University of Alberta

Insights into the nuclear localization of Scalloped.

by

Adam Cory Magico

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in
Molecular Biology and Genetics

Biological Sciences

©Adam Cory Magico
Fall, 2011
Edmonton, Alberta

Abstract

The *Drosophila melanogaster* protein Scalloped (Sd) is a member of the TEA/ATTS family of transcription factors, present throughout Eukarya. The protein is best known for its involvement in the development of the wing imaginal disc of *Drosophila*, although roles in nervous system, muscle (both cardiac and somatic), eye and leg development as well as the control of cellular proliferation have also been shown or predicted.   Scalloped itself lacks a transcriptional activation domain and thus is thought to rely on binding with transcription intermediary factors (TIFs) which have one or more of these domains. Under this paradigm it is thought that Sd facilitates nuclear localization and targeted DNA binding of the Sd/TIF complex, while the TIF activates the target gene(s).

In order to understand the mechanisms behind the nuclear translocation of Sd, it is demonstrated that a candidate bipartite nuclear localization signal is functional in S2 cells, and further that the region containing this signal is critical for Sd function during wing development.   Evidence for the presence of a nuclear export signal is also given. Finally, a broad region of the C-terminal domain of Sd is identified as also being required for the proper nuclear localization of the protein.

Although several TIFs of Sd have been identified, there are several lines of evidence which make it likely there are others yet to be discovered. Identifying these new factors would shed light on the function of Sd within whichever tissue a new TIF is discovered. With this in mind a candidate-gene approach was used to identify *Drosophila vestigial-like 4* (d*vgl-4*). Herein it is demonstrated that dVgl-4 contains two putative Sd interacting domains and is able to interact with Sd *in vitro* and *ex vivo*. It is also shown that this protein is able to act in a dominant negative fashion during wing development, and that there are likely two isoforms of mRNA expressed from this gene, and that the expression of each is likely under the control of different promoters. Finally, over-expression phenotypes are described for several tissues in order to begin elucidating a potential function for dVgl-4 in *Drosophila* development.

TABLE OF CONTENTS

CHAPTER ONE: GENERAL INTRODUCTION

CHAPTER TWO: IDENTIFICATION OF A CLASSICAL NUCLEAR LOCALIZATION SIGNAL IN SCALLOPED

CHAPTER THREE: IDENTIFYING A NOVEL BINDING PARTNER OF SCALLOPED

CHAPTER FOUR: GENERAL CONCLUSIONS AND FUTURE DIRECTIONS

SUMMARY OF FIGURES

SUMMARY OF TABLES

Summary of genes and abbreviations

A/P	anterior-posterior
A1-8	anterior segments 1-8
act	*actin*
antp	*antennapedia*
ap	*apterus*
ATTS	AbaA TEC-1 TEF-1 sequence
bcd	*bicoid*
bru-3	*bruno-3*
BSA	bovine serum albumin
cad	*caudal*
Crm1	*Chromatin Region Maintenance 1*
cNLS	classical nuclear localization signal
Co-IP	Co-Immunoprecipitation
cut	*ct*
Cy	*Curly*
D/V	dorsal-ventral
DAPI	4'-6-Diamidino-2-phenylindole
dcr-2	*dicer-2*
dE2F1	*Drosophila E2F1*
DHR96	*Drosophila Hormone receptor-like in 96*
dm	*diminutive* (aka *Drosophila myc; dmyc*)
Dl	*Delta*
Dll	*Distal-less*
d*Mef-2*	*Drosophila Mef-2*

dpp	*decapentaplegic*
DIAP1	*Drosophila inhibitor of apoptosis protein-1*
DTEF	*divergent TEF*
d*vgl-4*	*Drosophila vestigial-like 4*
eGFP	enhanced green fluorescent protein
eGFPx2	eGFPx2 + GST
egfr	*epidermal growth factor receptor*
emb	*embargoed*
en	*engrailed*
esg	*escargot*
ETF	*embryonic TEA domain-containing factor*
eve	*even-skipped*
Excl.	Excluded
fng	*fringe*
ftz	*fushi tarazu*
GR	*glucocorticoid receptor*
GST	glutathione-S-transferase
grk	*gurken*
h	*hairy*
H1-3	head segments 1-3
hb	*hunchback*
hh	*hedgehog*
h*PL*	human *placental lactogen*
hpo	*hippo*
Hsp90	*Heat shock protein 90*

Imp-α1-3	*Importin-α1-3*
Imp- β1	*Importin-β1*
inv	*invected*
kni	*knirps*
Kr	*Kruppel*
LB	Leptomycin B
MCAT	muscle-specific cytidine-adenosine-thymidine
Mef2	*Myocyte enhancer factor 2*
mRFP	monomeric red fluorescent protein
MyoD	*Myogenic differentiation*
N	*Notch*
NLS	nuclear localization signal
nos	*nanos*
NPC	nuclear pore complex
Nuc.	Nuclear
NES	nuclear export signal
Nups	nucleoporins
Or	*odor receptor*
Or59c	*Odorant receptor 59c*
Or85d	*Odorant receptor 59c*
Oregon-R	Ore[R]
ORN	olfactory receptor neuron
pal/myr	palmitoylation/myristoylation
qPCR	quantitative real-time PCR
ran	***RA**s-related **N**uclear protein*

RNAi	RNA interference
RTEF	*related to TEF-1*
run	*runt*
S2	Schneider 2
sd	*scalloped*
sens	*senseless*
siRNA	small-interfering RNA
sna	*snail*
SOP	sensory organ precursor
sal	*spalt*
SRF	*serum response factor*
stv	*starvin*
SV40	*Simian virus 40*
T1-3	thoracic segments 1-3
Tb	*Tubby*
TDU	*Tondu* (aka *vestigial-like 1*)
TEA	**TE**F-1/**TE**C-1 **A**baA
TEAD	TEA/ATTS DNA binding domain
TEC-1	*Transposon Enhancement Control-1*
TEF-1	*transcriptional enhancer factor-1*
TIF	transcriptional intermediary factor
tsh	*teashirt*
VID	Vestigial interacting domain
vg	*vestigial*
vg^{BE}	*vestigial* boundary enhancer

vgl 1-4	*vestigial-like1-4*
vg^{QE}	*vestigial* quadrant enhancer
vn	*vein*
Vps36	*Vacuolar protein sorting 36*
wg	*wingless*
w	*white*
y	*yellow*
Yap65	*Yes-associated protein-65* (*aka Yap1*)
yki	*yorkie*
YID	Yorkie interacting domain

Chapter One: General Introduction

The TEF-1 family of transcription factors

The grouping of the *transcriptional enhancer factor-1* (*TEF-1*) family of transcription factors is based on the TEA/ATTS DNA binding domains (TEAD) common to these proteins. The domain was described as a region of amino acid identity present in TEF-1 (*Homo sapiens*), Transposon Enhancement Control-1 (TEC-1, *Saccharomyces cerevisiae*) and AbaA (*Aspergillus nidulans*) (Andrianopoulos and Timberlake, 1991; Burglin, 1991), and it is these proteins for which the domain was named (TEA for **TE**F-1/**TE**C-1 **A**baA, and ATTS for **A**baA, **T**EC-1, **T**EF-1 **s**equence). Since the original classification in 1991, family members have been found throughout Eukarya. Indeed, four different classes of TEAD proteins have been found in vertebrates, and many organisms have several family members. Although many different naming schemes have been used for members of these four classes, a unified approach has been proposed in which the subfamilies are named *TEF-1*, *divergent TEF-1* (*DTEF-1*), *related to TEF-1* (*RTEF-1*) and *embryonic TEA domain-containing factor* (*ETF*), based on the first vertebrate example of each cloned (Xiao et al., 1991; Yasunami et al., 1995; Azakie et al., 1996; Stewart et al., 1994, 1998) and it is this scheme that will be used herein. While these proteins are about 70% identical overall, the TEAD is ~100% identical between the four classes (Yoshida, 2008). The TEAD was originally predicted to contain three α-helices (Davidson et al., 1988) and recent solution nuclear magnetic resonance spectroscopy work verifies this (Anbanandam et al., 2006). When the TEAD domain was first characterized, it was demonstrated that human TEF-1 could bind to Sph and GT-IIC binding sites in the early promoter of Simian virus 40 (Davidson et al., 1988). Later, it was also demonstrated that members of this family can interact with muscle-specific cytidine-adenosine-thymidine (MCAT) elements which

regulate the expression genes involved in muscle development (Farrance et al., 1992; Maeda et al., 2002b).

In vertebrates, the TEAD family of transcription factors has long been implicated in neural and muscle development (Shimizu et al., 1993; Chen et al., 1994; Maeda et al., 2002a; Milewski et al., 2004) and more recent work has shown that TEAD proteins act to control cellular proliferation in a variety of tissues via the Hippo (Hpo) pathway (Zhao et al., 2008; Ota and Sasaki, 2008). Although members of this family are generally widely expressed in a given organism, the ability of TEAD proteins to activate transcription is thought to rely on transcriptional intermediary factors (TIFs), which allow activation of a variety of downstream genes in a tissue specific manner, depending on the TIFs present (Hwang et al., 1993). One family of TIFs is the Vestigial-like (Vgl) family, indentified based on the presence of a TEAD protein interacting domain, called a TONDU (TDU) domain, which was originally identified in the human homologue of *Drosophila* Vestigial (Vg), TDU (Vaudin et al., 1999). In chicks, Vgl-2 responds to signalling by muscle differentiation factors (such as Myogenic differentiation, MyoD) and is expressed in developing and adult skeletal muscle tissue (Maeda et al., 2002a; Bonnet et al., 2010). Furthermore, *in vitro* experiments have demonstrated that that Vgl-2 interacts with RTEF-1, TEF-1 and Myocyte enhancer factor-2 (MEF2) which has long been known to be critical in animal muscle development (Olson et al., 1995; Maeda et al., 2002a). Intriguingly, RTEF-1 has a higher affinity for MCAT sites in the presence of Vgl-2 compared to when Vgl-2 is absent, while the opposite relationship is seen for TEF-1/Vgl-2 complexes (Chen et al., 2004a). This an important result, because in addition to being present in muscle gene promoters, MCAT elements are also present in placental gene promoters (e.g. human *placental lactogen,* h*PL*) which are known to be TEAD protein targets (Chen et al., 2004a). Thus, the ability of tissue specific TIFs to alter TEAD binding to target MCAT sequences provides one mechanism by which TIFs might regulate TEAD protein function. More evidence of the role of vgl-2 as a muscle specific TIF exists; *vgl-2* morpholinos interfere with the differentiation of skeletal muscle in chicks (Chen

2

et al., 2004a). Finally, both *vgl-2* and *DTEF-1* homologues are found in zebrafish muscle-cell progenitors

(Mann et al., 2007). Altogether, the evidence supports the idea that vertebrate Vgl-2 is a muscle specific

TIF of the TEAD proteins. There are other examples as well: Vgl-1 and Vgl-3 are virtually exclusive to the

placenta specific factors, and thus could be placental specific TIFs (Maeda et al., 2002a), while Vgl-4 has

been shown to be critical for cardiac muscle development and could be a heart-specific TIF (Chen et al.,

2004b). Another example is in mouse cell lines and embryos, where Yes associated protein-65 (Yap65)

acts together with mouse TEF-1 and ETF to mediate Hpo dependent proliferation (Zhao et al., 2008; Ota

and Sasaki, 2008). In total, it seems probable that the tissue specificity of TEAD proteins is conferred –

at least partially – by the array of cofactors present in a given tissue.

scalloped

Among invertebrates, the first TEAD family member identified was encoded by the *Drosophila*

melanogaster gene, *scalloped* (*sd*), which is X-linked and was cloned in 1991 (Campbell et al., 1991).

Drosophila is an excellent system to study the TEAD gene family for a variety of reasons. First, the

organism itself is a well-studied model system with many benefits. These include a published genome

(Adams et al., 2000), a short generation time (about 10 days at 25°C) and a wealth of genetic tools. For

example, many chromosomal markers exist as well as balancer chromosomes which are useful for

maintaining heterozygous mutations in the absence of selection. Furthermore, P-elements can be used

which allow for chromosomal integration of artificially constructed transgenes and the *UAS*-GAL4

system which can be used to drive transgene expression in a temporally and spatially specific manner

(Brand and Perrimon, 1993). There are also many pathways conserved between fruit flies and

vertebrates. Indeed, many signalling paradigms such as the Wingless (Wg) and Hedgehog (Hh) pathways

were originally discovered in *Drosophila* (Sharma and Chopra, 1976; Nüsslein-Volhard and Wieschaus, 1980).

Sd is itself a useful model for TEF-1 family function. Like other members of the family, Sd has a highly conserved TEAD (98% identical to that of TEF-1; Figure 1.1) and is 68% identical to TEF-1 throughout the rest of the protein (Campbell et al., 1992). Moreover, there is clearly functional redundancy, as TEF-1 is able to significantly rescue the wing phenotypes seen in sd^{ETX4} hypomorphs (Deshpande et al., 1997). In addition to the evidence for functional redundancy of the protein, it is apparent that the family is functioning in similar developmental pathways in both *Drosophila* and vertebrates. For instance, while Sd is best understood for its role in facilitating the development of the wing in *Drosophila* -which is quite specific to that organism - many other pathways for Sd function have been discovered. Indeed, like its vertebrate orthologues, Sd has been shown to have roles in muscle (both somatic and cardiac) development, neural development and cellular proliferation (Campbell et al., 1992; Srivastava and Bell, 2003; Garg et al., 2007; Goulev et al., 2008; Zhang et al., 2008; Deng et al., 2009). Another advantage to studying Sd is that, unlike in vertebrates, Sd is the only known TEF-1 family member in *Drosophila*. That said, *sd* is thought to code for several differentially spliced mRNAs (Figure 1.2; Campbell et al., 1991) and the product of each could function in unique ways. To date, the E21 mRNA is the only isoform whose product has been rigorously characterized and it is this protein product which is discussed herein.   Like the other members of the TEF-1 family, Sd has been shown to bind to MCAT elements and this binding also appears to be modulated by TIF cofactor binding, similar to RTEF-1 and TEF-1 as discussed previously. Indeed, as was the case with Vgl-2 and TEF-1, Vg is able to greatly reduce the ability of Sd to bind to MCAT sites. Moreover, when in complex with Vg, the binding preference of Sd changes from singlet so called "A sites", to doublet "B sites" (Halder and Carroll, 2001; Halder et al., 1998) .

The E21 mRNA isoform of *sd* codes for a 440 amino acid protein, which is known to have at least three (partially overlapping) domains and these domains are also present all of the other predicted isoforms (Figures 1.2 and 1.3). The first domain is the TEAD as noted previously. In addition to this DNA binding domain, two cofactors of Sd , Yorkie (Yki; the *Drosophila* homolog of Yap65) and Vg, have been shown to interact with the C-terminal domain of Sd (Simmonds et al., 1998; Paumard-Rigal et al., 1998; Goulev et al., 2008). More specifically, based on data involving the human homologs of Sd and Vg, Vg is thought to bind to a domain called the Vestigial interacting domain (VID) which lies within amino acids 220-344 (Vaudin et al., 1999). Moreover, a TDU domain has also been identified in Vg, and this domain (along with C- and N- terminal activation domains) has been shown to be necessary for Vg – and by extension Sd – function in wing development, supporting the idea that Sd lacks an activation domain and is thus unable to activate transcription in the absence of TIFs (Hwang et al., 1993; MacKay et al., 2003; Vaudin et al., 1999). The last domain is inferred from X-ray crystallography data which mapped the Yap65 interaction domains of TEF-1 and RTEF-1 (Li et al., 2010; Chen et al., 2010). The critical amino acids present in the Yap65 interacting domains of these proteins are also present in Sd and lie with the domain stretching from amino acids 267-435. Thus, it is plausible that this region contains the Yki-interacting domain of Sd. Sd has also been shown to interact with *Drosophila* Mef2 (dMef2), and the site of this interaction may constitute a fourth domain, though the binding site is currently uncharacterized (Deng et al., 2009).

Drosophila Development

The first step in the patterning of *Drosophila* occurs during oogenesis when the unfertilized egg is loaded with maternally provided mRNAs. Initially the anterior-posterior (A-P) and dorsal-ventral (D/V) polarity is established when *gurken* (*grk*) is expressed in the dorsal anterior corner of the developing

oocyte (Neuman-Silberberg and Schüpbach, 1993; Roth, 2003; Van De Bor et al., 2005).   This positional

information is then used to enrich the anterior and posterior poles of the embryo with maternally

loaded *bicoid* (*bcd*) and *nanos* (*nos*) mRNA, respectively, prior to fertilization and the activation of

zygotic transcription (Driever and Nüsslein-Volhard, 1988a, 1988b; St Johnston and Nüsslein-Volhard,

1992).   Furthermore, *dorsal* (*dl*) becomes enriched in the ventral side of the embryo (Steward, 1989; Ip

et al., 1991). Subsequent to fertilization, the developing embryo exists as a syncytium of nuclei, which

divide synchronously 13 times (Foe and Alberts, 1983). During the 14th cycle, the embryo begins the

process of cellularization. Additionally, between the 10th and 14th cycles, zygotic transcription begins in

earnest (Anderson and Lengyel, 1979). This transcription is regulated by the maternal factors that have

already established polarity within the embryo.   Indeed, the maternal *bcd*, *nos* and *dl* mRNAs are

translated and, due to their enrichment in particular regions of the embryo and the nature of the

syncytial embryo, establish anterior-posterior, posterior-anterior and ventral-dorsal gradients of

expression, respectively, of the three morpchogens.   Historically, the gradients were thought to be

based only on protein diffusion (Driever and Nüsslein-Volhard, 1988a), but both old and recent work

have demonstrated that *bcd* mRNA diffuses and that this diffusion is likely at least partially responsible

for the Bcd gradient (Frigerio et al., 1986; Spirov et al., 2009). Thus, it is possible this is also true for the

other two morphogens. Regardless of the mechanism(s) of gradient establishment, the information is

then used to further subdivide the developing embryo into 14 segments (three anterior which

eventually form the head, three thoracic which eventually form the legs, wings and halteres and 8

abdominal; Figure 1.4), in the following manner (DiNardo et al., 1994; Pick, 1998; Sanson, 2001):   First,

Bcd and Nos act to define the expression of two other maternally inherited mRNAs – *hunchback* (*hb*)

and *caudal* (*cad*) – such that Hb forms a gradient concentrated in the anterior portion of the developing

embryo, while the inverse is true of Cad. The expression of these genes leads into regulation of the

identical zygotic genes as well as downstream factors known as the gap genes, examples of which are

kruppel (*kr*), *knirps* (*kni*) and *hb* itself. The gap genes divide the embryo into coarse segments along the

A/P axis. Through the activity (both direct and indirect) of the gap proteins, the pair-rule genes (e.g.

even-skipped (*eve*), *fushi tarazu* (*ftz*), *hairy* (*h*) and *runt* (*run*)) further subdivide the embryo into seven

pairs of parasegments.   Each parasegment is then divided into an anterior, *engrailed* (*en*) expressing

region and a posterior, *wg* expressing region, establishing polarity for the subsequent development of

each segment (Figure 1.5). Finally, the homeotic transcription factors assign segmental identity to the

polarized segments. A classic example of a homeotic gene is *antennapedia* (*antp*).   The primary

function of Antp is to act as a switch between the leg and antennal developmental programs. Indeed,

one of the best known homeotic transformations involves gain-of-function mutations which cause

antennal discs to develop as legs (Gehring, 1967; Lewis et al., 1980). Conversely, loss-of-function

mutations cause a leg to antennal homeotic transformation (Struhl, 1981). It is important to note that

each segment also has anterior-posterior polarity, defined by the expression of Wg in the anterior half

and Engrailed (En) in the posterior half of each segment (Couso et al., 1993). After segmental

specification, gastrulation proceeds until the egg hatches and a first instar larva emerges approximately

24 hours after fertilization. This larva then undergoes two molts (each separated by roughly 24 hours of

feeding) progressing from the first to second to third larval instar. The third instar feeds for about 48

hours before leaving the food and beginning the five day process of pupariation. It is during pupariation

that the adult structures are formed, with the majority of the adult structures developing from clusters

of cells known as the imaginal discs, whose development begins in the embryonic ectoderm (Auerbach,

1936; Garcia-Bellido and Merriam, 1969; Wieschaus and Gehring, 1976), in a segment specific fashion

(Akam, 1987). From these discs the head, thorax, legs, wings, halteres and genitalia are formed – along

with cuticle that forms the majority of the body wall - while the cells of the histoblast nests make up the

abdominal epidermis.

Scalloped During Development

The *sd* locus was originally identified by mutational analysis in 1929 (Gruneberg, 1929).

However, once the gene was cloned, a thorough temporal and spatial analysis of the expression of the

gene was undertaken, using *sd*^{ETX4} which is an enhancer trap allele of *sd* (Campbell et al., 1991). In third

instar larvae, reporter expression was detected in almost all discs, including the wing, eye-antennal, leg

discs and genital discs. Furthermore, expression was observed in and around the optic lobes and in the

ventral nerve cord. Staining was also seen in the embryo, particularly in cells of the central and

peripheral nervous system (CNS and PNS). More recently, *sd* expression has also been detected in

embryonic somatic and cardiac muscle cells (Deng et al., 2009). The expression data of *sd* are

summarized in Table 1.1.

Consistent with the expression data described above Sd has been shown to be important for a

variety of developmental programs. Indeed, *sd* mutants have defects in cardiac development (Deng et

al., 2009). There is also evidence that Vg modulates flight muscle differentiation (Sudarsan et al., 2001).

Given that Sd is co-expressed with Vg in the precursors of flight muscle cells (Bernard et al., 2003), it is

possible that Sd is also involved in this process. That said, no *sd* mutants have been demonstrated to

have flight muscle defects. Mitotic clones of *sd*^{47M} (a larval lethal allele of *sd*) in the eye and leg disc

cause mispatterning and loss of eye bristles, as well as loss of distal leg tissue in the adults; while

dominant negative forms of Sd cause gross defects in the eyes and legs (Garg et al., 2007). Roles for

nervous system development have also been described, as *sd* mutants have defects in sensory bristles

(most notably in the wing margin but also in the adult eye; (Campbell et al., 1992; Srivastava and Bell,

2003; Garg et al., 2007). Also, Sd expression can modulate the expression of several *Odor receptor* (*Or*)

genes, and thus specify the identity of at least some olfactory receptor neurons (ORNs) which are

present in the olfactory organs (the antenna and maxillary palp) and whose identity is based on the

subset of *Or* genes expressed within them (Ray et al., 2008). Finally, Sd, along with Yki is required for cell proliferation in both the eyes and wings (Zhang et al., 2008; Goulev et al., 2008). It is important to note that no specific cofactor for Sd activity has, as of yet, been found in either the developing legs discs or the optic lobe. However, *yki* is broadly expressed (Chintapalli et al., 2007) and the protein may be a cofactor of Sd in either or both of these tissues. Still, verifying whether or not this is true, and identifying any other TIF(s) of Sd in these tissues remains an important issue.

Drosophila Wing Development

While progress is being made in understanding the role of Sd in the tissues noted previously, to date the majority of studies has focused on the role of Sd in the development of the wing imaginal disc of the third instar larvae of *Drosophila*. The wing of *Drosophila* has been an attractive model for studying Sd function, primarily because it is not necessary for the viability of laboratory stocks, and moreover, it is a highly patterned tissue which is sensitive to perturbations. Thus, it is worthwhile examining wing development in detail, along with what is known about the role of Sd in this tissue. In order to understand what follows, a fate map relating the regions of the developing wing disc that give rise to the corresponding regions of the adult wing is provided for reference as Figure 1.6 (Bryant, 1975).

In *Drosophila*, the cells of the wing primordia first arise due to their proximity to a group of cells in the thoracic segment which express both Wg (in a dorsal-ventral stripe) and Decapentaplegic (Dpp; perpendicular to the Wg stripe, i.e. along the A/P boundary) (Cohen et al., 1993; Campbell et al., 1993). Once specified as disc cells by these signals they proliferate and separate into two populations: ventral Distal-less (Dll) expressing cells (which will give rise to the leg discs in all three thoracic segments; T1, T2 and T3), and dorsal Dll-free cells (which will give rise to the wing and haltere discs in segments T2 and T3, respectively; (Fuse et al., 1996b)). While all three thoracic segments are competent to form wing

discs at this stage, the actions of the homeotic Sex comb reduced (Scr) and Ultrabithorax (Ubx) proteins repress or modify this formation in the first and third thoracic segments, respectively (Carroll et al., 1995; Weatherbee et al., 1998). Once the wing and haltere primordia segregate from the presumptive leg primordia, the presumptive wing discs begin to express Snail (Sna) and Escargot (Esg).   These transcription factors act to induce the expression of *vg,* which serves as the earliest known marker for the wing imaginal discs (Williams et al., 1991; Fuse et al., 1996a).

Wing disc development is a process that tightly couples growth (the wing disc expands from ~50 cells in first instar to ~50,000 cells by the late third instar) and patterning. This patterning begins in the embryo (as discussed above) and continues in first instar larvae, as cells of the wing disc maintain the embryonically derived expression of *en* in the posterior compartment. However, *wg* expression is lost during the migration of these cells from the leg disc primordia and is not seen at this time. Expression of *vg* is also maintained and is ubiquitous during this stage (Williams et al., 1993; Couso et al., 1993). During the second instar, the anterior/posterior (A/P) boundary is set up through the activity of Dpp which is expressed as a stripe on the anterior side of the A/P boundary, through the actions of Hh and the embryonically inherited En (Sanicola et al., 1995; Blair and Ralston, 1997). In addition to its role in specifying the A/P boundary, En, along with its co-expressed paralog Invected (Inv), is necessary to define the posterior compartment (Simmonds et al., 1995; Simmonds and Bell, 1998). Once the spatial characteristics of the wing are established, *wg* expression is re-initiated as a wedge in the ventral-anterior region of the disc and this is required, along with Dpp, to repress *teashirt* (*tsh*). The repression of *tsh* is likewise required for the proper development of the adult wing blade (Couso et al., 1993, 1995; Klein and Arias, 1998b; Wu and Cohen, 2002; Zirin and Mann, 2004). Another factor important in establishing polarity in the developing wing disc is Apterus (Ap), which has two functions: The first is to specify dorsal fate, while the second is to establish the dorsal/ventral (D/V) boundary (Diaz-Benjumea and Cohen, 1993; Blair et al., 1994; Williams et al., 1994; Couso et al., 1995; Milan et al., 2002). The

expression of *ap* (which is in the dorsal region of the wing disc, and overlaps the most dorsal region of wg expression) relies on Wg and Epidermal growth factor receptor (Egfr) signalling. Egfr is itself activated by a secreted ligand, Vein (Vn). However, while the expression of *ap* is controlled as described above, Vg is required for the proper activity of Ap (Couso et al., 1995; Wang et al., 2000; Delanoue et al., 2002). The role of Ap in defining the D/V boundary is mediated by the activation of Notch (N), through its ligands Serrate (Ser) and Delta (Dl), as well as a regulator of N activity, Fringe (Fng). It is known that Ap function is involved in regulating expression of s*er* and *fng*; furthermore the juxtaposition of dorsal Fng[+] cells with ventral Fng[-] cells is necessary for N activation along the D/V boundary (Couso et al., 1995; Kim et al., 1995), although it is not known how this occurs (Panin et al., 1997; Klein and Arias, 1998c; Delanoue et al., 2002; Milan and Cohen, 2003). Regardless, it is clear that this N activation, in conjunction with Wg signalling, is required to maintain *vg* expression along the D/V boundary, via the *vg* Boundary Enhancer (*vg*[BE]; Figure 1.7) (Couso et al., 1995; Kim et al., 1995, 1996; Klein and Arias, 1998b, 1999). By the end of the second instar, Vg is localized along the D/V boundary and, along with Wg (whose expression is refined to this region through the actions of N and Vg), is required for the specification of the wing margin (Liu et al., 2000; MacKay et al., 2003; Srivastava and Bell, 2003). In the third instar, the wing blade is specified by the activity of Vg which is activated in the pouch of the wing disc through the *vg* Quadrant Enhancer (*vg*[QE]; Figure 1.7). This enhancer is regulated by inputs from both the D/V (through margin localized Vg and Wg) and A/P (through Dpp signalling) boundaries but is not active in the wing margin, possibly due to repression via Egfr signalling along the D/V boundary (Klein and Arias, 1998b; Nagaraj et al., 1999; Guss et al., 2001). The pouch and D/V expression of Vg, along with Sd, is required for wing blade and margin fate. Also at this time, Wg is expressed in a θ pattern, forming a stripe across the D/V boundary, as well as marking the outline of the presumptive wing pouch (Figure 1.6; (Couso et al., 1994). It is the expression of Wg along the periphery of the nascent wing, along with Vg and other factors, that acts to specify and pattern the proximal and distal

wing hinge (Neumann and Cohen, 1997; Klein and Arias, 1998a, 1999; Casares and Mann, 2000; Azpiazu and Morata, 2000; Liu et al., 2000; Rodriguez et al., 2002; Kolzer et al., 2003; Whitworth and Russell, 2003).

Control of wing disc growth involves many of the same factors as patterning (reviewed in Neto-Silva et al., 2009). For example, both Dpp and Wg are required for patterning, as mentioned, and both can induce growth when ectopically expressed (Martín-Castellanos and Edgar, 2002; Giraldez and Cohen, 2003). However, in their absence the wing disc fails to grow properly and eventually is lost (Couso et al., 1993; Zecca et al., 1995). The mechanism of tissue loss is different for each; when Wg is absent, the pro-apoptotic gene *hid* is induced and the cells of the disc die. On the other hand, when Dpp signalling is absent, the cells of the disc delaminate from the epithelium and are killed due to the activation of the Jun-N-terminal kinase (JNK) stress pathway (Gibson and Perrimon, 2005; Shen and Dahmann, 2005). To some degree, the proliferative functions of Wg and Dpp have been shown to be effected by the actions of Diminutive (Dm; aka dMyc) (Johnston et al., 1999; Prober and Edgar, 2000; Duman-Scheel et al., 2004). Indeed, Dpp signalling can increase Dm levels, while Wg can inhibit *dm* expression at the D/V boundary thus inducing a transient cell cycle arrest. Dm affects cell proliferation through a mechanism known as cell competition; cells which express low levels of Dm relative to their neighbours will die, while those expressing higher levels of Dm will proliferate (de la Cova et al., 2004; Johnston, 2009). Dpp has also been shown to regulate Yki function via the Hpo/Wts pathway (Huang et al., 2005; Rogulja et al., 2008) and Yki is required for proliferation in the wing disc (Huang et al., 2005).

