae per genotype)
1:2	0.5	1000	500	2.5	0.97	6
		2000	1000	5	0.99	3
		1200	600	3	0.98	5
		1500	750	3.75	0.99	4
2:3	0.66	1000	666	1.67	0.96	11
		2000	1320	3.4	0.98	4
		1200	792	2.04	0.97	6
		1500	990	2.55	0.98	6
3:4	0.75	1000	750	1.25	0.95	18
		2000	1500	2.5	0.97	6
		1200	900	1.5	0.96	13
		1500	1125	1.875	0.96	9
4:5	0.8	1000	800	1	0.95	27
		2000	1600	2	0.96	8
		1200	960	1.2	0.96	20
		1500	1200	1.5	0.96	13
5:6	0.83	1000	833	0.833	0.95	39
		2000	1667	1.667	0.96	11
		1200	1000	1	0.95	27
		1500	1250	1.25	0.95	18

Table 3.1: A priori power calculations

Based on preliminary data, a range of plausible mean puncta brightness values were chosen at selected stoichiometric ratios. The effect size was calculated by G*Power using a fixed standard deviation of 200 intensity units. The α error probability was set at 0.05 and the desired power was set at 0.95. G*Power used these constants and the calculated effect sizes to calculate the actual power achieved and the required sample size for each combination of parameters.

Genotype 1	Genotype 2	Mean 1	Mean 2	SD 1	SD 2	Pooled SD	Effect size	n1	n2	Power achieved
EGFP-Dgo	EGFP-Dsh	582.97	797.52	188.77	291.62	242.98	0.833	21	19	**0.863**
EGFP-Dgo	EGFP-Pk	582.97	707.40	188.77	223.50	206.86	0.602	21	21	**0.607**
EGFP-Dgo	EGFP-Stbm^{attP40}	582.97	1865.85	188.77	634.18	462.83	2.772	21	20	1
EGFP-Dgo	Fz-EGFP	582.97	937.30	188.77	306.91	250.08	1.417	21	18	0.996
EGFP-Dgo	Stan-EGFP	582.97	1226.30	188.77	316.12	260.35	2.471	21	21	1
EGFP-Dsh	EGFP-Pk	797.52	707.40	291.62	223.50	258.02	0.349	19	21	**0.287**
EGFP-Dsh	EGFP-Stbm^{attP40}	797.52	1865.85	291.62	634.18	497.89	2.146	19	20	0.999
EGFP-Dsh	Fz-EGFP	797.52	937.30	291.62	306.91	299.15	0.467	19	18	**0.401**
EGFP-Dsh	Stan-EGFP	797.52	1226.30	291.62	316.12	304.76	1.407	19	21	0.997
EGFP-Pk	EGFP-Stbm^{attP40}	707.40	1865.85	223.50	634.18	470.69	2.461	21	20	1
EGFP-Pk	Fz-EGFP	707.40	937.30	223.50	306.91	265.10	0.867	21	18	**0.842**
EGFP-Pk	Stan-EGFP	707.40	1226.30	223.50	316.12	273.76	1.895	21	21	0.999
EGFP-Stbm^{attP40}	Fz-EGFP	1865.85	937.30	634.18	306.91	506.70	1.833	20	18	0.999
EGFP-Stbm^{attP40}	Stan-EGFP	1865.85	1226.30	634.18	316.12	497.17	1.286	20	21	0.992
Fz-EGFP	Stan-EGFP	937.30	1226.30	306.91	316.12	311.92	0.927	18	21	**0.882**

Table 3.4: Post-hoc power calculations

A simple pooled mean puncta brightness and standard deviation was calculated for each genotype. The effect size was calculated by G*Power using the means, the pooled standard deviation and the n numbers. The formula for the pooled standard deviation is shown on page 121. The α error probability was set at 0.05. Powers below the target value of 0.95 are highlighted in bold.

Chapter 4

Materials and Methods

Contents

Materials

Methods

Fly protocols

Molecular biology protocols

Materials

4.1 Antibodies

AP = affinity-purified, mAb = monoclonal

Rt = rat, Rb = rabbit, Ms = mouse

Rt3 anti-Stbm[N] DS18	Strutt and Strutt, 2008
Rb2 anti-Fz[Ex] AP Gly Nov04	Bastock and Strutt, 2007
Rt3 anti-Pk AP Gly Apr04	Strutt et al., 2013
Ms mAb anti-Fmi74 IgGs	Usui et al., 1999; purified by Sheffield BioServe
Rt3 anti-Dsh DS15	Strutt et al., 2006
Rb anti-GFP	Abcam
Ms mAb anti-β-Gal	Promega
Rb anti-β-Gal	Cappel
goat anti-Rb Cy2 111-225-144	Jackson ImmunoResearch
goat anti-Rt Cy2 112-225-167, preabsorbed against mouse	Jackson ImmunoResearch
donkey anti-Rt Cy2 715-225-150	Jackson ImmunoResearch
goat anti-Rt A568 A11077	Molecular Probes
goat anti-Ms A568 A11031	Molecular Probes
goat anti-Rb RRX 111-295-003	Jackson ImmunoResearch
donkey anti-Ms Cy5 715-175-151	Jackson ImmunoResearch
donkey anti-Ms A647 A31571	Alexa
goat anti-Rb Cy5 111-175-144	Jackson ImmunoResearch

4.2 Flies: core planar polarity mutants

dgo^{380} is a viable null; a deletion resulting from P element excision, which removes approximately one half of the 5' end of the gene, including some upstream sequence (Feiguin et al., 2001).

dsh^{1} is a viable hypomorph, induced by methyl-methanosulphonate by Fahmy in 1956 (Perrimon and Mahowald, 1987). A single amino acid substitution in the DEP domain (Axelrod et al., 1998).

dsh^{V26} , also called ***dsh***^{3}, is a lethal null consisting of a deletion of 573bp, which removes the central third of the gene (Klingensmith et al., 1994; Wehrli and Tomlinson, 1998).

fmi^{frz3}, also called ***stan***^{frz3}, is a viable hypomorph (Rawls and Wolff, 2003).

fmi^{E59} is a lethal null EMS-induced mutation, which causes a nonsense mutation in the ectodomain (Usui et al., 1999).

fmi^{E45} is a lethal null (Usui et al., 1999).

fmi^{192} is a lethal null (Rawls and Wolff, 2003).

Df(2R)ED2076 is a lethal deficiency covering *fmi* and *dgo* (Ryder et al., 2007).

pk-sple^{13} is a viable null affecting both the *pk* and *sple* isoforms (Gubb et al., 1999).

stbm^{6} is a commonly-used *stbm* null allele. Viable, generated by X-ray mutagenesis (Wolff and Rubin, 1998). Sequenced using JA24 and DS574 to confirm presence of mutation in stocks rescued by P[acman] Stbm and its derivatives. Consists of a 3bp deletion at the 242nd nucleotide followed by a 1bp insertion at the 246th nucleotide. This introduces a frameshift, which results in a stop codon at the 394th base pair. The cDNA is 1,753bp, so this knocks out transcription of the majority of the gene. Sequencing of the *stbm*^{6} mutation is complicated by the presence of a 9bp insertion polymorphism upstream of the mutation. This polymorphism is not present in our P[acman] Stbm constructs and its derivatives, or in our *w*^{1118} (OregonR) wild-type stock, but is present on our *stbm*^{6} and CyO chromosomes.

4.3 Flies: transgenic lines

4.3.1 Stocks made during this project

All transgenic constructs were inserted into *attP40* on 2R, unless otherwise indicated below. Likewise, unless otherwise indicated, all constructs were homozygous viable.

Ds = *dachsous*, Dgo = *diego*, Dsh = *dishevelled*, Fz = *frizzled*, Pk = *prickle*, Stan = *stan/fmi*, Stbm = *stbm/Vang*, EGFP = *enhanced green fluorescent protein*.

EGFP lines

w; FRT42 stbm^{6}; P[acman] EGFP-Stbm^{VK23}

w; P[acman] EGFP-Stbm^{attP40} FRT42 stbm^{6};;

w; P[acman] Stbm-LoxP-EGFP;; (no fluorescence)

w; P[acman] Stbm-EGFP-LoxP;;
w; P[acman] EGFP-Dgo FRT42 dgo^{380};;
w; P[acman] Dgo-LoxP-EGFP FRT42 dgo^{380};; (subtly abnormal localisation)
w; P[acman] Dgo-EGFP-LoxP FRT42 dgo^{380};;
yw dsh^{V26} FRT18; P[acman] EGFP-Dsh;;
w; P[acman] Dsh-LoxP-EGFP;; (no fluorescence)
yw dsh^{V26} FRT18; P[acman] Dsh-EGFP-LoxP;; (abnormal localisation)
w; Ds-EGFP;; (endogenous locus, published in Brittle et al., 2012)
w; EGFP-Pk;; (endogenous locus)
w;; Fz-EGFP (endogenous locus)
w; Stan-EGFP;; (endogenous locus)
w; P[acman] Stan-EGFP FRT40 fmi^{E59} / CyO;; (lethal)
w; P[acman] Stan-EGFP FRT40 fmi^{E45} / CyO;; (lethal)
w; P[acman] Stan-EGFP FRT40 fmi^{192} / CyO;; (lethal)
w; P[acman] Stan-EGFP FRT42 Df(2R)ED2076 / CyO;; (lethal)
(P[acman] Stan-EGFP is viable in the wild-type background)

Red and blue-emitting fluorophores

w; P[acman] Cerulean-Stbm FRT42 stbm^{6};;
w; P[acman] TagBFP-Stbm FRT42 stbm^{6};; (abnormal localisation)
w; P[acman] mCherry-Dgo FRT42 dgo^{380};;
w; P[acman] tdTomato-Dgo FRT42 dgo^{380};; (no fluorescence)
w; P[acman] Cerulean-Dgo FRT42 dgo^{380};;
w; P[acman] TagBFP-Dgo FRT42 dgo^{380};;
w;; P[acman] mCherry-Dsh^{attP2}
w;; P[acman] Cerulean-Dsh^{attP2}
w;; P[acman] TagBFP-Dsh^{attP2}

Other transgenic constructs

w; P[acman] Stbm FRT42 stbm^{6};;
w; P[acman] Dgo FRT42 dgo^{380};;
(No transformants obtained for P[acman] Dsh)
w;; pRK2 EGFP-LoxP-w^{+}-LoxP-Pk / TM2 (lethal, random P element insertion)
w; pRK2 Fz-EGFP-LoxP-w^{+}-LoxP;; (random P element insertion)
w;; pRK2 Stan-EGFP-LoxP-w^{+}-LoxP (random P element insertion)
w pRK2 Ds-EGFP-LoxP-w^{+}-LoxP;;; (random P element insertion)

4.3.2 Stocks obtained from other sources

Genotype	**Source**
yw / Y hs-hid; FLP, I-SceI, Sco / CyO hs-hid;;	Bloomington stock centre #25680
yw/ Y hs-hid;; FLP, I-SceI / TM3 hs-hid	Bloomington #25679
w; Gal4$^{477\ w-}$; TM2 / TM3	Bloomington #26258
w; Pin / CyO; Gal4$^{221\ w-}$	Bloomington #26259
yw Ubx-FLP; FRT42 P[w+, arm-LacZ];;	J.-P. Vincent
w; arm>Gal4;;	J.-P. Vincent
yw; P[w^{+}, UAS-Pk];;	Gubb et al., 1999
UAS-Fz	Adler et al., 1997
UAS-Fmi	Usui et al., 1999
w; FRT40;;	Xu and Rubin, 1993
yw; FRT42;;	Xu and Rubin, 1993
w;; FRT80b	Xu and Rubin, 1993

Assorted standard balancer stocks: various combinations of FM6, FM7 GFP, SM6a, CyO, CyO Act>GFP, TM2, TM3 Sb, TM6b, SM6A:TM6b

4.4 Plasmids, BACs and P[acman]s

Plasmids, BACs and P[acman] clones

pEGFP^{AmpR}	Clontech; Zhang et al., 1996
pmCherry	Clontech; Shaner et al., 2004
ptdTomato	Clontech; Shaner et al., 2004
pSK^{+} RpsL-kana	AddGene; Wang et al., 2009
pRK2^{AmpR}	Huang et al., 2008
Base vector for *in vivo* gene targeting donor constructs	
PL452-N-EGFP^{AmpR}	Addgene; Venken et al., 2008.
EGFP source vector for recombineering N-termini of proteins	
PL452-C-EGFP^{AmpR}	Addgene; Venken et al., 2008.
EGFP source vector for recombineering C-termini of proteins	
PL452-inverted C-EGFP	D. Strutt
BACR-12F04 (Pk)	BacPac resources
BACR-05B23 (Fz)	BacPac resources
BACR-08L17 (Fmi)	BacPac resources

BACR-48E08 (Ds)	BacPac resources
P[acman]$^{\mathrm{AmpR}}$-Stbm	(D. Strutt)

P[acman]$^{\mathrm{ChlorR}}$-Dgo (BacPac resources #CH322-99N23)
P[acman]$^{\mathrm{ChlorR}}$-Dsh (BacPac resources #CH322-187A06)
P[acman]$^{\mathrm{ChlorR}}$-Fmi (BacPac resources #CH321-13B09)
P[acman]$^{\mathrm{ChlorR}}$-Fmi (BacPac resources #CH321-66D09)
All other P[acman] constructs listed above were made during this project

PL452-N-mCherry and **PL452-N-tdTomato**

tdTomato was amplified by PCR from ptdTomato and mCherry from pmCherry. The primers were used to add a KpnI site to the 5' end and an NheI and a LoxP site to the 3' end of the fluorophore gene, removing the Stop codon in the process. PL452-N-EGFP and the fluorophore fragments were digested with NheI and KpnI (excising EGFP from the vector in the process) and gel-purified before ligation.

pRK2-EGFP and derivatives

See *in vivo* targeting protocol below

4.5 Software

G*Power v3.1

Faul et al., 2007. Available at the time of writing at http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/download-and-register

ImageJ 1.43u

Freeware, the National Institute of Health (NIH), Madison, Wisconsin, United States. Available at the time of writing at http://rsbweb.nih.gov/ij/

LOCI BioFormats

Freeware, the Laboratory for Optical and Computational Instrumentation. Available at the time of writing at http://loci.wisc.edu/software/bio-formats

MatLab

MATLAB Student Version and Image Processing Toolbox 7.12.0.635 (Release 2011a), The MathWorks Inc., Natick, Massachusetts, United States

Nikon NIS Elements AR 4.0

Shipped with the Nikon A1 confocal microscope

PackingAnalyzer 1.0

Used with kind permission of Benoît Aigouy.

Prism 6

GraphPad Software, La Jolla, California, United States. http://www.graphpad.com/scientific-software/prism/

Wordle

www.wordle.net

Methods

Fly protocols

4.6 Fly culture

Stable stocks were cultured in plastic vials, straight-sided plastic bottles or ICRF bottles with cotton wool or plastic foam stoppers on standard fly media. Crosses received supplemental yeast grains sprinkled on the surface of the media to encourage mating and larval growth. Weak stocks also received yeast grains, plus a segment of filter paper pushed into the media to allow the flies to more easily climb up out of the media. Stable stocks were maintained at 19°C and transferred to fresh media roughly every 14 days. Heat shocking was done by placing the vials or bottles into a water bath of the indicated temperature for the required amount of time.

The life cycle of *Drosophia melanogaster* is 10 days at 25°C, so the twentieth day is the first day when the second generation could emerge from a vial. Unless otherwise stated, all vials and bottles were cultured at 25°C and discarded on their twentieth day. The major exception to this is the initial *in vivo* targeting crosses described in Chapter 2 and in section 5.26.3 below, which use the *FLP, I-SceI* males – these flies do not survive at 25°C, so these crosses were kept at 19°C until the parents were removed. Parents of crosses were transferred to fresh media every 3-4 days. Unless crossing single flies, a minimum of three virgin females were crossed to a minimum of three males.

4.7 Balancing transformants

When transformants are received, the insertion must be recovered and balanced to make a stable stock. The transgenic flies often carry markers and additional transgenes on the X chromosome that were used in the transgenesis procedure, which should be removed before using these stocks in experiments. On arrival, non-virgin and male transgenic flies, identified by eye colour, were used to establish a 'back-up' stock. A single transformant male was crossed to appropriate balancer virgins, and in the next generation a single transformant male was again selected and crossed to balancer virgins. In the third generation, balanced males and virgin females were selected and crossed together to produce the final stable stock.

4.8 Mounting and imaging adult wings and legs

Approximately a dozen young adult males were collected into a fresh culture vial and left for a day or two, to allow them time to groom themselves clean of any detritus on the wings. The flies were then anaesthetised and forceps were used to remove the right wing as close to the hinge as possible. Taking the same wing from each fly facilitated placing the all of the wings the same way up on the slide. The wings were placed on a piece of blotting paper and the flies were discarded. A drop or two of isopropanol was dripped onto the wings from a glass pipette to dehydrate them. 14µl of Gary's Magic Mountant (GMM, 1:1 methysalicylate and Canada balsam) was placed onto the centre of a labelled glass slide. When the isopropanol had completely evaporated, the wings were transferred to the pool of GMM on the slide, taking care to avoid air bubbles. A 22x22mm coverslip was carefully lowered into place, again taking care to avoid bubbles. The slide was gently baked on a heat block overnight with a small weight (a 7ml screw-top bijou vial (Sterilin) filled with water) on top of the coverslip. Adult legs were removed and mounted using the same procedure.