Scalloped During Wing Development

A generally accepted model for Sd/Vg function is that each protein provides discrete functions to the complex. Specifically, Vg provides transcriptional activation, while Sd allows the complex to

12

localize to the nucleus and bind DNA (Halder et al., 1998; Paumard-Rigal et al., 1998; Simmonds et al., 1998; Srivastava et al., 2002; MacKay et al., 2003). There is evidence that the target specificity of DNA binding is conferred by the TEAD of Sd, but modified by Vg/Sd interactions (Halder and Carroll, 2001; Hwang et al., 1993). Several genes whose enhancers are bound and activated by the Sd/Vg complex, in combination with other signals, have been identified. These include *serum response factor* (*srf*), *cut* (*ct*), *Drosophila E2F1* (*dE2F1*), *Drosophila inhibitor of apoptosis protein-1* (*DIAP1*), *Or59c* and possibly *spalt* (*sal*) and *Or85d* (Jack and DeLotto, 1992; Guss et al., 2001; Halder and Carroll, 2001; Barrio and de Celis, 2004; Delanoue et al., 2004; Goulev et al., 2008; Zhang et al., 2008). Furthermore, *sd* and *vg* expression are auto-regulated and the Sd/Vg complex is necessary to maintain both vg^{QE} and *sd* transcription in the developing wing disc (Williams et al., 1993; Guss et al., 2001). The co-expression of Sd and Vg is necessary for proper patterning and proliferation of the blade and margin tissues of the wing (Delanoue et al., 2004) and misexpression of Vg is sufficient to induce ectopic wing tissue in competent tissues, which include most of the imaginal discs (Kim et al., 1996). It has been proposed that the presence of Sd – likely through its ability to complex with Vg – along with N and Wg signalling, is what confers competence to Vg induced wing formation in imaginal discs (Maves and Schubiger, 1998; Kurata et al., 2000; Baena-Lopez and Garcia-Bellido, 2003). For example, in the eye discs where Sd, N and Wg are expressed, but Vg normally is not, the ectopic expression of Vg leads to an outgrowth of wing tissue from the adult eyes (Simmonds et al., 1998). More specific to the patterning of the wing itself, it is also known that the Sd/Vg complex is required for either the establishment or maintenance of the D/V stripe of Wg expression required for proper wing patterning (Srivastava and Bell, 2003). It appears that Sd is also required for the proper function of Senseless (Sens) which in turn is vital for the specification of sense organ precursors (SOPs), which the bristles which line the margin of the adult wing eventually derive from (Srivastava and Bell, 2003). Finally, Sd complexes with Yki, and subsequently facilitates both the translocation of the complex to the nucleus and the binding of the complex to target DNA sites,

allowing for Yki to activate target genes which induce proliferation (an example of which is *diap1*) (Goulev et al., 2008; Zhang et al., 2008; Wu et al., 2008).

Scalloped as a model for the TEAD family

Understanding how the TEAD family functions is of great importance, since, as described above, family members participate in a variety of critical developmental processes such as neural genesis and muscle development. Moreover, now that the family is known to be involved in the Hippo pathway, it raises the possibility that TEAD family members may be critical for controlling proper cellular proliferation during development, which means the mis-regulation of this gene family may have implications in the development of cancers.

There are several reasons why it is advantageous to use Sd as a model for studying the TEAD family of proteins. First, as mentioned above, *Drosophila* is itself an excellent model system, with many genetic and molecular tools which can be utilized. Also, within *Drosophila*, *sd* is the only TEAD family member, which means there is no need to worry about functional redundancy. The high degree of protein identity seen between Sd and its vertebrate orthologues also implies that any insights gained into the structure and function of Sd are likely to be applicable to the other family members. Additionally, the fact that there is conservation of function and that (so far) all known cofactors of Sd are present in higher organisms (and vice versa; see the previous sections) implies that any new pathways and cofactors identified in *Drosophila* are likely to be present in vertebrates as well, and moreover, the information gained about them will also be applicable.

Research Focus

As described above, it has long been known that Sd is vitally important for wing development, and recent evidence for roles in other tissues and pathways have also been described. A functional dissection of the Sd protein has been carried out which attempted to assign functions to different regions of Sd, particularly in regards to wing development (Chow et al., 2004). Moreover, as expected for a transcription factor, Sd has been shown to be localized to the nucleus and is thought to be required for the proper nuclear localization of Vg (Halder et al., 1998; Simmonds et al., 1998; Srivastava et al., 2002). However, nothing is known in regard to how Sd itself translocates to the nucleus. Furthermore, it is not clear if all of the TIFs of Sd have been identified. In fact, it seems likely given the broad expression of Sd, that other unknown binding partners exist, besides the three described above (Vg, dMef-2, Yki). Based on the first of these two ideas, experiments to characterize regions which mediate the nuclear translocation of Sd are presented in Chapter Two. Indeed, the experiments presented therein show that Sd contains a bipartite nuclear localization signal which would explain the nuclear localization of Sd. However, the data described also provide evidence that Sd contains a nuclear export signal, and that the C-terminal region of Sd influences the localization of the protein. On the other hand, Chapter Three explores the possibility that an uncharacterized protein, CG10741, which contains two putative tandem TDU domains, is in fact an unidentified TIF of Sd. Finally, in Chapter Four, general conclusions are made and some insight into possible future directions is given.

Figure 1.1. Alignment of human TEAD proteins. The four human TEAD proteins (TEF-1, Related to TEF-1 (RTEF-1), divergent TEF-1 (DTEF-1) and Embryonic TEA domain-containing factor (ETF)) and the *Drosophila* Sd are shown aligned. Conserved residues are shaded black; the black bar running from amino acids 88-163 of Sd is the TEA domain; the gray bars running from amino acids 96-108, 127-139 and 145-162 of Sd represent α-helices one, two and three, respectively. Alignments were done using Jalview (Waterhouse et al., 2009).

```
hTEF-1    1 ME---PSS------------------------------------------WSGSESP-----AENMERMSD  21
hDTEF-1   1 ME---GTAGTIT-------------------------------------SNEWSSPTSPE----GSTASGGSQ  29
hRTEF-1   1 MA---SNS-----------------------------------------WNASSSP-----GEAREDGPE   21
hETF      1 MGEPRAGAALDD-------------------------------GSGWTGSEEG-----SEEGTGGSE       31
Sd        1 MKNITSSSTCSTGLLQLQNNLSCSELEVAEKTEQQAVGPGTIPSPWTPVNAGPPGALGSADTNGSM       66

hTEF-1   22 SADKPIDN-------------DAEGVWSPDIEQSFQEALAIYPPCGRRKIILSDEGKMYGRNELIA       74
hDTEF-1  30 ALDKPIDN-------------DAEGVWSPDIEQSFQEALAIYPPCGRRKIILSDEGKMYGRNELIA       82
hRTEF-1  22 GLDKGLDN-------------DAEGVWSPDIEQSFQEALAIYPPCGRRKIILSDEGKMYGRNELIA       74
hETF     32 GAGGDGGP-------------DAEGVWSPDIEQSFQEALAIYPPCGRRKIILSDEGKMYGRNELIA       84
Sd       67 VDSKNLDVGDMSDDEKDLSSADAEGVWSPDIEQSFQEALSIYPPCGRRKIILSDEGKMYGRNELIA      132

hTEF-1   75 RYIKLRTGKTRTRKQVSSHIQVLARRKSRDFHSKLK----DQTAKDKALQHMAAMSSAQIVSATAI      136
hDTEF-1  83 RYIKLRTGKTRTRKQVSSHIQVLARRKAREIQAKLK----DQAAKDKALQSMAAMSSAQIISATAF      144
hRTEF-1  75 RYIKLRTGKTRTRKQVSSHIQVLARKKVREYQVGIKAMNLDQVSKDKALQSMASMSSAQIVSASVL      140
hETF     85 RYIKLRTGKTRTRKQVSSHIQVLARRKSREIQSKLK----DQVSKDKAFQTMATMSSAQLISAPSL      146
Sd      133 RYIKLRTGKTRTRKQVSSHIQVLARRKLREIQAKIK--------------------VQFWQPGL      176

hTEF-1  137 HNKLGLPG-IPRPTFPGAPGFWPG-MIQTGQPGSSQDVKPFVQQAYPI-QPAVTAPIPGFEPASAP      199
hDTEF-1 145 HSSMALAR-GPGR--PAVSGFWQG-ALP-GQAGTSHDVKPFSQQTYAV-QPPL--PLPGFESPAGP      202
hRTEF-1 141 QNKFSPPSPLPQAVFSTSSRFWSSPPLLGQQPGPSQDIKPFAQPAYPI-QPPLPPTLSSYEP-LAP      204
hETF    147 QAKLGPTG--PQA--SELFQFWSG---GSGPPWNVPDVKPFSQTPFTLSLTPPSTDLPGYEPPQAL      205
Sd      177 QPST----------------------SQDFYDYSIKPFPQPPYPAGKTST--AVSGDETGIPP      215

hTEF-1  200 APS------VPAWQGRSIGTTKLRLVEFSAFLEQQRDPDSYNKHLFVHIGHANHSYSDPLLESVDI      259
hDTEF-1 203 APSP-SAPPAPPWQGRSVASSKLWMLEFSAFLEQQQDPDTYNKHLFVHIGQSSPSYSDPYLEAVDI      267
hRTEF-1 205 LPS--AAASVPVWQDRTIASSRLRLLEYSAFMEVQRDPDTYSKHLFVHIGQTNPAFSDPPLEAVDV      268
hETF    206 SPLPPPTPSPPAWQARGLGTARLQLVEFSAFVEPPDAVDSYQRHLFVHISQHCPSPGAPPLESVDV      271
Sd      216 SQLP--------WEGRAIATHKFRLLEFTAFMEIQRD-EIYHRHLFVQLGGK-PSFSDPLLETVDI      271

hTEF-1  260 RQIYDKFPEKKGGLKELFGKGPQNAFFLVKFWADLNCNIQ-DDAGA--------FYGVTSQYESSE      316
hDTEF-1 268 RQIYDKFPEKKGGLKDLFERGPSNAFFLVKFWADLNTNIE-DEGSS--------FYGVSSQYESPE      324
hRTEF-1 269 RQIYDKFPEKKGGLKELYEKGPPNAFFLVKFWADLNSTIQ-EGPGA--------FYGVSSQYSSAD      325
hETF    272 RQIYDKFPEKKGGLRELYDRGPPHAFFLVKFWADLNWGPSGEEAGAGGSISSGGFYGVSSQYESLE      337
Sd      272 RQIFDKFPEKSGGLKDLYEKGPQNAFYLVKCWADLNTDLTTGSETG------DFYGVTSQYESNE      330

hTEF-1  317 NMTVTCSTKVCSFGKQVVEKVETEYARFENGRFVYRINRSPMCEYMINFIHKLKHLPEKYMMNSVL      382
hDTEF-1 325 NMIITCSTKVCSFGKQVVEKVETEYARYENGHYSYRIHRSPLCEYMINFIHKLKHLPEKYMMNSVL      390
hRTEF-1 326 SMTISVSTKVCSFGKQVVEKVETEYARLENGRFVYRIHRSPMCEYMINFIHKLKHLPEKYMMNSVL      391
hETF    338 HMTLTCSSKVCSFGKQVVEKVETERAQLEDGRFVYRLLRSPMCEYLVNFLHKLRQLPERYMMNSVL      403
Sd      331 NVVLVCSTIVCSFGKQVVEKVESEYSRLENNRYVYRIQRSPMCEYMINFIQKLKNLPERYMMNSVL      396

hTEF-1  383 ENFTILLVVTNRDTQETLLCMACVFEVSNSEHGAQHHIYRLVKD                        426
hDTEF-1 391 ENFTILQVVTNRDTQETLLCIAYVFEVSASEHGAQHHIYRLVKE                        434
hRTEF-1 392 ENFTILQVVTSRDSQETLLVIAFVFEVTSEHGAQHHVYKLVKD                        435
hETF    404 ENFTILQVVTNRDTQELLLCTAYVFEVTSERGAQHHIYRLVRD                        447
Sd      397 ENFTILQVMRARETQETLLCIAYVFEVAAQNSGTTHHIYRLIKE                        440
```

17

Figure 1.2. Overview of *sd* mRNA isoforms. All predicted *sd* isoforms are shown along with their intron-exon structure, along with *cg8509*-RA (which is internal to the sd locus). The exons composing the TEA domain (orange box) as well as the Vestigial interacting domain (green box) are shown and are common to all of the isoforms. Modified from Tweedie et al., 2009.
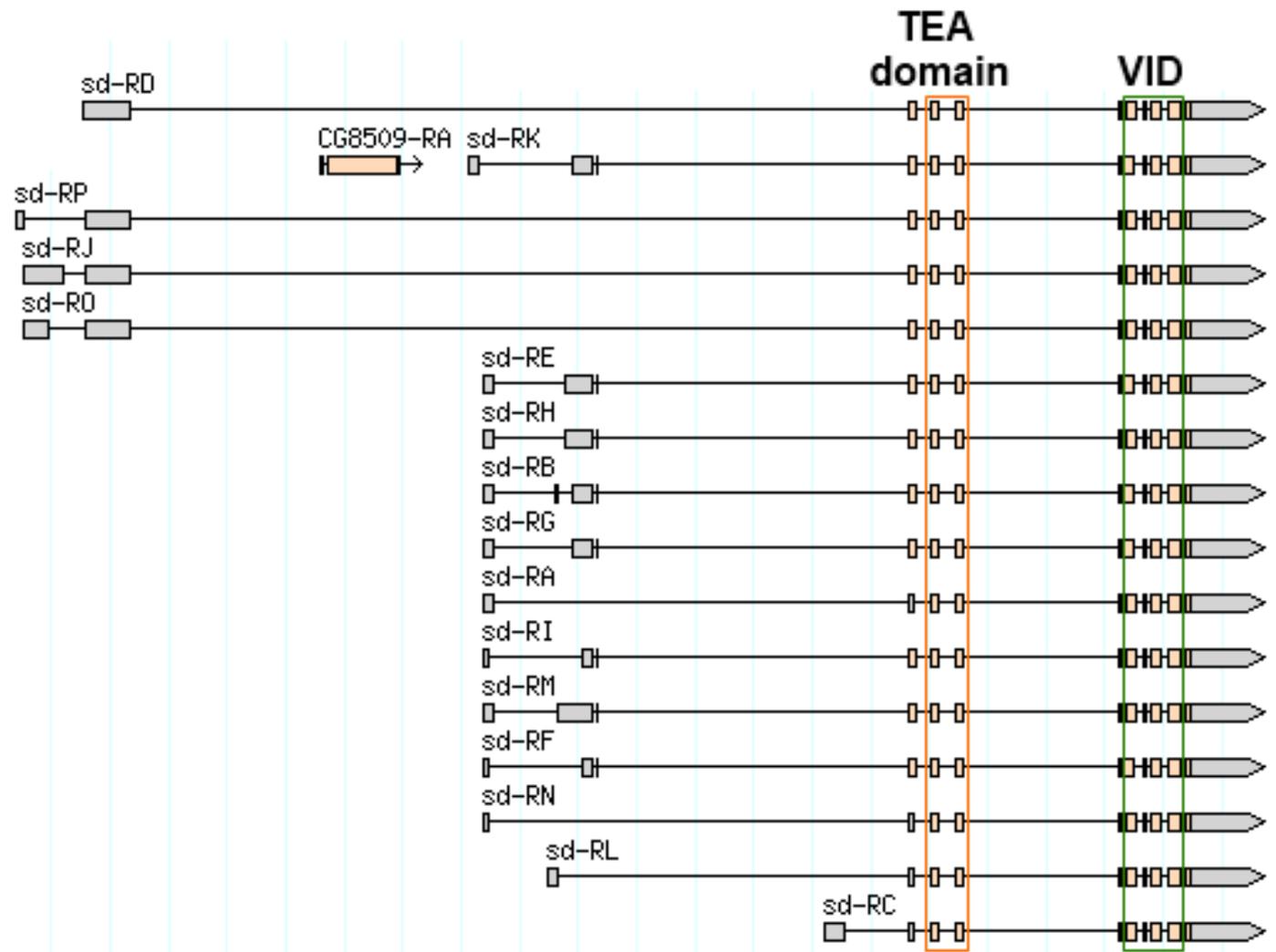
Figure 1.3. Schematic of the Sd protein. The three known domains of Sd are indicated. The TEA DNA binding domain and Vg interacting domain (VID) are as indicated. The black bar represents the extent of the predicted Yorkie interacting domain (YID), which is inferred based on the residues of TEF-1 and Related to TEF-1 which interact with Yes associated protein-65 (see text).
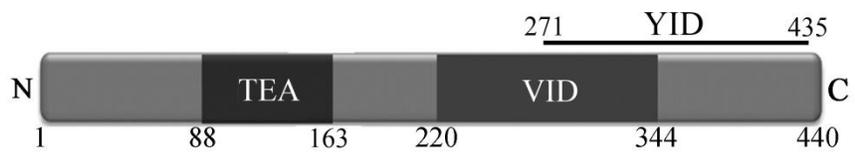
Figure 1.4. Fate map of *Drosophila* embryonic segments correlated to adult tissues. Simplified diagrams of a *Drosophila* embryo and adult are shown, with the adult structures colour-coded to show which embryonic segments they are derived from. H1-3 are head segments 1-3, T1-3 are thoracic segments 1-3 and A1-8 is abdominal segments 1-8. Modified from Sadava et al., 2009.

A H1-3 T1-3 A1-8 P
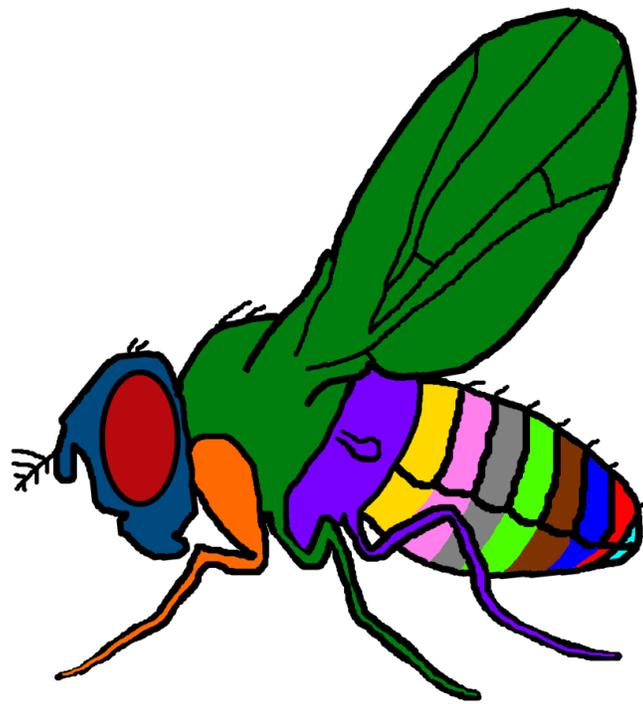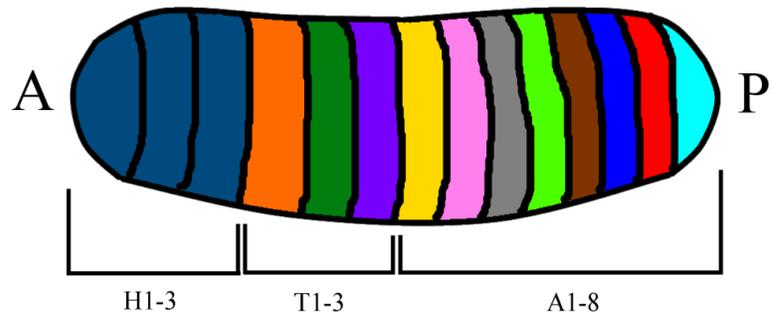
Figure 1.5. Schematic of early embryonic A/P patterning. Maternally inherited transcripts of *bicoid* (*bcd)* and *nanos* (*nos)* are translated and the proteins form anterior-posterior and posterior-anterior gradients, respectively. This leads to translation of *cad* transcripts which together with Bcd, regulate translation of maternal *hb* (*hunch-back*) transcripts and establish a pattern of Hb localization similar to Bcd. Hb induces the zygotic expression of gap genes, which further divide the embryo along the A/P axis. This causes further subdivision into parasegments due to the expression of the pair-rule genes, and then each parasegment is divided into A and P halves due to the effects of segment-polarity gene expression. Hb (M) is Hb translated using maternally inherited *hb* transcript; *hb* (Z) is zygotically transcribed *hb*. Modified from Griffiths, 2002.
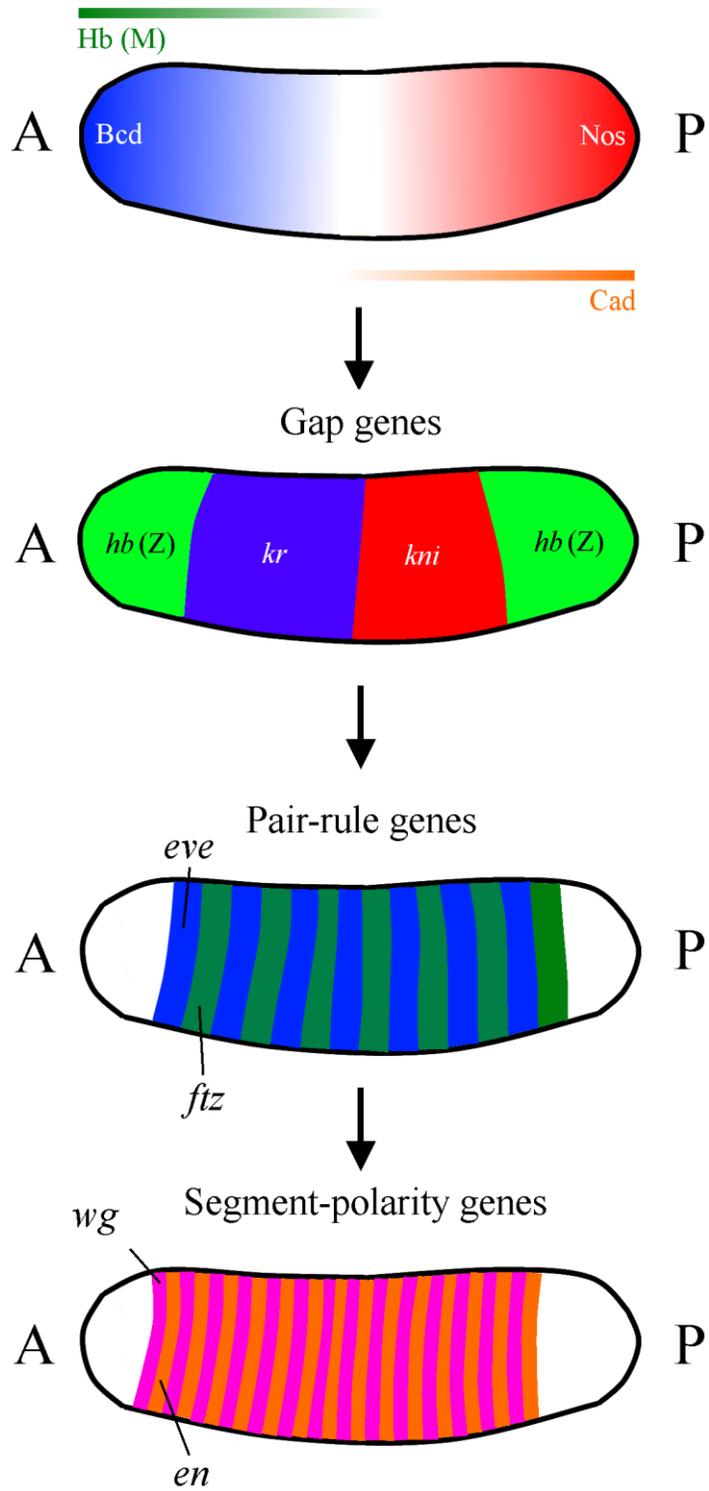
Figure 1.6. Simplified fate map of the wing imaginal disc. A) A diagram representing a third instar wing imaginal disc. Shown are the presumptive notum (blue), dorsal and ventral hinge (red and purple, respectively), dorsal and ventral blade (grey and green, respectively), pleura (pink) and margin (yellow). The wing pouch is inclusive of the area bounded by the presumptive dorsal and ventral hinge tissue. Modified from Bryant, 1975. B) A diagram of the dorsal surface of the adult wing, colour coded based on the regions of the disc that give rise to the corresponding tissues: the notum (blue), hinge (red), blade (grey) and margin (yellow).
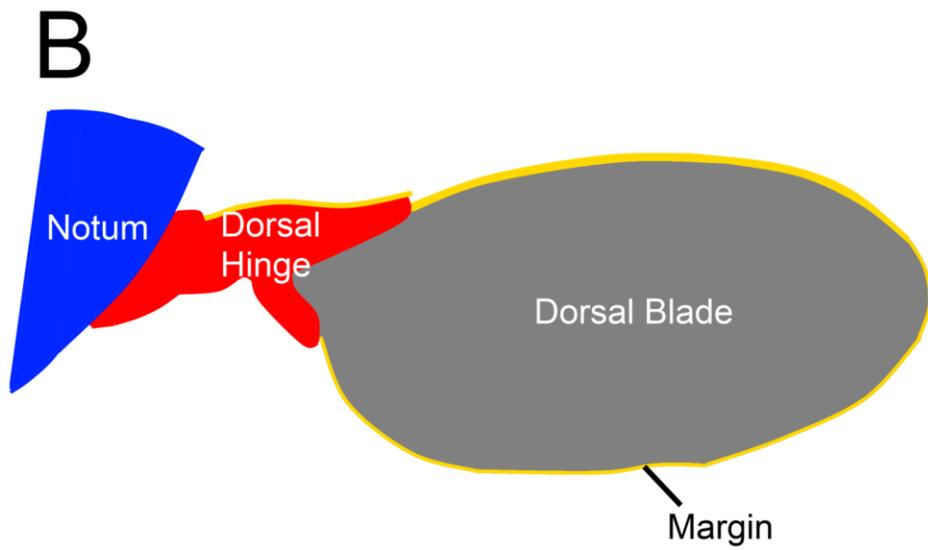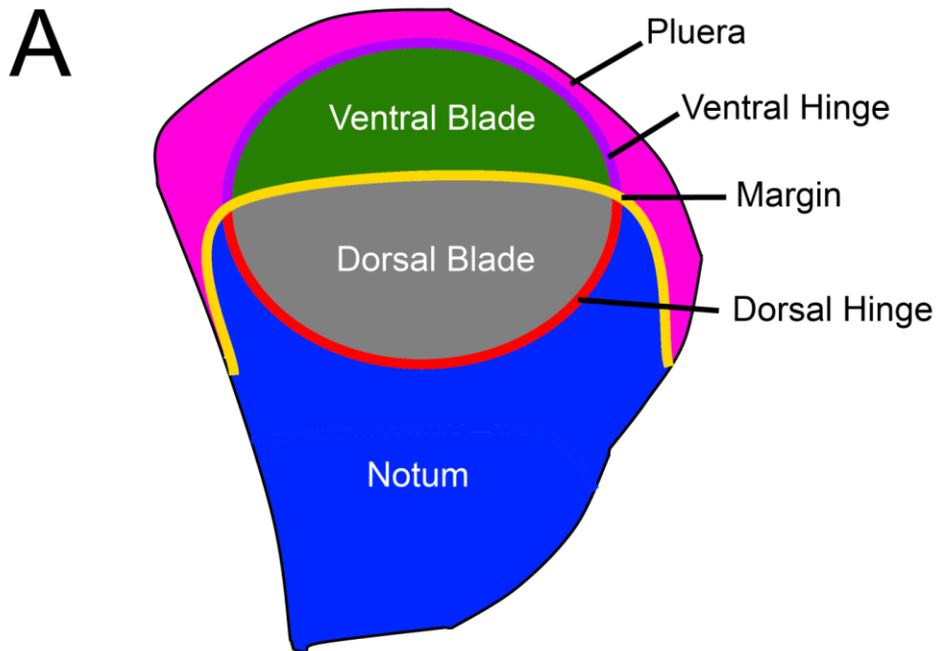
Figure 1.7. Overview of *vg* and *wg* expression in the wing imaginal disc of third instar larvae. Expression pattern of the *vg* Quadrant Enhancer (*vg*^{QE}; blue), the *vg* Boundary Enhancer (*vg*^{BE}; green), wingless (*wg;* red) and a region along the D/V axis where both the *vg*^{QE} and *wg* are transcriptionally active (yellow). See text for details.
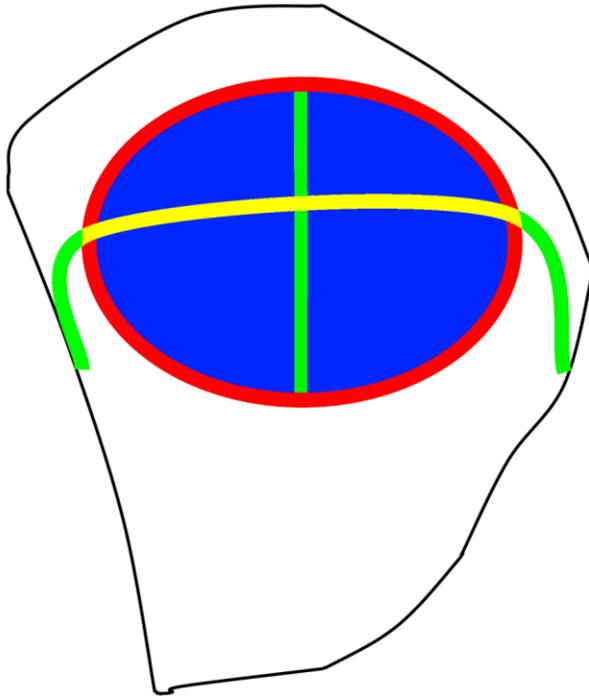
Table 1.1. Overview of regions of *sd* expression in *Drosophila* embryos and larva. Regions where *sd* transcript have been detected are indicated, along with the corresponding adult structure that is derived from the noted tissues (where applicable). MF is the morphogenetic furrow, and PNS is the peripheral nervous system. The data are a summary based on the results Campbell et al. 1992 and Deng et al. 2009.

Larval structure	Corresponding adult structure
Clypeolabrum discs	Labrum
Eye-antennal discs (behind MF)	Eyes and antennea
Wing discs	Wings
Haltere discs	Halteres
Leg discs (all thee pairs)	Legs
Genital disc	Genitilia
Optic lobe/ventral nerve cord	Central nervous system

Embryonic Structure	Corresponding adult structure
Somatic muscle precursors	N/A (larval muscle precursor)
Cardiac muscle precursors	Dorsal vessal
Anterior sense organs	N/A (larval anterior sense organ precursor)
PNS	N/A (larval PNS)
Supraesophageal ganglion	Brain

References

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., et al. (2000). The genome sequence of *Drosophila melanogaster*. Science *287*, 2185-2195.

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

Anbanandam, A., Albarado, D. C., Nguyen, C. T., Halder, G., Gao, X., and Veeraraghavan, S. (2006). Insights into transcription enhancer factor 1 (TEF-1) activity from the solution structure of the TEA domain. Proc. Natl. Acad. Sci. U.S.A *103*, 17225-17230.

Anderson, K. V., and Lengyel, J. A. (1979). Rates of synthesis of major classes of RNA in *Drosophila* embryos. Dev. Biol *70*, 217-231.

Andrianopoulos, A., and Timberlake, W. E. (1991). ATTS, a new and conserved DNA binding domain. The Plant Cell *3*, 747-748.

Auerbach, C. (1936). The development of the legs, wings and halteres in wild type and some mutant strains of *Drosophila*. Transcripts of the Royal Society of Edinburgh *58*, 787-816.

Azakie, A., Larkin, S. B., Farrance, I. K., Grenningloh, G., and Ordahl, C. P. (1996). DTEF-1, a novel member of the transcription enhancer factor-1 (TEF-1) multigene family. J. Biol. Chem *271*, 8260-8265.

Azpiazu, N., and Morata, G. (2000). Function and regulation of homothorax in the wing imaginal disc of *Drosophila*. Development (Cambridge, England) *127*, 2685-2693.

Baena-Lopez, L. A., and Garcia-Bellido, A. (2003). Genetic requirements of vestigial in the regulation of *Drosophila* wing development. Development (Cambridge, England) *130*, 197-208.

Barrio, R., and de Celis, J. F. (2004). Regulation of spalt expression in the *Drosophila* wing blade in response to the Decapentaplegic signaling pathway. Proceedings of the National Academy of Sciences of the United States of America *101*, 6021-6026.

Bernard, F., Lalouette, A., Gullaud, M., Jeantet, A. Y., Cossard, R., Zider, A., Ferveur, J. F., and Silber, J. (2003). Control of apterous by vestigial drives indirect flight muscle development in *Drosophila*. Developmental Biology *260*, 391-403.

Blair, S. S., Brower, D. L., Thomas, J. B., and Zavortink, M. (1994). The role of apterous in the control of dorsoventral compartmentalization and PS integrin gene expression in the developing wing of *Drosophila*. Development (Cambridge, England) *120*, 1805-1815.

Blair, S. S., and Ralston, A. (1997). Smoothened-mediated Hedgehog signalling is required for the maintenance of the anterior-posterior lineage restriction in the developing wing of *Drosophila*. Development (Cambridge, England) *124*, 4053-4063.

Bonnet, A., Dai, F., Brand-Saberi, B., and Duprez, D. (2010). Vestigial-like 2 acts downstream of MyoD activation and is associated with skeletal muscle differentiation in chick myogenesis. Mech. Dev *127*, 120-136.

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

Bryant, P. J. (1975). Pattern formation in the imaginal wing disc of *Drosophila melanogaster*: fate map, regeneration and duplication. J. Exp. Zool *193*, 49-77.

Burglin, T. R. (1991). The TEA domain: a novel, highly conserved DNA-binding motif. Cell *66*, 11-12.

Campbell, G., Weaver, T., and Tomlinson, A. (1993). Axis specification in the developing *Drosophila* appendage: the role of wingless, decapentaplegic, and the homeobox gene aristaless. Cell *74*, 1113-1123.

Campbell, S., Inamdar, M., Rodrigues, V., Raghavan, V., Palazzolo, M., and Chovnick, A. (1992). The scalloped gene encodes a novel, evolutionarily conserved transcription factor required for sensory organ differentiation in *Drosophila*. Genes & development *6*, 367-379.

Campbell, S. D., Duttaroy, A., Katzen, A. L., and Chovnick, A. (1991). Cloning and characterization of the scalloped region of *Drosophila melanogaster*. Genetics *127*, 367-380.

Carroll, S. B., Weatherbee, S. D., and Langeland, J. A. (1995). Homeotic genes and the regulation and evolution of insect wing number. Nature *375*, 58-61.

Casares, F., and Mann, R. S. (2000). A dual role for *homothorax* in inhibiting wing blade development and specifying proximal wing identities in *Drosophila*. Development (Cambridge, England) *127*, 1499-1508.

Chen, H., Maeda, T., Mullett, S. J., and Stewart, A. F. R. (2004a). Transcription cofactor Vgl-2 is required for skeletal muscle differentiation. Genesis *39*, 273-279.

Chen, H., Mullett, S. J., and Stewart, A. F. R. (2004b). Vgl-4, a novel member of the vestigial-like family of transcription cofactors, regulates alpha1-adrenergic activation of gene expression in cardiac myocytes. J. Biol. Chem *279*, 30800-30806.

Chen, L., Chan, S. W., Zhang, X., Walsh, M., Lim, C. J., Hong, W., and Song, H. (2010). Structural basis of YAP recognition by TEAD4 in the Hippo pathway. Genes & Development *24*, 290 -300.

Chen, Z., Friedrich, G. A., and Soriano, P. (1994). Transcriptional enhancer factor 1 disruption by a retroviral gene trap leads to heart defects and embryonic lethality in mice. Genes Dev *8*, 2293-2301.

Chintapalli, V. R., Wang, J., and Dow, J. A. T. (2007). Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. Nat Genet *39*, 715-720.

Chow, L., Berube, J., Fromont, A., and Bell, J. B. (2004). Ability of *scalloped* deletion constructs to rescue sd mutant wing phenotypes in *Drosophila melanogaster*. Genome / National Research Council Canada = Genome / Conseil national de recherches Canada *47*, 849-859.

Cohen, B., Simcox, A. A., and Cohen, S. M. (1993). Allocation of the thoracic imaginal primordia in the *Drosophila* embryo. Development *117*, 597-608.

Couso, J. P., Bate, M., and Martínez-Arias, A. (1993). A wingless-dependent polar coordinate system in *Drosophila* imaginal discs. Science *259*, 484-489.

Couso, J. P., Bishop, S. A., and Martinez Arias, A. (1994). The wingless signalling pathway and the patterning of the wing margin in *Drosophila*. Development *120*, 621-636.

Couso, J. P., Knust, E., and Martinez Arias, A. (1995). Serrate and wingless cooperate to induce vestigial gene expression and wing formation in *Drosophila*. Current Biology *5*, 1437-1448.

de la Cova, C., Abril, M., Bellosta, P., Gallant, P., and Johnston, L. A. (2004). *Drosophila* myc regulates organ size by inducing cell competition. Cell *117*, 107-116.

Davidson, I., Xiao, J. H., Rosales, R., Staub, A., and Chambon, P. (1988). The HeLa cell protein TEF-1 binds specifically and cooperatively to two SV40 enhancer motifs of unrelated sequence. Cell *54*, 931-942.

Delanoue, R., Legent, K., Godefroy, N., Flagiello, D., Dutriaux, A., Vaudin, P., Becker, J. L., and Silber, J. (2004). The *Drosophila* wing differentiation factor vestigial-scalloped is required for cell proliferation and cell survival at the dorso-ventral boundary of the wing imaginal disc. Cell death and differentiation *11*, 110-122.

Delanoue, R., Zider, A., Cossard, R., Dutriaux, A., and Silber, J. (2002). Interaction between *apterous* and early expression of v*estigial* in formation of the dorso-ventral compartments in the *Drosophila* wing disc. Genes to cells : devoted to molecular & cellular mechanisms *7*, 1255-1266.

Deng, H., Hughes, S. C., Bell, J. B., and Simmonds, A. J. (2009). Alternative requirements for Vestigial, Scalloped, and Dmef2 during muscle differentiation in *Drosophila melanogaster*. Mol. Biol. Cell *20*, 256-269.

Deshpande, N., Chopra, A., Rangarajan, A., Shashidhara, L. S., Rodrigues, V., and Krishna, S. (1997). The human transcription enhancer factor-1, TEF-1, can substitute for *Drosophila scalloped* during wingblade development. Journal of Biological Chemistry *272*, 10664-10668.

Diaz-Benjumea, F. J., and Cohen, S. M. (1993). Interaction between dorsal and ventral cells in the imaginal disc directs wing development in *Drosophila*. Cell *75*, 741-752.

DiNardo, S., Heemskerk, J., Dougan, S., and O'Farrell, P. H. (1994). The making of a maggot: patterning the *Drosophila* embryonic epidermis. Curr Opin Genet Dev *4*, 529-534.

Driever, W., and Nüsslein-Volhard, C. (1988a). A gradient of bicoid protein in *Drosophila* embryos. Cell *54*, 83-93.

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

Duman-Scheel, M., Johnston, L. A., and Du, W. (2004). Repression of dMyc expression by Wingless promotes Rbf-induced G1 arrest in the presumptive *Drosophila* wing margin. Proc. Natl. Acad. Sci. U.S.A *101*, 3857-3862.

Farrance, I. K., Mar, J. H., and Ordahl, C. P. (1992). M-CAT binding factor is related to the SV40 enhancer binding factor, TEF-1. J. Biol. Chem *267*, 17234-17240.

Foe, V. E., and Alberts, B. M. (1983). Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in *Drosophila* embryogenesis. J. Cell. Sci *61*, 31-70.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S., and Noll, M. (1986). Structure of the segmentation gene paired and the *Drosophila* PRD gene set as part of a gene network. Cell *47*, 735-746.

Fuse, N., Hirose, S., and Hayashi, S. (1996a). Determination of wing cell fate by the escargot and snail genes in *Drosophila*. Development *122*, 1059-1067.

Fuse, N., Hirose, S., and Hayashi, S. (1996b). Determination of wing cell fate by the escargot and snail genes in *Drosophila*. Development (Cambridge, England) *122*, 1059-1067.

Garcia-Bellido, A., and Merriam, J. R. (1969). Cell lineage of the imaginal discs in *Drosophila* gynandromorphs. J. Exp. Zool *170*, 61-75.

Garg, A., Srivastava, A., Davis, M. M., O'Keefe, S. L., Chow, L., and Bell, J. B. (2007). Antagonizing Scalloped With a Novel Vestigial Construct Reveals an Important Role for Scalloped in *Drosophila melanogaster* Leg, Eye and Optic Lobe Development. Genetics *175*, 659-669.

Gehring, W. (1967). [Formation of a complete mid leg with sternopleura in the antennae region by the mutant Nasobemia (Ns) of *Drosophila melanogaster*]. Arch Julius Klaus Stift Vererbungsforsch Sozialanthropol Rassenhyg *41*, 44-54.

Gibson, M. C., and Perrimon, N. (2005). Extrusion and death of DPP/BMP-compromised epithelial cells in the developing *Drosophila* wing. Science *307*, 1785-1789.

Giraldez, A. J., and Cohen, S. M. (2003). Wingless and Notch signaling provide cell survival cues and control cell proliferation during wing development. Development *130*, 6533-6543.

Goulev, Y., Fauny, J. D., Gonzalez-Marti, B., Flagiello, D., Silber, J., and Zider, A. (2008). SCALLOPED Interacts with YORKIE, the Nuclear Effector of the Hippo Tumor-Suppressor Pathway in *Drosophila*. Current Biology *18*, 435-441.

Griffiths, A. J. F. (2002). Modern genetic analysis: integrating genes and genomes (W.H. Freeman and Co.).

Gruneberg, H. (1929). Ein Beitrag zur Kenntnis der Rontgenmutationen des X-Chromosoms von *Drosophila melanogaster*. Biologisches Zentralblatt *49*, 680-694.

Guss, K. A., Nelson, C. E., Hudson, A., Kraus, M. E., and Carroll, S. B. (2001). Control of a genetic regulatory network by a selector gene. Science *292*, 1164-1167.

Halder, G., and Carroll, S. B. (2001). Binding of the Vestigial co-factor switches the DNA-target selectivity of the Scalloped selector protein. Development *128*, 3295-3305.

Halder, G., Polaczyk, P., Kraus, M. E., Hudson, A., Kim, J., Laughon, A., and Carroll, S. (1998). The Vestigial and Scalloped proteins act together to directly regulate wing-specific gene expression in *Drosophila*. Genes Dev *12*, 3900-3909.

Halder, G., and Carroll, S. B. (2001). Binding of the Vestigial co-factor switches the DNA-target selectivity of the Scalloped selector protein. Development (Cambridge, England) *128*, 3295-3305.

Halder, G., Polaczyk, P., Kraus, M. E., Hudson, A., Kim, J., Laughon, A., and Carroll, S. (1998). The Vestigial and Scalloped proteins act together to directly regulate wing-specific gene expression in *Drosophila*. Genes & development *12*, 3900-3909.

Huang, J., Wu, S., Barrera, J., Matthews, K., and Pan, D. (2005). The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the *Drosophila* Homolog of YAP. Cell *122*, 421-434.

Hwang, J. J., Chambon, P., and Davidson, I. (1993). Characterization of the transcription activation function and the DNA binding domain of transcriptional enhancer factor-1. The EMBO journal *12*, 2337-2348.

Ip, Y. T., Kraut, R., Levine, M., and Rushlow, C. A. (1991). The dorsal morphogen is a sequence-specific DNA-binding protein that interacts with a long-range repression element in *Drosophila*. Cell *64*, 439-446.

Jack, J., and DeLotto, Y. (1992). Effect of wing scalloping mutations on cut expression and sense organ differentiation in the *Drosophila* wing margin. Genetics *131*, 353-363.

Johnston, L. A., Prober, D. A., Edgar, B. A., Eisenman, R. N., and Gallant, P. (1999). *Drosophila* myc regulates cellular growth during development. Cell *98*, 779-790.

Johnston, L. A. (2009). Competitive interactions between cells: death, growth, and geography. Science *324*, 1679-1682.

Kim, J., Irvine, K. D., and Carroll, S. B. (1995). Cell recognition, signal induction, and symmetrical gene activation at the dorsal-ventral boundary of the developing *Drosophila* wing. Cell *82*, 795-802.

Kim, J., Sebring, A., Esch, J. J., Kraus, M. E., Vorwerk, K., Magee, J., and Carroll, S. B. (1996). Integration of positional signals and regulation of wing formation and identity by *Drosophila vestigial* gene. Nature *382*, 133-138.

Klein, T., and Arias, A. M. (1998a). Different spatial and temporal interactions between *Notch*, *wingless*, and *vestigial* specify proximal and distal pattern elements of the wing in *Drosophila*. Dev. Biol *194*, 196-212.

Klein, T., and Arias, A. M. (1998b). Different spatial and temporal interactions between *Notch*, *wingless*, and *vestigial* specify proximal and distal pattern elements of the wing in *Drosophila*. Developmental biology *194*, 196-212.

Klein, T., and Arias, A. M. (1998c). Interactions among Delta, Serrate and Fringe modulate Notch activity during *Drosophila* wing development. Development (Cambridge, England) *125*, 2951-2962.

Klein, T., and Arias, A. M. (1999). The vestigial gene product provides a molecular context for the interpretation of signals during the development of the wing in *Drosophila*. Development *126*, 913-925.

Kolzer, S., Fuss, B., Hoch, M., and Klein, T. (2003). Defective proventriculus is required for pattern formation along the proximodistal axis, cell proliferation and formation of veins in the *Drosophila* wing. Development (Cambridge, England) *130*, 4135-4147.

Kurata, S., Go, M. J., Artavanis-Tsakonas, S., and Gehring, W. J. (2000). Notch signaling and the determination of appendage identity. Proc. Natl. Acad. Sci. U.S.A *97*, 2117-2122.

Lewis, R. A., Wakimoto, B. T., Denell, R. E., and Kaufman, T. C. (1980). Genetic Analysis of the *Antennapedia Gene Complex* (*Ant-C*) and Adjacent Chromosomal Regions of *DROSOPHILA MELANOGASTER*. II. Polytene Chromosome Segments 84A-84B1,2. Genetics *95*, 383-397.

Li, Z., Zhao, B., Wang, P., Chen, F., Dong, Z., Yang, H., Guan, K., and Xu, Y. (2010). Structural insights into the YAP and TEAD complex. Genes & Development *24*, 235 -240.

Liu, X., Grammont, M., and Irvine, K. D. (2000). Roles for *scalloped* and *vestigial* in regulating cell affinity and interactions between the wing blade and the wing hinge. Developmental biology *228*, 287-303.

MacKay, J. O., Soanes, K. H., Srivastava, A., Simmonds, A., Brook, W. J., and Bell, J. B. (2003). An *in vivo* analysis of the *vestigial* gene in *Drosophila melanogaster* defines the domains required for Vg function. Genetics *163*, 1365-1373.

Maeda, T., Chapman, D. L., and Stewart, A. F. (2002a). Mammalian *vestigial-like 2*, a cofactor of TEF-1 and MEF2 transcription factors that promotes skeletal muscle differentiation. Journal of Biological Chemistry *277*, 48889-48898.

Maeda, T., Gupta, M. P., and Stewart, A. F. R. (2002b). TEF-1 and MEF2 transcription factors interact to regulate muscle-specific promoters. Biochemical and Biophysical Research Communications *294*, 791-797.

Mann, C. J., Osborn, D. P. S., and Hughes, S. M. (2007). Vestigial-like-2b (VITO-1b) and Tead-3a (Tef-5a) expression in zebrafish skeletal muscle, brain and notochord. Gene Expr. Patterns *7*, 827-836.

Martín-Castellanos, C., and Edgar, B. A. (2002). A characterization of the effects of Dpp signaling on cell growth and proliferation in the *Drosophila* wing. Development *129*, 1003-1013.

Maves, L., and Schubiger, G. (1998). A molecular basis for transdetermination in *Drosophila* imaginal discs: interactions between *wingless* and *decapentaplegic* signaling. Development (Cambridge, England) *125*, 115-124.

Milan, M., and Cohen, S. M. (2003). A re-evaluation of the contributions of Apterous and Notch to the dorsoventral lineage restriction boundary in the *Drosophila* wing. Development (Cambridge, England) *130*, 553-562.

Milan, M., Perez, L., and Cohen, S. M. (2002). Short-range cell interactions and cell survival in the *Drosophila* wing. Dev.Cell. *2*, 797-805.

Milewski, R. C., Chi, N. C., Li, J., Brown, C., Lu, M. M., and Epstein, J. A. (2004). Identification of minimal enhancer elements sufficient for Pax3 expression in neural crest and implication of Tead2 as a regulator of Pax3. Development *131*, 829-837.

Nagaraj, R., Pickup, A. T., Howes, R., Moses, K., Freeman, M., and Banerjee, U. (1999). Role of the EGF receptor pathway in growth and patterning of the *Drosophila* wing through the regulation of vestigial. Development (Cambridge, England) *126*, 975-985.

Neto-Silva, R. M., Wells, B. S., and Johnston, L. A. (2009). Mechanisms of growth and homeostasis in the *Drosophila* wing. Annu. Rev. Cell Dev. Biol *25*, 197-220.

Neumann, C., and Cohen, S. (1997). Morphogens and pattern formation. BioEssays : news and reviews in molecular, cellular and developmental biology *19*, 721-729.

Neuman-Silberberg, F. S., and Schüpbach, T. (1993). The *Drosophila* dorsoventral patterning gene gurken produces a dorsally localized RNA and encodes a TGF alpha-like protein. Cell *75*, 165-174.

Nüsslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. Nature *287*, 795-801.

Olson, E. N., Perry, M., and Schulz, R. A. (1995). Regulation of Muscle Differentiation by the MEF2 Family of MADS Box Transcription Factors. Developmental Biology *172*, 2-14.

Ota, M., and Sasaki, H. (2008). Mammalian Tead proteins regulate cell proliferation and contact inhibition as transcriptional mediators of Hippo signaling. Development *135*, 4059-4069.

Panin, V. M., Papayannopoulos, V., Wilson, R., and Irvine, K. D. (1997). Fringe modulates Notch-ligand interactions. Nature *387*, 908-912.

Paumard-Rigal, S., Zider, A., Vaudin, P., and Silber, J. (1998). Specific interactions between vestigial and scalloped are required to promote wing tissue proliferation in *Drosophila melanogaster*. Development genes and evolution *208*, 440-446.

Pick, L. (1998). Segmentation: painting stripes from flies to vertebrates. Dev. Genet *23*, 1-10.

Prober, D. A., and Edgar, B. A. (2000). Ras1 promotes cellular growth in the *Drosophila* wing. Cell *100*, 435-446.

Ray, A., van der Goes van Naters, W., and Carlson, J. R. (2008). A regulatory code for neuron-specific odor receptor expression. PLoS Biol *6*, e125.

Rodriguez, D., Terriente, J., Galindo, M. I., Couso, J. P., and Diaz-Benjumea, F. J. (2002). Different mechanisms initiate and maintain *wingless* expression in the *Drosophila* wing hinge. Development (Cambridge, England) *129*, 3995-4004.

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

Roth, S. (2003). The origin of dorsoventral polarity in *Drosophila*. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences *358*, 1317 -1329.

Sadava, D., Heller, H. C., Hillis, D. M., and Berenbaum, M. (2009). Life: The Science of Biology (W. H. Freeman).

Sanicola, M., Sekelsky, J., Elson, S., and Gelbart, W. M. (1995). Drawing a stripe in *Drosophila* imaginal disks: negative regulation of decapentaplegic and patched expression by engrailed. Genetics *139*, 745-756.

Sanson, B. (2001). Generating patterns from fields of cells. Examples from *Drosophila* segmentation. EMBO Rep *2*, 1083-1088.

Sharma, R. P., and Chopra, V. L. (1976). Effect of the *Wingless* (*wg1*) mutation on wing and haltere development in *Drosophila melanogaster*. Dev. Biol *48*, 461-465.

Shen, J., and Dahmann, C. (2005). Extrusion of cells with inappropriate Dpp signaling from *Drosophila* wing disc epithelia. Science *307*, 1789-1790.

Shimizu, N., Smith, G., and Izumo, S. (1993). Both a ubiquitous factor mTEF-1 and a distinct muscle-specific factor bind to the M-CAT motif of the myosin heavy chain beta gene. Nucleic Acids Res *21*, 4103-4110.

Simmonds, A. J., Brook, W. J., Cohen, S. M., and Bell, J. B. (1995). Distinguishable functions for engrailed and invected in anterior-posterior patterning in the *Drosophila* wing. Nature *376*, 424-427.

Simmonds, A. J., and Bell, J. B. (1998). A genetic and molecular analysis of an invectedDominant mutation in *Drosophila melanogaster*. Genome / National Research Council Canada = Genome / Conseil national de recherches Canada *41*, 381-390.

Simmonds, A. J., Liu, X., Soanes, K. H., Krause, H. M., Irvine, K. D., and Bell, J. B. (1998). Molecular interactions between Vestigial and Scalloped promote wing formation in *Drosophila*. Genes & development *12*, 3815-3820.

Spirov, A., Fahmy, K., Schneider, M., Frei, E., Noll, M., and Baumgartner, S. (2009). Formation of the bicoid morphogen gradient: an mRNA gradient dictates the protein gradient. Development *136*, 605-614.

Srivastava, A., and Bell, J. B. (2003). Further developmental roles of the Vestigial/Scalloped transcription complex during wing development in *Drosophila melanogaster*. Mechanisms of development *120*, 587-596.

Srivastava, A., MacKay, J. O., and Bell, J. B. (2002). A Vestigial:Scalloped TEA domain chimera rescues the wing phenotype of a *scalloped* mutation in *Drosophila melanogaster*. Genesis (New York, N.Y. : 2000) *33*, 40-47.

Srivastava, A., and Bell, J. B. (2003). Further developmental roles of the Vestigial/Scalloped transcription complex during wing development in *Drosophila melanogaster*. Mech. Dev *120*, 587-596.

St Johnston, D., and Nüsslein-Volhard, C. (1992). The origin of pattern and polarity in the *Drosophila* embryo. Cell *68*, 201-219.

Steward, R. (1989). Relocalization of the dorsal protein from the cytoplasm to the nucleus correlates with its function. Cell *59*, 1179-1188.

Stewart, A. F., Larkin, S. B., Farrance, I. K., Mar, J. H., Hall, D. E., and Ordahl, C. P. (1994). Muscle-enriched TEF-1 isoforms bind M-CAT elements from muscle-specific promoters and differentially activate transcription. J. Biol. Chem *269*, 3147-3150.

Stewart, A. F., Suzow, J., Kubota, T., Ueyama, T., and Chen, H. H. (1998). Transcription factor RTEF-1 mediates alpha1-adrenergic reactivation of the fetal gene program in cardiac myocytes. Circ. Res *83*, 43-49.

Struhl, G. (1981). A homoeotic mutation transforming leg to antenna in *Drosophila*. Nature *292*, 635-638.

Sudarsan, V., Anant, S., Guptan, P., Vijayraghavan, K., and Skaer, H. (2001). Myoblast Diversification and Ectodermal Signaling in *Drosophila*. Developmental Cell *1*, 829-839.

Tweedie, S. et al., 2009. FlyBase: enhancing *Drosophila* Gene Ontology annotations. *Nucleic Acids Research*, 37(Database), p.D555-D559.

Van De Bor, V., Hartswood, E., Jones, C., Finnegan, D., and Davis, I. (2005). gurken and the I Factor Retrotransposon RNAs Share Common Localization Signals and Machinery. Developmental Cell *9*, 51-62.

Vaudin, P., Delanoue, R., Davidson, I., Silber, J., and Zider, A. (1999). TONDU (TDU), a novel human protein related to the product of *vestigial* (*vg*) gene of *Drosophila melanogaster* interacts with vertebrate TEF factors and substitutes for Vg function in wing formation. Development *126*, 4807-4816.

Wang, S. H., Simcox, A., and Campbell, G. (2000). Dual role for *Drosophila* epidermal growth factor receptor signaling in early wing disc development. Genes & development *14*, 2271-2276.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-1191.

Weatherbee, S. D., Halder, G., Kim, J., Hudson, A., and Carroll, S. (1998). Ultrabithorax regulates genes at several levels of the wing-patterning hierarchy to shape the development of the *Drosophila* haltere. Genes Dev *12*, 1474-1482.

Whitworth, A. J., and Russell, S. (2003). Temporally dynamic response to Wingless directs the sequential elaboration of the proximodistal axis of the *Drosophila* wing. Developmental biology *254*, 277-288.

Wieschaus, E., and Gehring, W. (1976). Clonal analysis of primordial disc cells in the early embryo of *Drosophila melanogaster*. Dev. Biol *50*, 249-263.

Williams, J. A., Bell, J. B., and Carroll, S. B. (1991). Control of *Drosophila* wing and haltere development by the nuclear *vestigial* gene product. Genes Dev *5*, 2481-2495.

Williams, J. A., Paddock, S. W., and Carroll, S. B. (1993). Pattern formation in a secondary field: A hierarchy of regulatory genes subdivides the developing *Drosophila* wing disc into discrete subregions. Development *117*, 571-584.

Williams, J. A., Paddock, S. W., Vorwerk, K., and Carroll, S. B. (1994). Organization of wing formation and induction of a wing-patterning gene at the dorsal/ventral compartment boundary. Nature *368*, 299-305.

Wu, J., and Cohen, S. M. (2002). Repression of Teashirt marks the initiation of wing development. Development (Cambridge, England) *129*, 2411-2418.

Wu, S., Liu, Y., Zheng, Y., Dong, J., and Pan, D. (2008). The TEAD/TEF family protein Scalloped mediates transcriptional output of the Hippo growth-regulatory pathway. Dev. Cell *14*, 388-398.

Xiao, J. H., Davidson, I., Matthes, H., Garnier, J. M., and Chambon, P. (1991). Cloning, expression, and transcriptional properties of the human enhancer factor TEF-1. Cell *65*, 551-568.

Yasunami, M., Suzuki, K., Houtani, T., Sugimoto, T., and Ohkubo, H. (1995). Molecular characterization of cDNA encoding a novel protein related to transcriptional enhancer factor-1 from neural precursor cells. J. Biol. Chem *270*, 18649-18654.

Yoshida, T. (2008). MCAT Elements and the TEF-1 Family of Transcription Factors in Muscle Development and Disease. Arterioscler Thromb Vasc Biol *28*, 8-17.

Zecca, M., Basler, K., and Struhl, G. (1995). Sequential organizing activities of engrailed, hedgehog and decapentaplegic in the *Drosophila* wing. Development *121*, 2265-2278.

Zhang, L., Ren, F., Zhang, Q., Chen, Y., Wang, B., and Jiang, J. (2008). The TEAD/TEF Family of Transcription Factor Scalloped Mediates Hippo Signaling in Organ Size Control. Developmental Cell *14*, 377-387.

Zhao, B., Ye, X., Yu, J., Li, L., Li, W., Li, S., Yu, J., Lin, J. D., Wang, C., Chinnaiyan, A. M., et al. (2008). TEAD mediates YAP-dependent gene induction and growth control. Genes Dev *22*, 1962-1971.

Zirin, J. D., and Mann, R. S. (2004). Differing strategies for the establishment and maintenance of teashirt and homothorax repression in the *Drosophila* wing. Development *131*, 5683-5693.

Chapter Two: Identification of a classical nuclear localization signal in Scalloped[1]

Introduction to nuclear transport

One of the hallmarks of eukaryotes is the compartmentalization of their nuclear material within the nuclear envelope (which consists of an inner and outer membrane (D'Angelo and Hetzer, 2006). However, this compartmentalization creates a fundamental biological problem – namely, how is it possible for the genetic information present in the nucleus to be utilized by the molecular machinery of the cytoplasm, and conversely, how can the state of the cell be communicated to the nucleus in order to facilitate an appropriate transcriptional response?   In a general sense, these problems can be solved by allowing for the export of mRNAs from the nucleus and the import of proteins, such as transcription factors, into the nucleus. Indeed, many compounds are transported into and out of the nucleus, including various RNA species, proteins (both soluble and those targeted to the membrane), ions and small molecules like sugars, amino acids and nucleotides (Franke and Scheer, 1974; Sorokin et al., 2007). This transport occurs (often bilaterally) through large and complex protein channels present in the nuclear envelope, known as nuclear pore complexes (NPCs).

The NPC is a large complex ranging from 44 MDa in yeast to 60 MDa in vertebrates (Rout et al., 2000; Cressman et al., 2001). Although there is some variability in the structure of the NPC from organism to organism, generally it is composed of three segments: cytoplasmic fibrils, a central core (running between the two membranes of the nuclear envelope) and a nuclear basket consisting of nuclear filaments (Rout and Wente, 1994; Suntharalingam and Wente, 2003). These components each have an eight-fold radial symmetry, and given the large size of the NPC, a relatively scant variety of proteins called nucleoporins (Nups) compose the NPC (Rout et al., 2000; Cronshaw et al., 2002). In fact,

---

[1] A version of this chapter has been accepted for publication in PLoS ONE. A. C. Magico and J. B. Bell. Identification of a Classical Bipartite Nuclear Localization Signal in the *Drosophila* TEA/ATTS Protein Scalloped. All experiments were designed, conducted by and analyzed by A. C. Magico

only about 30 different Nups are found in the NPC, although each is typically present in multiples of

eight (due the symmetrical arrangement of the NPC), and in total 500-1000 Nups of these 30 species

make up the complex. The pore itself is plastic and roughly 45 nm in length, ranging in diameter from

about nine nm to 40 nm during passive and active transport (see below), respectively (Paine et al., 1975;

Keminer and Peters, 1999; Panté and Kann, 2002).