4.9 Staging pupae

When third instar larvae are ready to pupate they enter the 'wandering' phase, where they climb up the walls of the culture container. At the onset of pupation the anterior spiracles are everted and movement ceases, and within 30 to 60 minutes the white larval cuticle begins to 'tan' to an orange-brown colour to form the puparium. Pupal development times are given in 'hours after puparium formation' (hAPF). A cohort of pupae synchronised at the same developmental stage can thus be obtained simply by collecting white prepupae within a 30-minute time window. White prepupae were gently collected with a damp paintbrush and placed onto the walls of fresh fly culture vials containing fly media. The vials were then incubated at 25°C for the desired time.

When pupae must be sexed, the prepupae are placed dorsal side up on a dark surface under the dissecting microscope and covered in water. Under these conditions, the male imaginal testes are clearly visible about two-thirds of the way along the anterior-posterior axis as darker, clearer circles amongst the largely white and opaque body. After sexing, the pupae must be thoroughly dried with a tissue before being transferred to the vial for incubation. If the pupae are too wet they will stick so tightly to the vial wall that it will be impossible to remove them without killing them unless they are wetted, and wet pupae will not stick to double-sided tape for dissection.

4.10 Dissection of pupae for *in vivo* imaging of the pupal wing

One hour before the desired timing was reached (ie. after 27 hours' incubation for 28h APF pupae), the pupae were removed from the 25°C environment and dissected at room temperature (20±2°C). The pupae were gently dislodged from the vial with a dry paintbrush and transferred to the lid of a 90mm plastic Petri dish (Sterilin). Small rectangles of double-sided tape were placed around the edges of the Petri dish lid, and forceps were used to arrange the pupae on the tape. For a right-handed person, placing the pupae ventral side up with the anterior pointing to about 11 o'clock is the most convenient position for dissection. The pupae must also be rolled slightly around the AP axis in the clockwise direction, so that the right wing and legs rather than the ventral midline are uppermost.

The sharp tip of a fine pair of forceps was used to make a window in the cuticle over the developing wing, without damaging the wing tissue. This was most easily achieved by very gently rubbing the sharp tip of the forceps against the cuticle until a small hole was made. The forceps could then be used to gently pick open a slit in the cuticle, following the line between the two legs closest to the wing from posterior to anterior. The forceps were then used to peel back and remove the cuticle over the wing, ensuring that the tip of the wing remained restrained behind the cuticle at the posterior end of the window. Pupae in which the wing membrane was even slightly damaged during dissection were not used. Minor damage to the legs was occasional and acceptable.

The dissected pupae were dislodged from the tape with a drop of distilled water and transferred to the middle of the Petri dish lid. The twisted corner of a tissue (Kimtech) was used to dry the pupae. The exposed wing membrane was covered with a drop of halocarbon 700 oil (refractive index 1.414, Halocarbon Products Corporation), which prevents dessication, provides a reasonably close optical match for glass and assists in holding the pupa in the correct position for imaging. The oiled pupae were arranged in a 30mm glass-bottomed dish (IWAKI, Asahi Techno Glass Corporation) with the exposed wings placed as flat as possible against the glass. Imaging began approximately one hour after the start of dissection and continued for up to two, occasionally three hours; so all images used for quantification were from animals between X and X+3 hAPF. The distribution of core proteins in 20h APF pupae started to become visibly asymmetric at around 22hAPF, so 20h imaging sessions were restricted to a maximum of 90 minutes.

4.11 Confocal imaging and measurement of wing-membrane distance

Images of live pupal wings were taken using a Nikon A1 confocal microscope. Images were 1024x1024 pixels, using a 60x 1.4NA oil immersion objective and 3x zoom, resulting in a pixel size of 70x70nm. The pinhole was adjusted to 1.2 Airy units (AU) – usually the optimum is 1 AU, but due to the unconventional shape of the pinhole on this instrument 1.2 AU is more appropriate to achieve maximum confocality without sacrificing light transmission to the detector. 2x averaging was used, and a single image at a single focal plane was captured for each area to be imaged. At least two non-overlapping images were collected per pupa. Other imaging settings were as follows: laser transmission 18%, offset -10, gain 115, pixel dwell time 2.4 milliseconds.

The pupal wing was located and roughly focussed under brightfield illumination before switching to confocal mode. The uppermost region of wing epithelium was located using 1x zoom and a fast scanning mode to minimise exposure to laser light while selecting an area to capture. When such an area had been found, the zoom was changed to 3x and the focus was quickly adjusted. Settings were then changed to the capture mode described above and the image was acquired. The Z-position of the image in microns was recorded from the Z-stack setup utility. It was usually possible to repeat this process to acquire a second image from each pupa, depending on the position of the wing within the membrane. The cells of the wing hinge are much smaller than the rest of the wing; these areas were not imaged. Wing veins are also easily visible due to their smaller cells; these were avoided as much as possible because they are often slightly higher or lower in the Z axis than the surrounding wing cells, giving them a correspondingly higher or lower mean brightness. When veins were included in images, these cells were excluded from the puncta detection process as much as possible. After capture was completed, the focus was shifted up to the autofluorescent wing membrane to record its Z-position.

4.12 Single-fly genomic DNA extracts

Squishing buffer is 10mM Tris-HCl pH8.2, 1mM EDTA and 25mM NaCl in ddH_2O. For each extract, 1μl of 10mg/ml Proteinase K solution (Roche) was placed in a 1.5ml Eppendorf tube (Starstedt) on ice. A sacrificial adult fly was selected and anaesthetised before being transferred to the tube and thoroughly mashed with a P200 tip containing 50μl squishing buffer. The buffer was expelled into the tube after mashing the fly. The tube was then incubated for 25 minutes at 37°C, then 5 minutes at 95°C, chilled briefly on ice and centrifuged briefly to settle the solid remnants. 0.5μl of the resulting supernatant was used per 20μl PCR reaction. Extracts were stored at -20°C. When one has a choice in the matter, females are considered to give better results than males due to their larger average size.

Molecular biology protocols

4.13 *E. coli* culture

Cultures were grown in either Luria-Bertani broth (LB) in 142AS screw-top test tubes (Sterilin) or 15ml universal screw-top tubes (Sterilin/Thermo Scientific) in a shaking incubator at 37°C and 250rpm, or on LB agar plates in a 37°C static incubator, with the appropriate antibiotic(s). Tube lids were loosely fastened to allow air flow, but secured in place with adhesive tape so they were not shaken off in the incubator. The SW106 recombineering strain was always incubated at 32°C instead of 37°C. Liquid cultures for electroporation were grown in low-salt LB (standard LB recipe, omitting the NaCl) to prevent arcing in the electroporator. Plates and liquid cultures were typically inoculated in the afternoon or early evening and incubated overnight for use the following morning. Colonies were picked from plates using sterile wooden toothpicks or pipette tips onto fresh plates or into liquid cultures. A heat-sterilised wire loop was used to transfer liquid cultures or glycerol stocks onto agar plates and to spread bacteria across plates. Strains were preserved in glycerol for long-term storage by mixing the cell culture with glycerol in a 1:1 ratio in a cryovial, snap-frozen in a mixture of dry ice and ethanol, and stored at -80°C.

4.14 Restriction digestion

Restriction enzymes were obtained from New England BioLabs or Roche. Digestions were performed in 1.5ml Eppendorf tubes in the appropriate buffers according to the manufacturer's directions. 20μl reactions were incubated at 37°C for 1-2 hours, after which time heat-inactivatable enzymes were inactivated by incubating at 65°C for 15 minutes. When the restriction enzymes were not compatible, the reaction was conducted first with the enzyme with the lowest salt requirement and heat-inactivated if appropriate. The second enzyme and the appropriate buffer were then added, the volume made up to 30μl, and the incubation was repeated.

4.15 Standard PCR protocol

Polymerases used were BioTaq (Bioline) for diagnostic applications, and KOD hot-start DNA polymerase (Merck Millipore) for cloning. Reaction mixes were made according to the manufacturer's recommendations, primer stocks were diluted to 50μM and used at a concentration of 1:100.

The standard BioTaq cycling programme was 5 minutes at 95°C, [30 seconds at 95°C, 30 seconds at *A*°C, *E* seconds at 72°C] x30, 5 minutes at 72°C, hold at 12°C, where the annealing temperature *A* was typically 5°C lower than the lowest Tm of the primer pair used. The extension time *E* was calculated by allowing at least 30 seconds per kilobase to be amplified.

The standard KOD cycling programme was 2 minutes at 95°C, [20 seconds at 95°C, 10 seconds at *A*°C, *K* seconds at 70°C] x25, 1 minute at 70°C, hold at 12°C, where K is calculated by allowing 20 seconds per kb to be amplified.

4.16 DNA gel electrophoresis

Gels were typically 0.8% agarose in 0.5% TBE (45mM Tris-borate, 1mM EDTA), with the percentage of agarose increased or decreased respectively for separating particularly short or long DNA fragments. 2μl ethidium bromide (VWR International) per 35ml gel was used as a DNA dye. SYBRsafe (Invitrogen) was used instead of ethidium bromide for gels containing DNA to be gel-purified.

4.17 Purification of PCR products

PCR reactions were purified using the GeneJET kit (Thermo Scientific) according to the manufacturer's instructions. For gel purification, bands were excised from a SYBRsafe-stained agarose gel using a SafeImager light box and purified using the Fermentas gel purification kit according to the manufacturer's instructions.

4.18 Ligation

500ng of the vector and 1μg of the insert were restriction digested according to the protocol described above, with the addition of shrimp alkaline phosphatase (SAP) (NEB) to the vector reaction only. The digested fragments were gel purified using a kit (Fermentas) according to the manufacturer's instructions. 1μl of each eluate was run on a DNA gel to check the concentration of the recovered fragments. Ligations were conducted in 10μl volume, using 1μl T4 DNA ligase (Roche), 1μl 10x ligation buffer and 1μl 4mM ATP solution. 20ng of vector was used, and the insert was added in a 3:1 molar ratio of insert:vector (ie. if the insert and vector were the same size, 60ng insert was used with 20ng vector). ddH_2O was added to bring the total volume to 10μl. The ligation reaction was inclubated at 4°C overnight. 3μl of the ligation reaction was transformed by heat-shock into silver-efficiency α-select cells (BioLine).

4.19 Minipreps and midipreps

High copy number plasmids, such as pRK2 derivatives, were transformed by heat-shock into bronze-efficiency α-select cells (BioLine) and cultured using standard protocols to get a single colony growing in an appropriate volume of LB media at a reasonable density. Single-copy constructs such as P[acman] were grown in EPI300 cells; this procedure is described below in Step 9 of the recombineering protocol. Minipreps were carried out using the GeneJET kit (Thermo Scientific) and midipreps were done using the Qiagen kit, according to the manufacturer's instructions.

4.20 Transgenesis protocols

4.20.1 Recombineering

Resources and important notes:

- BACs (genomic DNA source) and P[acman] constructs (the vector to be modified) are available at BACPAC Resources (Venken et al., 2006). P[acman] and BAC constructs were restriction mapped before starting the protocol, as clones occasionally become rearranged.

- The template for the targeting DNA was the PL452 plasmid, which contains a LoxP-flanked *neo* selection cassette. PL452-N-EGFP and PL452-C-EGFP, for tagging the N- and C-termini of proteins respectively, were obtained from Addgene (Venken et al., 2008).
- The following homology arms were used for the PL452 EGFP cassettes:
    - PL452-N-EGFP 5' primer:
      ATGGTGAGCAAGGGCGAGGAGCTGTTC
    - PL452-N-EGFP 3' primer:
      ACTAGTGGATCCCCTCGAGGGACCTAATAAC
    - PL452-C-EGFP 5' primer:
      GCAGCCCAATTCCGATCATATTCAATAACC
    - PL452-C-EGFP 3' primer:
      CTTGTACAGCTCGTCCATGCCGAGAG
    - PL452-inverted C-EGFP 5':
      ATGGTGAGCAAGGGCGAGGAGCTGTTC
    - PL542-inverted C-EGFP 3':
      GCAGCCCAATTCCGATCATATTCAATAACC
- The cell line used for recombineering was SW106 (Frederick National Laboratory for Cancer Research), which carries the required phage-derived recombination machinery under a heat shock-released repressor, as well *Cre* for LoxP-mediated excision of the selection cassette.
- The SW cell lines must always be grown at 32°C, never at 37°C, so it often took more than 24 hours for colonies to reach a useful size. Incubating the cells at 37°C allows enough expression of the repressed recombinases to kill the cells. Even if the cells survive and manage to grow at 37°C they will have experienced strong selection for loss of the phage genes and will be useless for recombineering.
- The SW106 cells often grew very poorly in liquid culture when picked from a refrigerated plate, which was sometimes a problem when colonies had to be stored for a few days to allow time for testing potential recombinants. To avoid loss of precious colonies, working plates were kept at room temperature and candidate colonies were re-streaked to fresh plates as necessary. When plates were streaked on a Friday afternoon and incubated at room temperature in the lab over the weekend the colonies usually reached a good size by Monday morning.
- The targeting primers were made by Integrated DNA Technologies (IDT) using their Ultramer service. Additional primers were synthesised by IDT or Sigma.

- Antibiotics used are ampicillin at 100μg/ml, chloramphenicol at 12.5μg/ml and kanamycin at 100μg/ml.

Method:

1) Design recombineering primers

The DNA sequence around the desired insertion site in the fly genomic DNA of the target P[acman] construct was identified. Targeting primers A and B (see **Figure 2.1 A** and **B**) each had ~20nt homology to one end of the PL452 EGFP cassette and ~50nt homology to one side of the chosen insertion site in the genomic DNA. These primers were used with PL452-N-EGFP or PL452-C-EGFP to generate the targeting DNA shown in **Figure 2.1 C**. When tagging the N terminus, the targeting DNA has the structure [50bp 5' homology]-[EGFP-linker-LoxP-Kan^R-LoxP selection cassette]-[50bp 3' homology].

Primers A and B must result in a targeting DNA fragment that maintains the reading frame across the junction between EGFP and the gene of interest without introducing any Stop codons. It was sometimes necessary to add one or two nucleotides, particularly when tagging the C-terminus. The LoxP-Kan^R-LoxP cassette is removed in a later step to leave only one LoxP site, so it is the [EGFP-linker-LoxP-gene] fragment that needs to be read through. Primers C and D should lie in the unmodified genomic DNA just outside the 50bp homology regions, for amplifying and sequencing across the insertion site.

2) Make targeting DNA

Primers A and B were used to amplify from PL452-N-EGFP or PL452-C-EGFP as appropriate, using the proofreading polymerase KOD. 4 x 50μl reactions and 25 cycles yielded sufficient material for a couple of recombineering attempts. The PCR reactions were pooled, 2μl DpnI was added and the tube was incubated at 37°C for 2 hours. DpnI destroys the plasmid template but not the PCR product, as it only digests methylated DNA – ie. DNA of bacterial origin. The targeting DNA was gel purified (Fermentas kit), eluted in ddH_2O, and the final concentration was checked. The optimum amount of DNA for recombineering is 200-300ng/μl in a maximum 3μl volume.

3) Transform P[acman] construct into recombineering cells

Two days in advance of transformation, SW106 cells were streaked on a nonselective agar plate and incubated >24h at 32°C. A single colony was picked into 1.5ml low-salt LB (standard LB recipe, omitting the NaCl) (no antibiotic) and placed in a shaking incubator at 32°C and 250rpm overnight. 400μl of this overnight culture was added to 20ml LB (per transformation) in a sterile conical flask and incubated for 2-4 hours at 32°C and 250rpm, until the culture reached an OD600 of 0.4-0.7.

The cells were made electrocompetent by pelleting 15ml of culture per transformation in a 15 ml plastic Falcon tube in a chilled centrifuge, removing the supernatant and gently resuspending in 1ml chilled ddH_2O before topping up to 15ml again. The cells were pelleted and resuspended in this manner two more times. After the final resuspension in 1ml ddH_2O, the cells were transferred to a chilled 1.5ml Eppendorf tube, briefly spun down and the supernatant removed, leaving approximately equal volumes of cells and water in a total volume of <100μl.

A few ng of P[acman] DNA was added and the cells were gently resuspended, transferred to a 0.1 cm electroporation cuvette (BioRad) and electroporated using the 'EC1' setting of a BioRad MicroPulser (1.8kV). 1ml LB was immediately added, the cells were transferred to a clean 142AS culture tube, and incubated at 32°C, 100rpm for 1.5 hours to allow cells to recover and express the antibiotic resistance gene. The cells were then gently pelleted and a pipette was used to carefully remove any mucous-like layer from the top of the pellet (this appeared to consist largely of dead cells). The majority of the supernatant was aspirated and the remaining cell pellet was carefully resuspended and spread in a gradient across an agar plate containing the appropriate antibiotic to select for the P[acman] (ampicillin or chloramphenicol). The plate was Incubated >24h at 32°C.

4) Recombineer targeting DNA into P[acman] construct

Step 3 was repeated using a single SW106 colony carrying the P[acman] construct, with the following additions and modifications. Once the culture had reached the required OD600, the flask was heat shocked at 42°C for 15 minutes in a shaking waterbath, then chilled for 2-3 minutes in ice water with agitation before proceeding to pelleting and washing the cells. The heat shock releases repression of the recombinase gene expression. Before electroporation, 200-300ng of targeting DNA was added to the cells. After recovery, the transformed cells were spread on a kanamycin plate and incubated >24h at 32°C, or until normal-sized colonies appeared.

5) Select correctly-recombineered colonies

A kanamycin, an ampicillin and a chloramphenicol plate were divided into six numbered sectors each. Six candidate colonies were transferred from the recombineering plate to the kanamycin plate, one colony per sector, and to the corresponding sectors of the chloramphenicol and ampicillin plates, using a fresh sterile toothpick for each colony. A sterile wire loop was used to spread cells from each colony across the whole sector. The plates were incubated >24h at 32°C.

Correctly-recombined colonies grew on kanamycin and on the antibiotic corresponding to the original P[acman] construct, but not on the third plate. For problematic recombinations, colonies were additionally screened by PCR at this point using primers C and D. For each colony to be tested, an Eppendorf was prepared containing 400μl ddH_2O. Using a P200 set to <50μl (to avoid contamination of pipette barrel) with a fresh tip, the colony to be tested was picked and pipetted up and down in the water. 1μl of this suspension was used per 20μl PCR reaction. The suspensions were stored at 4°C while the PCR ran. To recover the bacteria for further liquid culture, the tubes were gently centrifuged and most of the water was carefully aspirated. The cells were then gently resuspended in the remaining small volume of liquid and transferred directly to a 142AS tube containing LB and the appropriate antibiotic.

6) Remove kanamycin selection cassette

One colony was picked from each correctly-recombined sector of the kanamycin plate into 1.5ml LB plus kanamycin and incubated overnight at 32°C and 250rpm. The following morning, 100μl of each overnight culture was transferred to 5ml LB (no antibiotic) in a fresh tube. These were incubated at 32°C and 250rpm for 3 hours and then L-arabinose (Sigma-Aldritch) was added to a final concentration of 0.1%. Arabinose induces expression of Cre recombinase in the recombineering cells, which causes the excision of the floxed kanamycin selection cassette.

The cultures were incubated at 32°C and 250rpm for a further 1 hour. One loop of bacteria from each candidate culture was then streaked onto a sector of an ampicillin or chloramphenicol plate as appropriate, depending on the resistance of the P[acman] construct used. Plates were incubated >24h at 32°C.

7) Select and preserve correctly recombineered colonies

A colony was streaked from each sector onto kanamycin, ampicillin and chloramphenicol plates as in step 6. The desired colonies were now the ones which only had the antibiotic resistance conferred by the P[acman] construct, and had lost the kanamycin resistance. Overnight cultures were set in 2.5ml LB plus the appropriate P[acman] antibiotic and incubated at 32°C and 250rpm overnight. The following morning, 1ml of each overnight culture was used to make a glycerol stock.

8) Recover the modified P[acman] construct and sequence the DNA

The remaining 1.5ml overnight cultures were used to make rough minipreps. P[acman] is a large, single-copy plasmid, so if a miniprep column is used then the yield is usually very low and dilute. Instead, the instructions in a commercial miniprep kit were followed until the point where the neutralised supernatant should be transferred to a column. This supernatant was instead centrifuged again and removed to a fresh tube. 750μl isopropanol was added, the tube was left to stand at room temperature for 10 minutes and then spun hard for 10 minutes. The isopropanol was removed and replaced with 1ml 70% ethanol, the tube was spun hard for 5 minutes and then the ethanol was removed. The tubes were air dried for 5 minutes before adding 50μl ddH_2O. The tubes were left to stand for a few minutes before centrifuging briefly and removing the supernatant to a fresh tube.

Primers C and D were used to amplify the modified region from the rough minipreps. The amplified fragment was purified using a commercially available kit and sequenced using the same primers. Sequences were carefully checked for point mutations or other aberrations.

9) Transfer the sequenced construct to EPI300 cells and make a high copy number miniprep

EPI300 cells (TransforMax, via epicentre) switch to using a plasmid's high-copy *oriV* on application of CopyControl solution (shipped with the cells) or L-arabinose. 1-2μl of the rough P[acman] miniprep was added to 25μl EPI300 cells and electroporated. 1ml LB was immediately added, and the cells were incubated at 37°C and 100rpm for one hour before spreading on a plate containing the appropriate antibiotic and incubating at 37°C overnight. The following day, a single colony was picked into 2.5ml LB containing the appropriate antibiotic in a 15ml universal tube and incubated at 37°C and 250rpm overnight. The following morning, 1ml of the overnight culture was used to make a glycerol stock.

9ml LB, plus the appropriate antibiotic, plus L-arabinose to a final concentration of 0.1% or CopyControl solution according to the manufacturer's instructions was added to the remaining overnight culture and incubated at 37°C and 250rpm for 5 hours. The P[acman] construct was recovered by miniprep kit (GeneJET) according to the manufacturer's instructions, with the following modifications. The 10ml culture was pelleted in a 2ml Eppendorf tube in five batches, double the usual volume of resuspension, lysis and neutralisation solutions were used, and the supernatant was loaded onto the column in two batches. Elution buffer was heated to 75°C in a water bath prior to use. DNA from this miniprep was sent to the *Drosophila* transgenesis company for microinjection.

4.20.2 Preparation of pRK2-EGFP constructs for *in vivo* gene targeting

pRK2-EGFP

EGFP was amplified by PCR from pEGFP (Clontech), using the primers to add an NdeI site followed by a Gly/Ser linker to the 5' end and a KpnI site to the 3' end. pRK2 and the amplified NdeI-linker-EGFP-KpnI fragment were digested using NdeI and KpnI and ligated together. pRK2-EGFP includes the Stop codon of EGFP.

For the *in vivo* targeting, appropriate homology arms of approximately 3kb each had to be added to pRK2-EGFP to target the recombination to the desired locus. This was achieved by amplifying the desired fragment from an appropriate BAC, using the primers to add suitable restriction sites and, where necessary, extra nucleotides to maintain the reading frame of the final construct.

For *dachsous*, a 5' homology arm consisting of 3,026bp of the final exon of *ds*, terminating immediately before the Stop codon, was amplified using primers JA16 and JA17, which add an EcoRI and an NheI site respectively. JA17 contains an additional 2bp to maintain the reading frame of the final construct. This homology arm and the pRK2-EGFP vector were digested with EcoRI and NheI and ligated together, to add the 5' homology arm to pRK2-EGFP upstream of the linker-EGFP sequence. The 3' homology arm was 3,038bp of the 3' UTR, starting immediately after the Stop codon of *ds*, generated using JA18 and JA19, digested and ligated into the pRK2-EGFP Ds 5' arm construct immediately downstream of the final LoxP site of the pRK2 EGFP-LoxP-w^+-LoxP, to give the final targeting construct ready for injection. All coding DNA in the homology arms, including the linker-EGFP sequence, was sequenced before injection.

JA16 EcoRI **Ds** C-term LH arm Fwd:
5' CGGAATTCCGAGATCAGTGAGAGTCTTTCCGTGG
JA17 NheI **Ds** C-term LH arm Rev:
5' CCATATGGCTAGCTCCATCCGTGTCCCCACA TTTCCCCTC
JA18 SpeI **Ds** 5' UTR RH arm Fwd:
5' GCGCACTAGTCGAGCACTGGAGCTGTATATCCATCC
JA19 XhoI **Ds** 5' UTR RH arm Rev:
5' CCGGCTCGAGTTCCTTTGGCCAGCCGCTGAGTTAAG

The pRK2-EGFP construct for Fz was generated using the same method but with these primers:
JA12 NheI **Fz** C-term LH arm Fwd:
5' CCATATGGCTAGCCATTTGTGTGGAATTGTCACCCAGACC
JA13 NdeI **Fz** C-term LH arm Rev:
5' GGAATTCCATATGGACGTACGCCTGCGCCCGGGTC
JA14 AscI **Fz** 3' UTR RH arm Fwd:
5' CAAGGCGCGCCTATGAGACGGGTCCGGCGGGC
JA15 XhoI **Fz** 3' UTR RH arm Rev:
5' CCGGCTCGAGCTGCAATTTGTTGCCAAGGAGCTTAGC

The 5' homology arm of the Stan-EGFP construct was a little more complicated because the 31bp final exon of Stan was re-coded and copied to the 3' end of the preceding exon.

DS515 NotI **Stan** LH arm Fwd:
GCTCTAGAGCGGCCGCGGATAACAGCCCATGAAATTCCGCCACC
DS516 NheI **Stan** LH arm Rev (recoded exon in bold, 3' end of preceeding exon in lowercase):
GGCTCGAGGCTAGCCC**GACGGTTGTTTCATCGTCGTCGATGTTGCGT**tcggtatccgtgat gcttgtcagtgg

Pk was tagged at the N-terminus, so the EGFP sequence was added to the 5' homology arm by SOE PCR (splicing by overlap extension) and ligated into the original pRK2 vector. The 5' homology arm was first generated by PCR as for the other constructs, using the PK 5' UTR as a template, including the ATG of the first Pk exon and adding a NotI site at the 5' end. At the 3' end, 24bp of the start of EGFP (excluding the ATG codon) was added instead of a restriction site. EGFP was amplified from pEGFP, using the primers to add 24bp homology to the 3' end of the homology arm at the 5' end of EGFP and a KpnI site to the 3' end. These two fragments were then used as the templates for a third PCR. The overlapping homology of these fragments primed them onto each other, splicing the two fragments together and resulting in NotI-Pk 5'UTR-ATG-EGFP-KpnI. The 3' homology arm for Pk was generated in the usual way, starting at the first codon after the ATG of the first exon of Pk, adding a 5' SpeI site and 3' AscI site. The 3' homoloy arm was ligated into pRK2 first, as the 5' homology arm contained a SpeI site.

JA34 NotI Pk LH arm Fwd:
CATAGCGGCCGCCTACTACGCCGACTACCCGGAAC
JA35 **Pk** LH arm EGFP splice Rev: (reverse complement of JA36)
GAACAGCTCCTCGCCCTTGCTCACCATGTTTGCGGAGCGGAGGATGTATCC
JA36 **Pk** LH arm EGFP splice Fwd: (reverse complement of JA36)
GGATACATCCTCCGCTCCGCAAACATGGTGAGCAAGGGCGAGGAGCTGTTC
JA37 KpnI **Pk** LH arm rev:
CCGGTACCCTTGTACAGCTCGTCCATGCCG
JA38 SpeI **Pk** first exon RH arm Fwd:
GCGCACTCTAGTCGATACCCCAAATCAAATGCC
JA39 AscI **Pk** first exon RH arm Rev:
GTTGGCGCGCCCTGGCCCAACTTGAGAGGAAGC

4.20.3 *In vivo* gene targeting

The pRK2-EGFP vector containing the appropriate homology arms was transformed into *Drosophila* via P element-mediated transgenesis to generate several insertions at random genomic loci. These transformants were balanced using the protocol described above to generate the donor lines. For targeting, it is important to use a donor line where the donor construct is not inserted on the same chromosome as the locus that will be targeted. The targeting events are identified by eye colour and during the screening step they are phenotypically indistinguishable from the original donor chromosome. Movement of the red eye rescue gene from the donor chromosome to the targeted chromosome is the first test for a correct targeting event.

The donor virgins must be crossed to the appropriate *heat shock-hid* line to generate targeting-competent offspring. The appropriate line to use depends on the chromosome being targeted: yw/Y$^{\text{hs-hid}}$;; FLP, I-SceI/TM3$^{\text{hs-hid}}$ for targeting the second chromosome, or yw/Y$^{\text{hs-hid}}$; FLP, I-SceI/CyO$^{\text{hs-hid}}$;; for targeting the third chromosome. Crossing the donor virgins to the *FLP, I-SceI* males produces offspring carrying both the donor chromosome and *FLP, I-SceI*, which will be mosaic due to the patchy excision of the targeting DNA. The *FLP* and *I-SceI* genes excise and linearise the targeting DNA from the donor chromosome, allowing it to potentially find and recombine with its homologous genomic locus to produce the desired transgenic fly.

The *FLP, I-SceI* males used in this initial cross facilitate the collection of mosaic virgin females by passing *hs-hid* to their sons (on the Y chromosome) and any offspring that carry the balancer chromosome rather than the *FLP, I-SceI* chromosome. Expression of *hid* kills the flies by inducing apoptosis[1], so the heat shock results in the death of all male offspring and all non-mosaic offspring thus leaving only the desired mosaic females. Note that the *FLP, I-SceI* stocks, particularly the males, do not survive at 25°C, probably due to leaky *hid* expression. These stocks were consequently cultured at 19°C with added yeast grains, and a large piece of filter paper in the media to assist the flies' survival. Under these culture conditions these stocks were generally healthy, as long as the number of males in the stocks was monitored carefully when transferring the adults to fresh media.

To generate the mosaic, targeting-competent virgins, a large number of donor females (50-100 or more) were crossed in bottles to as many *FLP, I-SceI* males as could be collected, until the sex ratio approached 1:1 – this was necessary, as the males are not terribly healthy. The bottles were incubated at 19°C while the parents were laying. Once larvae had started to emerge the adults were transferred to new media every two days to continue laying. The bottles were heat shocked every day for 2 hours at 38°C from one day after removal of the parents until the first pupae appeared, after which the

---

[1] *hid* stands for *head involution defective*. This gene is also called *Wrinkled* (*W*), and was first published in the delightfully-named paper "Mutations observed in *Drosophila* stocks taken up into the stratosphere", V. Jollos 1936.

bottles were incubated at 25°C. The only offspring from this heat-shocked cross were the mosaic females, which will all be virgins due to the larval death of their male siblings.

The efficiency of targeting DNA excision in these virgins can be assessed by the degree of mosaicism in the eye, as the targeting DNA carries a *GMR>white* rescue gene that gives a strong red eye colour in a *white* genetic background. The crossing and heat shocking schedule described above usually resulted in a population of adult females with almost entirely white eyes with just a few reddish ommatidia remaining, indicating that the targeting cassette was being excised efficiently. However, targeting was also achieved using females with a red:white ratio of closer to 50:50 by eye area.

The mosaic virgins were then crossed to generate several thousand offspring for screening. The targeting DNA also contains *UAS>Rpr* to reduce the rate of false positives. When the targeting DNA recombines with its homologous locus, *UAS>Rpr* is lost. Crossing the mosaic virgins to males carrying ubiquitously-expressed *Gal4* thus kills any offspring which retain *UAS>Rpr*, either through non-excision or incorrect reintegration of the targeting DNA. Occasional *UAS>Rpr* 'escapers' were seen – these are red-eyed flies which have the appearance of a pharate adult, and usually die young with uninflated wings and a largely unpigmented body. The donor lines were routinely tested for *Rpr* function before use in the targeting protocol by crossing to the Gal4-expressing line and checking that all offspring were white-eyed. Healthy escapers were very rarely seen in these initial donor chromosome test crosses.

For the actual screening crosses, three to six mosaic females were crossed with three *Gal4* males in numbered vials and incubated at 25°C. Up to 80 such vials were set for each experiment, and the parents were flipped into fresh vials every 3-4 days until either all the females had died or they stopped producing larvae. Using this schedule it was estimated that each female produced an average of 200 offspring, resulting in up to 64,000 screenable offspring per experiment.

The vast majority of offspring from the screening crosses were not 'hits', and have white eyes – it is much easier for the targeting DNA, and its associated red eye rescue gene, to be excised than it is to be re-integrated into the genome. Targeting candidate flies are easily identified by their bright red eyes, and these individuals are collected and crossed singly to an appropriate balancer stock. Only lines in which the red eye colour segregates to the targeted chromosome were kept, all others were discarded.

Table 4.1 below shows the efficiency of targeting for the four targeted constructs made in the course of this project.

Table 4.1: ***in vivo*** **gene targeting screening results**

Construct	Mosaic virgins crossed	Flies screened	Number of w+ flies	Number of w+ fertile and on targeted chromosome	Flies screened per hit
Fz-EFGP	121	14,500	19	14	1,040
EGFP-Pk	65	12,000*	24	17	700
Stan-EGFP	198	23,800	18	17	1,400
Ds-EFGP	516	62,000	16	8	7,750

*Screening stopped due to large number of verified hits obtained.

The final step of the gene targeting protocol was to remove the red eye marker gene. When tagging the N-terminus of a protein this is essential, otherwise the construct cannot be read through from the EGFP sequence into the gene's cDNA. For the C-tagged constructs, EGFP expression was observed while the *white* rescue gene was still in place (and this expression was used to verify the different targeted lines), but the marker could potentially have been affecting mRNA stability or processing, so it is good practice to remove it. This is done by crossing a male from the targeted stock to virgins expressing *Cre* on the X chromosome. *Cre* works on LoxP sites in the same way as *FLP* works on FRT sites, and the *GMR>white* rescue cassette is duly excised. To establish stable white-eyed targeted stocks, five white-eyed males (technically mosaic for *GMR>white* excision) are crossed singly to balancer flies, and then a single male offspring is backcrossed to the balancers. This two-generation scheme ensures removal of *Cre* from the final stock. This process often seems to result in lethality or sterility, which is why five white-eyed males are chosen to establish lines for each stock to be flipped.

4.21 MatLab script used for image processing

Black = functions and variables.