As noted above, there is species to species variation in structure of NPCs. Thus, it is not

surprising that there is variation in the Nups that compose the NPC as well. Indeed, many Nups

identified in yeast show very weak identity with those identified in vertebrates, and vice versa (Vasu and

Forbes, 2001; Suntharalingam and Wente, 2003). That said, based on position within the NPC as well as

neighbor interactions, the majority of Nups appear to have orthologs in both vertebrates and yeast

(reviewed in (Sorokin et al., 2007). Nups can be divided into three classes, based on their sequence and

function. The most common class of Nup (by percent of NPC mass) are the FG-Nups, which are named

for amino acid sequence repeats present within the protein (FG, FXFX, GLFG or FN), and are generally

distributed evenly at the nuclear and cytoplasmic faces of the NPC, although the asymmetries that do

exist are thought to be important for transport (see below). The other two classes of are the

transmembrane Nups, which fix the NPC to the nuclear membranes, and those that contain α-solenoids

(α-helices) and β-propellers (β-sheets), which are thought to anchor the Nups of the inner and outer

membranes together (Rout et al., 2000; Cronshaw et al., 2002; reviewed in Devos et al., 2006; Sorokin et

al., 2007).

Transport via the NPC is mediated by one of the two mechanisms alluded to previously. The

first is via passive diffusion through the NPC, which allows the passage of small metabolites, ions and

proteins smaller than approximately 40-50 kDa, although the efficiency of transport decreases as the

protein size approaches these upper-limits (Macara, 2001; Tran and Wente, 2006). Passive diffusion

does not require energy, and allows these small molecules which are not much larger than 6 nm in diameter to transverse the NPC without requiring interactions with the Nups (Görlich and Kutay, 1999). However, even relatively small compounds such as tRNAs and histones are transported actively, rather than passively, suggesting that passage through the NPC is generally a regulated process (Zasloff, 1983; Jäkel et al., 1999). The second mechanism which allows translocation through the NPC is active transport and is dependent on interactions with a subset of the Nups.

While active and passive transport through the NPC facilitates the movement of various proteins across the nuclear envelope, there are also examples of proteins that have domains capable of retaining them in a given location, thereby retarding transport via the NPC. These are known as retention signals and come in two flavours – nuclear and cytoplasmic – which can act to retain a protein in the given compartment under certain conditions. For instance, Glucocorticoid receptor (GR) is a steroid hormone receptor transcription factor which regulates genes implicated in a variety of processes, including glucose homeostasis, lipid metabolism and cancer (reviewed in Charmandari et al., 2004; Chrousos et al., 2004). GR is known to transport actively into the nucleus, yet in the absence of steroid hormone, GR is retained in the cytoplasm; likely due to interactions with a Heat shock protein 90 (Hsp90) containing complex which retains the protein in the cytoplasm. Once a ligand binds, GR is released from the complex and allowed to enter the nucleus actively. On the other hand, nuclear GR is retained there in the absence of a ligand, even though the naïve (ligand unbound) receptor is normally targeted for nuclear export. This nuclear retention is due to a specific nuclear retention signal, which may facilitate specific interactions with 14-3-3σ protein (Chintapalli et al., 2007). Cytoplasmic and nuclear retention signals provide an additional layer of regulation that acts on some proteins which undergo active transport. The following three sections will focus on active transport.

Active nuclear transport

Although there are alternate pathways (discussed below), the majority of translocation events are thought to be mediated by proteins known as Karyopherins (Görlich and Kutay, 1999). These proteins can be specific for import (Importins), export (Exportins) or both (Transportins). There are two types of Karyopherins: Karyopherins-α and Karyopherins-β. The number of Karypherins present in a given species is variable (for instance, the number of Karyopherins-α varies from three in *Drosophila* to six in humans and the number of Karyopherins-β varies from six in *Caenorhabditis elegans* to 20 in humans (Máthé et al., 2000; Török et al., 1995; Küssel and Frasch, 1995); reviewed in Weis, 2003 and Tejomurtula et al., 2009, but there are certainly far fewer than there are proteins which undergo active transport. Thus, a given Karyopherin must be able to recognize and transport many different proteins. This occurs via recognition sequences known as nuclear localization signals (NLSs) and nuclear export signals (NESs) which are found within cargo proteins. Karyopherins-β can either interact directly to cargo allowing for transport, or they can interact with Karyopherins-α which in turn interact with cargo proteins directly (i.e. they act as adapters of Karyopherins-β mediated transport) into or out of the nucleus.

Nuclear import

The stereotypical example of nuclear import involving Karyopherins-α/β (Importins-α/β) and NLSs is the Ran-dependent model, an overview of which is shown as Figure 2.2. In the α/β model, an Importin-α binds to an NLS enriched for basic residues (which is called a classical NLS, or cNLS) in a cargo protein, and then Importin-α is in turn bound by Importin-β1 (Sorokin et al., 2007). This whole complex then interacts – via Importin-β1 – with the FG repeats of FG-Nups present in the cytoplasmic fibrils of the NPC (Wu et al., 1995; Yokoyama et al., 1995; Adam and Adam, 1994; Görlich et al., 1994; 1995). There is

evidence that the interaction between Importin-β and the FG-Nups is dependent on binding the small-GTPase **RA**s-related **N**uclear protein-GTP (Ran-GTP), which is present at low levels in the cytoplasm (Shah et al., 1998). However, this is paradoxical since the presence of Ran-GTP is known to interfere with the stability of the Importin-α/β complex (Floer and Blobel, 1996; Lounsbury and Macara, 1997a). One explanation for these conflicting data that has been proposed is that Ran-GTP is quickly hydrolyzed to Ran-GDP upon binding the cNLS-cargo protein/Importin-α/Importin-β1/FG-Nup complex (due to Ran protein's intrinsic but normally slow GTPase activity, which is enhanced by accessory proteins present in the complex that accelerate the GTP->GDP conversion) (Lounsbury and Macara, 1997b; Lonhienne et al., 2009). Ran-GDP, along with associated factors, is then translocated through the channel into the nucleus. The precise mechanism by which this occurs is unknown, but it appears that the complex passes from cytoplasmic Nups to nuclear Nups, which only interact with Ran-GDP (Panté and Aebi, 1996; Moore and Blobel, 1994; Nehrbass and Blobel, 1996). Once in the nucleus, GDP is quickly exchanged to GTP, allowing for dissociation from the nuclear Nups. Furthermore, the presence of Ran-GTP also causes Importin-α to release its cNLS containing cargo protein into the nucleus, thus achieving the ultimate goal of nuclear translocation (Rexach and Blobel, 1995).

It should be noted that the process described above is similar to when Importins-β mediate interactions with the cargo protein directly, through a non-classical NLS. However, cNLSs have a relatively well defined character, making it easier to identify proteins which contain them. The cNLS itself is defined based on similarity to the first member of this class of NLS discovered, the Simian virus 40 (SV40) T-antigen NLS (Lanford and Butel, 1984; Kalderon et al., 1984). These signals are found in two forms – monopartite and bipartite – which have one and two clusters of basic amino acids with the following consensus sequences: [K(K/R)X(K/R)] and [(R/K)$_2$X$_{\sim 10}$(R/K)$_{>3/5}$], respectively (Robbins, 1991; Dingwall and Laskey, 1991; Chelsky et al., 1989), although recent work has shown that the length of the spacer in a bipartite signal can be as large as 29 amino acids (Lange et al., 2010).

In *Drosophila*, there are three known members of the Importin-α (Imp-α) family: Imp-α1,2 and 3 (Török et al., 1995; Küssel and Frasch, 1995; Máthé et al., 2000; Mason et al., 2002). Based on the results of rescue experiments, the three Imp-α proteins are generally functionally redundant, although specialized roles in gametogenesis have been found for Imp-α1 and Imp-α2. However, neither of those proteins are essential for survival (Mason et al., 2002; Gorjánácz et al., 2002; Ratan et al., 2008). On the other hand, Imp-α3 is required for larval survival and development of larval and adult structures (Máthé et al., 2000).

Some proteins are known to be transported actively without requiring Importins-β to mediate the process. Rather, these are able to interact with Nups directly. One example is β-catenin which has been shown to interact directly with cytoplasmic filaments (Fagotto et al., 1998). Moreover, this study demonstrated that the interaction is likely at the same sites as Importin-β1 binds to, since the presence of Importin-β1 can inhibit the import of β-catenin. Finally, while the mechanism of translocation is unknown, transport of β-catenin can occur in both a RanGTP-dependent and independent fashion, and it has been demonstrated that interactions Smad3 and Smad4 promote nuclear translocation (Fagotto et al., 1998; Zhang et al., 2010).

Nuclear export

In a similar fashion to NLSs, nuclear export signals (NESs) are recognized by specific exportin-β proteins which shuttle proteins though the NPC and into the cytoplasm. However, in this case it is RanGTP that associates with the NES/exportin-β complex in the nucleus, which is subsequently transported to the cytoplasm, and the cargo released upon GTP hydrolyzing to GDP (Bischoff et al., 1994; Lindsay et al., 2001). The best characterized exportin-β is Chromatin Region Maintenance 1 (Crm1), and the single *Drosophila* ortholog is encoded by *embargoed* (*emb*). Crm1 recognizes

48

hydrophobic NESs that are typically L/I rich, with a classical consensus of (LX{2,3}[LIVMF]X{2,3}LX[LI])

(Bogerd et al., 1996); however, a variety of exportins and NESs exist (reviewed in (Macara, 2001;

Sorokin et al., 2007). Furthermore, there are many examples of functional Crm1 dependent NESs that

do not fit this pattern. For example, when this consensus was originally derived, an NES that was known

to deviate from this pattern had already been discovered in the equine infectious anemia virus Rev

protein (Meyer et al., 1996). Recently, Kusugi *et al* tested a large set of artificially generated NESs for

their ability to facilitate Crm1 mediated nuclear export and used these results to generate six classes of

consensus sequences (1a-d, 2 and 3; Table 2.1), which were then compared to experimentally derived

signals (Kosugi et al., 2008 and see the NES database at NESbase: www.cbs.dtu.dk/databases/NESbase/;

la Cour et al., 2003).

In the remainder of this chapter, experiments are shown which provide compelling evidence

that Sd contains a bipartite cNLS. Additionally, further evidence that is largely consistent with the

presence of an NES and some data which starts to unravel a broader role of the C-terminal domain of Sd

in regulating the nuclear translocation of the protein are also presented.

Results

Sd contains a putative NLS matching the classic bipartite sequence, which is conserved in many TEAD

family members.

Using *in silico* analysis, an NLS fitting the consensus of the bipartite family of signals (see

introduction) which could account for the theorized ability of Sd to translocate itself and its binding

partners to the nucleus was previously identified (Srivastava et al., 2002; Robbins, 1991). The sequence

of this signal is RKQVSSHIQVLARRKLR, which is a close match to the classical consensus of

[(R/K)$_2$X$_{\sim10}$(R/K)$_{>3/5}$] mentioned above (Figure 2.3A; Robbins, 1991). Moreover, the amino acids

comprising this putative NLS are highly conserved among TEAD family members from species within

both *Choanozoa* and *Animalia* (Figure 2.3B).

The NLS within Sd is sufficient to target an eGFP reporter to the nucleus.

In order to confirm the function of the putative NLS of Sd, I elected to tag the protein with an

eGFP reporter and express the fusion proteins (under the control of a heat shock driver) in *Drosophila* S2

cells. The results of the experiments listed below are summarized in Table 2.2. When eGFP is

expressed alone, diffuse signal is observed throughout the cytoplasm and nucleus of the cells, with

~61% of the total signal located in the nuclei of cells, on average (Figure 2.4A). This is likely because the

small size of eGFP (~27kDa) enables it to pass through the NPC via passive diffusion. It has been

previously shown that a chimeric protein consisting of amino acids 63-211 of Sd and full-length Vg is

able to substitute for endogenous Sd function during wing development (Srivastava et al., 2002). This,

combined with the presence of the predicted bipartite sequence within this stretch of amino acids,

implied that this region of Sd is sufficient to permit nuclear translocation of the complex. To verify this,

we expressed a reporter construct containing a fragment of Sd which contained both the TEAD and the

putative NLS signal (TEA-eGFP; amino acids 88-174). In this case over 90% of the signal is nuclear in S2

cells (Figure 2.4B). Extending this further, amino acids 143-163 (the predicted NLS extended by two

amino acids on either side) were also sufficient to strongly target eGFP (NLS-eGFP) to the nucleus (88%

nuclear; Figure 2.4C). The large increase in nuclear signal compared to eGFP alone, suggests that these

fusion peptides are being translocated much more efficiently. However, these two fusion peptides are

both smaller than 40kDa, so it is also possible that nuclear retention, rather than nuclear translocation,

has been increased. To eliminate this possibility, we also tested ability of the TEAD, the NLS and the

TEAD lacking the NLS (amino acids 88-144) to drive eGFPx2 + GST (hereafter referred to as simply

eGFPx2) to the nucleus. Unlike eGFP alone, this tag is very large (94KDa) and is almost completely

excluded from the nucleus (Figure 2.4D and see (Chan et al., 2007). As before, both the TEAD and NLS of

Sd are able to shift the localization of this tag to the nucleus (TEA-eGFPx2 and NLS-eGFPx2; Figures 2.4E

and F), giving 79% and 60% nuclear signal, respectively. Conversely, the TEAD lacking the NLS failed to

drive the protein tag (TEAΔNLS-eGFPx2) to the nucleus, as less than 20% of the observed signal was

nuclear (Figure 2.5G). As a general observation, I noted that eGFP and NLS-eGFP appeared to be able to

localize to the nucleolus, while all other constructs tested (including those described below) were largely

excluded from this region.

The NLS is necessary for the proper nuclear localization of Sd as well as efficient Importin-α3 binding.

When expressed in S2 cells, eGFP-Sd shows very strong nuclear localization (Figure 2.5A). When

the NLS was either deleted (Sd ΔNLS, Figure 2.5B) or the six basic amino acids (R145, K146, R157, R158,

K159 and R161), identified in Figure 2.3A, were mutated to asparagines (Sd mNLS^{N+C}; Figure 2.5C) the

ratio of nuclear signal to total signal is reduced to less than 50%, compared to greater than 90% for

intact Sd (Table 2.2). This provides evidence that the identified NLS is required for the proper

localization of Sd.

Extending this analysis, tagged Sd isoforms were generated where only the N-terminal basic

amino acids (R145 and K146), or the C-terminal basic amino acids (R157, R158, K159 and R161) are

mutated to asparagines (Sd mNLS^{N} and Sd mNLS^{C}, respectively). When the N-terminal amino acids are

mutated, a small but significant (p<0.001) increase in cytoplasmic signal is observed (Figure 2.5D) and

the nuclear fraction is reduced to ~80% (Table 2.2). Conversely, mutating the C-terminal basic amino

acids results in diffuse localization of the eGFP signal to both the nucleus and cytoplasm (Figure 2.5E).

The magnitude of mis-localization is similar to that seen when the entire NLS is deleted or both clusters

of basic amino acids are mutated, with less than 50% of the total signal seen in the nucleus (Table 2.2).

Surprisingly, regardless of which method of NLS disruption was employed, a significant fraction (>40%; Table 2.2) of signal was still observed in the nucleus of expressing cells.

As mentioned previously, Imp-α3 appears to be generally required throughout development and so I elected to test both the ability of this protein to bind Sd, and whether this binding was dependent on the NLS of Sd. To do this 3xFLAG-tagged Sd or Sd mNLS^{N+C} were expressed in S2 cells and tested for the ability to co-immunoprecipitate (Co-IP) endogenous Imp-α3. A mock transfection was also done using water. While Imp-α3 was detected in the lysate of all three types of transfected cells, only 3xFLAG-Sd and, to a much lesser extent, 3xFLAG-Sd mNLS^{N+C} were able to Co-IP Imp-α3 (Figure 2.5F).

Discrete regions within the C-terminal domain of Sd act to facilitate or repress nuclear localization.

There are many examples of proteins which contain multiple signals/regions which influence (in both a positive and negative fashion) the localization of the protein (for examples see Ylikomi et al., 1992; Weber et al., 1998; Zheng et al., 2005; Knapp et al., 2009). Given our results, it was hypothesized this might be true for Sd as well. To test this, a complete series of ~50 aa deletions of Sd was generated and assayed for the ability to drive eGFP to the nucleus (Figure 2.6A). Three deletions (Sd Δ1-56, Sd Δ51-102 and Sd Δ199-248) which in all cases leave the NLS intact, showed a small decrease in the ratio of nuclear to cytoplasmic signal of ~7-9%, relative to full-length Sd (Figure 2.6A, rows 2,3 and 6 compared to row 1). As the deletions are significant in size, this minor perturbation is likely due to overall changes to the tertiary structure of the deletion molecules, rather than the disruption of specific signals. A fourth construct, deleting the N-terminus portion of Sd up to the NLS was also tested (Sd Δ1-142, Figure 2.6A row 16). In this case the localization was reduced further relative to the other N-terminal deletions (70.4% nuclear vs. 85.1% and 83.2% for Sd Δ1-56 and Sd Δ51-102, respectively). However, this reduction of ~24% relative to wildtype is still less severe than those seen for deletions encompassing the NLS or the C-terminal domain of Sd (see below). Additionally, disrupting both the NLS

and C-terminal domain, but leaving the TEAD otherwise intact, essentially abolishes all signal in the nucleus (Figure 2.6A, rows 17-20 and see below).

The five other deletions (Sd Δ101-149, Sd Δ246-300, Sd Δ301-355, Sd Δ354-400 and Sd Δ392-440) all had a greatly reduced nuclear signal relative to cytoplasmic signal, as compared to full-length Sd (ranging from a 40% reduction with Sd Δ354-400 to a 67% reduction with Sd Δ246-300; Figure 2.6A rows 4,5 and 7-10). The first, Sd Δ101-149, disrupts the NLS of Sd, lending further support to the notion that this domain is required for Sd localization. The other four deletions either disrupt the Vestigial interacting domain, (VID, Sd Δ246-300 and Sd Δ301-355) or the remainder of the C-terminal domain of Sd (Sd Δ354-400 and Sd Δ392-440). A small 20 amino acid deletion at the C-terminus of Sd is also able to reduce the ratio of nuclear signal to total signal by 65%, relative to full length Sd (Sd Δ421-440, Figure 2.6A row 13). These data show that large portions of the C-terminal domain of Sd, including the VID, are necessary for Sd to direct the eGFP tag to the nucleus of S2 cells. However, this domain cannot direct eGFP to the nucleus alone since both Sd Δ348-440 and Sd Δ1-400 are located predominantly in the cytoplasm. Interestingly, mutating the seven critical basic amino acids of the NLS in conjunction with each of the four large deletions in the C-terminus (Sd mNLS^{N+C} Δ246-300, Sd mNLS^{N+C} Δ301-355, Sd mNLS^{N+C} Δ354-400 and Sd mNLS^{N+C} Δ392-440) results in a phenotype considerably stronger than that when only the NLS is mutated or only the deletions are present.   Indeed, three of these constructs were exclusively cytoplasmic in all cells studied, while the fourth, Sd mNLS^{N+C} Δ354-400, was exclusively cytoplasmic >80% of the time and showed a diffuse localization in the remainder of the cells examined (Figure 2.6A rows 17-20 and compare to Figure 2.5C and Table 2.2). Additionally, two known alleles of *sd*, *sd*^{68L} and *sd*^{11L} previously mapped to the C-terminal coding region of *sd* (Srivastava et al., 2004) were generated as eGFP fusion constructs and expressed in S2 cells. The mutant fusion proteins generated both localized strongly to the nucleus (data not shown).

Contrary to the deletion results detailed above, Sd molecules truncated just downstream of the beginning of the VID or roughly half-way into the VID (Sd Δ229-440 and Sd Δ294-440) locate strongly to the nucleus (>90% nuclear signal), even though they lack the more C-terminal portions of the molecule shown to be important via the previously described deletion analysis (data not shown and Figure 2.6A row 11, respectively). An additional series of truncations was generated to further narrow down potential signals in this last region. As mentioned above, Sd Δ348-440 showed a mis-localization phenotype, with less than 41% of the signal being nuclear (Figure 2.6A row 12). Truncations further C-terminal to amino acid 347 (Sd Δ374-440 and Sd Δ401-440) also had a strong mis-localization phenotype (data not shown). These results imply that one or more regions within amino acids 294-348 interfere with nuclear localization in some fashion, at least in the absence of the remainder of the C-terminus. Consistent with these results, a construct containing the majority of these amino acids (Sd Δ1-300) shows strong cytoplasmic signal with only 35.5% nuclear signal on average and almost half of the cells showing nuclear exclusion of the eGFP signal (Figure 2.6A row 15). However, it should be mentioned that the previously mentioned internal deletion Sd Δ301-355, is largely localized to the cytoplasm, yet also deletes the majority of this region. Representative cells for the described phenotypes are shown as Figures 2.6B-E.

One potential flaw in the previous analysis is that the deletions generated may have an impact on protein structure and/or stability and therefore the changes in localization seen may be a secondary effect of the deletions, rather than a primary effect due to the removal of targeting signals. While it is impossible to rule out this possibility completely, there are a few lines of evidence to counter this line of reasoning: First, two deletions (Sd Δ301-355 and Sd Δ392-440) were tested with a C-terminal GFP tag, rather than an N terminal tag. No significant difference in localization between the C-tagged forms and the N-tagged form were seen (data not shown). Secondly, unstable proteins which are abundantly expressed would be expected to form aggregates known as inclusion bodies (reviewed in Markossian

and Kurganov, 2004). While a small amount of aggregation is seen, the relative levels appear to be low, especially given that eGFP alone is known to aggregate readily. Thus, it seems unlikely that the distributions of signal seen in the Sd deletions is simply due to a properly folded eGFP moiety being size-excluded from the nucleus due to a bulky misfolded Sd isoform anchoring it, even though the properly folded isoform would still be able to mediate nuclear translocation.

The region antagonizing Sd nuclear localization contains a putative NES and is responsive to Leptomycin B.

Based on the results described above, amino acids 294-347 of Sd act to inhibit nuclear localization in some fashion. Within this stretch of amino acids, there is a region with an abundance of hydrophobic residues (11/16 residues, not including K), beginning at V332 and ending at V347 (Figure 2.7A). Although the identity of the residues differs slightly between family members, this hydrophobic region is also present in TEAD proteins from *Choanozoa* and *Animalia.* The consensus of this region contains hydrophobic residues in 10/16 positions total, and these residues align with those in Sd with the exception of residue I339. This residue is hydrophobic in only 4/11 of the species examined (Figure 2.7B). The hydrophobic region of Sd can be aligned with four of the NES classes (1a, 1b, 1d and 3), while the consensus sequence aligns with three of the NES classes (1a, 1b and 3) described by Kusugi *et al* (Figure 2.7C and see Table 2.1; (Kosugi et al., 2009).

To test the possibility that this region contains a NES, a small peptide which includes the putative NES region (Q325 to E352) was fused N-terminally to eGFP (NES-eGFP) and expressed in S2 cells. This caused the average nuclear fraction to be reduced by ~26%, relative to eGFP alone. Moreover, contrary to eGFP, which never showed nuclear exclusion, the NES-eGFP expressing cells examined showed nuclear exclusion of the eGFP tag (Figure 2.7D) 25% of the time. The other

distributions seen were also quantified and tabulated in Table 2.2. Compared to eGFP which showed an

enrichment of nuclear signal 80.0% of the time, this distribution was observed in only 22.7% of the NES-

eGFP expressing cells. Finally, 55.3% of NES-eGFP cells showed more diffuse localization, compared to

20.0% for eGFP alone. Altogether, although NES-eGFP had a range of phenotypes, some of which

overlapped eGFP, the presence of the hydrophobic region of Sd generally decreased the amount of

nuclear signal observed and resulted in nuclear exclusion in many cases.

Leptomycin B (LB) is a potent inhibitor of Crm1 dependent nuclear export (Kudo et al., 1998;

Bogerd et al., 1998). Thus, we tested the ability of this chemical to influence the sub-cellular trafficking

of NES containing constructs (Figure 2.7E). When LB is added to cells expressing eGFP alone, no

significant change in localization is seen. Similarly, Sd Δ301-355 and Sd Δ294-440 (which lack the NES

described above) do not show a response to LB treatment. On the other hand, the NES-eGFP construct

is responsive to LB, as are deletion constructs which are lacking the NLS but contain the NES (Sd ΔNLS

and Sd Δ1-300). Furthermore Sd isoforms which contain both the NLS and NES, but are disrupted more

C-terminally to the NES (Sd Δ348-440 and Sd Δ392-440) are also rescued by the addition of LB.

3xFLAG-PMSD and SD mNLS^{N+C} are potent dominant-negative forms of sd and cannot substitute for wild-

type Sd in wing development.

To test for the necessity of Sd nuclear localization *in vivo*, a Sd protein that contains a Yes

palmitoylation/myristoylation (pal/myr) signal as well as a Fyn linker sequence appended to the N-

terminal domain of Sd (PMSD) was constructed. This sequence is known to target eGFP to the plasma

membrane and endosomes (McCabe and Berthiaume, 1999). As Figures 2.8A and B demonstrate, fusing

this sequence to Sd and a monomeric red fluorescent protein (mRFP) tag likewise targets this fusion

protein to these same locations, rather than the nucleus as is the case for Sd lacking the (pal/myr) signal.

Two transgenic lines (3-2 and 4-1) each containing a flag-tagged form of this construct (*UAS-3xFLAG-*

PMSD) were generated, and the transgene was expressed under the control of a *sd*-GAL4 driver. In

these crosses, 76 and 111 progeny were scored, respectively. The majority of the progeny of the first

cross were females (45%) or males (34%) which inherited a balancer chromosome, rather than the

transgene. The remaining 21% of the flies were females with greatly reduced wings and halteres

(Figures 2.8D), relative to an Oregon-R (Ore^R) fly (Figure 2.8C). No non-balancer male progeny were

observed. In the second cross, 29% and 21% of the progeny were females or males, respectively, which

inherited the balancer chromosome. Furthermore, 27% of the progeny were females with greatly

reduced wings and halteres similar to those seen when the 3-2 line was used. Contrary to the 3-2 line,

the 4-1 line also yielded male progeny with this phenotype. These flies accounted for 23% of the total

progeny. Transgenic flies containing a flag-tagged *UAS-3xFLAG-SD mNLS*$^{N+C}$ transgene were also

generated. A similar range of progeny phenotypes was also seen when a *UAS-3xFLAG-SD mNLS*$^{N+C}$ was

expressed using the *sd* driver. Again two lines were used, A (39 progeny of the *sd*-GAL4 cross scored)

and B (62 progeny of the *sd*-GAL4 cross scored). When line A was used, the distribution of progeny

females with the balancer, progeny males with the balancer and progeny females with reduced

wing/haltere tissue (Figure 2.8E) was 50%, 42% and 8%, respectively. No non-balancer male flies were

observed. The equivalent distribution observed when using line B was 45%, 19% and 32%. In this case

males with the wing/haltere phenotype were seen 3% of the time. None of the progeny from any of the

four crosses had any obvious defects outside those observed in the wing and haltere.

Over-expression of wildtype Sd is able to cause strong wing phenotypes in an otherwise

wildtype background. However, in *sd* mutants which have a strong wing phenotype (*sd*^{58d}; Campbell et

al., 1992) this same construct is also able to significantly restore wing development when driven with *vg*-

GAL4 (Chow et al., 2004).   While both *UAS-3xFLAG-PMSD* and *UAS-3xFLAG-SD mNLS*$^{N+C}$ have a strong

dominant negative effect in wildtype flies, as shown above, neither is able to rescue the wings of *sd*^{58d}

flies when driven with *vg*-GAL4 (data not shown).

The SV40 large T-antigen NLS is the prototypical classic NLS (cNLS) and is known to be able to

direct eGFP to the nucleus (Cressman et al., 2001; Kalderon et al., 1984; Lanford and Butel, 1984). As

such, we tested to see if this NLS was able to rescue our Sd NLS mutants by generating transgenic lines

which contained a 3xFLAG-SV40NLS-Sd mNLS^{N+C} transgene. While the addition of this signal was able to

increase the amount of eGFP-Sd mNLS^{N+C} found in the nucleus of S2 cells from ~44% to 68%, no change

in the *in vivo* dominant negative phenotypes were seen, and this isoform of Sd was still unable to rescue

sd^{58d} mutants (data not shown).

Western analysis of Sd reveals the presence of two bands, one of which is phosphatase sensitive.

As shown above, Sd contains an NLS as well as possibly an NES and moreover, the C-terminal

domain of Sd is also important for the nuclear localization but there is no evidence for an additional NLS

in that domain. Together these data raise the possibility that Sd may shuttle between the nucleus and

cytoplasm, and therefore that there may be regulation of this process. One manner in which NLSs and

NESs may be regulated is via post-translational modification of the protein (reviewed in (Sorokin et al.,

2007). These modifications may either alter the ability of Importins/Exportins to bind to their cognate

sequences directly (e.g. by sterically interfering with binding) or indirectly (e.g. by causing

conformational changes which cover or uncover the NLS or NES). To determine whether this might be

true of Sd as well, FLAG-Sd was expressed *ex vivo*, purified using anti-FLAG beads and analyzed by

Western blot on a low-bis acrylamide gel. Under these conditions, two bands were observed for Sd – a

smaller one and a slightly shifted larger band, raising the possibility that the larger band is a post-

translationally modified form of Sd (Figure 2.9). Further evidence that this is indeed the case, and that

the nature of the modification is phosphorylation was the fact that this band was sensitive to λ

phosphatase treatment (Figure 2.9). Indeed, *in silico* prediction programs (e.g. NetPhos; Blom et al.,

1999) predict a multitude of phosphorylation sites spread more or less evenly throughout the protein.

However, when the 50 amino acid deletions noted in Figure 2.6 were tested in a similar fashion to determine which region of Sd was necessary for the presumed phosphorylation, none of them lacked the second, higher molecular weight, band.

Discussion

The data presented show that a the previously predicted putative NLS of Sd is indeed functional. Both eGFP and eGFPx2-GST are targeted to the nucleus by the NLS of Sd, even though the latter is too big to undergo passive diffusion into the nucleus. Based on the sequence of the NLS, and the fact that this sequence facilitates Imp-α3 binding, this signal is likely a member of the bipartite family of cNLSs. Moreover, although mutating the N-terminal basic amino acids in the signal only has a minor effect on the strength of the signal, this is consistent with typical bipartite signals, where the N-terminal cluster of basic amino acids is less critical then the C-terminal cluster (Kosugi et al., 2009). To our knowledge, this is the first such signal that has been confirmed to be functional within a TEAD containing protein. However, the signal is well-conserved and it is plausible that it is also functional in other representatives of this widespread and important family of transcription factors.