Pink = strings, for display or to specify options within a function

Green = comments, not part of the code.

```
clear all
clc

% Open original image: use bfopen function from Loci Bioformats.
imdata = bfopen;

% It's fun to time the process! Usually around 10 seconds for the
first image and about 5 seconds for subsequent images.
tic;

disp('Postprocess is working, please wait.')

% Bioformats opens the image as a rather complicated 1x4 cell array
which itself contains several cell arrays. We need to extract the
actual image data from this unwieldy beast.

originalimage = imdata{1};
originalimage = originalimage{1};

% Pre-processing steps:
% Convert from 16-bit to 8-bit (my images are 12-bit, but MatLab can
only handle 8-bit and 16-bit, nothing in between. Not sure whether
PackingAnalyzer can handle 16 bits. 8-bit is sufficient for all the
processing stuff and there's no need to use more bits than we need.
BUT. Can't convert directly to 8-bit using uint8, as this just clips
all pixel values greater than 255 to 255, without rescaling them. The
best solution I've % found is to manually rescale the pixel values
before conversion (4095 in 12-bit is equal to 255 in 8-bit, 4095/255 =
0.0625).
Iorig = 0.0625 .* originalimage;
Iorig = uint8(Iorig);

% 1) Linear contrast expansion to stretch brightness range, saturating
top 0.1% and bottom 0.1% of data. Has little effect on bright images,
but makes dim images much easier to analyse. Doing this before the
noise reduction steps preserves brightness resolution, which I think
is more important for dim images than amplifying the noise. Usually
recommend to contrast adjust AFTER noise removal, as noise removal
decreases overall brightness of image.
I = imadjust(Iorig, stretchlim(Iorig, [0.001 0.999]));

% 2) Gentle Gaussian smoothing to even out the noise:
I = imfilter(I, fspecial('Gaussian', 5, 0.75));

% 3) Median filter to further reduce noise while preserving edges:
I = medfilt2(I, [5 5]);

% Sorting out filename and path:
impath = imdata{1};
impath = impath{2};
dot = strfind(impath, '.');

% Tell MatLab where the membrane and cytoplasm are:

% Load cell skeleton image:
skelpath = impath(1:dot-1);
skelpath(dot:dot+21) = '_pp/handCorrection.png';
skel = imread(skelpath);

% Convert skeleton from m * n * 3 RGB .png to logical:
skel = rgb2gray(skel);
skel = logical(skel);
```

```
% Packing Analyzer adds a border to the skel image, which we need to
remove so that we can fill in the holes to get a mask of the cells
(otherwise imfill (used later) will fill the entire image).
skel = imclearborder(skel);

% Dilate skeleton to width of membrane:
skelfixed = imdilate(skel, strel('disk', 5, 0));

% Clean up skeleton: 'diag' operation eliminates 8-connectivity of bg,
so prevents small loops forming when subsequently thinned. Thinning of
dilated skel plus removal of spurs sorts out any dodgy cell vertices
rather nicely. This step is essentially a rather coarse-grained
closing operation.
skelfixed = bwmorph(skelfixed, 'diag');
skelfixed = bwmorph(skelfixed, 'thin', Inf);
skelfixed = bwmorph(skelfixed, 'spur', Inf);

% Re-dilate to get membrane mask:
memb = imdilate(skelfixed, strel('disk', 5, 0));

% Now fill in the cells and subtract memb, this leaves just the
cytoplasm.
cyto = imfill(memb, 'holes');
cyto = cyto & ~memb;
cyto = bwmorph(cyto, 'clean');

% Image registration: splits the image into 16 slices, based on
brightness (as far as I can tell, this doesn't try to put the same
number of pixels in each bin)
X = grayslice(I, 16);
figure, imshow(X, colormap(jet(16))); colorbar

% *** CHANGE THIS VALUE TO YOUR PIXEL AREA ***
pixelarea = 0.0049;


% ***   Puncta detection.   ***
%
% Start loop here...

    for i = 15:-1:7;

currentslice = X==i;
currentslice = bwmorph(currentslice, 'bridge');

% Detect objects:
OBJECTS = bwconncomp(currentslice, 8);

% Make label matrix for the detected objects:
LABELS = labelmatrix(OBJECTS);

% Measure properties of the detected objects:
PROPS = regionprops(OBJECTS, 'Area', 'Eccentricity', 'EulerNumber',
'Perimeter');

% Keep objects with less than or equal to one hole. The Euler number
is the number of objects in the image, minus the number of holes in
the objects. Puncta with no holes = 1. Puncta with >1 hole have E<0.
keep1 = find([PROPS.EulerNumber] >=0);
wip1 = ismember(LABELS, keep1); % wip = knitting term, 'Work In
Progress'
```

```
% Circularity measure: Perfect circle has value of 1. Very non-
circular things tend towards 0. Values greater than 1 are possible
because of the way that regionprops calculates perimeters, and this
has a disproportionate effect on small objects. If all your objects
with circ >>1 are very small then this is ok! Ref:
http://www.mathworks.co.uk/matlabcentral/newsreader/view_thread/292521
circs = (4 * pi * [PROPS.Area]) ./ ([PROPS.Perimeter].^2);
keep2 = find(circs >0.7);
wip2 = ismember(LABELS, keep2);

% Results will be dumped into cell matrix P:
P{i} = wip1 & wip2;

    end

% Collate results (there must be a prettier way to do this):
puncta = P{15} | P{14} | P{13} | P{12} | P{11} | P{10} | P{9} | P{8} |
P{7};

% Clean up puncta image: Remove any puncta which fall outside the
membrane mask. Join any features separated by just 1 pixel, remove
single-pixel points, and fill any holes. Finally, remove any puncta
which are still smaller than 9 pixels.
puncta = puncta & memb;
puncta = bwmorph(puncta, 'bridge');
puncta = bwmorph(puncta, 'clean');
puncta = imfill(puncta, 'holes');
puncta = bwmorph(puncta, 'close');
puncta = imfill(puncta, 'holes');
puncta = bwareaopen(puncta, 9);
figure, imshow(puncta);

% Save puncta image, in case it's needed later.
ppath = impath(1:dot-1);
ppath(dot:dot+10) = '_puncta.png';
imwrite(puncta, ppath, 'png');


% *** Other stats not used: ***

% Eccentricity is sort of a measure of circularity (which will not
work well on objects of just a few pixels...), a circle has 0
eccentricity and a very elliptical shape has an E approaching 1.
% keep3 = find([PROPS.Eccentricity]<0.87);
% wip3 = ismember(LABELS, keep3);

% Perimeter:area ratio: round things should have a P/A of <2ish. For
% circles, P/A decreases as A and P increase.
% keep4 = find(([PROPS.Perimeter] ./ [PROPS.Area]) <1.4);
% wip4 = ismember(LABELS, keep4);



% *** Puncta stats ***

% bwconncomp detects and labels every discrete patch of foreground
pixels.
PUNCTA = bwconncomp(puncta);

% regionprops extracts the requested properties of each connected
component (=feature) into a structure array. Adding the
'originalimage' argument makes it automatically use the pixel values
from original raw image, so these values are all 'clean' - taken
```

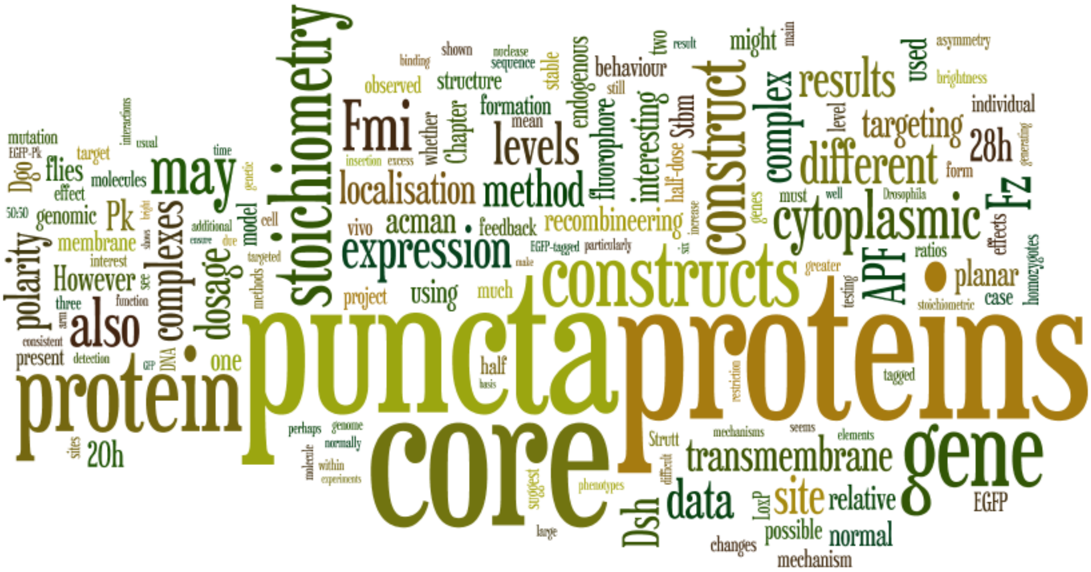
```
directly from the original unprocessed confocal data. Structure arrays
must then be converted back to plain matrices for export. Nb. Can't
interface directly between MatLab and Excel on a Mac.
Parea = regionprops(PUNCTA, 'Area');
Parea = cell2mat(struct2cell(Parea));
Pareatotal = pixelarea * sum(Parea(:));
Pareastd = pixelarea * std(Parea(:));
Parea = pixelarea * mean(Parea(:));

% Pperimeter = regionprops(PUNCTA, 'Perimeter');
% Peccentricity = regionprops(PUNCTA, 'Eccentricity');

Pmean = regionprops(PUNCTA, originalimage, 'MeanIntensity');
Pmean = cell2mat(struct2cell(Pmean));
Pmeanstd = std(Pmean(:));
Pmean = mean(Pmean(:));

Pmax = regionprops(PUNCTA, originalimage, 'MaxIntensity');
Pmax = cell2mat(struct2cell(Pmax));
Pmaxstd = std(single(Pmax(:)));
Pmax = mean(Pmax(:));

Pmin = regionprops(PUNCTA, originalimage, 'MinIntensity');
Pmin = cell2mat(struct2cell(Pmin));
Pminstd = std(single(Pmin(:)));
Pmin = mean(Pmin(:));

Pn = PUNCTA.NumObjects;
punctastats = [Parea, Pareastd, Pmean, Pmeanstd, Pmax, Pmaxstd, Pmin,
Pminstd, Pn];



% *** Cytoplasm stats: ***

CYTO = bwconncomp(cyto);

Carea = regionprops(CYTO, 'Area');
Carea = cell2mat(struct2cell(Carea));
Careatotal = pixelarea * sum(Carea(:));
Careastd = pixelarea * std(Carea(:));
Carea = pixelarea * mean(Carea(:));

% Cperimeter = regionprops(CYTO, 'Perimeter');
% Ceccentricity = regionprops(CYTO, 'Eccentricity');

Cmean = regionprops(CYTO, originalimage, 'MeanIntensity');
Cmean = cell2mat(struct2cell(Cmean));
Cmeanstd = std(Cmean(:));
Cmean = mean(Cmean(:));

Cmax = regionprops(CYTO, originalimage, 'MaxIntensity');
Cmax = cell2mat(struct2cell(Cmax));
Cmaxstd = std(single(Cmax(:)));
Cmax = mean(Cmax(:));

Cmin = regionprops(CYTO, originalimage, 'MinIntensity');
Cmin = cell2mat(struct2cell(Cmin));
Cminstd = std(single(Cmin(:)));
Cmin = mean(Cmin(:));

Cn = CYTO.NumObjects;
```

```
cytostats = [Carea, Careastd, Cmean, Cmeanstd, Cmax, Cmaxstd, Cmin,
Cminstd, Cn];


% *** Membrane stats ***
% Subtract outer ring from memb so each cell has the same amount of
membrane around it. Edge of membrane is defined by skel, which runs
through the centre of the membrane. This is badly explained, sorry.
Membrane stats INCLUDE the puncta. Membrane is a single fused object.

membrane = imclearborder(~skelfixed, 4);

% If you want membrane stats per cell then remove '+ skelfixed' from
this line:
membrane = membrane - cyto + skelfixed;

MEMBRANE = bwconncomp(membrane);

Marea = regionprops(MEMBRANE, 'Area');
Marea = pixelarea * cell2mat(struct2cell(Marea));

% Mperimeter = regionprops(MEMBRANE, 'Perimeter');
% Meccentricity = regionprops(MEMBRANE, 'Eccentricity');

Mmean = regionprops(MEMBRANE, originalimage, 'MeanIntensity');
Mmean = cell2mat(struct2cell(Mmean));

Mmax = regionprops(MEMBRANE, originalimage, 'MaxIntensity');
Mmax = cell2mat(struct2cell(Mmax));

Mmin = regionprops(MEMBRANE, originalimage, 'MinIntensity');
Mmin = cell2mat(struct2cell(Mmin));

Mn = MEMBRANE.NumObjects;
membranestats = single([Marea, 0, Mmean, 0, Mmax, 0, Mmin, 0, Mn]);



summary = [punctastats, cytostats, membranestats];


% The following code writes the results to a tab-delimited text file,
which can then be copied or imported into Excel. Don't fiddle with
this please, it took ages to figure out!

writepath = impath(1:dot);
writepath(dot+1:dot+3) = 'txt';

data = fopen(writepath, 'w');
fprintf(data, '%s', writepath);
fprintf(data, '\n');

cols = {'PArea', 'PStDev', 'PMean', 'PStDev', 'PMax', 'PStDev',
'PMin', 'PStDev', 'Pn', 'CArea', 'CStDev', 'CMean', 'CStDev', 'CMax',
'CStDev', 'CMin', 'CStDev', 'Cn', 'MArea', 'MStDev', 'MMean',
'MStDev', 'MMax', 'MStDev', 'MMin', 'MStDev', 'Mn',};
fprintf(data, '%9s\t %9s\t %9s\t %9s\t %9s\t %9s\t %9s\t %9s\t %9s\t
%9s\t %9s\t %9s\t %9s\t %9s\t %9s\t %9s\t %9s\t %9s\t %9s\t %9s\t
%9s\t %9s\t %9s\t %9s\t %9s\t %9s\t %9s\n', cols{:});
```

```
myformat = '%9.4f \t %9.4f \t %9.4f \t %9.4f \t %9.4f \t %9.4f \t
%9.4f \t %9.4f \t %9.4f \t %9.4f \t %9.4f \t %9.4f \t %9.4f \t %9.4f
\t %9.4f \t %9.4f \t %9.4f \t %9.4f \t %9.4f \t %9.4f \t %9.4f \t
%9.4f \t %9.4f \t %9.4f \t %9.4f \t %9.4f \t %9.4f';
rows = {'Puncta'; 'Cyto'; 'Membrane'};

fprintf(data, myformat, summary);

fclose(data);


% End timer.
toc
```

Chapter 5

Discussion



Contents

5.1 Introduction

The primary aim of this project was to elucidate the relative stoichiometry of the core proteins in puncta in the *Drosophila* pupal wing. To achieve this goal, EGFP-tagged core protein constructs were generated using two transgenesis techniques that should preserve endogenous expression levels. An image quantification method was devised and validated by testing on different genetic dosages of the EGFP-tagged core proteins. The stoichiometry of the core proteins at 28h after puparium formation (APF) and 20h APF was then determined. This discussion will attempt to put these results into the wider context of the planar polarity field and highlight their strengths and limitations.

5.2 Transgenesis methods for endogenous expression levels in *Drosophila*

Chapter 2 described the steps taken to try to ensure endogenous expression and normal behaviour of the EGFP-tagged core proteins. Recombineering involved adding the fluorophore gene to a genomic rescue construct, which was then inserted into the fly genome. Gene targeting was achieved by generating a donor construct that was inserted into the fly genome and then mobilised to recombine with the desired locus.

While both methods yielded satisfactory results, each had their pros and cons. The speed and efficiency of recombineering was ideal for generating many different tagged versions of a gene, but the final product was genetically sub-optimal. Injection of a genomic rescue construct meant that all of the genes contained within the construct were duplicated in the transgenic flies, and normal dosage was only restored for the gene of interest. No overt phenotypes were observed in any lines that could be attributed to such duplications, but it is possible that some subtle effects were overlooked.

The P[acman] constructs used represent a tiny proportion of the genome, but when recombineering it would be wise to thoroughly inspect the additional genes included in the chosen P[acman] construct to check for potential problems. For example, if a gene at the edge of the rescue construct is truncated there is the possibility that expression of this truncated gene could cause dominant-negative effects. Overexpression phenotypes are unlikely from the doubling of gene dosage unless an affected gene happens to be particularly sensitive to overexpression. None of the P[acman] constructs caused planar polarity phenotypes when in the wild-type background. When a gene is deliberately overexpressed using the UAS-Gal4 system the level of expression is generally many-fold higher than endogenous.

P[acman] constructs were chosen or designed on the basis that they included the sequence upstream and downstream of the gene of interest, at least up to the next verified gene (i.e. one with experimental data to support its existence, rather that one with only a GC number and a computer prediction). This method does not necessarily result in the inclusion of all regulatory elements for the gene of interest, as genes can lie in the introns of other genes and enhancer elements can lie many kilobases away from their targets. Additionally, the gene annotation in FlyBase may be inaccurate and is constantly subject to review. The gene of interest may also have unidentified additional exons. For example, the annotation of *fmi* was updated after the EGFP construct was made to include a non-coding exon almost 30kb upstream of the start of the first coding exon. Before this update there were roughly 50kb upstream of *fmi* until the next gene. The P[acman] construct used for Fmi was very large and fortunately included this new exon, but this illustrates the danger of choosing how much flanking DNA to take into the construct.

When recombineering, it is also necessary to select a genomic landing site. Controlling the insertion site is desirable, because one can select a site that is known to perform well and not suffer from position effects. Standardising the insertion site for a range of constructs also standardises any position effects, giving more internally consistent results. However, good expression is not the same as endogenous expression. The P[acman] constructs are relatively large and should be protected from the effects of any enhancers around the insertion site, but this is not guaranteed, as enhancers can work across long stretches of DNA. Markstein et al. (2008) used the gypsy retrovirus insulator sequence to suppress the effects of any neighbouring enhancers at insertion sites. Using this genetic element with a luciferase reporter, they found that expression levels were greatly increased above the normal level seen from several *attP* sites, including *attP40*, which was used extensively in this project. They conclude that Drosophila genes are generally controlled by repressive elements, which the insulator suppresses. The insulator could therefore also be preventing the normal control of gene expression. On balance, it may be advisable to avoid introduction of additional genetic elements when attempting to preserve endogenous expression levels.

The location of the landing site relative to the gene(s) of interest must also be considered, as the inserted transgene must be balanced or recombined with a null mutant for the gene of interest to restore normal gene dosage. Overall, recombineering is an efficient and powerful technique, but necessitates some fussing over the final fly stocks.

By contrast, the *in vivo* targeting protocol provides an equally powerful method for manipulation of a gene, but generating the final fly lines is much more labour-intensive. This means that fewer variations of the same construct can be produced in a given timeframe, so if the construct does not behave as expected it is much more difficult to try again with a modified donor construct. Gene targeting can also be applied to a gene of any size. Recombineering is well-suited to genes of less than 20kb, but is more difficult for larger genes. As shown for Fmi, it is possible to obtain transformants for recombineered constructs of up to 120kb, but, for this construct, the additional technical challenge of working with such large pieces of DNA dwarfed the benefits. The four targeted constructs generated in the course of this project all mobilised and recombined relatively painlessly. However, this is likely to be more due to good fortune than anything else, as colleagues have experienced greater difficulty with targeting protocols.

In conclusion, the targeting protocol gives a better, cleaner genetic result and should be favoured wherever possible. The extra time invested in a targeting protocol is offset by the lack of complications from landing site selection and gene dosage. However, if there is any question of whether the intended construct will function *in vivo*, or multiple different versions are required, then recombineering should be used instead.

Recombineering and gene targeting were chosen for this project because these methods allow preservation of the endogenous gene structure and its genomic context. Both of these can have severe effects on the expression level of a transgenic construct, and great care was taken to ensure that the expression and behaviour of the tagged gene and its protein produce was disturbed as little as possible. It was shown in Chapter 2 that the resultant tagged proteins were able to function grossly as normal, with no polarity phenotypes observed in the transgenic flies.

Despite this, the C-terminal tagged recombineered core proteins showed varying degrees of abnormal localisation. Since the constructs were designed to avoid this, and the alternatively-tagged versions based on the same parental P[acman] construct were able to localise normally, it is reasonable to suppose that these effects were caused by some problem specific to these particular constructs. For Stbm-LoxP-EGFP and Dsh-LoxP-EGFP, which did not fluoresce, moving the LoxP site to a position downstream of the EGFP stop codon resulted in constructs that expressed a fluorescent product. However, the Dsh-EGFP-LoxP construct had very abnormal localisation, accumulating in bright cytoplasmic puncta as well as at the usual apicolateral location. The Dgo-LoxP-EGFP construct had only a very subtle defect in puncta morphology, but this was also resolved by moving the LoxP site.

It seems that in each case having the LoxP site in the junction between the target gene and the fluorophore had some interfering effect on the protein or its expression. Since the N-tagged constructs did not appear to be affected by the presence of the LoxP site it is unlikely that some inherent property of the LoxP RNA or its amino acid translation was responsible for this effect.

To avoid such complications in future, it might be better to use a counterselection technique to insert the fluorophore without leaving any extraneous sequences behind. Wang et al. (2009) published a modified recombineering protocol that uses two selectable markers in the targeting DNA. As for normal recombineering, the presence of the targeting cassette can be selected for using a kanamycin resistance gene. The second marker, *RpsL*, can be selected against, as it prevents growth of colonies on streptomycin media even in a cell line that is inherently resistant to streptomycin. Most of the *E. coli* strains used for molecular biology are resistant to streptomycin, including the SW106 recombineering cells.

Homology arms are added to the counterselection targeting cassette to direct it to recombine with the locus to be modified. The resulting targeting DNA is recombineered into the P[acman] construct as usual. A colony that is both kanamycin resistant and streptomycin sensitive is selected. This colony is used for a second round of recombineering, where a ‘repair’ DNA fragment (or synthesised oligonucleotide), which encodes the desired changes to be made to the construct, is transformed into the cells. Recombination of the repair DNA with the P[acman] makes the desired change while simultaneously removing the counterselection cassette. The correctly recombined colonies are now kanamycin sensitive and have regained their original streptomycin resistance. The counterselection method allows precise genetic modifications without leaving any extraneous sequence behind in the finished construct, so would permit LoxP-free constructs to be made by recombineering.

A LoxP site is present in the targeted constructs too, but these all localised normally. On the other hand, there were no alternative versions of these constructs available for comparison, so any subtle abnormalities of protein localisation might have been overlooked. In Fz-EGFP and Stan-EGFP, the LoxP site is downstream of the fluorophore stop codon, so based on the above recombineering results it would not be expected to cause any problems. However, the EGFP-Pk construct has the LoxP site in between the fluorophore and the core protein. Then again, this is an N-terminal tagged construct, and the N-tagged P[acman] constructs all behaved normally. All constructs additionally contained a Gly/Ser linker sequence between the fluorophore and the core protein gene, but this was common to all constructs and did not appear to affect expression or localisation.

Roy and Hart (2010) report an increase in unrelated lethal mutations in their *in vivo* gene targeting experiments, suggesting that this method facilitates generalised random mutagenesis at other genomic loci. This is something that should be monitored in gene targeting protocols. The hits obtained in the four targeted constructs made in this project varied in their rates of acquisition of lethality, but viable targeted chromosomes were obtained for every construct. Additional viable non-specific mutations may also be present on these chromosomes, but no overt defects were observed in the flies.

In conclusion, recombineering and gene targeting are both useful methods for generating endogenously-expressed transgenic constructs. These methods are relatively straightforward to use, and, particularly for recombineering, do not take significantly longer to produce than traditional transgenes under heterologous promoters. To ensure experimental results that are physiologically relevant it might be advantageous to use these protocols as standard. Many results show that each core proteins reacts to the relative levels of the others (e.g. Strutt et al., 2013), so it may be beneficial to do more experiments with endogenously-expressed core protein constructs.

Beyond preserving the endogenous gene structure and the upstream and downstream sequences it is difficult to suggest any other means that might ensure endogenous expression behaviour of fusion proteins. In most cases, the identity of the promoter and any associated enhancer sequences is unknown. Another way to preserve normal fusion protein behaviour would be to take even more care over placement of the fluorophore within the target protein, or to make several different constructs with different positioning of the tag relative to the target protein. However, the effectiveness of this method is limited by what is known about the molecular function(s) of the target protein in question.

5.2.1 Recent updates to recombineering and gene targeting

Some upgrades to and variations of the recombineering protocol have recently been published, any of which could be useful for further work. Mutagenesis via serial small mismatch recombineering (MSSMR) combines retrieval of genomic DNA from a BAC into a P[acman] construct while simultaneously making a mutation (Jacobs et al., 2011). Left and right homology arms for the retrieval are added to an empty P[acman] vector by traditional restriction enzyme based-cloning. The region to be mutated is amplified from the BAC by PCR, using the primers to make the desired changes and to add restriction sites to each end. This 'mutation arm' is cloned into the P[acman] construct in between the homology arms. Two rounds of retrieval then take place. The P[acman] is first cut using the restriction enzyme on one side of the mutation arm and retrieves the matching section of genomic DNA by recombination between the mutation arm, the homology arm and the BAC. The process is then repeated using the restriction site at the other end of the mutation arm to retrieve the other section of genomic DNA.

A set of three techniques utilising sequence-specific zinc-finger or transcriptional activator-like effector (TALE) based nucleases, termed 'genome editing' may also prove useful (reviewed in Gaj et al., 2013). The nuclease has a non-specific DNA cutting module and a sequence-specific DNA-binding module that targets the activity of the nuclease to a specific location. By altering the DNA-binding module, combining elements from a tested library of zinc finger modules that target specific codons, the nuclease can be targeted to any genomic locus in a controllable manner. When combined with an oligo or targeting construct encoding the desired mutations, the targeted double-strand breaks induced by the sequence-specific nuclease can result in homologous recombination at higher frequency than the usual homologous recombination mediated methods. Successful genome editing by CRISPR (clustered regulatory interspaced short palindromic repeats), which uses a targeting RNA to direct the activity of Cas9, a CRISPR-associated nuclease, has recently been reported in *Drosophila* (Gratz et al., 2013).

Gao et al. (2008) describe a method combining homologous recombination with ΦC31-mediated site-specific integration suitable for generating many different modifications at a single locus. An *attP* site is introduced to the target locus by *in vivo* gene targeting using the same method described in Chapter 2, which is then used as a landing site for plasmids encoding the desired mutations. Huang et al. (2009) report satisfactory results using this method, but as it alters the gene structure more than just adding in EFGP it may have unintended consequences for gene expression.

5.3 Fluorophore expression *in vivo*

As discussed in Chapter 2, EGFP is a good fluorophore for *in vivo* applications due to its very stable barrel structure, high quantum yield and favourable folding kinetics. The performance of the image quantitation method and potential improvements are also discussed in Chapter 2. The main improvement in fluorophore selection would be to use a monomeric mutant of EGFP, to eliminate any possibility of fluorophore dimerisation affecting the results. Dimerisation of GFP can be ablated by a single amino acid substitution (Zacharias et al., 2002). Although the rough calculations shown in Chapter 2 suggest that fluorophore dimerisation is unlikely to have a large effect on the results, it is always best to avoid such issues in the first place.

One issue that should perhaps be considered is whether fusing another protein to EGFP affects its brightness. If this effect exists, it could happen to different degrees with different proteins attached, which would make comparisons between EGFP on different proteins invalid due to the unmeasurable variation in EGFP brightness. However, the GFP molecule is a very stable barrel structure and the fluorochrome lies within the barrel (Ormö et al., 1998). It is therefore conceptually unlikely that attaching something to the end of the barrel-forming chain would affect the fluorescence. The linker sequence that was included in every construct between the EGFP and the core protein should allow enough flexibility in the protein:GFP junction that the GFP molecule will not be affected by its core protein partner. Quenching due to many EGFPs being present in close proximity has not been observed (Geng et al., 2008).

While the EGFP constructs were satisfactory, the constructs that used Cerulean, TagBFP, tdTomato and mCherry were disappointing. While the mCherry constructs, particularly mCherry-Dgo, appeared to localise normally, the signal was too weak to give high resolution images. In the case of TagBFP, the signal was generally strong but abnormal localisation was seen. Cerulean is a mutation of GFP that has a very similar structure, so with more efficient excitation these constructs may yet be useful. It may also be possible to improve the red-emitting constructs, using different red fluorophores. In particular, mApple has a similar extinction coefficient to mCherry, with a higher quantum yield and an emission maximum close to the 561 nm laser line commonly found on confocal microscopes (Shaner et al., 2004; Day and Davidson, 2009).

With bright constructs in two different colours, further interesting studies would be possible. FRET (Förster resonance energy transfer) can occur between two compatible fluorophores in close proximity; the energy emitted when one fluorophore is excited is transferred to the other and makes it fluoresce (Clegg, 2009). FRET experiments could confirm which core proteins are in close contact at various developmental stages and perhaps reveal more about the structure of the core complex. Experiments to study colocalisation or cotrafficking of core proteins *in vivo* would also benefit from two or even three-colour imaging, especially when used with a super-resolution microscope. Even the poor-quality dual-colour image shown in **Figure 2.11 G**, **H** and **I** reveals an interesting aspect of Fz and Dgo localisation (or lack thereof) in cytoplasmic spots. This is an area with much scope for further study that is currently restricted by the lack of suitable coloured constructs.

5.4 Image quantification

The method developed to measure the puncta seems robust. The pre-processing steps outlined in Chapter 3 greatly facilitate puncta detection, although the script could be improved with some minor adjustments. The data for each genotype fit the nonlinear regression model well, which solved the wing-membrane distance problem. The stoichiometric ratios are the same at the three wing-membrane distances used for interpolation at 28h APF, so the point at which the graph is interpolated within this range does not affect the results. The major remaining question is whether the MatLab script detects and measures a representative portion of each individual punctum and the puncta in the image as a whole.

The results from Chapter 3 suggest that this is the case, as the puncta measurement produced coherent data. An interesting test of the puncta measurement method, which was not conducted due to time restraints, would be to investigate the effect of changing the puncta detection parameters on the measured stoichiometry. The current puncta detection method is quite stringent. A more lenient (and probably more accurate) method, which would be more technically challenging to implement, would be to define the edge of the punctum by the point at which its fluorescence equals the local background.

The imaging method used did not give sufficient resolution to properly elucidate the physical structure of puncta. The material at the periphery of puncta may represent more diffuse material consisting of various ratios of individual molecules that are being added or removed from the main core of the punctum. If this is the case then the edges of puncta may have different stoichiometric ratios to the stable material in the centres.

5.5 Implications of the dosage testing

The hand-measured Pk preliminary data showed that half-dose puncta were the same size as in the EGFP-Pk homozygotes, but half as bright. Since this manual puncta detection method is more reliable than the automated detection, these measured puncta areas should be accurate. This result was duplicated in the automated measurement of EGFP-Pk and EGFP-Pk / *pk-sple*^{13} images, where the half-dose puncta were half as bright as those of the homozygotes and the same as the puncta of 50:50 flies, which have half EGFP-Pk and half endogenous Pk. The same effect was also observed in the EGFP-Dgo dosage testing.

This must mean that the puncta in these half-dose flies are somehow half as ‘dense’ as those in the homozygotes. These flies display the normal polarised core protein localisation at 28h APF. This suggests that there is indeed less Pk and Dgo in the half-dose wings than in the homozygotes, implying that there is no mechanism for increasing cellular Pk or Dgo levels when the gene dosage is halved. Strutt et al. (2013) showed that Pk is usually subject to constant degradation in the proteasome, so it seems that this process is not ameliorated by halving the gene dosage. It would be interesting to see whether Dgo experiences the same kind of turnover. Within this range, the absolute protein levels in puncta therefore seem unimportant for establishing planar polarity, implying that the stoichiometric ratios rather than the absolute levels of the core proteins is the more important property for planar polarity.

It would also be interesting to investigate whether the levels of the other core proteins in puncta in half-dose flies are correspondingly reduced, or whether the stoichiometry changes. This would have implications for our understanding of the mechanisms of core complex and puncta formation. Half-dose Fz-EGFP puncta had a mean intensity between that of the homozygotes and the 50:50 flies. Assuming that the half-dose flies have half the amount of Fz protein of the homozygotes or 50:50 flies, this suggests that the half-dose flies increase the proportion of their Fz protein pool that incorporates into puncta. Perhaps Fz is a limiting factor for puncta formation, or the levels of Fz in puncta are actively controlled according to some feedback mechanism (such as the proposed Pk, Dgo and Dsh-based mechanism). The localisation of cytoplasmic proteins in puncta could be a passive process that is not actively regulated, which would explain the lack of effect of total gene dosage.

The EGFP-Stbm dosage data should be re-done using the *attP40* insertion instead of the *VK23* line to see what happens with this construct. The mean intensity of half-dose Stan-EGFP puncta is more similar to that of the 50:50 puncta than the homozygotes, but more data should be gathered to confirm this.

5.6 The stoichiometry of the core planar polarity complex at 28h APF

The stoichiometry results must be interpreted with the following assumptions in mind. The EGFP-tagged core proteins were presumed to express and behave normally, as demonstrated by the functional controls and comparisons of constructs in different sites, with different fluorophores and tagged at different ends. The MatLab script appears to detect and measure puncta appropriately, because the dosage testing results are largely as predicted and generally replicate the hand-measured data. The nonlinear regression accounts for much of the variation in the data, so the interpolated values are assumed to be representative of the actual mean brightness of the puncta. The stoichiometry of core proteins in puncta is additionally assumed to be the same as that of the individual complexes.

This last assumption is the least sound, as puncta may contain some core proteins that are not bound into complexes, but even if this is not the case, the relative stoichiometry of the core proteins in puncta is still an interesting line of inquiry. However, this is probably a reasonable assumption, as the FRAP data show that there is a larger stable fraction of molecules in the puncta than in non-puncta membrane for Fz and Fmi (Strutt et al., 2011). The cytoplasmic proteins Dgo, Dsh and Pk have a lower stable fraction in puncta than Fz and Fmi (S. Warrington, unpublished data), suggesting that a greater proportion of these proteins present in the puncta are not stably bound into complexes, or that they dissociate from the complex and are turned over more rapidly than Fz and Fmi.

Additionally, all core proteins are present in puncta and the excess is transported away from the membrane, at least for Fmi (Strutt et al., 2011), implying that only what is needed is maintained at the puncta. Many studies also show that in the absence of one of the core proteins the localisations of the others are disrupted (discussed in the main introduction), so it is not unreasonable to suppose that this might also be the case at the level of the individual complexes. It can be assumed that the vast majority of individual complexes exist in their final, stable configuration, that core protein complexes in puncta are stable, and that the relative stoichiometry of the core proteins in puncta is the same as that of the individual complexes. Further FRAP experiments using the constructs generated in this project can add more data to this aspect of the puncta behaviour.

The working hypothesis for the stoichiometry of the core proteins complex was that it consisted of two molecules of Fmi with one molecule each of the other five core proteins (see **Figure 1.2**). This was the simplest model that fitted the available data. However, the most striking result from the stoichiometry measurements is that Fmi is not the most abundant core protein in puncta, instead being present at the same level as Fz. The results of the stoichiometry experiment show that the core complex is likely to be more complicated than the initial model predicted.