As mentioned, the NLS of Sd shows homology to the classically defined bipartite family. However, the sequence is not consistent with a more refined consensus derived by Kosugi *et al* (2009). These researchers compared published NLS sequences to randomly generated artificial sequences which were assayed for their ability to direct eGFP to the nuclei of various cell lines. In this way they generated two consensus sequences: $KRX_{10-12}K(K/R)X(K/R)$ and $KRX_{10-12}K(K/R)(K/R)$. Even though the NLS of Sd (RKQVSSHIQVLARRKLR) is similar to both of these patterns, it is unique in that RK, rather than

KR, is found at the N-terminal portion of the signal and furthermore R, rather than K, is found at the first position of the C-terminus. Thus, the NLS of Sd is a novel member of the bipartite family of cNLSs.

It has been previously speculated that mutant forms of Sd, which retain the ability to interact with Vg and other co-factors but lack the ability to enter the nucleus or bind DNA, act in a dominant negative fashion by titrating the binding partners of Sd This in turn reduces the amount of these co-factors available to interact with endogenous Sd (Garg et al., 2007; Simmonds et al., 1998; Chow et al., 2004).(Chow et al., 2004)(Chow et al. 2004) We have reinforced this idea by expressing isoforms of Sd which are either targeted to the cytoplasmic membrane and endosomes (3xFLAG-PMSD) or have a mutated NLS (3xFLAG-Sd mNLS^{N+C}). Both these isoforms act as strong dominant negative forms of Sd during wing development, implying they are still able to interact and titrate endogenous Vg. However, neither is able to substitute for endogenous Sd in a sd^{58d} mutant background, demonstrating that a critical function is impaired in both isoforms of Sd. In the case of 3xFLAG-PMSD, the protein has not been altered in any way, thus it is unlikely that anything other than the protein's sub-cellular localization has changed. By extension, the fact that 3xFLAG-Sd mNLS^{N+C} gives identical phenotypes to 3xFLAG-PMSD and that the NLS is clearly functional in S2 cells strongly suggests that localization is similarly impaired *in vivo*. Contrary to this, the SV40 NLS is not able to rescue the function of Sd mNLS^{N+C} *in vivo*, even though it can rescue localization *in vitro*. We do not believe these results are incompatible for three reasons. First, the magnitude of rescue in S2 cells was significant, but not complete. Therefore, it is possible that no effect is seen phenotypically. Second, our data are consistent with the notion that the sub-cellular localization of Sd is regulated in some fashion. Thus, the SV40 tagged form of Sd may still not be localizing to the nucleus at the correct times. Finally, the mutations fall within the DNA binding domain of Sd, and thus might have secondary effects on the protein's ability to function *in vivo*.

In addition to identifying a cNLS in Sd, we also identified an NES which likely relies on Crm1 to facilitate nuclear export, which together with the presence of the NLS we identified, implies that there is a switch between nuclear and cytoplasmic forms of Sd and that the protein may be capable of shuttling between the two domains. Furthermore, our data indicate that the domain C-terminal to the NES (amino acids 353-440) must have at least one other signal which facilitates nuclear import. *In silico* analysis did not identify any other regions which resemble an NLS, and the C-terminal domain of Sd is not sufficient to target an eGFP tag to the nucleus, so it is unlikely that another NLS exists within this domain of Sd. Rather, all available evidence suggests that this domain is responsible for protein-protein interactions, since two of the three known cofactors of Sd (Yki and Vg) are known to bind to this domain (Simmonds et al., 1998; Goulev et al., 2008). The binding site of the third Sd-binding protein (dMef2) has not been elucidated (Deng et al., 2009). This would also help to explain why Sd is still partially able to locate to the nucleus when the NLS is disrupted. Likely, this domain allows Sd to bind a co-factor which is able to translocate to the nucleus. It is quite possible that one of the other proteins is endogenous Sd, since Sd is known to dimerize and there is evidence that Sd transcripts are enriched in S2 cells (Chintapalli et al., 2007, and our unpublished data). It is worth noting that neither Yki nor Vg have a predicted NLS. Furthermore, Yki is completely cytoplasmic in the absence of Sd (Goulev et al., 2008; Zhang et al., 2008). Most evidence suggests that Vg requires Sd for nuclear localization, yet it shows some nuclear accumulation when expressed alone in S2 cells (Halder et al., 1998; Simmonds et al., 1998; Srivastava et al., 2004). This is likely due to endogenous Sd, rather than the presence of an NLS in Vg. However, the Mef2 family is known to contain an NLS and dMef2 transcripts are present in S2 cells (Borghi et al., 2001; Chintapalli et al., 2007). Taken together, it seems likely that the C-terminal domain of Sd modulates the nuclear localization of the protein by binding accessory factors that either facilitate nuclear transport directly, and/or alter the function of the localization signals of Sd. This idea is

a novel one for a TEAD containing protein, and given the high sequence similarity of proteins of this family, has implications for the regulation of TEAD proteins in other organisms.

Two alleles of *sd*, *sd*[68L] and *sd*[11L], have been mapped to the 3' coding region of the gene. These alleles cause the lethal mutations Y355N and H433L, respectively (Srivastava et al., 2004). The first causes a reduction in Vg nuclear localization in *sd*[68L] flies, even though the product of this mutant allele is able to interact with Vg *in vitro*. The second lies within the region deleted in Sd Δ421-440, which we have shown to be important for nuclear localization. Thus, we hypothesized that one or both might be involved in the nuclear localization of Sd. However, both Sd[11L] and Sd[68L] are able to strongly direct an eGFP tag to the nucleus of S2 cells (data not shown). This implies that neither mutation directly impacts the nuclear localization of Sd. However, these results do reinforce the idea that the C-terminal domain has functions in addition to those already described.

In summary, herein evidence has been presented which indicates that the sub-cellular localization of Sd is dependent on multiple signals, including at least one bipartite cNLS and possibly an NES as well. Furthermore, the domain C-terminal to the NES of Sd is also important for trafficking the protein. While it seems likely that this is mediated by the ability of this domain to facilitate binding to cofactors, rather than direct binding to importins and exportins (although we cannot rule this possibility out), the mechanism by which this occurs is yet to be determined.

Materials and Methods

Construct design- Internal deletions were generated using inverse PCR followed by blunt-end ligation prior to cloning. Substitution mutations (mutations to the *sd* NLS coding sequence) were

generated either by inverse PCR with non-overlapping primers, followed by blunt-end ligation prior to cloning, or by using inverse PCR with primers containing partially overlapping 5' ends, followed by *Dpn*I treatment and transformation into *E. coli* (modified from Fisher and Pei, 1997). Deletions, point mutations, the TEA coding sequence and the NLS coding sequence were cloned into pENTR using the pENTR/D-TOPO kit (Invitrogen Life Technologies). These constructs were then subsequently subcloned into pHGW (N-terminal eGFP), pHWG (C-terminal eGFP), pHFW (N-terminal 3xFLAG) or pTFW (N-terminal 3xFLAG, pUAST based transformation vector) using LRII recombinase (Invitrogen Life Technologies) according to the Murphy lab protocols (www.ciwemb.edu/labs/murphy/Gateway%20vectors.html#_References). In order to make C-terminal GFPx2-GST tagged proteins pMT/v5(A)+eGFPx2-GST was used (described in (Chan et al., 2007). To clone into this vector, *Kpn*I restriction sites were appended to the NLS, TEA and the TEA ΔNLS coding domains using PCR amplification. These sites were then used for cloning 5' to the tags. Oligonucleotides were used to append the palmitoylation, myristoylation and a linker domain to the *sd* coding sequence in order to generate PMSD, which was subsequently cloned in pENTR and subcloned into the monomeric red fluorescent protein (mRFP) tagging vector, pHRW. Oligonucleotides were also used to add the SV40 NLS coding sequence (which translates to PKKKRKV) into the *Not*I site of pENTR+Sd mNLS^{N+C}. Routine PCRs were done with PlatinumTaq HIFI, while inverse PCRs were done with either Pfx^{50} or AccuPrime Pfx^{50} (all from Invitrogen Life Technologies).

Drosophila Stocks- sd, *PMsd-mRFP*, *sd mNLS^{N+C}* and *SV40-sd mNLS^{N+C}* were cloned into pTFW for subsequent micro-injection. The first was injected as described previously (Rubin and Spradling, 1983), into *y w*; *Δ2-3/Sb* embryos. The other two injections were performed commercially (BestGene). At least two independent lines for each injection were generated. All crosses were performed at room temperature. *y w*; *Δ2-3/Sb* was a gift from A. Simmonds.

Cell culture- S2 cells were obtained from Invitrogen Life Technologies. The cells were cultured in

HyQ CCM3 (HyClone) at room temperature and 0.6 µg of the desired plasmids were transfected using

Cellfectin (Invitrogen Life Technologies) according to the manufacturer's directions. In order to drive

expression of GFP tagged constructs, the cells were heat-shocked @ $37^{o}C$ for 40 minutes, approximately

36 h after transfection. pMT/v5(A) based constructs were induced by adding 0.4mM $CuSO_4$, 24 h after

transfection.   Induced cells were collected 38 hours post-transfection, washed, fixed in 2%

paraformeldehyde and stained with DAPI diluted to a final concentration of 1 µg/ml. PBS was used as a

buffer for all manipulations. The cells were mounted in PBS for imaging and coverslips sealed with

VALAP (1:1:1 mixture of vasoline, lanolin and parafin wax (North, 2006). For Leptomycin B treatment,

cells were incubated with 25 nM of the chemical for 2 h prior to heat-shock.

Cells were imaged on a Zeiss 510 confocal microscope, using the appropriate filters for eGFP,

mRFP and 4',6-diamidino-2-phenylindole (DAPI). To minimize potential cross-talk between channels,

scans were done sequentially.   Images were initially imported and analyzed in ImageJ (Abramoff et al.,

2004). Subsequently Adobe Illustrator CS3 10.0 was used for final assembly (annotations and

adjustments to brightness and contrast). Microsoft Excel 2007 was used to perform two-sample t-tests

assuming unequal variance in order to test for statistical differences between the mean nuclear

localizations.

Quantification of nuclear signal was done determining the total cellular signal and the nuclear

signal using ImageJ. Cells were than normalized for both cytoplasmic and nuclear size. Finally, the

normalized nuclear signal was divided by the normalized total signal to get the percent nuclear signal.

The percent nuclear signal was then arbitrarily assigned to one of four categories: Nuclear denotes

cells that contain exclusively or almost exclusively nuclear signal (>80% nuclear signal). Diffuse Nuclear

includes cells which show predominant expression in the nucleus along with varying degrees of

cytoplasmic signal (79-58% nuclear signal). Diffuse is for cells with signal approximately evenly

distributed between the nucleus and cytoplasm or slightly enriched in the cytoplasm (57-36% nuclear

signal). Excluded categorizes those cells which have exclusive or almost exclusive cytoplasmic signal

(<35% nuclear signal).

Co-immunoprecipitations- pHFW + *sd* and pHFW + *sd mNLS^{N+C}* were transiently transfected and

induced in S2 cells as described above. A mock transfection was also done with water. Instead of fixing

the cells, they were lysed in RIPA buffer (50 mM Tris-HCl, 150 mM NaCl, 1% NP-40, 0.5% deoxycholic

acid, 0.1% SDS) containing Complete Protease Inhibitor Cocktail (Roche) for 15 min on ice. The lysed

cells were then harvested and the lysate incubated with αFLAG M2 Affinity Gel (Sigma-Aldrich) for two

hours at 4°C. The affinity beads were extracted and diluted into standard 4x SDS protein loading buffer.

Equal amounts of 3xFLAG-Sd and 3xFLAG-Sd mNLS^{N+C} protein were loaded and separated on a 10% poly-

acrylamide gel. Blotting was on Hybond ECL (GE Healthcare) with subsequent analysis using either anti-

FLAG (Sigma-Aldrich) or anti-Importin-α3 (Máthé et al., 2000) as primary antibodies. Detection was with

horseradish peroxidase-labelled anti-mouse or anti-rabbit secondary antibodies (Invitrogen), both at

1:50000, and the SuperSignal Substrate Western Blotting kit (Pierce).

Analysis of post-translational modifications of Sd- 3xFLAG-Sd was expressed and purified as in

the Co-ip experiments detailed above , except that 30 mM Sodium Pyrophosphate, 0.5 mM DTT, 10 mM

Sodium Orthovanadate and 50 mM Sodium Flouride were added to inhibit phosphatase activity in the

cell lysate. Additionally, subsequent to purification, the purified protein was either treated with λ

phosphatase (New England Biolabs) in buffer, or with buffer alone. These samples were then analyzed

by Western blot on a low-bis (119:1 acrylamide:bis-acrylamide) gel using αFLAG to detect the fusion

protein.

Alignments- Jalview (Waterhouse et al., 2009) was used to align TEAD containing sequences

identified through BLASTp (www.ncbi.nlm.nih.gov, except EGL-44) searches using the Sd protein

sequence as the query. EGL-44 was identified using wormbase (www.wormbase.org,WS204, July 29th

2009).

Figure 2.1. Simplified schematic of nuclear pore complex. The nuclear pore complex spans the inner and outer nuclear membranes and consists of cytoplasmic filaments, a nuclear core (through which the central pore spans) and nuclear filaments. Modified from Sorokin et al., 2007.

Central pore

Central core

Cytoplasmic fibrils

Outer nuclear membrane

Inner nuclear membrane

Nuclear basket

Nuclear filament
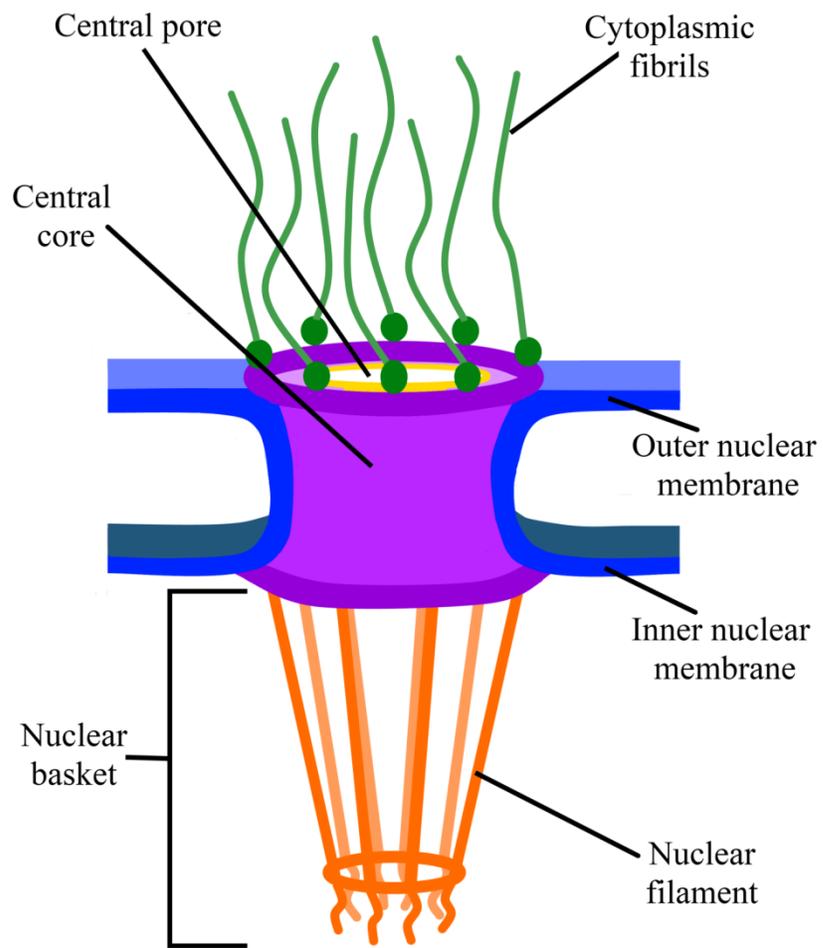
Table 2.1. Consensus of experimentally determined NES sequences. Shown are the three

classes (the first of which is composed of four subclasses) of consensus sequences demonstrated to be

competent to act as NESs in a yeast-based assay (Kosugi et al., 2008). X, X2, X3 are any one, two or

three amino acids, respectively; Φ is L, I, V, M, F, C, W, A or T, but no more than one Φ can be C, T, A or

W. A higher proportion of L and I indicates a stronger likelihood that a given sequence that matches one

of the indicated consensuses is a true NES.

Class	Consensus
1a	Φ-X3-Φ-X2-Φ-X-Φ
1b	Φ-X2-Φ-X2-Φ-X-Φ
1c	Φ-X3-Φ-X3-Φ-X-Φ
1d	Φ-X2-Φ-X3-Φ-X-Φ
2	Φ-X-Φ-X2-Φ-X-Φ
3	Φ-X2-Φ-X3-Φ-X2-Φ

Figure 2.2. Simplified overview of Importin-α/β mediated nuclear import. Importin-α binds to the NLS of a cytoplasmic cargo protein. Subsequent to this, Importin-β and Ran-GDP also enter the complex. This complex is then competent to travel through the nuclear pore and into the nucleus. Once in the nucleus, Ran-GTP replaces the Ran-GDP bound to Importin-β, allowing for Importin-α and the cargo protein to disassociate from the complex. Modified from Sorokin et al., 2007.
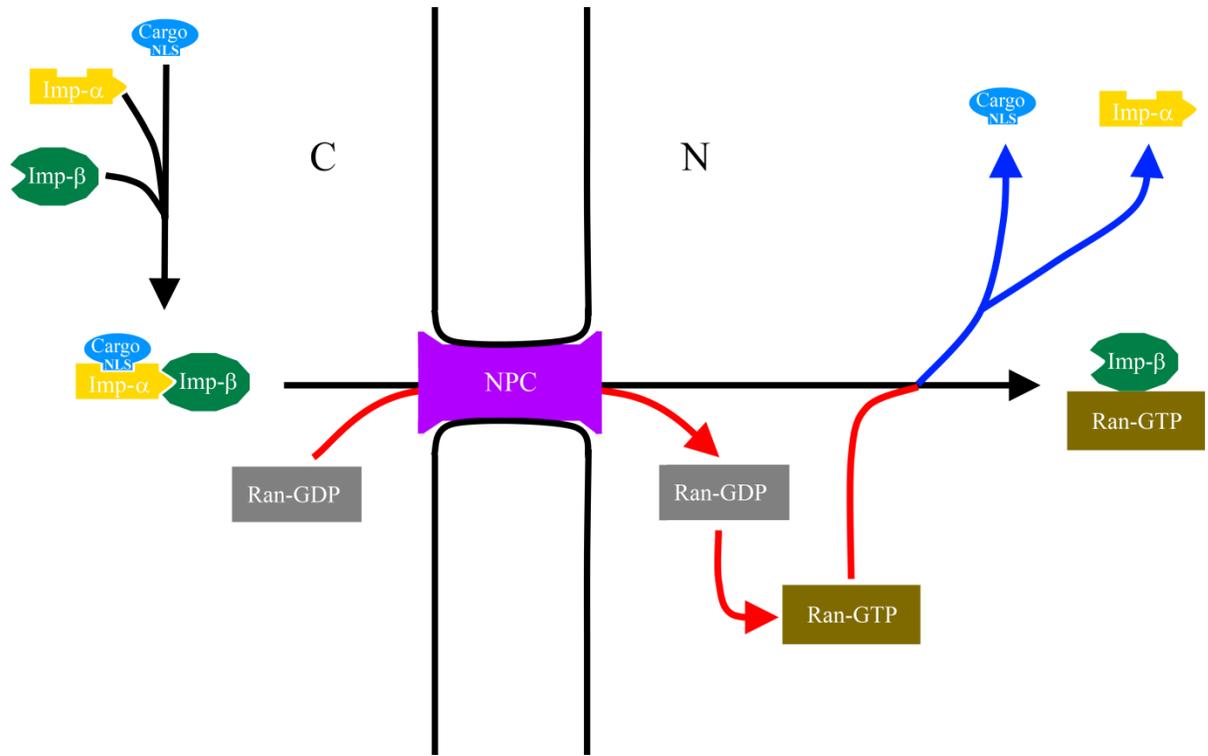
Figure 2.3. Identification of a putative bipartite NLS. (A) A schematic diagram of Sd. Sd contains two known functional domains, the TEA (DNA binding) domain and the Vestigial interacting domain (VID), as shown. At the C-terminus of the TEA domain, there is a 17 amino acid stretch from R145 to R161 which closely matches the consensus classic bipartite NLS sequence. (B) The region corresponding to the bipartite NLS shows strong identity with a variety of TEAD proteins from both animals and *Choanozoa* protists. Arrowheads mark the sites of the two N-terminal and five C-terminal residues known to be important for the bipartite sequence. 'X' marks the 10 intervening amino acids lying between the two termini. A '+' indicates a basic residue (L/R) lies at one of the N- or C-terminal critical sites in the consensus sequence of the aligned TEAD proteins. The dark shading indicates identity with the consensus, while the lighter shading indicates similarity.

Table 2.2. Quantification of the cellular distribution of the eGFP tagged peptides. The eGFP

fusion constructs from Figures 2.4A-G, 2.5A-E and Figure 2.7D were assayed for the percentage of eGFP

signal seen in the nuclei of the expressing cells; see materials and methods. (S.E.M) is the standard error

of the mean. A † denotes a construct with diffuse or nuclear excluded signal (<58% nuclear signal). N is

the total number of cells measured from at least two independent transfections. The next four columns

represent four arbitrary localization patterns along with the mean nuclear signal each grouping

represents. For each peptide, the percentage of cells that fall into one of the four categories is

indicated. The means of the experimental constructs TEA-eGFP, NLS-eGFP and NES-eGFP are statistically

different from their control (eGFP) at $p < 0.001$. Likewise, NLS-eGFPx2, TEA-eGFPx2 and TEAΔNLS-

eGFPx2 are significantly different from eGFPx2, at $p < 0.0001$. Finally the mean of the control eGFP-Sd

was significantly different from the four reporter constructs in which the NLS was mutated, at $p < 0.001$.

Nuc. = Nuclear. Excl. = Excluded.

		Average %Nuc./Total	%Nuc.	%Diffuse Nuc.	%Diffuse	%Excl.
Construct (KDa)	N	(S.E.M.)	(>80%)	(79-58%)	(57-36%)	(<35%)
eGFP (29.7)	20	61.1 (0.9)	0.0	80.0	20.0	0.0
TEA-eGFP (39.6)	25	94.6 (0.9)	100.0	0.0	0.0	0.0
NLS-eGFP (32)	34	88.2 (1.2)	85.3	14.7	0.0	0.0
eGFP-Sd (78.7)	32	92.5 (0.6)	100.0	0.0	0.0	0.0
eGFPx2 (82.9)†	25	22.5 (1.2)	0.0	0.0	8.0	92.0
NLS-eGFPx2 (86.0)	19	60.8 (2.8)	4.2	62.5	29.2	4.2
TEA-eGFPx2 (94.0)	19	78.8 (2.4)	57.9	42.1	0.0	0.0
TEAΔNLS-eGFPx2 (91.9)†	19	14.8 (0.7)	0.0	0.0	0.0	100.0
eGFP-SDΔNLS (76.6)†	31	42.6 (1.1)	0.0	3.2	83.9	12.9
eGFP-SD mNLS^N (78.5)	37	79.6 (1.6)	51.4	43.2	5.4	0.0
eGFP-SD mNLS^C (78.5)†	38	46.9 (0.8)	0.0	0.0	100.0	0.0
eGFP-SD mNLS^{N+C} (78.5)†	35	44.1 (1.3)	0.0	2.9	91.4	5.7
NES-eGFP (32.5)†	44	45.1 (2.1)	0.0	22.7	52.3	25.0

Figure 2.4. The NLS of Sd directs an eGFP tag to the nucleus. (A-G) Localization of the indicated

eGFP reporter tagged peptides in transiently transfected in S2 cells with DAPI stained nuclei and

visualized via confocal microscopy. A^1-G^1 are the green (eGFP) channels. A^2-G^2 are the blue (DAPI)

channels. A^3-G^3 are the green and blue channels (merge). Hatched lines indicate the boundary of cells,

as determined by the extent of the weak cytoplasmic signal. Percentages indicate the percent nuclear

signal relative to total signal measured in the given cell. (A) eGFP. When eGFP is expressed alone,

diffuse expression is seen throughout the cell, including the nucleus. (B) TEA-eGFP. A fragment of Sd

stretching from amino acids 88-178 (which includes the entire TEA/NLS domain) shows almost exclusive

reporter activity within the nucleus of the expressing cells. (C) NLS-eGFP. Amino acids 143-163 of Sd

(which includes the NLS and two flanking amino acids on either side) drives reporter expression to the

nucleus. (D) eGFPx2 + HA (referred to hereafter as eGFPx2). eGFPx2 expression is excluded from the

nucleus. (E) TEA-eGFPx2. A TEA-eGFPx2 fusion is primarily nuclear. (F) eGFPx2 + NLS. This construct is

found throughout the cell, but is enriched in the nucleus. (G) TEAΔNLS-eGFPx2. When the NLS is

removed from the TEA domain, it is no longer able to direct the tag to the nucleus.
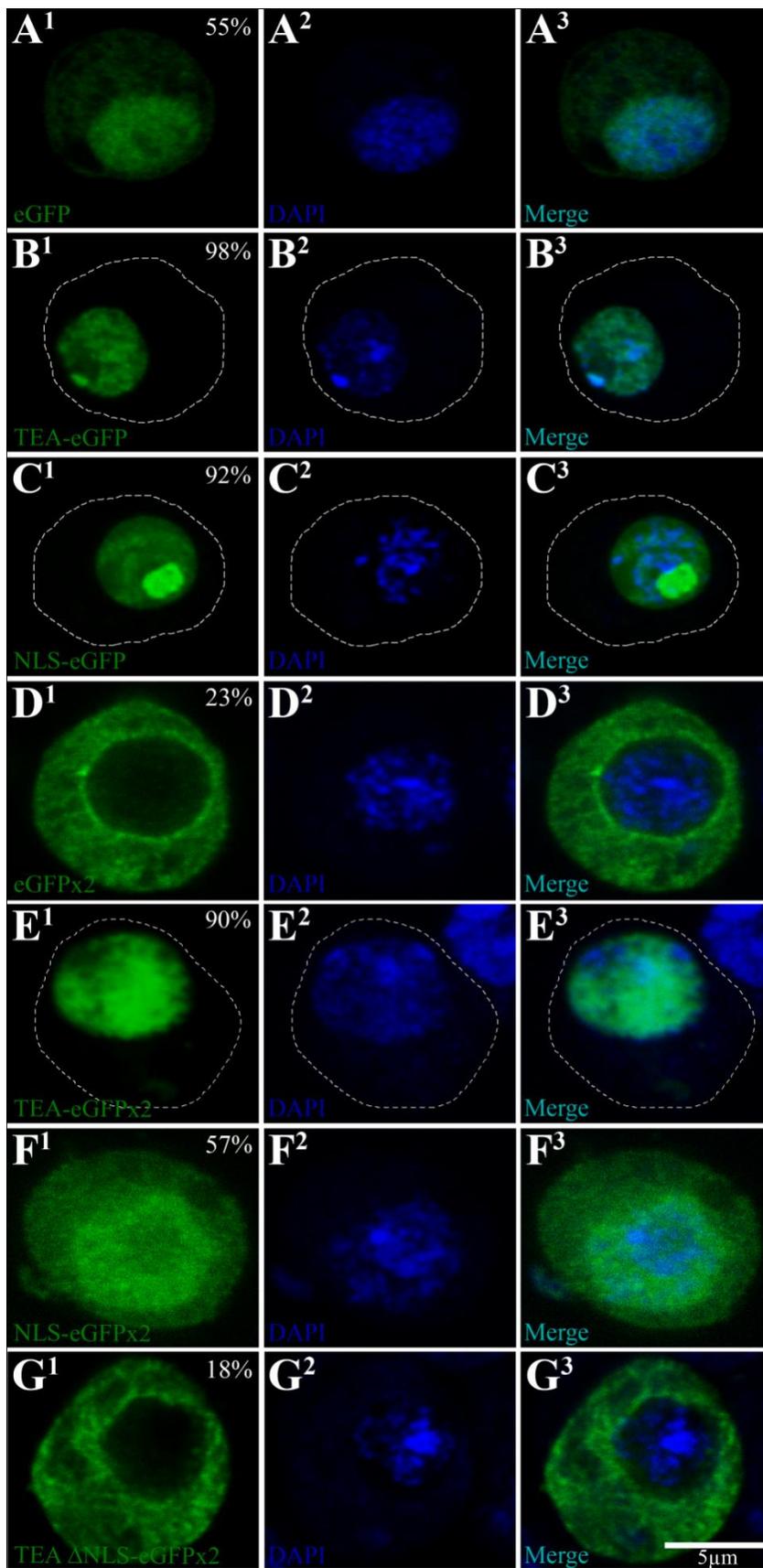
Figure 2.5. The intact NLS is necessary for proper nuclear translocation and Importin-α3 binding. (A-E) Localization of the indicated eGFP reporter tagged proteins in transiently transfected in S2 cells with DAPI stained nuclei and visualized via confocal microscopy.   See legend for Figures 2.4A-G for details. (A) eGFP-SD. When Sd is expressed in S2 cells, reporter activity is predominantly nuclear. (B) eGFP-SD ΔNLS. Deleting amino acids 143-163 of Sd disrupts its localization and leads to diffuse reporter activity throughout both the nucleus and cytoplasm. (C) eGFP-SD mNLS^{N+C}. Mutation of the six basic amino acids identified as being critical in the consensus bipartite sequence (see Figure 2.3) to N causes disruption of localization similar to that seen when the NLS is deleted. (D) eGFP-SD mNLS^{N}. When the two N-terminal basic amino acids are mutated to N, a lesser disruption of the nuclear signal is observed (compare to A). (E) eGFP-SD mNLS^{C}. Sd with the four C-terminal basic amino acids mutated to N drives diffuse localization of the eGFP reporter, similar to that seen for SD ΔNLS and SD mNLS^{N+C}. (compare to panels B and C, respectively).   (F) Co-IP of Sd and Imp-α3. Cells expressing 3xFLAG-Sd, 3xFLAG-Sd mNLS^{N+C} as well as cells mock transfected with water alone were lysed, immunoprecipitated with αFLAG beads and analyzed via western blotting. Detection was with anti-FLAG or anti-Imp-α3. Detection with αFLAG ensures expression of the two tagged proteins is approximately equal. The lysate of all cells had a strong Imp-α3 signal. Imp-α3 co-immunoprecipitated strongly with 3xFLAG-Sd, while only weakly with 3xFLAG-Sd mNLS (N+C). The mock transfected cells showed almost no Imp-α3 signal after immunoprecipitation, controlling for the specificity of the anti-FLAG beads.

A¹ eGFP-SD — 90% — A² DAPI — A³ Merge

B¹ eGFP-SD ΔNLS — 43% — B² DAPI — B³ Merge

C¹ eGFP-SD mNLS^{N+C} — 42% — C² DAPI — C³ Merge

D¹ eGFP-SD mNLS^N — 80% — D² DAPI — D³ Merge

E¹ eGFP-SD mNLS^C — 48% — E² DAPI — E³ Merge — 5μm

F

WB:αIMPα3

WB:αFLAG

3x FLAG-Sd   3x FLAG-Sd mNLS^{N+C} Mock

3x FLAG-Sd   3x FLAG-Sd mNLS^{N+C} Mock

IP: αFLAG

Lysate
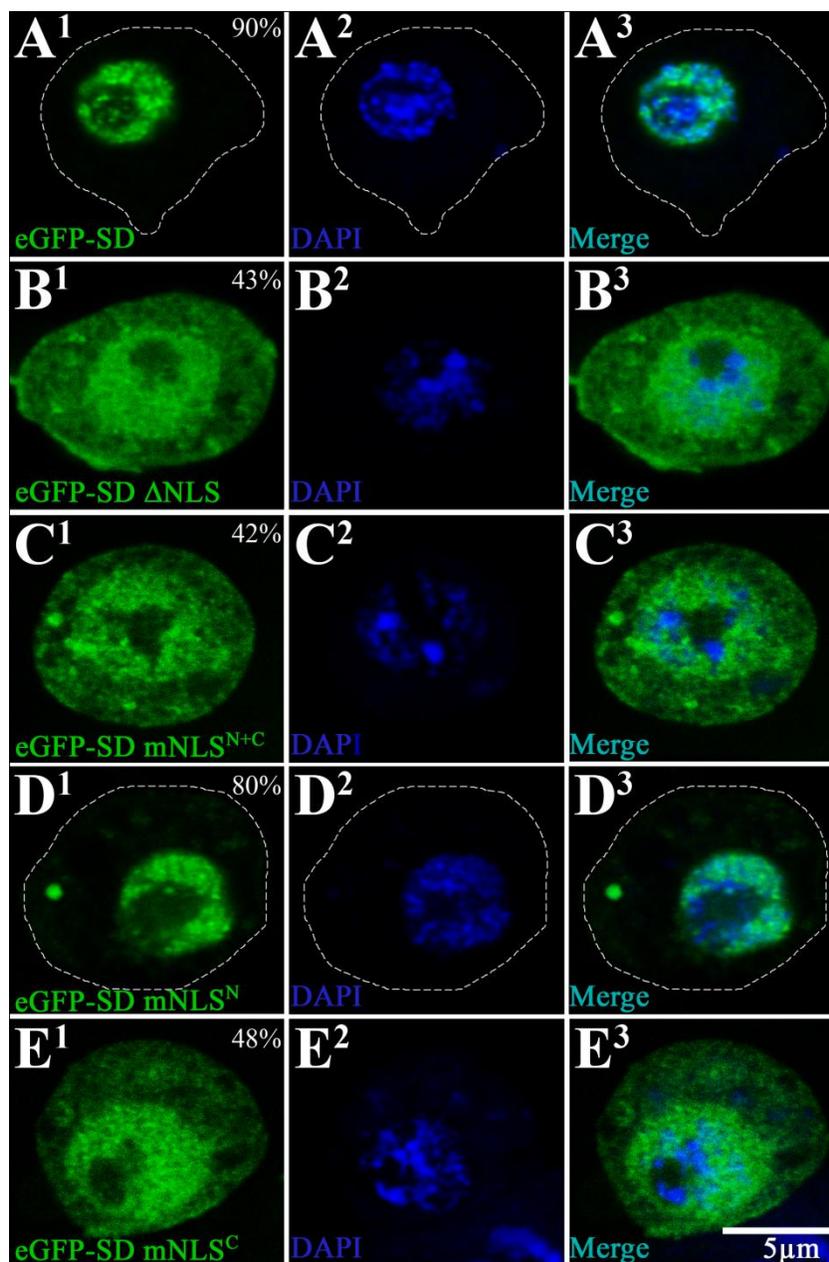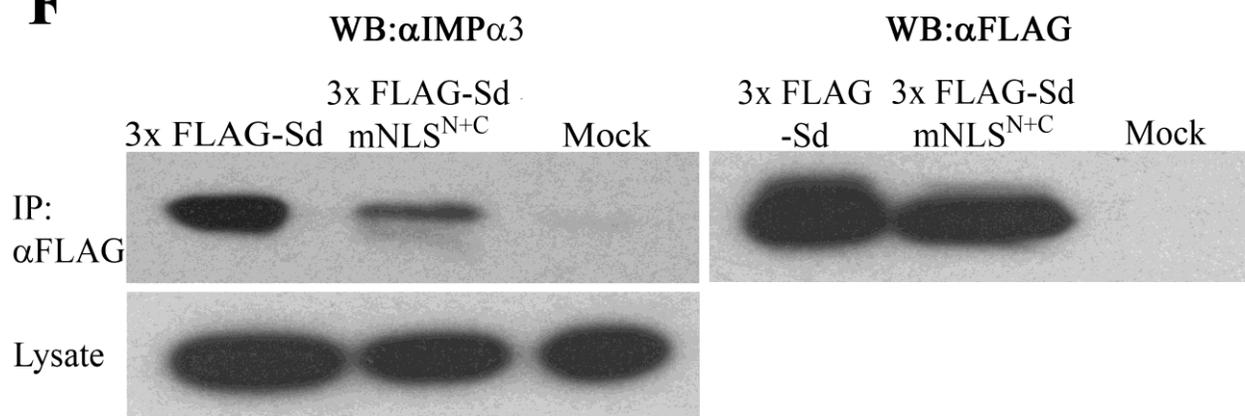
Figure 2.6. The C-terminal domain can act to both repress and facilitate the nuclear localization of Sd. A series of internal deletions and truncations of Sd were generated, expressed with a fused N-terminal eGFP marker in S2 cells and assayed for cellular distribution. (A) Schematic of the various Sd isoforms generated along with a summary table of the localization experiments. The domains of Sd are as described in Figure 2.3A. 'mNLS^{N+C}' is described in Figure 2.5C. For a description of the table, see Table 2.2. (B-E) Representative cells showing 81%, 65%, 47% and 31% nuclear signal (B, C, D and E, respectively). See Figure 2.4A-G for details.

A

| Row	eGFP-construct (KDa)	Schematic	N	Average %Nuc./Total (S.E.M.)	%Nuc. (>80%)	%Diffuse Nuc. (79-58%)	%Diffuse (57-36%)	%Excl. (<35%)
1	Sd (78.7)		32	92.5 (0.6)	100.0	0.0	0.0	0.0
2	Sd Δ1-56 (72.9)		42	85.1 (0.9)	81.0	19.0	0.0	0.0
3	Sd Δ51-102 (73.4)		39	83.2 (1.0)	74.4	25.6	0.0	0.0
4	Sd Δ101-149 (72.8)		38	48.0 (1.1)	0.0	5.3	92.1	2.6
5	Sd Δ150-200 (72.6)		41	46.5 (1.0)	0.0	7.3	92.7	0.0
6	Sd Δ199-248 (72.9)		40	85.6 (1.0)	80.0	20.0	0.0	0.0
7	Sd Δ246-300 (72.2)		41	30.9 (1.0)	0.0	0.0	26.8	73.2
8	Sd Δ301-355 (72.5)		24	53.1 (1.8)	0.0	25.0	75.0	0.0
9	Sd Δ354-400 (72.5)		39	55.2 (1.8)	0.0	43.6	51.3	5.1
10	Sd Δ392-440 (74.8)		27	34.3 (1.4)	0.0	0.0	48.1	51.9
11	Sd Δ294-440 (63)		29	90.6 (1.0)	96.6	3.4	0.0	0.0
12	Sd Δ348-440 (69.1)		28	40.8 (1.7)	0.0	3.6	75.0	21.4
13	Sd Δ421-440 (77.1)		30	32.2 (1.4)	0.0	0.0	33.3	66.7
14	Sd Δ1-400 (35.1)		24	46.3 (1.2)	0.0	4.2	95.8	0.0
15	Sd Δ1-300 (45.8)		35	35.5 (1.1)	0.0	0.0	51.4	48.6
16	Sd Δ1-142 (64.6)		29	70.4 (1.8)	17.2	75.9	6.9	0.0
17	Sd mNLS^{N+C}Δ246-300 (72.0)		28	23.0 (0.8)	0.0	0.0	0.0	100.0
18	Sd mNLS^{N+C}Δ301-355 (72.2)		29	20.4(0.7)	0.0	0.0	0.0	100.0
19	Sd mNLS^{N+C}Δ354-400 (72.3)		26	28.3 (1.2)	0.0	0.0	15.4	84.6
20	Sd mNLS^{N+C}Δ392-440 (74.6)		19	24.3 (1.0)	0.0	0.0	0.0	100.0
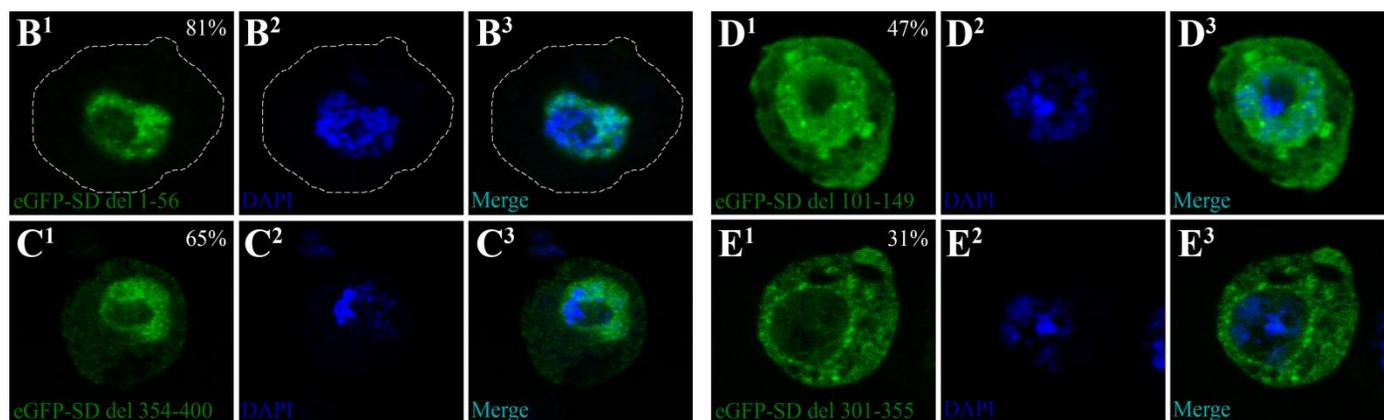
Figure 2.7. Sd contains a sequence at amino acids 332-347 which resembles an NES, and increases the cytoplasmic fraction of a fused eGFP tag in a leptomycin B (LB) sensitive manner. (A) Schematic of Sd with the putative NES marked. The domains of Sd are described in Figure 2.3A. Hydrophobic residues are underlined. The open and closed arrowheads mark the boundaries of the region intact in SD Δ344-440 (which directs eGFP to the cytoplasm) and missing in SD Δ294-440 (which directs eGFP to the nucleus), as seen in Figure 2.6A, rows 12 and 11, respectively. (B) Alignment of several TEAD proteins. Dark shading indicates hydrophobic residues L, I, V, M and F, while light shading indicates hydrophobic residues C, W, A or T, with the first group being generally more favourable to NES function (Kosugi et al., 2008). (C) Overview of four NES classes in comparison to the Sd hydrophobic sequence, and the comparable TEAD consensus.   In the second column, the consensus of the four given classes of NES, derived from a comparison to natural and synthetic NESs (see text and Table 2.1 for details) are indicated. In the third column, the Sd hydrophobic region is aligned to fit these patterns, and if possible, the fourth column shows the equivalent residues from the TEAD protein consensus. Within the Sd or TEAD containing protein consensus sequence, an underline represents a hydrophobic residue. A bolded and enlarged hydrophobic residue is one that is compatible with the associated NES class pattern. (D) Sd amino acids 330-347 (which includes two amino acids N-terminal to V332 and two amino acids C-terminal to K345) were fused N-terminal to eGFP and assayed for spatial distribution. A representative cell showing nuclear exclusion of the fusion protein is shown. Figure 2.4A-G for details. (E) Nuclear fraction of eGFP tagged constructs in LB treated and untreated S2 cells. Isoforms of Sd which contained the NES (NES, Sd ΔNLS, Sd Δ348-440 and Sd Δ392-440), had an increased nuclear fraction in LB treated cells, relative to untreated cells. eGFP alone, and Sd fragments in which the NES was deleted (Sd Δ301-355 and Sd Δ294-440), did not show a significant increase in nuclear localization after LB treatment. N is >15 for all conditions. * indicates a significant difference at P < 0.001. Error bars are the standard error of the mean.

A

VVLVCSTIVCSFGKQV
332 ——————————— 347

N · TEA/NLS · VID · C
1 · 88 · 161 · 220 · 344 · 440
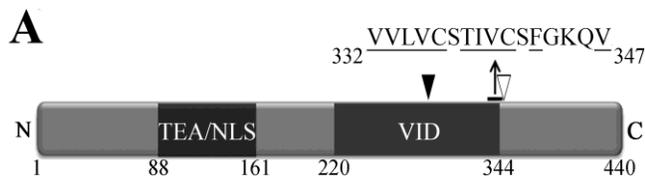
B

Consensus	M	V	I	T	C	S	T	K	V	C	S	F	G	K	Q	V
Sd (*D. melanogaster*)	V	V	L	V	C	S	T	I	V	C	S	F	G	K	Q	V
Undescribed (*A. gambiae*)	M	V	I	T	C	S	T	K	V	C	S	F	G	K	Q	V
TEF-1 (*H. sapiens*)	M	T	V	T	C	S	T	K	V	C	S	F	G	K	Q	V
TEAD-1 (*M. musculus*)	M	T	V	T	C	S	T	K	V	C	S	F	G	K	Q	V
TEF-1 (*X. laevis*)	M	T	I	T	C	S	T	K	V	C	S	F	G	K	Q	V
Tead-1 (*D. rerio*)	M	T	I	T	C	S	T	K	V	C	S	F	G	K	Q	V
TEAD-4 (*G. gallus*)	M	V	I	T	C	S	T	K	V	C	S	F	G	K	Q	V
EGL-44 (*C. elegans*)	F	Q	L	K	V	S	T	M	A	C	S	F	G	N	Q	A
Undescribed (*M. brevicollis*)	M	V	V	E	I	S	M	C	A	I	Q	L	G	K	P	V
AbaA (*A. nidulans*)	I	L	L	S	K	P	T	S	N	L	Y	Q	A	P	P	Q
TEC-1p (*S. cerevisiae*)	V	V	P	R	S	A	T	V	T	Q	L	Q	S	R	P	V

C

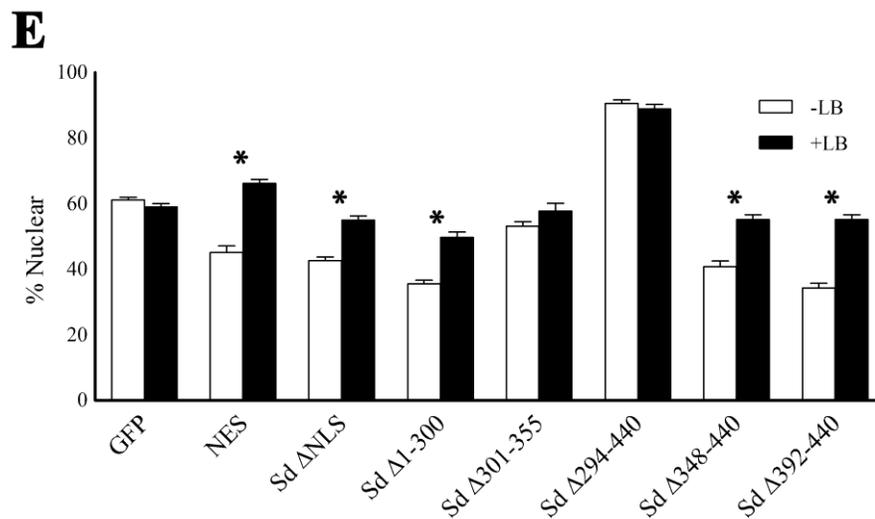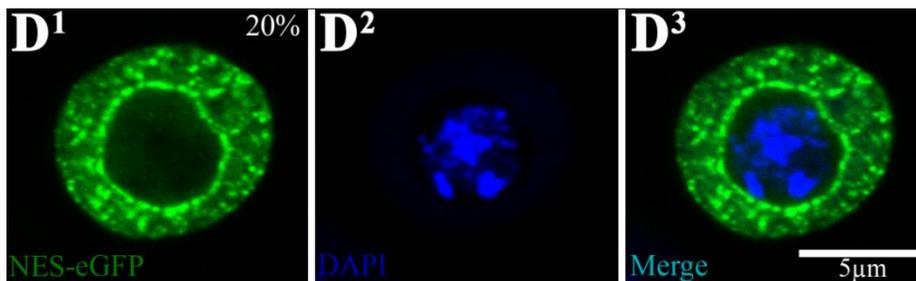| NES Class	Consensus Sequence	Sd Sequence	Equivalent TEAD Consensus Sequence
1a	Φ-3X-Φ-2X-Φ-X-Φ	**V**VLV**C**ST**I**VCS**F**GKQV	N/A
		VV**L**VCS**T**IV**C**S**F**GKQV	M**I**TCS**T**KV**C**S**F**GKQV
1b	Φ-2X-Φ-2X-Φ-X-Φ	**V**VL**V**CS**T**IVCSFGKQV	**M**VI**T**CS**T**KV**C**SFGKQV
		V**V**LV**C**ST**I**V**C**SFGKQV	MVI**T**CS**T**KV**C**S**F**GKQV
1d	Φ-2X-Φ-3X-Φ-X-Φ	**V**VL**V**CS**T**IV**C**SFGKQV	N/A
2	Φ-X-Φ-2X-Φ-X-Φ	V**V**L**V**CS**T**IV**C**SFGKQV	M**V**I**T**CS**T**K**V**CSFGKQV
		VV**L**V**C**ST**I**V**C**SFGKQV	N/A

E

84

Figure 2.8. During wing development, 3xFLAG-PMSD and 3xFLAG-SD mNLS^{N+C} act as dominant negative forms of Sd. (A and B) Localization of the indicated mRFP reporter tagged proteins in transiently transfected in S2 cells with DAPI stained nuclei and visualized via confocal microscopy. See Figure 2.4A-G for details. (A) Sd-mRFP expression. Sd strongly localizes an mRFP tag to the nucleus. (B) PMSD-mRFP expression. Sd tagged with a N-terminal palmitoylation/myristoylation sequence (PMSD) and C-terminal mRFP tag shows strong localization to the cytoplasmic membrane of S2 cells. (C-E) Light micrographs of flies with the indicated genotypes. (C) Wildtype Oregon-R (Ore^R) fly. (D-E) Males containing either *UAS-3xFLAG-PMSD* or *UAS-3xFLAG-SD mNLS^{N+C}* (see Figure 2.5C) inserted on the 2nd chromosome and balanced over *CyO* were crossed (two independent lines/insert) to virgin females homozygous for *sd*-GAL4 and the resultant progeny were scored. Insets are magnified views of the wing tissue. Scale bars are 1mm (D-E) or 0.1mm (D and E insets). Arrows indicate the wing, while arrowheads indicate the haltere. (D) Female fly containing *UAS-3xFLAG-PMSD* under the control of *sd*-GAL4. Almost no wing or haltere tissue is present. (E) Female fly containing *UAS-3xFLAG-mNLS^{N+C}* under the control of *sd*-GAL4. Again, virtually no wing or haltere tissue is present.

SD-mRFP

DAPI

Merge

PMSD-mRFP

DAPI

Merge

Ore^R

sd-GAL4/X; UAS-*3xFLAG-PMSD*/+

sd-GAL4/X; UAS-*3xFLAG-SDmNLS^{N+C}*/+
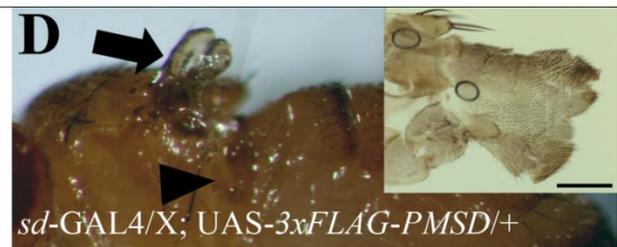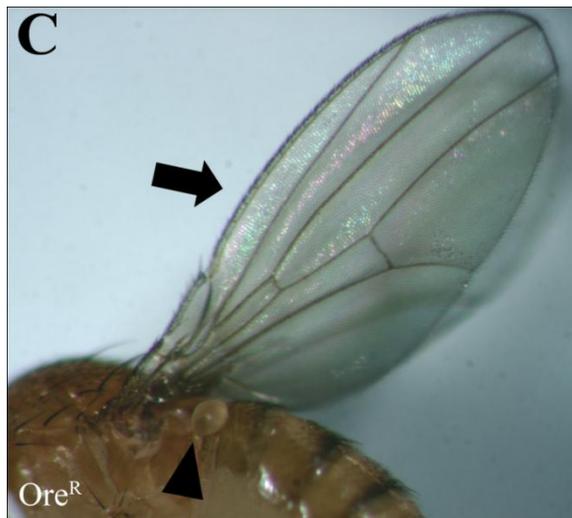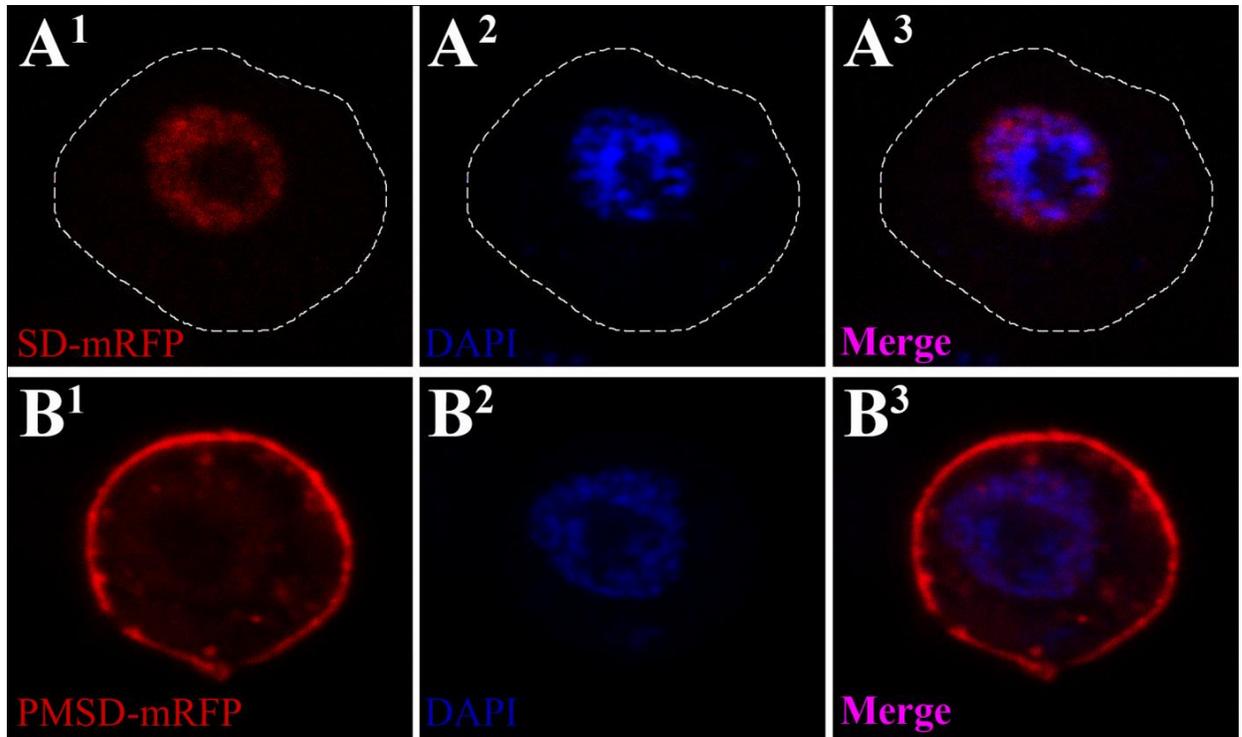
Figure 2.9. Sd shows two isoforms when analyzed by Western blot – one which is sensitive to λ

phosphatase. 3xFLAG-Sd was expressed *ex vivo* and purified using mouse αFLAG beads. The purified

protein was then treated either with λ phosphatase in buffer, or with buffer alone and then both

samples were analyzed by Western blot on a low-bis acrylamids gel. In the absence of phosphatase two

bands are seen, one running slightly higher than the other (closed arrowhead). In the presence of

phosphatase, the upper band is absent (open arrowhead). Detection was with αFLAG. PP'tase = λ

phosphatase.

WB:αFLAG

3xFLAG-Sd
-PP'tase

3xFLAG-Sd
+ PP'tase
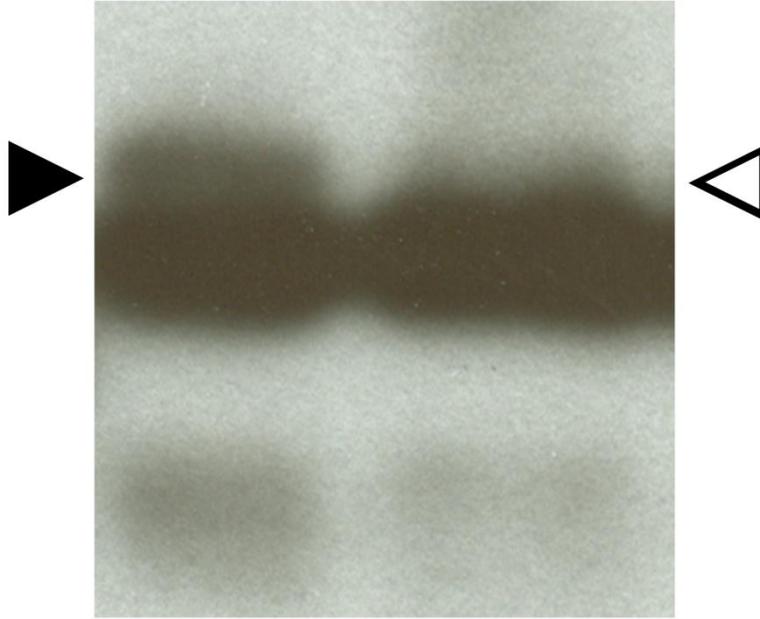
References

Abramoff, M., Magelhaes, P., and Ram, S. (2004). Image Processing with ImageJ. Biophotonics International *11*, 36-42.

Adam, E. J., and Adam, S. A. (1994). Identification of cytosolic factors required for nuclear location sequence-mediated binding to the nuclear envelope. J. Cell Biol *125*, 547-555.

Bischoff, F. R., Klebe, C., Kretschmer, J., Wittinghofer, A., and Ponstingl, H. (1994). RanGAP1 induces GTPase activity of nuclear Ras-related Ran. Proc. Natl. Acad. Sci. U.S.A *91*, 2587-2591.

Blom, N., Gammeltoft, S., and Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol *294*, 1351-1362.

Bogerd, H. P., Echarri, A., Ross, T. M., and Cullen, B. R. (1998). Inhibition of human immunodeficiency virus Rev and human T-cell leukemia virus Rex function, but not Mason-Pfizer monkey virus constitutive transport element activity, by a mutant human nucleoporin targeted to Crm1. J. Virol *72*, 8627-8635.

Bogerd, H., Fridell, R., Benson, R., Hua, J., and Cullen, B. (1996). Protein sequence requirements for function of the human T-cell leukemia virus type 1 Rex nuclear export signal delineated by a novel *in vivo* randomization-selection assay. Mol. Cell. Biol. *16*, 4207-4214.

Borghi, S., Molinari, S., Razzini, G., Parise, F., Battini, R., and Ferrari, S. (2001). The nuclear localization domain of the MEF2 family of transcription factors shows member-specific features and mediates the nuclear import of histone deacetylase 4. J. Cell. Sci *114*, 4477-4483.

Campbell, S., Inamdar, M., Rodrigues, V., Raghavan, V., Palazzolo, M., and Chovnick, A. (1992). The scalloped gene encodes a novel, evolutionarily conserved transcription factor required for sensory organ differentiation in *Drosophila*. Genes & development *6*, 367-379.

Chan, W. M., Shaw, P. C., and Chan, H. Y. E. (2007). A green fluorescent protein-based reporter for protein nuclear import studies in *Drosophila* cells. Fly *1*, 340-342.

Charmandari, E., Kino, T., and Chrousos, G. P. (2004). Glucocorticoids and their actions: an introduction. Ann. N. Y. Acad. Sci *1024*, 1-8.

Chelsky, D., Ralph, R., and Jonak, G. (1989). Sequence requirements for synthetic peptide-mediated translocation to the nucleus. Mol. Cell. Biol *9*, 2487-2492.

Chintapalli, V. R., Wang, J., and Dow, J. A. T. (2007). Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. Nat Genet *39*, 715-720.

Chow, L., Berube, J., Fromont, A., and Bell, J. B. (2004). Ability of scalloped deletion constructs to rescue sd mutant wing phenotypes in *Drosophila melanogaster*. Genome *47*, 849-859.

Chrousos, G. P., Charmandari, E., and Kino, T. (2004). Glucocorticoid action networks--an introduction to systems biology. J. Clin. Endocrinol. Metab *89*, 563-564.

la Cour, T., Gupta, R., Rapacki, K., Skriver, K., Poulsen, F. M., and Brunak, S. (2003). NESbase version 1.0: a database of nuclear export signals. Nucleic Acids Res *31*, 393-396.

Cressman, D. E., O'Connor, W. J., Greer, S. F., Zhu, X.-S., and Ting, J. P.-Y. (2001). Mechanisms of Nuclear Import and Export That Control the Subcellular Localization of Class II Transactivator. J Immunol *167*, 3626-3634.

Cronshaw, J. M., Krutchinsky, A. N., Zhang, W., Chait, B. T., and Matunis, M. J. (2002). Proteomic analysis of the mammalian nuclear pore complex. The Journal of Cell Biology *158*, 915 -927.

D'Angelo, M. A., and Hetzer, M. W. (2006). The role of the nuclear envelope in cellular organization. Cell. Mol. Life Sci *63*, 316-332.

Deng, H., Hughes, S. C., Bell, J. B., and Simmonds, A. J. (2009). Alternative requirements for Vestigial, Scalloped, and Dmef2 during muscle differentiation in *Drosophila melanogaster*. Mol. Biol. Cell *20*, 256-269.