If all the assumptions are correct and the stoichiometry ratios can be taken at face value then some interesting numbers emerge. Based on these results, an individual core complex contains two molecules each of the cytoplasmic proteins Pk, Dgo and Dsh, three each of Fmi and Fz, and five Stbms. Fmi strongly prefers to form homodimers and these numbers would give 1.5 dimers, so let them be doubled again to give six Fmi (three homodimers) with six Fz molecules and ten Stbms, then four each of Pk, Dgo and Dsh. The core protein complex would thus contain 34 subunits.

However, given the numerous assumptions and sources of noise in the data, it is safer to draw more generalised qualitative conclusions. This still presents interesting results. For example, the cytoplasmic proteins are less abundant in puncta than the transmembrane ones. There is also a marked difference in the behaviour in the half dosage between the cytoplasmic and the transmembrane proteins – when the total gene dosage is reduced, the cytoplasmic protein puncta have the same brightness as in the normal dosage, whereas the brightness of the transmembrane protein puncta is increased. This behaviour may reflect these two groups having distinct different functions in the establishment of planar polarity.

Several of the models described in the introduction predict that the transmembrane proteins form the basis of the complex and the cytoplasmic proteins have a role in stabilising or clustering complexes of the same polarity. If the cytoplasmic proteins somehow form bridges between sub-complexes of transmembrane proteins then this could account for their lower abundance. The concept of a core protein complex would then have to be adjusted to a more raft-like structure, with sub-complexes of the transmembrane proteins linked together by a network of the cytoplasmic proteins. If this is indeed the function of the cytoplasmic proteins then each might be expected to have multiple interaction sites with their transmembrane partners.

5.7 Changes in stoichiometry between 20h and 28h APF

At 20h APF the stoichiometry shows a marginally more uniform pattern than at 28h APF, but from **Figure 3.15 E**, it can be seen that as the wing ages the stoichiometric ratios shift. The changes are not dramatic and the overall relationships between the six proteins are maintained. Pk and Dgo are present at the same level, with a greater amount of Fz, Stan and Dsh, and even more Stbm, although not as much Stbm as at 28h APF. The main difference is that the levels of the cytoplasmic proteins are decreased relative to the transmembrane proteins from 20h to 28h APF. This is consistent with the model that the cytoplasmic proteins are instrumental in the feedback mechanisms that are thought to establish asymmetry, or cluster complexes together. Over time, the action of the core proteins might result in the accumulation of the transmembrane proteins into the puncta. A study of the changes in stoichiometry with a finer-grained time course, covering as much of pupal wing development as possible, may be informative.

This relative increase in Dsh at 20h APF is interesting. At 28h APF the Dsh level is closer to that of the other cytoplasmic proteins Pk and Dgo. Dsh is a particularly interesting core protein, because it has a qualitatively different subcellular distribution to the others. As well as being present in puncta it also shows a distinctive fuzzy cytoplasmic signal and the edges of the puncta often appear less well-defined than with other core proteins. This localisation has been observed with other Dsh fusion proteins as well as the EGFP-Dsh construct used in this project. Additionally, Dsh is backed up by Dgo at the distal edge of the cell, whereas its proximal counterpart Pk acts alone. Since Dsh is known to be involved in *wingless* signalling it may be that this localisation reflects this secondary aspect of the cellular function of Dsh. However, the post hoc power calculations suggest that this dataset is not able to clearly distinguish changes in Dsh relative to Dgo, Pk or Fz. More Dsh data may add more power and present a more solid picture of the stoichiometry of Dsh in relation to the rest of the core complex.

While the ratio of cytoplasmic to transmembrane core proteins changes, the amounts of Stbm, Fz and Fmi do not change relative to each other. This is consistent with the model that the interactions between the transmembrane core proteins are self-assembling and do not necessarily require the presence of the cytoplasmic proteins. In a *pk* or *dgo* mutant the puncta still assemble, although they lose their asymmetric localisation,

At 20h APF the wing tissue is growing and undergoing morphogenetic processes. It was shown in Chapter 3 that the cell size and shape is more variable than at 28h APF, and the familiar hexagonal lattice of cells is largely absent. The puncta at this earlier stage are more numerous and slightly smaller than at 20h APF. Between 20h and 28h APF the puncta thus become larger, fewer and asymmetrically distributed. This is consistent with the hypothesis that puncta represent clusters of complexes. The smaller puncta at 20h could cluster on the proximal-distal axis while being removed from the anterior and posterior cell edges.

Intriguingly, some very low-quality time lapse movies taken while gathering the 20h images suggest that puncta at this stage travel laterally along the membranes, an observation which was not replicated at 28h APF (data not shown). Preliminary super-resolution imaging also clearly shows core protein puncta streaming laterally along cell membranes in both directions (D. Strutt, unpublished data). This behaviour of puncta could be an important physical mechanism for puncta formation and generation of asymmetry and should be investigated in greater depth.

5.8 Core complex stoichiometry implications

At both of the measured timepoints the levels of Fz and Fmi are not significantly different from each other, while those of Stbm are significantly higher. From the perspective of Fmi homodimers, this means that there are two Fz molecules for every distally-localised Fmi molecule, and three or more Stbms for every proximally-localised Fmi. If Fz and Stbm bind directly to Fmi then Fmi must have multiple binding sites for both Fz and Stbm. While Fmi is a large molecule, the majority of its bulk is in the extracellular domain, particularly the cadherin repeats, which form the basis of the proposed homophilic interactions, and is thus not necessarily accessible to Fz and Stbm. It would be interesting to understand the exact nature of the molecular interactions between Fz, Stbm and Fmi, as it is difficult to formulate a realistic configuration for the core complex without this information. A simple model would be that Fmi binds a Fz dimer, in which case Fmi may have just one binding site for such a dimer. Data from Carron et al. (2013) suggest that *Xenopus* XFz3 may dimerize *in vivo*, so this should perhaps be investigated in *Drosophila* too.

The excess of Stbm was unexpected but not entirely inconsistent with published data, particularly the le Garrec model (2006). These authors proposed a feedback mechanism whereby Fz and Stbm compete for binding to Fmi, which would require Fmi to be the limiting factor, as reflected in the stoichiometry. Furthermore, the le Garrec model also posits that the concentration of 'activated' Fz limits complex formation, with no corresponding restriction on Stbm, again consistent with the observed stoichiometry. However, by 28h APF the core proteins are already planar polarised and stably localised, so any feedback mechanisms involved in the establishment of polarity should have already played their part. On the other hand, Burak and Shraiman (2009) emphasise the importance of constant flux for the maintenance of stable asymmetric protein localisation, and newly synthesised core proteins must presumably also find their way to the correct edge of the cell. Perhaps the same mechanisms that polarise the core proteins in the first place continue to be active until the emergence of the prehair and associated loss of core protein asymmetry.

The published models all consider protein interactions in the feedback loops in general terms, as there is not enough data available to permit more detailed modelling of protein binding and unbinding. If these stoichiometry results are incorporated into a computer simulation of planar polarity then interesting properties of the system may emerge in the interactions between the core proteins.

5.9 Control of core protein levels, stoichiometry and polarity

Several studies have shown that the levels of core proteins at the membrane are both interdependent and actively regulated (see introduction). This may mean that there is a certain level of tolerance of changes in core protein expression level and the total cellular core protein pool. However, it would be interesting to see whether the stoichiometry of the puncta is affected by overexpression or mutation of one of the core proteins. Overexpression of most of the core proteins increases the levels of the others at the membrane, so the complexes might be expected to still form in the usual stoichiometric ratios. In the absence of *pk*, *dgo* or *dsh* the remaining core proteins still go to the membrane, but they lose their asymmetric localisation. In the absence of one of the six core proteins it would not be possible to form the complex in the usual stoichiometry because one of the components would be missing, so it would be interesting to see whether the remaining proteins still assemble in the usual ratios.

The decrease in cytoplasmic core proteins relative to the transmembrane core proteins from 20h to 28h APF may suggest that an excess of the cytoplasmic factors arrives at the membrane, and that molecules which do not bind into complexes must be removed. The recent findings for Pk and Dsh in the two recent Strutt et al. papers (2013a and 2013b) support this idea. An excess of the cytoplasmic proteins at the membrane may cause too much non-polarised clustering of complexes.

There may also be a functional separation between the cellular protein levels and the distribution of the protein within the cells, or active regulation of protein levels and localisation. This certainly seems to be the case for the transmembrane core proteins. The dosage testing showed that their mean puncta brightness is greater in the half dose than in the 50:50 genotype. This suggests that when the total gene dosage is halved, either expression from the remaining copy is increased above normal levels, or a greater proportion of the protein pool is recruited into the puncta, implying that the transmembrane proteins are limiting for core complex formation. The transmembrane core proteins could be actively recruited to the puncta, where they are received by an excess of the cytoplasmic factors.

In particular, Strutt et al. (2013) revealed a role for ubiquitination and neddylation in controlling levels of planar polarity proteins at cell membranes. Ubiquitin is a small protein that is attached to other proteins to target them for degradation. The authors found that knocking out *Cullin-3*, or its binding partners *diablo* and *kelch*, caused an increase of all six core proteins at apicolateral membranes, despite Cullin-3 acting specifically only on Dsh. This shows that an increase in just Dsh is sufficient to cause increased membrane localisation of the other five proteins. This also suggests that, like Pk, Dsh is usually constantly targeted for degradation.

5.10 The role of puncta

It is still not clear why the individual core complexes aggregate into puncta (if this is indeed what the puncta are). Clustering could be important for the putative feedback mechanisms, perhaps to make the system more robust to noise. If the core proteins complexes were distributed evenly along the edges of the cells then they might be too diffuse to be effective, or it might be more difficult for cells to communicate their polarisation vector to their neighbours. A possible model for puncta formation is that small puncta form stochastically, resulting from the local amplification of a small initial asymmetry by the feedback mechanism. This feedback mechanism, or some inherent aggregative property of the core complexes, could then locally deplete additional core proteins from the neighbouring membrane resulting in the observed discrete punctate structures. However, puncta formation is clearly important for planar polarity, as loss of puncta is accompanied by loss of polarity,

5.11 Summary

The main goals of the project were achieved. A functional EGFP-tagged construct, with normal subcellular localisation, was generated for each core protein. A satisfactory method was devised for detecting and measuring core protein puncta in confocal images and a method was also implemented to account for the relationship between image brightness and wing-membrane distance. The stoichiometry and dosage testing data support the core complex hypothesis, but not the current idea of the core complex structure.

Overall, it seems that the absolute expression levels of the core proteins are not crucial for puncta formation and establishment of planar polarity, and that core complexes self-assemble in the correct stoichiometry. There is a marked difference in stoichiometric behaviour between the cytoplasmic and transmembrane core proteins, perhaps reflecting their different roles in the planar polarity mechanism. The change in stoichiometry between 20h and 28h APF may represent the stabilisation of the planar polarity system, or a switch from 'active polarisation' to a 'polarity maintenance' mode. These results will help further understanding of the establishment and maintenance of asymmetric protein localisation and the molecular basis of planar polarity. Further observations of stoichiometry at different timepoints and in core protein mutant backgrounds will expand knowledge of puncta formation and asymmetry.

References

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

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* signaling. 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.

Aigouy, B., Farhadifar, R., Staple, D.B., Sagner, A., Röper, J.-C., Jülicher, F., and Eaton, S. (2010). Cell flow reorients the axis of planar polarity in the wing epithelium of *Drosophila*. Cell *142*, 773-786.

Ambegaonkar, A.A., Pan, G., Mani, M., Feng, Y., and Irvine, K.D. (2012). Propagation of Dachsous-Fat planar cell polarity. Curr Biol *22*, 1302-1308.

Amonlirdviman, K., Khare, N.A., Tree, D.R.P., Chen, W.-S., Axelrod, J.D., and Tomlin, C.J. (2005). Mathematical modeling of planar cell polarity to understand domineering nonautonomy. Science *307*, 423-426.

Antic, D., Stubbs, J.L., Suyama, K., Kintner, C., Scott, M.P., and Axelrod, J.D. (2010). Planar cell polarity enables posterior localization of nodal cilia and left-right axis determination during mouse and *Xenopus* embryogenesis. PLoS One *5*, e8999.

Axelrod, J.D. (2001). Unipolar membrane association of Dishevelled mediates Frizzled planar cell polarity signaling. Genes Dev *15*, 6.

Axelrod, J.D. (2009). Progress and challenges in understanding planar cell polarity signaling. Semin Cell Dev Biol *20*, 964-971.

Axelrod, J.D., Miller, J.R., Shulman, J.M., and Moon, R.T. (1998). Differential recruitment of Dishevelled provides signaling specificity in the planar cell polarity and Wingless signaling pathways. Genes Dev *12*, 2610-2622.

Bartel, D.P. (2009). MicroRNAs: Target recognition and regulatory functions. Cell *136*, 215-233.

Bastock, R., and Strutt, D. (2007). The planar polarity pathway promotes coordinated cell migration during *Drosophila* oogenesis. Development *134*, 3055-3064.

Bastock, R., Strutt, H., and Strutt, D. (2003). Strabismus is asymmetrically localised and binds to Prickle and Dishevelled during *Drosohila* planar polarity patterning. Development *130*, 3007-3014.

Bellaiche, Y., Mogila, V., and Perrimon, N. (1999). I-*Sce*I endonuclease, a new tool for studying DNA double-strand break repair mechanisms in *Drosophila*. Genetics *152*, 1037-1044.

Bennett, V., and Chen, L. (2001). Ankyrins and cellular targeting of diverse membrane proteins to physiological sites. Curr Opin Cell Biol *13*, 61-67.

Bhanot, P., Brink, M., Samos, C.H., 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*, 225-231.

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

Bjarnadottír, T.K., Gloriam, D.E., Hellstrand, S.H., Kristansson, H., Fredriksson, R., and Schiöth, H.B. (2012). Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. Genomics *88*, 263-273.

Boutros, M., Mihaly, J., Bouwmeester, T., and Mlodzik, M. (2000). Signaling specificity by Frizzled receptors in *Drosophila*. Science *288*, 1825-1828.

Boutros, M., Paricio, N., Strutt, D.I., and Mlodzik, M. (1998). Dishevelled activates JNK and discriminates between JNK pathways in planar polarity and *wingless* signaling. 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.

Brittle, A., Thomas, C., and Strutt, D. (2012). Planar polarity specification through asymmetric subcellular localization of Fat and Dachsous. Curr Biol *22*, 907-914.

Burak, Y., and Shraiman, B.I. (2009). Order and stochastic dynamics in *Drosophila* planar cell polarity. PLoS Comp Biol *5*.

Carreira-Barbosa, F., Concha, M.L., Takeuchi, M., Ueno, N., Wilson, S.W., and Tada, M. (2003). Prickle 1 regulates cell movements during gastrulation and neuronal migration in zebrafish. Development *130*, 4037-4036.

Carron, C., Pascal, A., Dijane, A., Boucaut, J.-C., Shi, D.-L., and Umbhauer, M. (2003). Frizzled receptor dimerization is sufficient to activate the Wnt/β-catenin pathway. J Cell Sci *116*, 2541-2550.

Casal, J., Lawrence, P.A., and Struhl, G. (2006). Two separate molecular systems, Dachsous/Fat and Starry night/Frizzled, act independently to confer planar cell polarity. Development *133*, 4561-4572.

Castro, J.P., and Carareto, C.M.A. (2004). *Drosophila melanogaster* P transpoable elements: mechanisms of transposition and regulation. Genetica *121*, 107-118.

Chae, J., Kim, M.-J., Goo, J.H., Collier, S., Gubb, D., Charlton, J., Adler, P.N., and Park, W.J. (1999). The *Drosophila* tissue polarity gene *starry night* encodes a member of the protocadherin family. Development *126*, 5421-5429.

Chalfie, M., and Kain, S.R. (2005). The three-dimensional structure of green fluorescent proteins and its implications for function and design. In Green fluorescent protein: properties, applications and protocols, pp. 77-78.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science *263*, 802-805.

Chang, C.W., and Mycek, M.A. (2009). Improving precision in time-gated FLIM for low-light live-cell imaging. Paper presented at: Progress in Biomedical Optics and Imaging - Proceedings of SPIE (Munich).

Chen, Y., Müller, J.D., Ruan, Q., and Gratton, E. (2002). Molecular brightness characterization of EGFP in vivo by fluorescence fluctuation spectroscopy. Biophys J *82*, 133-144.

Chiu, C.-S., Jensen, K., Sokolova, I., Wang, D., Li, M., Deshpande, P., Davidson, N., Mody, I., Quick, M.W., Quake, S.R., *et al.* (2002). Number, density and surface/cytoplasmic distribution of GABA transporters at presynaptic structures of knock-in mice carrying GABA transporter subtype 1-green fluorescent protein fusions. The Journal of Neuroscience *22*, 10251-10266.

Chiu, C.-S., Kartalov, E., Unger, M., Quake, S.R., and Lester, H.A. (2001). Single-molecule measurements calibrate green fluorescent protein surface densities on transparent beads for use with 'knock-in' animals and other expression systems. J Neurosci Methods *105*, 55-63.

Cho, B., and Fischer, J.A. (2011). Ral GTPase promotes asymmetric Notch activation in the *Drosophila* eye in reponse to Frizzled/PCP signaling by repressing ligand-independent receptor activation. Development *138*, 1349-1359.

Cho, E., and Irvine, K.D. (2004). Action of *fat*, *four-jointed*, *dachsous* and *dachs* in distal-to-proximal wing signaling. Development *131*, 4489-4500.