Devos, D., Dokudovskaya, S., Williams, R., Alber, F., Eswar, N., Chait, B. T., Rout, M. P., and Sali, A. (2006). Simple fold composition and modular architecture of the nuclear pore complex. Proceedings of the National Academy of Sciences of the United States of America *103*, 2172 -2177.

Dingwall, C., and Laskey, R. A. (1991). Nuclear targeting sequences--a consensus? Trends Biochem. Sci *16*, 478-481.

Fagotto, F., Glück, U., and Gumbiner, B. M. (1998). Nuclear localization signal-independent and importin/karyopherin-independent nuclear import of beta-catenin. Curr. Biol *8*, 181-190.

Fisher, C. L., and Pei, G. K. (1997). Modification of a PCR-based site-directed mutagenesis method. BioTechniques *23*, 570-571, 574.

Floer, M., and Blobel, G. (1996). The nuclear transport factor karyopherin beta binds stoichiometrically to Ran-GTP and inhibits the Ran GTPase activating protein. J. Biol. Chem *271*, 5313-5316.

Franke, W. W., and Scheer, U. (1974). Pathways of nucleocytoplasmic translocation of ribonucleoproteins. Symp. Soc. Exp. Biol, 249-282.

Görlich, D., and Kutay, U. (1999). Transport between the cell nucleus and the cytoplasm. Annu. Rev. Cell Dev. Biol *15*, 607-660..

Görlich, D., Prehn, S., Laskey, R. A., and Hartmann, E. (1994). Isolation of a protein that is essential for the first step of nuclear protein import. Cell *79*, 767-778.

Görlich, D., Vogel, F., Mills, A. D., Hartmann, E., and Laskey, R. A. (1995). Distinct functions for the two importin subunits in nuclear protein import. Nature *377*, 246-248.

Garg, A., Srivastava, A., Davis, M. M., O'Keefe, S. L., Chow, L., and Bell, J. B. (2007). Antagonizing Scalloped With a Novel Vestigial Construct Reveals an Important Role for Scalloped in *Drosophila melanogaster* Leg, Eye and Optic Lobe Development. Genetics *175*, 659-669.

Gorjánácz, M., Adám, G., Török, I., Mechler, B. M., Szlanka, T., and Kiss, I. (2002). Importin-alpha 2 is critically required for the assembly of ring canals during *Drosophila* oogenesis. Dev. Biol *251*, 271-282.

Goulev, Y., Fauny, J. D., Gonzalez-Marti, B., Flagiello, D., Silber, J., and Zider, A. (2008). SCALLOPED Interacts with YORKIE, the Nuclear Effector of the Hippo Tumor-Suppressor Pathway in *Drosophila*. Current Biology *18*, 435-441.

Halder, G., Polaczyk, P., Kraus, M. E., Hudson, A., Kim, J., Laughon, A., and Carroll, S. (1998). The Vestigial and Scalloped proteins act together to directly regulate wing-specific gene expression in *Drosophila*. Genes & development *12*, 3900-3909.

Jäkel, S., Albig, W., Kutay, U., Bischoff, F. R., Schwamborn, K., Doenecke, D., and Görlich, D. (1999). The importin beta/importin 7 heterodimer is a functional nuclear import receptor for histone H1. EMBO J *18*, 2411-2423.

Küssel, P., and Frasch, M. (1995). Pendulin, a *Drosophila* protein with cell cycle-dependent nuclear localization, is required for normal cell proliferation. J. Cell Biol *129*, 1491-1507.

Kalderon, D., Roberts, B. L., Richardson, W. D., and Smith, A. E. (1984). A short amino acid sequence able to specify nuclear location. Cell *39*, 499-509.

Keminer, O., and Peters, R. (1999). Permeability of single nuclear pores. Biophys. J *77*, 217-228.

Knapp, A. A., McManus, P. M., Bockstall, K., and Moroianu, J. (2009). Identification of the nuclear localization and export signals of high risk HPV16 E7 oncoprotein. Virology *383*, 60-68.

Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M., and Yanagawa, H. (2009). Six Classes of Nuclear Localization Signals Specific to Different Binding Grooves of Importin α. Journal of Biological Chemistry *284*, 478 -485.

Kosugi, S., Hasebe, M., Tomita, M., and Yanagawa, H. (2008). Nuclear export signal consensus sequences defined using a localization-based yeast selection system. Traffic *9*, 2053-2062.

Kudo, N., Wolff, B., Sekimoto, T., Schreiner, E. P., Yoneda, Y., Yanagida, M., Horinouchi, S., and Yoshida, M. (1998). Leptomycin B inhibition of signal-mediated nuclear export by direct binding to CRM1. Exp. Cell Res *242*, 540-547.

Lanford, R. E., and Butel, J. S. (1984). Construction and characterization of an SV40 mutant defective in nuclear transport of T antigen. Cell *37*, 801-813.

Lange, A., McLane, L. M., Mills, R. E., Devine, S. E., and Corbett, A. H. (2010). Expanding the definition of the classical bipartite nuclear localization signal. Traffic *11*, 311-323.

Lindsay, M. E., Holaska, J. M., Welch, K., Paschal, B. M., and Macara, I. G. (2001). Ran-binding protein 3 is a cofactor for Crm1-mediated nuclear protein export. J. Cell Biol *153*, 1391-1402.

Lonhienne, T. G., Forwood, J. K., Marfori, M., Robin, G., Kobe, B., and Carroll, B. J. (2009). Importin-beta is a GDP-to-GTP exchange factor of Ran: implications for the mechanism of nuclear import. J. Biol. Chem *284*, 22549-22558.

Lounsbury, K. M., and Macara, I. G. (1997a). Ran-binding protein 1 (RanBP1) forms a ternary complex with Ran and karyopherin beta and reduces Ran GTPase-activating protein (RanGAP) inhibition by karyopherin beta. J. Biol. Chem *272*, 551-555.

Lounsbury, K. M., and Macara, I. G. (1997b). Ran-binding protein 1 (RanBP1) forms a ternary complex with Ran and karyopherin beta and reduces Ran GTPase-activating protein (RanGAP) inhibition by karyopherin beta. J. Biol. Chem *272*, 551-555.

Máthé, E., Bates, H., Huikeshoven, H., Deák, P., Glover, D. M., and Cotterill, S. (2000). Importin-alpha3 is required at multiple stages of *Drosophila* development and has a role in the completion of oogenesis. Dev. Biol *223*, 307-322.

Macara, I. G. (2001). Transport into and out of the Nucleus. Microbiol. Mol. Biol. Rev. *65*, 570-594.

Markossian, K. A., and Kurganov, B. I. (2004). Protein folding, misfolding, and aggregation. Formation of inclusion bodies and aggresomes. Biochemistry Mosc *69*, 971-984.

Mason, D. A., Fleming, R. J., and Goldfarb, D. S. (2002). *Drosophila melanogaster* Importin {alpha}1 and {alpha}3 Can Replace Importin {alpha}2 During Spermatogenesis but Not Oogenesis. Genetics *161*, 157-170.

McCabe, J. B., and Berthiaume, L. G. (1999). Functional roles for fatty acylated amino-terminal domains in subcellular localization. Molecular biology of the cell *10*, 3771-3786.

Meyer, B., Meinkoth, J., and Malim, M. (1996). Nuclear transport of human immunodeficiency virus type 1, visna virus, and equine infectious anemia virus Rev proteins: identification of a family of transferable nuclear export signals. J. Virol. *70*, 2350-2359.

Moore, M. S., and Blobel, G. (1994). Purification of a Ran-interacting protein that is required for protein import into the nucleus. Proc. Natl. Acad. Sci. U.S.A *91*, 10212-10216.

Nehrbass, U., and Blobel, G. (1996). Role of the nuclear transport factor p10 in nuclear import. Science *272*, 120-122.

North, A. J. (2006). Seeing is believing? A beginners' guide to practical pitfalls in image acquisition. J. Cell Biol *172*, 9-18.

Paine, P. L., Moore, L. C., and Horowitz, S. B. (1975). Nuclear envelope permeability. Nature *254*, 109-114.

Panté, N., and Aebi, U. (1996). Sequential binding of import ligands to distinct nucleopore regions during their nuclear import. Science *273*, 1729-1732.

Panté, N., and Kann, M. (2002). Nuclear pore complex is able to transport macromolecules with diameters of about 39 nm. Mol. Biol. Cell *13*, 425-434.

Ratan, R., Mason, D. A., Sinnot, B., Goldfarb, D. S., and Fleming, R. J. (2008). *Drosophila* importin alpha1 performs paralog-specific functions essential for gametogenesis. Genetics *178*, 839-850.

Rexach, M., and Blobel, G. (1995). Protein import into nuclei: association and dissociation reactions involving transport substrate, transport factors, and nucleoporins. Cell *83*, 683-692. Robbins, J. (1991). Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: Identification of a class of bipartite nuclear targeting sequence. Cell *64*, 615-623. Available at:

Rout, M. P., Aitchison, J. D., Suprapto, A., Hjertaas, K., Zhao, Y., and Chait, B. T. (2000). The yeast nuclear pore complex: composition, architecture, and transport mechanism. J. Cell Biol *148*, 635-651.

Rout, M. P., and Wente, S. R. (1994). Pores for thought: nuclear pore complex proteins. Trends Cell Biol *4*, 357-365.

Rubin, G. M., and Spradling, A. C. (1983). Vectors for P element-mediated gene transfer in *Drosophila*. Nucl. Acids Res. *11*, 6341-6351.

Shah, S., Tugendreich, S., and Forbes, D. (1998). Major binding sites for the nuclear import receptor are the internal nucleoporin Nup153 and the adjacent nuclear filament protein Tpr. J. Cell Biol *141*, 31-49.

Simmonds, A. J., Liu, X., Soanes, K. H., Krause, H. M., Irvine, K. D., and Bell, J. B. (1998). Molecular interactions between Vestigial and Scalloped promote wing formation in *Drosophila*. Genes & development *12*, 3815-3820.

Sorokin, A. V., Kim, E. R., and Ovchinnikov, L. P. (2007). Nucleocytoplasmic transport of proteins. Biochemistry Moscow *72*, 1439-1457.

Srivastava, A., MacKay, J. O., and Bell, J. B. (2002). A Vestigial:Scalloped TEA domain chimera rescues the wing phenotype of a scalloped mutation in *Drosophila melanogaster*. Genesis (New York, N.Y. : 2000) *33*, 40-47.

Srivastava, A., Simmonds, A. J., Garg, A., Fossheim, L., Campbell, S. D., and Bell, J. B. (2004). Molecular and functional analysis of scalloped recessive lethal alleles in *Drosophila melanogaster*. Genetics *166*, 1833-1843.

Suntharalingam, M., and Wente, S. R. (2003). Peering through the Pore: Nuclear Pore Complex Structure, Assembly, and Function. Developmental Cell *4*, 775-789. Available at: [Accessed 22:06:29].

Török, I., Strand, D., Schmitt, R., Tick, G., Török, T., Kiss, I., and Mechler, B. M. (1995). The overgrown hematopoietic organs-31 tumor suppressor gene of *Drosophila* encodes an Importin-like protein accumulating in the nucleus at the onset of mitosis. J. Cell Biol *129*, 1473-1489.

Tejomurtula, J., Lee, K.-B., Tripurani, S. K., Smith, G. W., and Yao, J. (2009). Role of importin alpha8, a new member of the importin alpha family of nuclear transport proteins, in early embryonic development in cattle. Biol. Reprod *81*, 333-342.

Tran, E. J., and Wente, S. R. (2006). Dynamic Nuclear Pore Complexes: Life on the Edge. Cell *125*, 1041-1053.

Vasu, S. K., and Forbes, D. J. (2001). Nuclear pores and nuclear assembly. Curr. Opin. Cell Biol *13*, 363-375.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-1191.

Weber, F., Kochs, G., Gruber, S., and Haller, O. (1998). A Classical Bipartite Nuclear Localization Signal on Thogoto and Influenza A Virus Nucleoproteins. Virology *250*, 9-18.

Weis, K. (2003). Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. Cell *112*, 441-451.

Wu, J., Matunis, M. J., Kraemer, D., Blobel, G., and Coutavas, E. (1995). Nup358, a cytoplasmically exposed nucleoporin with peptide repeats, Ran-GTP binding sites, zinc fingers, a cyclophilin A homologous domain, and a leucine-rich region. J. Biol. Chem *270*, 14209-14213.

Ylikomi, T., Bocquel, M. T., Berry, M., Gronemeyer, H., and Chambon, P. (1992). Cooperation of proto-signals for nuclear accumulation of estrogen and progesterone receptors. EMBO J *11*, 3681-3694.

Yokoyama, N., Hayashi, N., Seki, T., Panté, N., Ohba, T., Nishii, K., Kuma, K., Hayashida, T., Miyata, T., and Aebi, U. (1995). A giant nucleopore protein that binds Ran/TC4. Nature *376*, 184-188.

Zasloff, M. (1983). tRNA transport from the nucleus in a eukaryotic cell: carrier-mediated translocation process. Proceedings of the National Academy of Sciences of the United States of America *80*, 6436 - 6440.

Zhang, L., Ren, F., Zhang, Q., Chen, Y., Wang, B., and Jiang, J. (2008). The TEAD/TEF Family of Transcription Factor Scalloped Mediates Hippo Signaling in Organ Size Control. Developmental Cell *14*, 377-387.

Zhang, M., Wang, M., Tan, X., Li, T.-F., Zhang, Y. E., and Chen, D. (2010). Smad3 prevents beta-catenin degradation and facilitates beta-catenin nuclear translocation in chondrocytes. J. Biol. Chem *285*, 8703-8710.

Zheng, C., Brownlie, R., Babiuk, L. A., and van Drunen Littel-van den Hurk, S. (2005). Characterization of the Nuclear Localization and Nuclear Export Signals of Bovine Herpesvirus 1 VP22. J. Virol. *79*, 11864-11872.

Chapter Three: Identifying a novel binding partner of Scalloped[2]

Sd interacting proteins

As noted in the general introduction (see Chapter 1, page 2), Sd requires a TIF in order to function properly as a transcription factor. While several TIFs (e.g. Vg, Yki, dMEF2) have been identified, there are Sd expressing tissues in which no TIF has yet been identified, such as the leg discs and optic lobes. Additionally, two TIFs, Vg and Yki, are required for wing development while two others, Vg and dMEF2, are required for muscle development. As such, a given tissue may require multiple TIFs and thus even tissues where TIFs have been identified may have yet undiscovered binding partners of Sd. In vertebrates, several TIFs are known. Probably the best characterized are the Vgl family – Vgl-1 (or Tondu; TDU) and Vgl-2-4. Vgl-1, 2 and 3 each contains a single well-conserved Sd interacting domain (called a TDU domain), while Vgl-4 contains two tandem domains (Figure 3.1) (Vaudin et al., 1999; Maeda et al., 2002; Chen et al., 2004). Interestingly, when *hvgl-4* was identified, a *Drosophila* gene with unknown function – *cg10741* – was also identified as potentially coding for a protein with two TDU domains (Chen et al., 2004). While hVgl-4 and CG10741 (hereafter referred to as *Drosophila* Vgl-4; dVgl-4) show very little similarity outside the TDU domains, each of the putative TDU domains in dVgl-4 have a corresponding TDU domain of high similarity in hVgl-4 (i.e. TDU1 of dVgl-4 and hVgl-4 are very similar and TDU2 of dVgl-4 and hVgl-4 are also very similar; Figure 3.1).

---

[2] All experiments in this chapter were designed, conducted by and analyzed by A. C. Magico, except for the GST pull-down experiment which was designed and conducted by A. J. Simmonds.

Vgl-4

To date, only two studies have presented data pertaining to Vgl-4in vertebrates. The first worked on

hVgl-4 (Chen et al., 2004). In that study, they found that hVgl-4 is expressed primarily in heart, brain and

kidney tissue culture cells and is able to interact with both hTEF-1 and hMEF-2. Interestingly, the two

TDU domains of hVgl-4 were shown to interact differentially with these two cofactors. Specifically,

TDU1 mediates the interaction with hTEF-1, while TDU2 mediates the interaction with hMEF-2. They

suggested that hVgl-4 could therefore act as a bridge between the two proteins. Perhaps the most

interesting discovery was that hVgl-4 interfered with hTEF-1 mediated transcriptional activation,

suggesting that hVgl-4 might act as a negative regulator of hTEF-1 mediated transcription. Finally, they

found that the protein is able to shuttle between the nucleus and cytoplasm and identified a putative

NES signal in the protein (which is not present in dVgl-4).

The second study of *vgl-4* was done in zebrafish (Faucheux et al., 2010). In this study they

identified zebrafish homologues for all four *vgl* genes, and determined their temporal and spatial

expression patterns. Zebrafish *vgl-4* is inherited maternally in the zygote followed by ubiquitous

expression throughout the developing embryo.

d*vgl-4*

The d*vgl-4* gene lies on chromosome 3L, between 70B2 and 70B3. There are two predicted transcripts –

RA and RB – which code for predicted proteins of sizes 535 (dVgl-4 PA) and 332 amino acids (dVgl-4 PB),

respectively. These two putative proteins have differing N-termini, but both contain the predicted TDU

domains mentioned above, although they are not in close proximity as they are in hVgl-4 (Figure 3.2).

Outside of the TDU domains, no predicted functional domains have yet been detected. Only one

mutant phenotype for d*vgl-4* is known: discoloration of the notum has been observed when double stranded RNA against d*vgl-4* is generated under control of *pannier* (*pnr*)-GAL4, in order to generate an RNA interference (RNAi) response against the endogenous transcript (Mummery-Widmer et al., 2009). RNAi is a method to specifically reduce the expression of an endogenous gene in eukaryotic cells (Liu and Paroo, 2010; Carthew and Sontheimer, 2009). In *Drosophila*, RNAi relies on the exogenous expression of double-stranded mRNA (typically generated by expressing an inverted repeat using sequences specific to a gene of interest). The double-stranded mRNA is processed into short 21-23 nt fragments called small interfering RNA (siRNA) by the enzyme Dicer-2 (Bernstein et al., 2001). Dcr-2 along with these fragments is then assembled into a superstructure known as the RNA-induced silencing complex (RISC; Pham et al., 2004). This enzyme then guides these fragments to complementary mRNA sequences (which would thus be present in the transcript of the gene being targeted for silencing) and these mRNA molecules are degraded, thus reducing or abolishing gene expression prior to translation (Hammond et al., 2000; Schwarz et al., 2004).

In the remainder of this chapter, I show that dVgl-4 contains two probable TDU domains, based on sequence conservation between dVgl-4 and hVgl-4. Consistent with this idea, *in vitro* and *ex vivo* data is presented which indicates that dVgl-4 is able to interact with Sd. An allele of dVgl-4 is also analyzed and shown to be hypomorphic for one of two predicted isoforms of d*vgl-4*, raising the possibility that the two isoforms are controlled by independent promoters. Finally, attempts to generate further alleles of d*vgl-4* are detailed.

Results

Investigating d*vgl-4 expression*

Genome-wide screens of gene expression have shown weak d*vgl-4* expression in the PNS of stage 13-16 embryos (Tomancak et al., 2002). This expression partially overlaps with Sd expression during that time (Campbell et al., 1992). In order to examine protein expression, both in the embryos and the larvae, a custom peptide antibody directed against amino acids TKWRRERRQRSAGY (Figure 3.1) was manufactured. To assay the ability of this antibody to detect dVgl-4, 3xFLAG-dVgl-4 was expressed in cell culture. The cells expressing the fusion construct were then lysed and the lysate was analyzed by western blot using either anti-FLAG or anti-dVgl-4. In both cases, a protein running at approximately 50 KDa was detected (Figure 3.3A), although the anti-dVgl-4 antibody also detected a band at ~43 KDa. Next, a transgenic line containing *UAS*-d*vgl-4* was crossed to a line containing a *patch* (*ptc*)-GAL4 driver, which drives GAL4 expression along the A/P boundary of the wing disc. Immunostaining of the resultant third instar larval wing discs with anti-dVgl-4 showed a *ptc* pattern of expression, as expected (Figure 3.3B). However, while the antibody could reliably detect ectopically expressed dVgl-4, no reliable and specific signal was observed in any late stage embryos or in 3rd instar larvae although earlier stages were not tested.

Ex vivo and in vitro analysis of d*Vgl-4 and Sd interactions*

Given the presence of the tandem TDU domains, it is reasonable to expect that dVgl-4 is capable of interacting with Sd. To test this hypothesis, a Myc-tagged Sd was co-expressed with 3xFLAG-tagged dVgl-4 in S2 cell culture and vice versa. The cells were lysed and the 3xFLAG-dVgl-4 or 3xFLAG-Sd proteins were immunoprecipitated using anti-FLAG beads. The precipitate was then analyzed for the

99

presence of Myc-Sd or Myc-dVgl-4 via Western blot. Consistent with the ability of the two proteins to

interact, the Myc-tagged proteins were detected in the immunoprecipitate of the 3xFLAG-tagged

proteins in both combinations (Figure 3.4A). Interestingly, Myc-dVgl-4 is detected as a doublet,

although 3xFLAG dVgl-4 is not. Further evidence of this interaction was provided by Andrew Simmonds

who, in collaboration, demonstrated that *in vitro* translated GST-tagged Sd is able to pull down

radiolabelled dVgl-4 (Figure 3.4B). There is also evidence (albeit weak) of a genetic interaction, since

using the *UAS*/GAL4 system (Brand and Perrimon, 1993) a *UAS-3xFLAG-dvgl-4* transgene can be driven in

the wing disc by *sd*-GAL4 and this results in almost complete abolishment of the adult wing (Figure 3.9A

and see below). Analysis of the expression of eGFP tagged dVgl-4 in S2 cells also revealed that the

protein localizes to the nucleus (Figure 3.5).

*Characterization of a P-element insertion into the promoter region of d*vgl-4

To date, there are no described alleles of d*vgl-4*. However, there are several p-element

insertions both 5' and 3' of the open reading frame. One of those insertions – pBac{RB}CG10741^{e01789},

generated as part of the Exelixis collection (Thibault et al., 2004) – consists of a piggyBac transposable

element inserted 124 bp upstream of the predicted *dvgl-4* RB transcriptional start site. This site was

confirmed by sequencing from the 3' end of the p-element insertion. Flies homozygous for this insertion

show a marked disruption in both the tergite bristles of the abdomen and the overall abdominal

pigmentation patterning (Figure 3.6A). Moreover, analysis by real-time quantitative polymerase chain

reaction (qPCR) demonstrated that the expression of the RB mRNA isoform is almost undetectable in

homozygous d*vgl-4*^{e01789} third instar larvae, but present in heterozygous animals of the same

developmental age; however, expression of the RA isoform is normal (Figure 3.6B). In order to verify

that the phenotype was specific for the insertion, homozygous d*vgl-4*^{e01789} flies were crossed to flies

carrying one of two deficiencies (Df(3L)Exel6119 and the much larger deficiency Df(3L)ED4502), both of

which uncover the d*vgl-4* locus. Unfortunately, d*vgl-4* [e01789] hemizygotes showed no phenotype,

although both deficiencies showed the expected larval lethal phenotype when transheterozygous with a

mutant allele of *starvin* (*stv*) which is close to the d*vgl-4* locus.

*d*vgl-4 *loss-of-function*

In an effort to generate loss-of-function data for d*vgl-4*, three approaches were used. The first

was to use RNAi (see above) against d*vgl-4*. In order to generate siRNA specific for dvgl-4, three lines

were used, one generated by the Vienna *Drosophila* RNAi Centre (VDRC; Dietzl et al., 2007), and two

additional lines were created by generating transgenic flies containing the vector pWIZ with an inverted

repeat of exon three of the d*vgl-4* open reading frame. In both cases, expression of the inverted repeat

of d*vgl-4* is under the control of the *UAS*-GAL4 system. A wide variety of GAL4 drivers were used to

express the RNAi (including *sd*-GAL4, *vg*-GAL4, *mef*-GAL4 and *pnr*-GAL4; see Table 3.1 for the complete

list), however no obvious phenotypes were observed for any driver, including *pnr*-GAL4, which has been

previously shown to give a notal phenotype (see introduction). However, it is important to note that the

VDRC RNAi transgene line used in this case did not generate a phenotype when driven with *pnr*-GAL4 in

the Knoblich lab screen mentioned in the introduction (Mummery-Widmer et al., 2009) either; rather

the phenotype they observed was only present when using a different VDRC strain that is no longer

available (from either the VDRC or the Knoblich lab itself).

The second approach to generating loss-of-function data was to attempt to generate a defined

deletion between Exelixis pBAC insertion sites (Figure 3.7A). This method involves recombination

between two FRT sites inserted on two homologous chromosomes, which generates a chromosome in

which the region between the FRT sites is deleted (Thibault et al., 2004; Parks et al., 2004). Two lines

containing FRT insertion sites near d*vgl-4* – pBac{RB}CG10741[e01789] (which is the source of the d*vgl-4*

[e01789] allele as mentioned) and pBac{WH}f01796 – were chosen and would generate an excision of d*vgl-4*

as well as four other genes: *spt20*, *Vacuolar protein sorting 36 (Vps36)*, *Liprin-β* and *cg10710*, using this

method (Figure 3.7B). This is an improvement over the smallest deficiency currently available which

uncovers d*vgl-4*, Df(3L)Exel6119, and eliminates 13 genes in addition to d*vgl-4*, including *starvin* (*stv*)

and *bruno-3* (*bru-3*) – which are larval and embryonic lethals, respectively, when homozygous null.

However, it was not possible to generate the smaller deletion, since the supposed pBac{WH}f01796

containing stock received from the VDRC had lost the insertion, and all other available combinations

delete either *stv* or *bru-3* or both.

The third approach attempted was to generate a null allele using ends-out recombination (Gong

and Golic, 2003; Figure 3.8). This method involves generating a transgenic animal containing a P-

element with a 5' region of homology to d*vgl-4* and a 3' region of homology to d*vg-4*, with a marker – in

this case *white*[+] (*w*[+]) – between them. A series of crosses are used to excise (via FLPAse treatment to

generated recombination between flanking FRT sites) and linearize (via I-*Sce*I treatment to cut at

flanking I-sites) the two regions of homology along with the marker, and test for the reintegration of this

genetic element into the genome. Ideally, the reintegration event is by homologous recombination

between the two regions of homology and their corresponding sequences at the d*vgl-4* locus, thus

replacing the coding sequence of the d*vgl-4* gene with the *w*[+] marker, and generating a null allele.

Following FLPase and I-*Sce*I treatment, females which had survived the heatshock used to drive the

FLPase and I-SceI were crossed to w males. Adult progeny of these crosses (representing the products

of roughly 10,000 gametes) were screened for the presence of red- or mosaic-eyed flies which were

isolated and the chromosome of P-element integration mapped. The original P-element insertion was in

chromosome II, thus those insertion events that mapped to chromosome III (of which there were almost

500) represented re-integration events (potentially, but not necessarily, specific to d*vgl-4*) of the excised

P-element and were selected for further analysis. Of those, over 120 were screened molecularly (via PCR) for the loss of the d*vgl-4* locus, and these and the remaining lines were also screened genetically by crossing the flies containing a balanced putative null allele to flies carrying a balanced deficiency uncovering the d*vgl-4* locus (Df(3L)Exel6119). The molecular screening did not detect any null d*vgl-4* alleles, and similarly, no fly transheterozygous for the excision/insertion event and the 3L deficiency showed reduced viability or obvious defects in the adult mutants.

*d*vgl *gain-of-function*

In addition to attempting to generate loss-of-function phenotypes, *UAS-3xFLAG-dvgl-4* was exogenously expressed with a variety of drivers in order to generate phenotypic data related to over-expression.   As mentioned previously, expression driven by *sd*-GAL4 leads to extreme loss of wing tissue (Figure 3.9A) and pupal lethality. Indeed, no *sd>3xFLAG-dvgl-4* males survive as adults. However, other than the wing defects, no phenotypes are observed in other tissues of the surviving females, including those of known *sd*-GAL4 expression, such as the eyes and legs. Among those animals that die as pupae, there is seldom any identifiable tissue present, although occasionally the heads and eyes are recognizable. Strong phenotypes are also seen when *pnr*-GAL4, *actin* (*act*)-GAL4 and *escargot* (*esg*)-GAL4 are used to express the transgene (Figures 3.9B and C and data not shown). The driver *pnr*-GAL4 was tested because, as noted previously, an RNAi phenotype has been observed using this driver. Expression using this driver was primarily pupal lethal. The few adult flies seen had severely disrupted notal tissue and disorganization of the notal bristles (Figure 3.9B) and similar phenotypes were seen in pupae that failed to eclose. On the other hand, *act*-GAL4 was used to test for phenotypes in a broader range of tissue. The result of using this driver was primarily pupal lethality, but the few surviving adults had wings that appeared as masses of bulbous tissue, and also had disrupted notal tissue (Figure 3.9C),

but no defects in other *sd*-expressing tissues (such as the eyes and legs) were observed. Finally, *esg*-

GAL4 was of interest because it drives expression in the histoblast nests, which develop into the adult

abdomen – the region disrupted in the d*vgl-4*^{e01789} mutant. When this driver was used, disruption of

abdominal patterning, similar to that in the d*vgl-4*^{e01789} mutant was seen (data not shown).

Discussion

The difficulties in generating loss-of-function data for d*vgl-4* have considerably hindered analysis

of its role in development and at this point it is not possible to say with certainty that this gene is

important for *Drosophila* development. Indeed, the inability of three separate lines (made with two

different transgenes) of *UAS*-RNAi argues that this gene is dispensable for development. However, a

single line of RNAi has been used to show a phenotype in notal development when driven by *pnr*-GAL4,

as mentioned (Mummery-Widmer et al., 2009). This raises the possibility that the position of insertion

of the RNAi transgene in the *Drosophila* genome is influencing the double-stranded mRNA expression

levels, and thus the efficiency of gene knockdown. However, given that only one out of four lines shows

this phenotype it is also possible that the phenotype is due to a secondary effect in which the insertion

of the transgenic element is interfering with the expression of an unrelated gene. Unfortunately, since

this line of transgenic animals is no longer available, it is impossible to determine definitively which

possibility is correct. That said, there are ways to potentially overcome the position effect of the

insertion site, if that is indeed the culprit behind the lack of phenotype. The most direct way would be

to generate new lines containing the RNAi transgene. This could be done by reintroducing the Δ2-3

transposase to mobilize the P-element and allow for a new reintegration event. This would likely

generate second-site mutations though (due to imprecise excision of the P-element) and so it would be

superior to generate new transgenic animals by re-injecting the RNAi containing P-element construct

into embryos containing the Δ2-3 transposase. Even better would be to remake the RNAi transgene in a

vector compatible with the φC31 integrase system (Groth et al., 2004; Bischof et al., 2007). This method

allows for targeted integration of a transgene of interest into one of several sites in the *Drosophila*

genome, and thus allows for more consistent transgene expression. Beyond testing novel insertion

sites, it might also be possible to increase the strength of the RNAi effect in three ways. The first would

be to introduce a *UAS*-GAL4 element into the final animal containing the *UAS*-RNAi and driver-GAL4.