Choi, K.-W., and Benzer, S. (1994). Rotation of photoreceptor clusters in the developing *Drosophila* eye requires the *nemo* gene. Cell *78*, 125-136.

Clark, H.F., Brentrup, D., Schneitz, K., Bieber, A., Goodman, C.S., and Noll, M. (1995). *dachsous* encodes a member of the cadherin superfamily that controls imaginal disc morphogenesis in *Drosophila*. Genes Dev *9*, 1530-1542.

Classen, A.-K., Anderson, K., I., Marois, E., and Eaton, S. (2005). Hexagonal packing of *Drosophila* wing epithelial cells by the planar cell polarity pathway. Dev Cell *9*, 805-817.

Clegg, R.M. (2009). Förster resonance energy transfer - FRET: what is it, why do it, and how it's done, Vol 33 (Elsevier).

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.

Collier, S., Lee, H., Burgess, R., and Adler, P.N. (2005). The WD40 repeat protein Fritz links cytoskeletal planar polarity to Frizzled subcellular localization in the *Drosophila* epidermis. Genetics *169*, 2035-2045.

Copeland, N.G., Jenkins, N.A., and Court, D.L. (2001). Recombineering: A powerful new tool for mouse functional genomics. Nat Rev Genet *2*, 769-779.

Cormack, B.P., Valdivia, R.H., and Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). Gene *173*, 33-38.

Curtin, J.A., Quint, E., Tsipouri, V., Arkell, R.M., Cattanach, B., Copp, A.J., Henderson, D.J., Spurr, N., Stanier, P., Fisher, E.M., *et al.* (2003). Mutation of *Celsr1* disrupts planar polarity of inner ear hair cells and causes severe neural tube defects in the mouse. Curr Biol *13*, 1129-1133.

Darken, R.S., Scola, A.M., Rakeman, A.S., Das, G., Mlodzik, M., and Wilson, P.A. (2002). The planar polarity gene strabismus regulates convergent extension movements in Xenopus. EMBO J *21*, 976-985.

Das, G., Jenny, A., Klein, T.J., Eaton, S., and Mlodzik, M. (2004). Diego interacts with Prickle and Strabismus/Van Gogh to localize planar cell polarity complexes. Development *131*, 4467-4476.

Day, R.N., and Davidson, M.W. (2009). The fluorescent protein palette: tools for cellular imaging. Chemical Society Reviews *38*, 2287-2921.

Devenport, D., and Fuchs, E. (2008). Planar polarizaion in embryonic epidermis orchestrates global asymmetric morphogenesis of hair follicles. Nat Cell Biol *10*, 1257-1268.

DiAntonio, A., Burgess, R.W., Chin, A.C., Deitcher, D.L., Scheller, R.H., and Schwartz, T.L. (1993). Identification and characterization of *Drosophila* genes for synaptic vesicle proteins. The Journal of Neuroscience *13*, 4924-4935.

Dundr, M., McNally, J.G., Cohen, J., and Misteli, T. (2002). Quantitation of GFP-fusion proteins in single living cells. Journal of Structural Biology *140*, 92-99.

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

Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in *Drosophila*. Genome Biology *5*.

Farkas, G., Leibovitch, B.A., and Elgin, S.C.R. (2000). Chromatin organisation and transcriptional control of gene expression in *Drosophila*. Gene *253*, 117-136.

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioural and biomedical sciences. Behavior Research Methods *39*, 175-191.

Feiguin, F., Hannus, M., Mlodzik, M., and Eaton, S. (2001). The ankyrin repeat proten Diego mediates Frizzled-dependent planar polarization. Dev Cell *1*, 93-101.

Fischer, E., Legue, E., Doyen, A., Nato, F., Nicolas, J.-F., Torres, V., Yaniv, M., and Pontoglio, M. (2006). Defective planar cell polarity in polycystic kidney disease. Nat Genet *38*, 21-23.

Gaj, T., Gersbach, C.A., and Barbas III, C.F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol *31*, 397-405.

Gao, B., Song, H., Bishop, K., Elliot, G., Garrett, L., English, M.A., Andre, P., Robinson, J., Sood, R., Minami, Y.*, et al.* (2011). Wnt signaling gradients establish planar cell polarity by inducing Vangl2 phosphorylation through Ror2. Dev Cell *20*, 163-176.

Gao, G., McMahon, C., Chen, J., and Rong, Y.S. (2008). A powerful method combining homologous recombination and site-specific recombination for targeted mutagenesis in *Drosophila*. Proc Natl Acad Sci USA *105*, 13999-14004.

Gao, Q., Gilbert, D., Heiner, M., Liu, F., Maccagnola, D., and Tree, D.R.P. (2013). Multiscale modeling and analysis of planar cell polarity in the Drosophila wing. IEEE/ACM Transactions on Computational Biology and Bioinformatics *10*, 337-351.

Geng, J., Baba, M., Nair, U., and Klionsky, D.J. (2008). Quantitative analysis of autophagy-related protein stoichiometry by fluorescence microscopy. J Cell Biol *181*, 129-140.

Geng, J., and Klionsky, D.J. (2010). Determining Atg protein stoichiometry at the phagophore assembly site by fluorescence microscopy. Autophagy *6*, 144-147.

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737-741.

Golic, K.G., and Lindquist, S. (1989). The FLP recombinase of yeast catalyzes site-specific recombination in the *Drosophila* genome. Cell *59*, 499-409.

Gong, W.J., and Golic, K.G. (2003). Ends-out, or replacement, gene targeting in *Drosophila*. Proc Natl Acad Sci USA *100*, 2556-2561.

Goodrich, L.V., and Strutt, D.I. (2011). Principles of planar polarity in animal development.

Development *138*, 1877-1892.

Graf, E.R., Daniels, R.W., Burgess, R.W., Schwartz, T.L., and DiAntonio, A. (2009). Rab3 dynamically controls protein composition at active zones. Neuron *64*, 663-677.

Gratz, S.J., Cummings, A.M., Nguyen, J.N., Hamm, D.C., Donohue, L.K., Harrison, M.M., Wildonger, J., and O'Connor-Giles, K.M. (2013). Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics *194*, 1029-1035.

Grebe, M. (2004). Ups and downs of tissue and planar polarity in plants. Bioessays *36*, 719-729.

Groth, A.C., Olivares, E.C., Thyagarajan, B., and Calos, M.P. (2004). A phage integrase directes efficient site-specific integration in human cells. Proceedings of the Academy of Natural Sciences of the United States of America *97*, 5595-6000.

Gubb, D., and García-Bellido, A. (1982). A genetic analysis of the determination of cuticular polarity during development in *Drosophila melanogaster*. Journal of Embryology and Experimental Morphology *68*, 37-57.

Gubb, D., Green, C., Huen, D., Coulson, D., Johnson, G., Tree, D.R.P., Collier, S., and Roote, J. (1999). The balance between isoforms of the Prickle LIM domain protein is critical for planar polarity in *Drosophila* imaginal discs. Genes Dev *13*, 2315-2327.

Guo, N., Hawkins, C., and Nathans, J. (2004). Frizzled6 controls hair patterning in mice. Proc Natl Acad Sci USA *101*, 9277-9281.

Hanratty, W.P., and Ryerse, J.S. (1981). A genetic melanotic neoplasm of *Drosophila melanogaster*. Dev Biol *83*, 238-249.

Harrison, D.A., Binari, R., Nahreini, T.S., and Perrimon, N. (1995). Activation of a *Drosophila* Janus kinase (JAK) causes hematopoietic neoplasia and developmental defects. The EMBO Journal *14*, 2857-2865.

Hashimoto, M., Shinohara, K., Wang, J., Ikeuchi, S., Yoshiba, S., Meno, C., Nonaka, S., Takada, S., Hatta, K., Wynshaw-Boris, A.*, et al.* (2010). Planar polarization of node cells determines the rotational axis of node cilia. Nat Cell Biol *12*, 170-176.

Heim, R., and Tsien, R.Y. (1996). Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. Curr Biol *6*, 178-182.

Hirota, Y., Meunier, A., Huang, S., Shimozawa, T., Yamada, O., Kida, Y.S., Inoue, M., Ito, T., Kato, H., Sakaguchi, M.*, et al.* (2010). Planar polarity of multiciliated ependymal cells involves the anterior migration of basal bodies regulated by non-muscle myosin II. Development *137*, 3037-3046.

Hochleitner, E.O., Kastner, B., Fröhlich, T., Schmidt, A., Lührmann, R., Arnold, G., and Lottspeich, F. (2004). Protein stoichiometry of a multiprotein complex, the human spliceosomal U1 small nuclear ribonucleoprotein. The Journal of Biological Chemistry *280*, 2536-2642.

Hoskins, R.A., Nelson, C.R., Berman, B.P., Laverty, T.R., George, R.A., Ciesiolka, L., Naeemuddin, M., Arenson, A.D., Durbin, J., David, R.G.*, et al.* (2000). A BAC-based physical map of the major autosomes of Drosophila melanogaster. Science *287*, 2271-2274.

Huang, J., Zhou, W., Dong, W., Watson, A.M., and Hong, Y. (2009). Directed, efficient, and versatile modifications of the *Drosophila* genome by genomic engineering. Proc Natl Acad Sci USA *106*, 8284-8289.

Huang, J., Zhou, W., Watson, A.M., Jan, Y.-N., and Hong, Y. (2008). Efficient ends-out gene targeting in *Drosophila*. Genetics *180*, 703-707.

Itoh, K., Jacob, J., and Sokol, S.Y. (1998). A role for *Xenopus* Frizzled 8 in dorsal development. Mechanisms of Development *74*, 145-157.

Jacobs, J.S., Hong, X., and Eberl, D.F. (2011). A "mesmer"ising new approach to site-directed mutagenesis in large transformation-ready constructs: Mutagenesis via Serial Small Mismatch Recombineering. Fly *5*, 162-169.

Jain, R.K., Joyce, P.B.M., Molinete, M., Halban, P.A., and Gorr, S.-U. (2001). Oligomerization of green fluorescent protein in the secretory pathway of endocrine cells. The Biochemical Journal *350*, 645-649.

Jenny, A., Darken, R.S., Wilson, P.A., and Mlodzik, M. (2003). Prickle and Strabismus form a functional complex to generate a correct axis during planar cell polarity signaling. The EMBO Journal *22*, 4409-4420.

Jenny, A., Reynolds-Kenneally, J., Das, G., Burnett, M., and Mlodzik, M. (2005). Diego and Prickle regulate Frizzled planar cell polarity signalling by competing for Dishevelled binding. Nat Cell Biol *7*, 691-697.

Joglekar, A.P., Bouck, D.C., Molk, J.N., Bloom, K.S., and Salmon, E.D. (2006). Molecular architecture of a kinetochore-microtubule attachment site. Nat Cell Biol *8*, 581-585.

Jones, K.H., Liu, J., and Adler, P.N. (1996). Molecular analysis of EMS-induced *frizzled* mutations in *Drosophila melanogaster*. Genetics *142*, 205-215.

Keller, R. (2002). Shaping the vertebrate body plan by polarizes embryonic cell movements. Science *298*, 1950-1954.

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.

Kibar, Z., Torban, E., McDearmid, J.R., Reynolds, A., Berghout, J., Mathieu, M., Kirillova, I., De Marco, P., Merello, E., Hayes, J.M.*, et al.* (2007). Mutations in *VANGL1* associated with neural-tube defects. The New England Journal of Medicine *356*, 1432-1437.

Kibar, Z., Vogan, K.J., Groulx, N., Justice, M.J., Underhill, A., and Gros, P. (2001). *Ltap*, a mammalian homologue of *Drosophila Strabismus/Van Gogh*, is altered in the mouse neural tube mutant Loop-tail. Nat Genet *28*, 251-255.

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.

Kornau, H.-C., Schenker, L.T., Kennedy, M.B., and Seeburg, P.H. (1995). Domain interactions between NMDA receptor subunits and the postsynaptic density protein PSD-95. Science *269*, 1737-.

Krasnow, R.E., Wong, L.L., and Adler, P.N. (1995). *dishevelled* is a component of the frizzled signaling pathway in *Drosophila*. Development *121*, 4095-4102.

Krishnan, A., Almén, M.S., Fredriksson, R., and Schiöth, H.B. (2012). The origin of GPCRs: identification of mammalian like *Rhodopsin*, *Adhesion*, *Glutamate* and *Frizzled* GPCRs in fungo.

PLoS One *7*, e29817.

Lawrence, P.A. (1966). Gradients in the insect segment: the orientation of hairs in the milkweed bug *Oncopeltus fasciatus*. The Journal of Experimental Biology *44*, 607-620.

Lawrence, P.A., and Casal, J. (2013). The mechanisms of planar cell polarity, growth and the Hippo pathway: some known unknowns. Dev Biol *377*, 1-8.

Lawrence, P.A., Struhl, G., and Casal, J. (2007). Planar cell polarity: one or two pathways? Nat Rev Genet *8*, 555-563.

Le Garrec, J.-F., Lopez, P., and Kerszberg, M. (2006). Establishment and maintenance of planar epitheial cell polarity by asymmetric cadherin bridges: A computer model. Dev Dyn *235*, 235-246.

Lee, E., Salic, A., Krüger, R., Heinrich, R., and Kirschner, M.W. (2003). The roles of APC and axin derived from experimental and theoretical analysis of the Wnt pathway. PLoS Biol *1*, e10.

Lin, Y.-Y., and Gubb, D. (2009). Molecular dissection of Drosophila Prickle isoforms distinguishes their essential and overlapping roles in planar cell polarity. Dev Biol *325*, 386-399.

Lindqvist, M., Horn, Z., Bryja, V., Schulte, G., Papchristou, P., Ajima, R., Dyberg, C., Arenas, E., Yamaguchi, T.P., Lagercrantz, H.*, et al.* (2010). Vang-like protein 2 and Rac1 interact to regulate adherens junctions. J Cell Sci *123*, 472-483.

Linkert, M., Rueden, C.T., Allan, C., Burel, J.-M., Moore, W., Patterson, A., Loranger, B., Moore, J., Neves, C., MacDonald, D.*, et al.* (2010). Metadata matters: access to image data in the real world. J Cell Biol *189*, 777-782.

Locke, M. (1959). The cuticular pattern in an insect, Rhodnius prolixus stål. J Exp Biol *36*, 459-477.

Lu, B., Usui, T., Uemura, T., Jan, L., and Jan, Y.-N. (1999). Flamingo controls the planar polarity of sensory bristles and asymmetric division of sensory organ precursors in *Drosophila*. Curr Biol *9*, 1247-1250.

Mahoney, P.A., Weber, U., Onofrechuk, P., Blessman, H., Bryant, P.J., and Goodman, C.S. (1991). The *fat* tumour suppressor gene in *Drosophila* encodes a novel member of the cadherin gene superfamily. Cell *67*, 853-868.

Mao, Y., Rauskolb, C., Cho, E., Hu, W.-L., Hayter, H., Minihan, G., Katz, F.N., and Irvine, K.D. (2006). Dachs: An unconventional myosin that functions downstream of Fat to regulate growth, affinity and gene expression in *Drosophila*. Development *133*, 2539-2551.

Markstein, M., Pitsouli, C., Villalta, C., Celniker, S.E., and Perrimon, N. (2008). Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. Nat Genet *40*, 476-483.

Mason, D.W., and Williams, A.F. (1980). The kinetics of antibody binding to membrane antigens in solution and at the cell surface. Biochem J *187*, 1-20.

Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L., and Lukyanov, S.A. (1999). Fluorescent proteins from nonbioluminescent *Anthozoa* species. Nat Biotechnol *17*, 969-973.

Maurer-Stroh, S., Koranda, M., Benetka, W., Schneider, G., and Sirota, F.L. (2007). Towards complete sets of farnesylated and geranylgeranylated proteins. PLoS Comp Biol *3*, 634-648.

McGill, M.A., McKinley, R.F.A., and Harris, T.J.C. (2009). Independent cadherin-catenin and

Bazooka cluters interact to assemble adherens junctions. J Cell Biol *185*, 787-796.

Melan, M.A., and Sluder, G. (1992). Redistribution and differential extraction of soluble proteins in permeabilized cultures cells. Implications for immunofluorescence microscopy. J Cell Sci *101*, 731-743.

Merzlyak, E.M., Goedhart, J., Shcherbo, D., Bulina, M.E., Shcheglov, A.S., Fradkov, A.F., Gaintzeva, A., Lukyanov, K.A., Lukyanov, S., Gadella, T.W.J.*, et al.* (2007). Bright monomeric red fluorescent protein with an extended fluorescence lifetime. Nat Methods *4*, 555-557.

Mitchell, B., Stubbs, J.L., Huisman, F., Taborek, P., Yu, C., and Kintner, C. (2009). The PCP pathway instructs the planar orientation of ciliated cells in the *Xenopus* larval skin. Curr Biol *19*, 924-929.

Mogridge, J. (2004). Using light scattering to determine the stoichiometry of protein complexes. Methods in Molecular Biology *261*, 113-118.

Mohr, S.E., Dilon, S.T., and Boswell, R.E. (2001). The RNA-binding protein Tsunagi interacts with Mago Nashi to establish polarity and localize *oskar* mRNA during *Drosophila* oogenesis. Genes Dev *15*, 2886-2899.

Morgan, T.H. (1910). Sex limited inheritance in *drosophila*. Science *32*, 120-122.

Morrow, G., Inaguma, Y., Kato, K., and Tanguay, R.M. (2000). The small heat shock protein Hsp22 of Drosophila melanogaster is a mitochondrial protein displaying oligomeric organization. The Journal of Biological Chemistry *275*, 31204-31210.

Murphy, K.C. (1998). Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. J Bacteriol *180*, 2063-2071.

Nakamura, M., Kiefer, C.S., and Grebe, M. (2012). Planar polarity, tissue polarity and planar morphogenesis in plants. Curr Opin Plant Biol *15*, 593-600.

Niedenthal, R.K., Riles, L., Johnston, M., and Hegemann, J.H. (1996). Green fluorescent protein as a marker for gene expression and subcellular localization in budding yeast. Yeast *12*, 773-786.

Niethammer, M., Kim, E., and Sheng, M. (1996). Interaction between the C terminus of NMDA receptor subunits and multiple members of the PSD-95 family of membrane-associated guanylate kinases. The Journal of Neuroscience *16*, 2157-2163.

Nübler-Jung, K. (1987). Insect epidermis: disturbance of supracellular tissue polarity does not prevent the expression of cell polarity. Roux's Archives of Developmental Biology *196*, 286-289.

Nübler-Jung, K., Bonitz, R., and Sonnenschein, M. (1987). Cell polarity during wound healing in an insect epidermis. Development *100*, 163-170.

Ormö, M., Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y., and Remington, S.J. (1996). Crystal structure of the Aequoria victoria Green Fluorescent Protein. Science *273*, 1392-1395.

Park, M., and Moon, R.T. (2002). The planar cell-polarity gene stbm regulates cell behaviour and cell fate in vertebrate embryos. Nat Cell Biol *4*, 20-25.

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. Mechanisms of Development *45*, 127-137.

Perrimon, N., and Mahowald, A.P. (1987). Multiple functions of segment polarity genes in *Drosophila*. Dev Biol *119*, 587-600.

Petrášek, J., and Friml, J. (2009). Auxin transport routes in plant development. Development *136*, 2675-2688.

Pfeiffer, B.D., Ngo, T.-T.B., Hibbard, K.L., Murphy, C., Jenett, A., Turman, J.W., and Rubin, G.M. (2010). Refinement of tools for targeted gene expression in *Drosophila*. Genetics *186*, 735-755.

Phillips Jr, G.N. (1997). Structure and dynamics of green fluorescent protein. Curr Opin Struct Biol *7*, 821-827.

Piepho, H. (1955). Über die Ausrichtung der Schuppenbälge und Schuppen am Schmetterlingsrumpf. Naturwissenschaften *42*, 22.

Povelones, M., Howes, R., Fish, M., and Nusse, R. (2005). Genetic evidence that *Drosophila frizzled* controls planar cell polarity and Armadillo signaling by a common mechanism. Genetics *171*, 1643-1654.

Qin, X., Ahn, S., Speed, T.P., and Rubin, G.M. (2007). Global analysis of mRNA translational control during earyly *Drosophila* embryogenesis. Genome Biology *8*, R63.

Rizzo, M.A., Springer, G.H., Granada, B., and Piston, D.W. (2004). An improved cyan fluorescent protein variant useful for FRET. Nat Biotechnol *22*, 445-449.

Rogulja, D., Rauskolb, C., and Irvine, K.D. (2008). Morphogen control of wing growth through the Fat signaling pathway. Dev Cell *15*, 309-321.

Rong, Y.S., and Golic, K.G. (2000). Gene targeting by homologous recombination in *Drosophila*. Science *288*, 2013-2018.

Rousset, R., Mack, J.A., Wharton Jr., K.A., Axelrod, J.D., Cadigan, K.M., Fish, M.P., Nusse, R., and Scott, M.P. (2001). *naked cuticle* targets *dishevelled* to antagonize Wnt signal transduction. Genes Dev *15*, 658-671.

Roy, S., and Hart, C.M. (2010). Targeted gene replacement by homologous recombination in *Drosophila* stimulates production of second-site mutations. Fly *4*, 12-17.

Rubin, G.M., and Spradling, A.C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. Science *218*, 348-253.

Rulifson, E.J., Wu, C.-H., and Nusse, R. (2000). Pathway specificity by the bifunctional receptor Frizzled is determined by affinity for Wingless. Mol Cell *6*, 117-126.

Russ, J.C. (2011). The image processing handbook, 6 edn (CRC Press).

Ryder, E., Ashburner, M., Bautista-Llacer, R., Drummond, J., Webster, J., Johnson, G., Morley, T., Sang Chan, Y., Blows, F., Coulson, D.*, et al.* (2007). The DrosDel deletion collection: a *Drosophila* genomewide chromosomal deficiency resource. Genetics *177*, 615-629.

Saburi, S., Hester, I., Fischer, E., Pontoglio, M., Eremina, V., Gessler, M., Quaggin, S.E., Harrison, R., Mount, R., and McNeill, H. (2008). Loss of Fat4 disrupts PCP signaling and oriented cell division and leads to cystic kidney disease. Nat Genet *40*, 1010-1015.

Sagner, A., Merkel, M., Aigouy, B., Gaebel, J., Brankatschk, M., Jülicher, F., and Eaton, S. (2012). Estabilshment of global patterns of planar polarity durign growth of the *Drosophila* wing epithelium. Curr Biol *22*, 1296-1301.

Sawa, H., Lobel, L., and Horvitz, H.R. (1996). The *Caenorhabditis elegans* gene *lin-17*, which is required for certain asymmetric cell divisions, encodes a putative seven-transmembrane protein similar to the *Drosophila* Frizzled protein. Genes Dev *10*.

Schamberg, S., Houston, P., Monk, N.A.M., and Owen, M.R. (2009). Modelling and Analysis of Planar Cell Polarity. Bull Math Biol, 1-36.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat Methods *9*, 671-675.

Schotta, G., and Reuter, G. (2000). Controlled expression of tagged proteins in *Drosophila* using a new modular P-element vector system. Mol Gen Genet *262*, 916-920.

Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N.G., Palmer, A.E., and Tsien, R.Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. Nat Biotechnol *22*, 1567.

Shaner, N.C., Patterson, G.H., and Davidson, M.W. (2007). Advances in fluorescent protein technology. J Cell Sci *120*, 4247-4260.

Shapiro, L., and Weis, W.I. (2009). Structure and biochemistry of cadherins and catenins. Cold Spring Harbor Perspectives in Biology *1*, a003053.

Sharan, S.K., Thomason, L.C., Kuznetsov, S.G., and Court, D.L. (2009). Recombineering: A homologous recombination-based method of genetic engineering. Nature Protocols, 206-223.

Sharma, P., and McNeill, H. (2013). Regulation of long-range planar cell polarity by Fat-Dachsous signaling. Development *140*, 3869-3881.

Shimada, Y., Usui, T., Yanagawa, S.-i., Takeichi, M., and Uemura, T. (2001). Asymmetric colocalisation of Flamingo, a seven-pass transmembrane cadherin, and Dishevelled in planar cell polarization. Curr Biol *11*, 859-863.

Shimada, Y., Yonemura, S., Ohkura, H., Strutt, D., and Uemura, T. (2006). Polarized transport of Frizzled along the planar microtubule arrays in *Drosophila* wing epithelium. Dev Cell *10*, 209-222.

Shimomura, O., Johnson, F.H., and Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequoria*. Journal of Cellular and Comparative Physiology *59*, 223-239.

Sidow, A. (1996). Gen(om)e duplications in the evolution of early vertebrates. Curr Opin Genet Dev *6*, 715-722.

Simons, M., and Mlodzik, M. (2008). Planar cell polarity signaling: from fly development to human disease. Annu Rev Genet *42*, 517-540.

Smits, A.H., Jansen, P.W.T.C., Poser, I., Hyman, A.A., and Vermeulen, M. (2013). Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Res *41*, e28.

Song, H., Hu, J., Chen, W., Elliot, G., Andre, P., Gao, B., and Yang, Y. (2010). Planar cell polarity breaks bilateral symmetry by controlling ciliary positioning. Nature *466*, 378-382.

Songyang, Z., Fanning, A.S., Fu, C., Xu, J., Marfatia, S.M., Chishti, A.H., Crompton, A., Chan, A.C., Anderson, J.M., and Cantley, L.C. (1997). Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science *275*, 73-77.

Strapps, W.R., and Tomlinson, A. (2001). Transducing properties of *Drosophila* Frizzled proteins. Development *128*, 4829-2835.

Struhl, G., Casal, J., and Lawrence, P.A. (2012). Dissecting the molecular bridges that mediate the function of Frizzled in planar cell polarity. Development *139*, 3665-3674.

Strutt, D. (2008). The planar polarity pathway. Curr Biol *18*, R898-R902.

Strutt, D. (2009). Gradients and the specification of planar polarity in the insect cuticle. Cold Spring Harbor Perspectives in Biology *1:5*, a000489.

Strutt, D., Madder, D., Chaudhary, V., and Artymiuk, P.J. (2012). Structure-function dissection of the Frizzled receptor in Drosophila melanogaster suggests different mechanisms of action in planar polarity and canonical Wnt signaling. Genetics *192*, 1295-1313.

Strutt, D., and Strutt, H. (2007). Differential activities of the core planar polarity proteins during *Drosophila* wing patterning. Dev Biol *302*, 181-194.

Strutt, D., and Strutt, H. (2009). Asymetric localisation of planar polarity proteins: Mechanisms and consequences. Semin Cell Dev Biol *20*, 957-963.

Strutt, D., Weber, U., and Mlodzik, M. (1997). The role of RhoA in tissue polarity and Frizzled signalling. Nature *387*, 292-295.

Strutt, D.I. (2001). Asymmetric localization of Frizzled and the establishment of cell polarity in the *Drosophila* wing. Mol Cell *7*, 376-375.

Strutt, H., Price, M.A., and Strutt, D. (2006). Planar polarity is regulated by casein kinase Iε in *Drosophila*. Curr Biol *16*, 1329-1336.

Strutt, H., Searle, E., Thomas-MacArthur, V., Brookfield, R., and Strutt, D. (2013). A Cul-3-BTB ubiquitylation pathway regulates junctional levels and asymmetry of core planar polarity proteins. Development *140*, 1693-1702.

Strutt, H., and Strutt, D. (2002). Nonautonomous planar polarity patterning in *Drosophila*: Dishevelled-independent functions of Frizzled. Dev Cell *3*, 851-863.

Strutt, H., and Strutt, D. (2008). Differential stability of Flamingo protein complexes underlies the establishment of planar polarity. Curr Biol *18*, 1555-1564.

Strutt, H., Thomas-MacArthur, V., and Strutt, D. (2013). Strabismus promotes recruitment and degradation of farnesylated Prickle in Drosophila melanogaster planar polarity specification. PLoS Genet *9*, e1003654-e1003654.

Strutt, H., Warrington, S., and Strutt, D. (2011). Dynamics of core planar polarity protein turnover and stable assembly into discrete membrane subdomains. Dev Cell *20*, 511-525.

Stumpf, H.F. (1967). Über die Lagebestimmung der Kutikularzonen innerhalb eines Segmentes von *Galleria mellonella*. Dev Biol *16*, 144-167.

Subach, O.M., Gundorov, I.S., Yoshimura, M., Subach, F.V., Zhang, J., Grüenwald, D., Souslova, E.A., Chudakov, D.M., and Verkhusha, V.V. (2008). Conversion of red fluorescent protein into a bright blue probe. Chem Biol *15*, 1116-1124.

Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J., and Stahl, F.W. (1983). The double-strand-break repair model for recombination. Cell *33*, 25-35.

Takaine, M., and Mabuchi, I. (2007). Properties of actin from the fission yeast *Schizosaccharomyces pombe* and interaction with fission yeast profilin. The Journal of Biological Chemistry *282*, 21683-21694.

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* signaling to establish both cell polarity and cell identity. Development *120*, 347-360.

Thorpe, H.M., and Smith, M.C.M. (1998). In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invetase family. Proceedings of the Academy of Natural Sciences of the United States of America *95*, 5505-5510.

Tree, D.R.P., Ma, D., and Axelrod, J.D. (2002). A three-tiered mechanism for regulation of planar cell polarity. Semin Cell Dev Biol *13*, 217-224.

Tree, D.R.P., Shulman, J.M., Rousset, R., Scott, M.P., Gubb, D., and Axelrod, J.D. (2002). Prickle mediates feedback amplification to generate asymmetric planar cell polarity signaling. Cell *109*, 371-381.

Usui, T., Shima, Y., Shimada, Y., Hirano, S., and Burgess, R.W. (1999). Flamingo, a seven-pass transmembrane cadherin, regulates planar cell polarity under the control of Frizzled. Cell *98*, 585-595.

Veeman, M.T., Slusarkski, D.C., Kaykas, A., Hallagan Louie, S., and Moon, R.T. (2003). Zebrafish Prickle, a modulator of noncanonical Wnt/Fz signaling, regulates gastrulation movements. Curr Biol *13*, 680-685.

Venken, K.J.T., and Bellen, H.J. (2007). Transgenesis upgrades for *Drosophila melanogaster*. Development *143*, 3571-3584.

Venken, K.J.T., Carlson, J.W., Schulze, K.L., Pan, H., He, Y., Spokony, R., Wan, K.H., Koriabine, M., de Jong, P.J., White, K.P.*, et al.* (1999). Versatile P[acman] BAC libraries for transgenesis studies in *Drosophila melanogaster*. Nat Methods *6*, 431-434.

Venken, K.J.T., He, Y., Hoskins, R.A., and Bellen, H.J. (2006). P[acman]: A BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. Science *314*, 1747-1751.

Villano, J.L., and Katz, F.N. (1995). *four-jointed* is required for intermediate growth in the proximal-distal axis in *Drosophila*. Development *121*, 2767-2777.

Vinson, C.R., and Adler, P.N. (1987). Directional non-cell autonomy and the transmission of polarity information by the *frizzled* gene of *Drosophila*. Nature *329*, 549-551.

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.

Wang, S., Zhao, Y., Leiby, M., and Zhu, J. (2009). A new positive/negative selection scheme for precise BAC recombineering. Mol Biotechnol, 110-116.

Wasserscheid, I., Thomas, U., and Knust, E. (2007). Isoform-specific interaction of Flamingo/Starry Night with excess Bazooka affects planar cell polarity in the *Drosophila* wing. Dev Dyn *236*, 1064-1071.

Waters, J.C. (2009). Accuracy and precision in quantitative fluorescence microscopy. The journal of Cell Biology *185*, 1135-1148.

Wehrli, M., and Tomlinson, A. (1998). Independent regulation of anterior/posterior and equatorial/planar polarity in the Drosophila eye; evidence for the involvement of Wnt signaling in the equatorial/polar axis. Development *125*, 1421-1432.

Wiedenmann, J., Ivanchenko, S., Oswald, F., Schmitt, F., Röcker, C., Salih, A., Spindler, K.-D., and Nienhaus, G.U. (2004). EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. Proc Natl Acad Sci USA *101*, 15905-15910.

Wigglesworth, V.B. (1940). Local and general factors in the development of "pattern" in Rhodnius prolixus (Hemiptera). J Exp Biol *17*, 180-200.

Winter, C.G., Wang, B., Ballew, A., Royou, A., Karess, R., Axelrod, J.D., and Luo, L. (2001). *Drosophila* Rho-associated kinase (Drok) links Frizzled-mediated planar cell polarity signaling to the actin cytoskeleton. Cell *105*, 81-91.

Wison, P.D. (2004). Polycystic kidney disease. The New England Journal of Medicine *351*, 151-164.

Wolff, T., and Rubin, G.M. (1998). *strabismus*, a novel gene that regulates tissue polarity and cell fate decisions in *Drosophila*. Development *125*, 1149-1159.

Wong, L.L., and Adler, P.N. (1993). Tissue polarity genes of Drosophila regulate the subcellular location for prehair initiation in pupal wing cells. The journal of Cell Biology *123*, 209-221.

Wu, J., and Mlodzik, M. (2008). The Frizzled extracellular domain is a ligand for Van Gogh/Stbm during nonautonomous planar cell polarity signaling. Dev Cell *15*, 462-469.

Wu, J.-Q., and Pollard, T.D. (2005). Counting cytokinesis proteins globally and locally in fission yeast. Science *310*, 310-314.

Xu, T., and Rubin, G.M. (1993). Analysis of genetic mosaics in developing and adult Drosophila tissues. Development *117*, 1223-1237.

Yanagawa, S.-i., van Leeuwen, F., Wodarz, A., Klingensmith, J., and Nusse, R. (1995). The Dishevelled protein is modified by Wingless signaling in Drosophila. Genes Dev *9*, 1087-1097.

Yang, Z. (2008). Cell polarity signaling in *Arabidopsis*. Annu Rev Cell Dev Biol *24*, 551-575.

Yu, D., Ellis, H.M., Lese, E.-C., Jenkins, N.A., Copeland, N.G., and Court, D.L. (2000). An efficient recombination system for chromosome engineering in *Escherichia coli*. Proceedings of the Academy of Natural Sciences of the United States of America *97*, 5979-5983.

Zacharias, D.A., Violin, J.D., Newton, A.C., and Tsien, R.Y. (2002). Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. Science *296*, 913-916.

Zeidler, M.P., Perrimon, N., and Strutt, D. (1999). The *four-jointed* gene is required in the *Drosophila* eye for ommatidial polarity specification. Curr Biol *9*, 1363-1372.

Zhang, F.L., and Casey, P.J. (1996). Protein prenylation: Molecular mechanisms and functional consequences. Annu Rev Biochem *65*, 241-269.

Zhang, G., Gurtu, V., and Kain, S.R. (1996). An enhanced Green Fluorescent Protein allows sensitive detection of gene transfer in mammalian cells. Biochem Biophys Res Commun *227*, 707-711.