This would cause a feed-back which would cause increased expression of the *dvgl-4* RNAi transgene.

Another method would be to add a *UAS-dcr*2 element into the final test animal, since it is thought that

by up-regulating the RNAi processing machinery in this fashion, the strength of the gene knock-down by

RNAi increases (Rousset et al., 2010). As a final option, the driver-GAL4/*UAS-dvgl-4* RNAi combination

could be assembled together in a fly carrying a heterozygous deficiency for *dvgl-4* (e.g. Df(3L)Exel6119).

In principal it would even be possible to combine all three approaches, although the genetics would be

complicated.

The fact that no phenotype was generated when the product of the ends-out recombination

was crossed to a deficiency uncovering d*vgl-4* could be due to multiple factors. First, it is possible that

no specific re-integration of the P-element occurred, and thus no null allele was generated. Typically, of

500 re-integration events, 0.1 – 5% of them (or 0.5 – 25 of the re-integrations) would be expected to be

specific (Huang et al., 2008). Thus, it is certainly possible that *dvgl-4* represents a locus resistant to the

homologous recombination necessary to generate the null allele, and thus falls on the low end of that

frequency. It is also possible that a null was generated, but did not show a phenotype when carried over

a deficiency. This could be due to functional redundancy, or perhaps be because the gene does not

actually code for a functional protein or perhaps the phenotype is only manifest under certain

conditions. For instance, a null allele of the gene *Drosophila Hormone receptor-like in 96* (*DHR96*)

generated by ends-in recombination showed no phenotype on standard media, but was homozygous lethal on cholesterol depleted media (King-Jones et al., 2006).

The abdominal phenotype of the allele d*vgl-4*[e01789] is most likely due to a second-site mutation on the third chromosome based on the inability of deficiencies for d*vgl-4* to recapitulate the phenotype when transheterozygous with d*vgl-4*[e01789]. That said, qPCR analysis of the P-element insertion did reveal something interesting about the d*vgl-4* locus. Namely, it appears the RA and RB mRNA isoforms are likely under the control of alternative promoters. Alternative promoter elements can drive the expression of different isoforms of the same gene in a spatially and temporally distinct fashion, in order to further refine the expression patterns of the differing isoforms (Ayoubi and Van De Ven, 1996). Evidence for this is given by the fact that insertion of the P-element into the 5' first exon of d*vgl-4* RB almost completely abolishes the expression of the RB isoform while the RA isoform is unaffected and, as shown in Figure 3.10, the first exon of the RB isoform is at the 5' end of the gene, but the first exon of the RA isoform is interior to the gene and not common to the RB isoform. Based on this, it seems unlikely that both isoforms are expressed under the control of the same promoter and then subject to differential splicing, but instead, the evidence suggests that each isoform is transcribed independently under the control of its own promoter element.

While no loss-of-function data was generated, there is evidence that d*vgl-4* is capable of coding for a functional protein product. Indeed, the RB mRNA isoform codes for a protein product which, consistent and as predicted by experiments with hVgl-4, is able to interact with Sd both *in vitro* and *ex vivo*. Thus it is likely that one or both the predicted TDU domains are indeed functional and are responsible for mediating this interaction. This is not surprising given that the TDU domains of dVgl-4 are very similar to both those of hVgl-4 and the TDU consensus, implying functional conservation. Furthermore, the protein is nuclear in S2 cells, compatible with a possible function as a modulator of Sd

transcriptional activity, an idea based on the assumption that dVgl-4 is orthologous to hVgl-4. Also

consistent with this is the fact that dVgl-4 expression in the wing disc using an *sd*-GAL4 driver leads to a

strong dominant-negative effect. Furthermore, most of the progeny of this cross die as pupae, which

could be due to interference with a critical developmental role of Sd – perhaps somatic or cardiac

muscle development, for instance. On the other hand, the surviving flies have no defects in their eyes,

heads or legs, which implies that dVgl-4 is unable to interfere with Sd in these tissues, at least when

driven with a *sd*-GAL4 driver. The fact that dVgl-4 expression driven in the notum via a *pnr*-GAL4 results

in extreme disruption of that tissue is also consistent with the notion that the gene codes for a product

that is functional, and moreover, correlates well with the observation made previously that expression

of an inverted repeat of part of d*vgl-4* causes notal discoloration. Thus, these are two pieces of

evidence that the protein is functional within the notum, though the inability to verify the RNAi data

weakens this argument. Abdominal defects are also seen when an *esg*-GAL4 driver is used to express

UAS-d*vgl-4* in the precursors of the adult abdomen – the histoblast nests. Finally, expression with *act*-

GAL4 recapitulates the phenotypes seen with the first two drivers (namely wing and notal defects as

well as pupal lethality), but curiously, does not show the same abdominal phenotype observed when the

esg-GAL4 driver is used. In total, it is clear that dVgl-4 is capable of functioning in a variety of

developmental contexts, and would be predicted to interact with Sd in order to mediate some

developmental process. That said, what that developmental process or processes might be is unknown

at this time and is subject to future investigation (discussed in Chapter Four).

Materials and Methods

Cell culture experiments- See Chapter 2, page 61.

Antibody generation and detection- The antibody was made commercially by Invitrogen as follows: the peptide sequence TKWRRERRQRSAGY was synthesized and used to make a Rabbit anti-dVgl-4 antibody. Bleeds from two rabbits were subsequently peptide purified prior to shipping the final purified antibodies. For Western blotS, the antibody was used at a 1:1000 dilution. Detection was with 1:50000 horse radish peroxidise (HRP) conjugated anti-rabbit secondary antibody (Invitrogen) and the SuperSignal Substrate Western Blotting kit (Pierce).

Construct design and cloning- Cloning into pHFW (N-terminal 3xFLAG), pTFW (N-terminal 3xFLAG) and pHMW (N-terminal Myc tag) was as per Chapter 2, page 60. To generate an RNAi construct, exon three of *dvgl-4* (which is present in both the RA and RB mRNA isoforms) was PCR amplified with *Xba*I sites appended to each end. *Xba*I generates sticky ends compatible with both *Avr*II and *Nhe*I. These are the sites found 5' and 3', respectively, to a *w* intron found within the pWIZ vector used to make the transgenic RNAi lines (Lee and Carthew, 2003). Cutting the vector with *Avr*II and *Nhe*I and the PCR product with *Xba*I allowed for the insertion of exon three of d*vgl-4 w* intron in the forward (5' to 3') direction, and downstream of the *w* intron in the inverse (3' to 5') direction. Thus, upon expression, the *w* intron is spliced out, and the inverted repeat of exon three forms a double-stranded RNA. In order to generate the construct containing the inverted repeat, it was necessary to use SURE2 cells (Stratagene) as a host strain, since it was extremely prone to rearrangements when expressed in standard DH5α cells. For ends-out recombination, the p[w35] vector was used. A region 5' to the transcriptional start site of d*vgl-4* ("A") was PCR amplified with *Bam*HI sites appended and cloned into the *Bam*HI sites of p[w35]. A region 3' to the transcriptional start site ("B") was likewise generated with *Sph*I sites appended and cloned into the corresponding sites in p[w35]+ d*vgl-4* A , giving p[w35]+d*vgl-4* AB.

Drosophila stocks- Df(3L)Exel6119, Df(3L)ED4502, pBac{RB}CG10741[e01789], *w1118*, *pnr*-gal4, *esg*-GAL4, were acquired from Bloomington (stock numbers 7598, 8097, 17985, 6326 and 3039, respectively).

pBac{WH}f01796 was from the Exelixis stock collection, *act*-GAL4 was a gift from J. Locke and *Mef2-*

GAL4 was a gift from H. Deng. For ends-out recombination *y*[1] *w*[*]; p{70FLP}11 P{70I-SceI}2B

noc[*Sco*]/*CyO*, *S*[2] and yw; Flp; *Sb*/TM6 *Tb* flies were used (both from Bloomington). The *UAS-3xFLAG-*

sd lines are described in Chapter two, page 21 and *UAS-3xFLAG*-d*vgl-4* was cloned in a similar fashion.

The injections done to generate the *UAS*-3xFLAG-dvgl-4 and p[w35] + d*vgl-4* AB transgenic lines (two

each) were performed commercially by Bestgene.   The two pWIZ + d*vgl-4* RNAi transgenic lines were

made as per Chapter Two, page 61.

Co-immunoprecipitation of dVgl-4 and Sd- 3xFLAG-Sd/dVgl-4 were expressed and purified as per Chapter

2, page 23. In this case, blotting of the Myc-labelled Sd/dVgl-4 proteins was with 1:500 mouse anti-Myc

(Invitrogen). Detection was with 1:50000 HRP conjugated anti-mouse secondary antibody (Invitrogen)

and the SuperSignal Substrate Western Blotting kit (Pierce).

GST pull-down experiment- This experiment was performed by Andrew Simmonds as follows: Sd-GST

fusion proteins were expressed in *E. coli* (Rosetta 2(DE3), Novagen) and purified according to the

manufacturer's directions (GE Biotech). Vg and dVgl-4 were S35 labeled *in vitro* using the TNT-coupled

in vitro transcription-translation system (Promega). For the *in vitro* binding assay, 3-6 μl of S^{35}-labeled

probe protein was incubated with 2 μg of immobilized GST fusion protein in 500 μl of buffer (20 mM Tris

pH 7.6, 100 mM NaCl, 0.5 mM EDTA, 10% glycerol, 1% Tween-20) containing 0.25% bovine serum

albumin (BSA) and protease inhibitor cocktail for 2h at 4°C. The beads were washed six times in 500 μl of

the same buffer and the bound proteins were resolved by SDS-PAGE and analyzed by autoradiography.

Ends-out recombination-   Virgin females containing p[w35] + d*vgl-4* AB inserted on the second

chromosome were crossed to *y*[1] *w*[*]; p{70FLP}11 P{70I-SceI}2B *noc*[*Sco*]/*CyO*, *S*[2]. Third instar larvae

from this cross were then heat-shocked @ 37°C for 30 min to generate the linear d*vgl-4* AB fragment.

Adult females were then crossed to yw; FLP; *Sb*/TM6 *Tb* males in order to facilitate the removal of any

remaining p[w35] +d*vgl-4* AB constructs which had not excised during the heatshock step. Red- or

mosaic-eyed progeny were then crossed to w^{1118} flies in order to map the w^+ insertion site. Finally, lines

which had a 3rd chromosome insertion site were balanced and crossed to Df(3L)Exel6119 or analyzed by

PCR (using primers interior and exterior to the predicted w^+ insertion site).

Figure 3.1. Tondu (TDU) domain consensus from the human and *Drosophila* Vg/Vgl family. A)

Alignment of the core TDU domains from the four human Vgl proteins and *Drosophila* Vg and Vgl-4.

Dark shading shows amino acids that are conserved in 50% (4/8) of the shown TDU domains. Grey

shading indicates those residues that are conserved in 25% (2/8) of the TDU domains. TDU1 is the most

N-terminal TDU domain in both hVgl-4 and dVgl-4. TDU2 is the most C-terminal TDU domain in both

hVgl-4 and D-Vgl-4. B) Logo map (Schneider and Stephens, 1990) of the core consensus TDU domain

sequence. For each of the 11 positions of the core TDU domain, the probability of any given residue

being found is indicated. The logo map was generated using WebLogo 3.0 (Crooks et al., 2004).

A

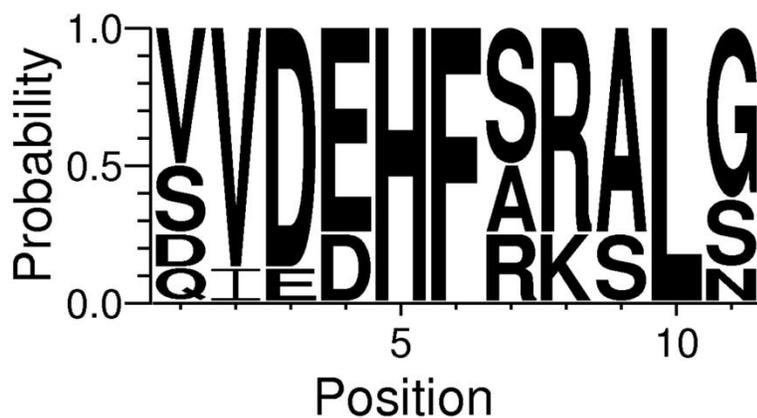|	
hVgl-1	VVDEHFSRALS
hVgl-2	VVDEHFSRALS
hVgl-3	VVDEHFSRALG
Vg	QVDEHFSRALN
hVgl-4 (TDU1)	VVEEHFRRSLG
dVgl-4 (TDU1)	DIDEHFRRSLG
hVgl-4 (TDU2)	SVDDHFAKALG
dVgl-4 (TDU2)	SVDDHFAKALG

B

Figure 3.2. Protein sequences of hVgl-4 and the two dVgl-4 isoforms. In the three protein sequences shown, the black shading indicates the N-terminal Tondu-1 (TDU1) domain and the C-terminal TDU2 domain. Grey shading indicates the variable N-terminal regions of the two dVgl-4 isoforms, PA and PB (which are the products of the RA and RB mRNA isoforms, respectively). The underlined residues are those against which the rabbit anti-dVgl-4 antibody was raised.

hVgl-4

METPLDVLSRAASLVHADDEKREAALRGEPRMQTLPVASALSSHRTGPPPISPSKRKFSMEPGDEDLDCDNDHVSKMS
RIFNPHLNKTANGDCRRDPRERSRSPIERAVAPTMSLHGSHLYTSLPSLGLEQPLALTKNSLDASRPAGLSPTLTPGERQN
RPSVITCASAGARNCNLSHCPIAHSGCAAPGPASYRRPPSAATTCDPVVEEHFRRSLGKNYKEPEPAPNSVSITGSVDDH
FAKALGDTWLQIKAAKDGASSSPESASRRGQPASPSAHMVSHSHSPSVVS

dVgl-4 RA

MALRLDYRCLLDAFEDYYYHKEIQRLVAETAGGATATSPASSASSASSTASISSASCSSGPSTSSIVSSAASSHGSLAQVAT
ARAAAALADQQALASQRAMFYNVQHPQQLEQLHALQAESGNQQMHPQANADPNASSMANSLLWQPWRDLQQA
AAMHHQLYRQQQQQLQLHSEMRATSKVLTTTKWRRERRQRSAGYQPHEAGNSERERERERDREDRDMDSPIDMSV
TTGALKQRASPPPPYREPLPGTNYAANSRPSVITQAPPKREPPEQAHSTDMAMCDIDEHFRRSLGENYAALFAKKSPTP
TPTPTPSPSGTPKQQVSPLAYGLPSSTSTAASQHYQQQRSPLAKSGWVILEPESLQPELPPPQEEPLPLSLALHRTQTPPS
PPPSATGSAPALPTAVSQVMEAAVAGRRILDTPHHTPPRYNTPPPPPPAYGIAGTTVVAPTLTPTPTNPTPSQIPTPTP
SMPAIIRVKAEPGLAAVAASSTQTPPASPTSSTNISIFTKTEASVDDHFAKALGETWKKLQGGHKE

dVgl-4 RB

METALDVLSRAATMVQNNPSEMRATSKVLTTTKWRRERRQRSAGYQPHEAGNSERERERERDREDRDMDSPIDMS
VTTGALKQRASPPPPYREPLPGTNYAANSRPSVITQAPPKREPPEQAHSTDMAMCDIDEHFRRSLGENYAALFAKKSPT
PTPTPTPSPSGTPKQQVSPLAYGLPSSTSTAASQHYQQQRSPLAKSGWVILEPESLQPELPPPQEEPLPLSLALHRTQTPP
SPPPSATGSAPALPTAVSQVMEAAVAGRRILDTPHHTPPRYNTPPPPPPAYGIAGTTVVAPTLTPTPTNPTPSQIPTPT
PSMPAIIRVKAEPGLAAVAASSTQTPPASPTSSTNISIFTKTEASVDDHFAKALGETWKKLQGGHKE

Figure 3.3. A dVgl-4 antibody detects ectopic dVgl-4 expression. A) 3xFLAG-Vgl-4 was expressed in S2 cells and the lysate was subjected to Western blot analysis. Primary detection was with anti-FLAG (left) or anti-dVgl-4 (right). Both detected a protein running at ~68 KDa, and the anti-dVgl-4 also detected a band at ~55 KDa. B) Wing discs were dissected from third instar larva expressing *UAS-3xFLAG-dvgl-4* under the control of *ptc*-GAL4 and immunostained using rabbit anti-dVgl-4 as the primary antibody. A *ptc* pattern of expression along the A/P axis is seen. The secondary antibody used was goat anti-rabbit conjugated with alkaline phosphatase, which was detected using the alkaline phosphatase substrate BCIP/NBT.

A

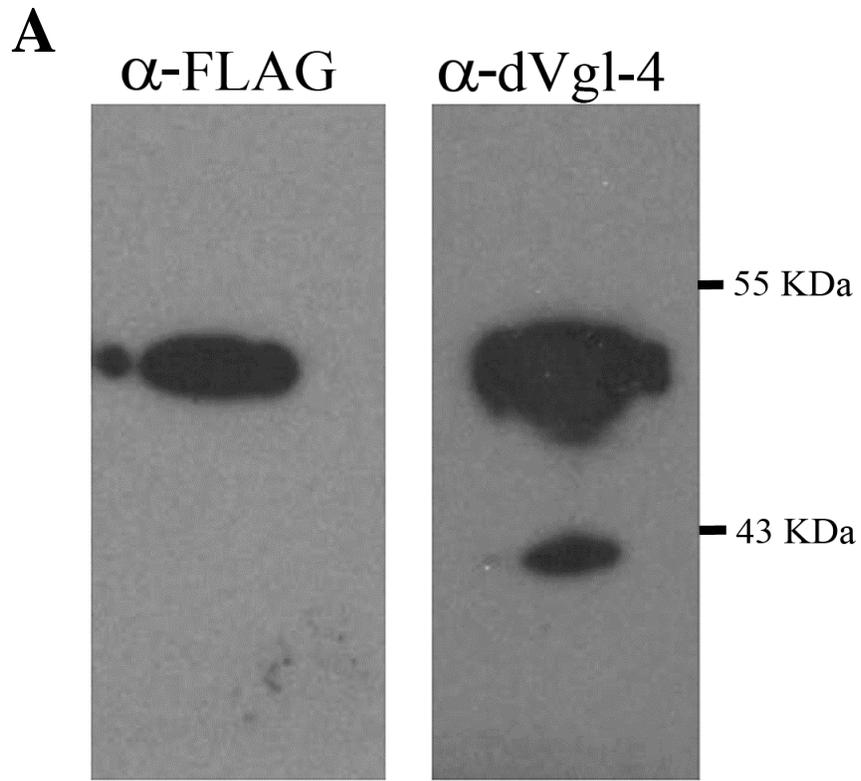α-FLAG   α-dVgl-4

55 KDa

43 KDa

B

Figure 3.4. Sd and dVgl-4 interact in both *ex vivo* and *in vitro* assays. A) Myc-Sd was expressed

either alone or with 3xFLAG-dVgl-4, and Myc-dVgl-4 was likewise expressed either alone or with

3xFLAG-Sd. The cells from the four transfection experiments were then lysed and purified using anti-

FLAG beads. The immunoprecipitated lysates were then analyzed by Western blot, and primary

detection was with either anti-FLAG or anti-Myc. B) *In vitro* translated and radio-labelled dVgl-4 and Vg

(as a positive control) were ran over a column containing either bound GST-Sd, or not (negative control).

A sample of radio-labelled Vg and dVgl-4 prior to loading ("Load"), eluted protein from the columns

("Eluate") which either contained GST-Sd or not and the flow through ("Flow-through") from each

column were then analyzed by SDS-PAGE. When GST-Sd is present in the column, almost all of the

loaded Vg and dVgl-4 probe proteins bind prior to elution. Conversely, most of the radio-labelled

protein probes are present in the flow-through when run over a column lacking GST-Sd.
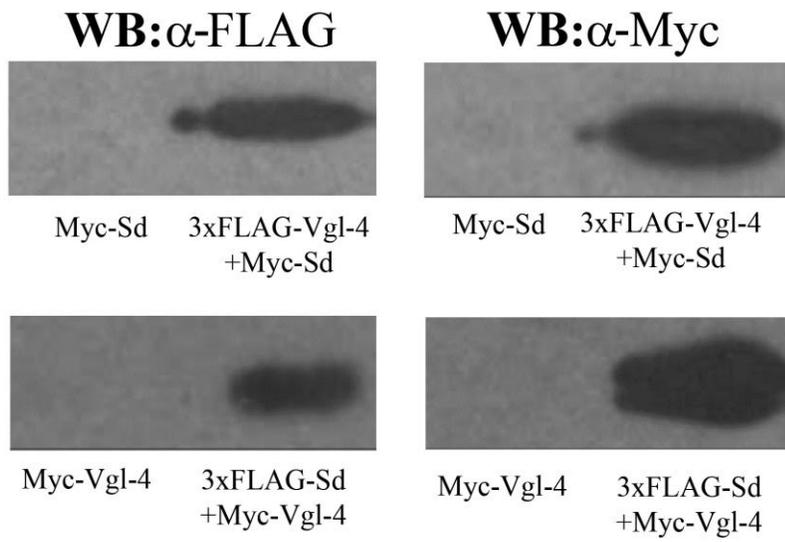
A

IP:α-FLAG

WB:α-FLAG   **WB:α-Myc**

Myc-Sd   3xFLAG-Vgl-4   Myc-Sd   3xFLAG-Vgl-4
          +Myc-Sd   +Myc-Sd

Myc-Vgl-4   3xFLAG-Sd   Myc-Vgl-4   3xFLAG-Sd
             +Myc-Vgl-4   +Myc-Vgl-4
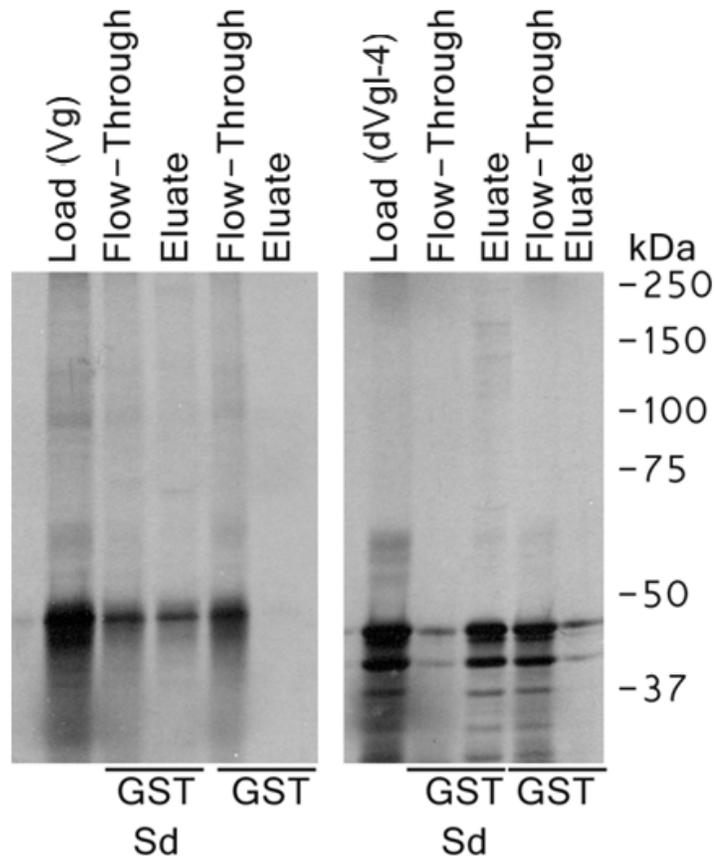
B

Figure 3.5. dVgl-4 localizes to the nucleus of S2 cells. eGFP-dVgl-4 was expressed under the

control of heatshock in DAPI stained S2 cells. A) Green (eGFP) channel. B) Blue (DAPI) channel. C)
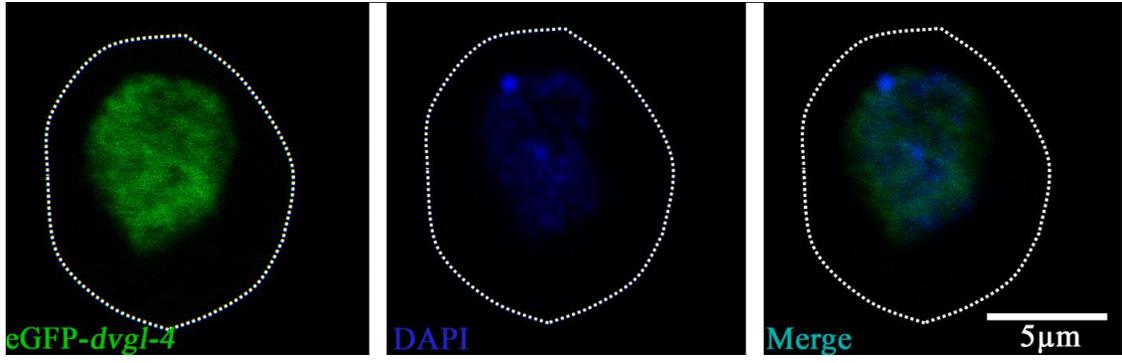
Merge. The hatched line marks the outline the cell.

eGFP-*dvgl-4*   DAPI   Merge   5μm

Figure 3.6. The expression of d*vgl-4* RB is virtually abolished in homozygous d*vgl-4*^{e01789}

insertion mutants. A) Phenotype of a female d*vgl-4*^{e01789} homozygote (right) compared to a female

d*vgl-4*^{e01789} /*TM6 Tb* heterozygote (left). The arrow indicates a break in the abdominal pigmentation

and disorganization of the tergite bristles. The wings were removed from both flies to aid in the

visualization of the abdomen. B) Quantitative real-time PCR analysis of the RA and RB vgl-4 isoform

expression levels in homozygous d*vgl-4*^{e01789} mutants relative to w^{1118} flies. RNA was extracted from

whole third instar w^{1118} and d*vgl-4*^{e01789} larvae. These samples were then analyzed for the expression

levels of the RA and RB transcripts in the mutant larvae relative to the w^{1118} larvae (which are arbitrarily

assigned a value of 1.00 for expression level). For clarity, the fold change relative to w^{1118} larvae is

indicated numerically above each bar. Error bars are the 95% confidence intervals.

$yw;\ dvgl\text{-}4^{e01789}/TM6\ Tb$ $yw;\ dvgl\text{-}4^{e01789}/dvgl\text{-}4^{e01789}$
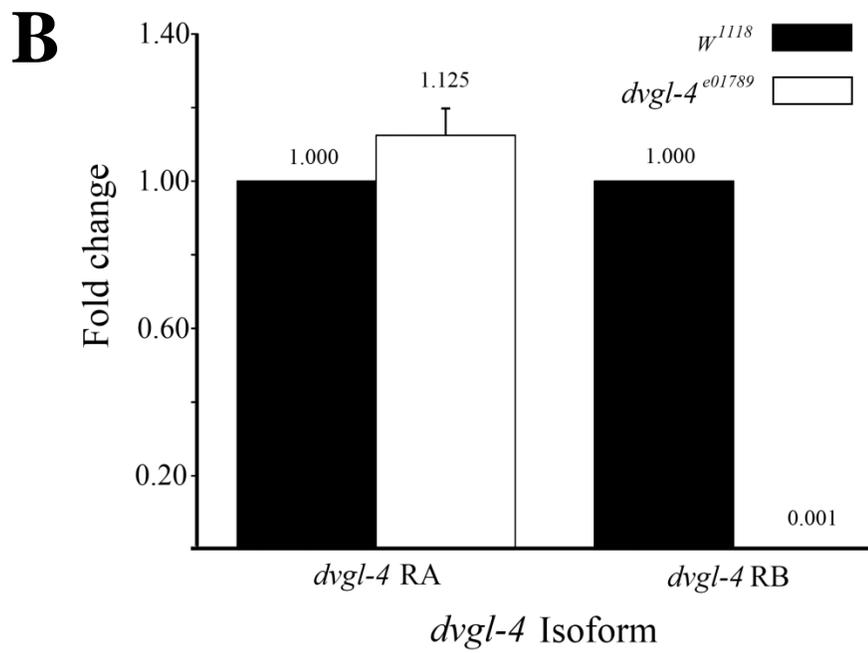
Table 3.1. Overview of GAL4 drivers used to drive *UAS*-d*vgl-4* RNAi. Each was used to generate

transheterozygous flies containing the driver and three different *UAS*-d*vgl-4* RNAi lines (one from the

VDRC; *UAS*-d*vgl-4* RNAi [VDRC], and two generated in-house; *UAS*-d*vgl-4* RNAi[pWIZ-1] and *UAS*-d*vgl-4* RNAi[pWIZ-2]). The general regions of third instar larval expression are indicated. As a whole, the drivers also are

expressed in a wide-variety of embryonic tissues.

Driver	Expressed in third instar larval
sd-GAL4	Imaginal discs, optic lobe
vg-GAL4	Wing disc
act5c-GAL4	Ubquitous or nearly so
pnr-GAL4	Wing and eye-antennal imaginal discs
heatshock-GAL4	Ubiquitous, or nearly so, when exposed to elevated temperatures (~37°C)
wg-GAL4	Imaginal discs, optic lobe
ey-GAL4	Eye-antennal imaginal disc
Dll-GAL4	Imaginal discs, optic lobe
en-GAL4	Wing and leg imaginal discs
mef-GAL4	Somatic, viceral and cardiac muscle
dpp-GAL4	Imaginal discs
esg-GAL4	Histoblast nests, imaginal discs

Figure 3.7. Overview of generation of a d*vgl-4* deficiency using the Exelixis methodology. A) In order to generate Exelixis deletions, two strains containing P-element mediated FRT insertion sites are needed. When the two strains are crossed and the FRT sites are trans-heterozygous, treatment with FLPase will generate a chromatid with the region between the two sites (in this case d*vgl-4*) deleted, as well as a chromatid containing a duplication of the region. (Parks et al., 2004). B) This diagram illustrates the genomic region between the FRT sites pBac{RB}CG10741[e01789] and pBac{WH}f01796. FLPase mediated FRT recombination between the two sites would yield a chromatid lacking d*vgl-4*, *Spt20*, *Vacuolar protein sorting 36* (*Vps-36*), *Liprin-β* and *cg10710*.

A

FRT 1 *dvgl-4*

FRT 1 *dvgl-4*

FLPase mediated recombination

dvgl-4 FRT 2

dvgl-4 FRT 2

FRT 1 *dvgl-4*

dvgl-4 FRT 1/2 *dvgl-4*

FRT 1/2

dvgl-4 FRT 2

B

dvgl-4 *Vps36* *cg10710*
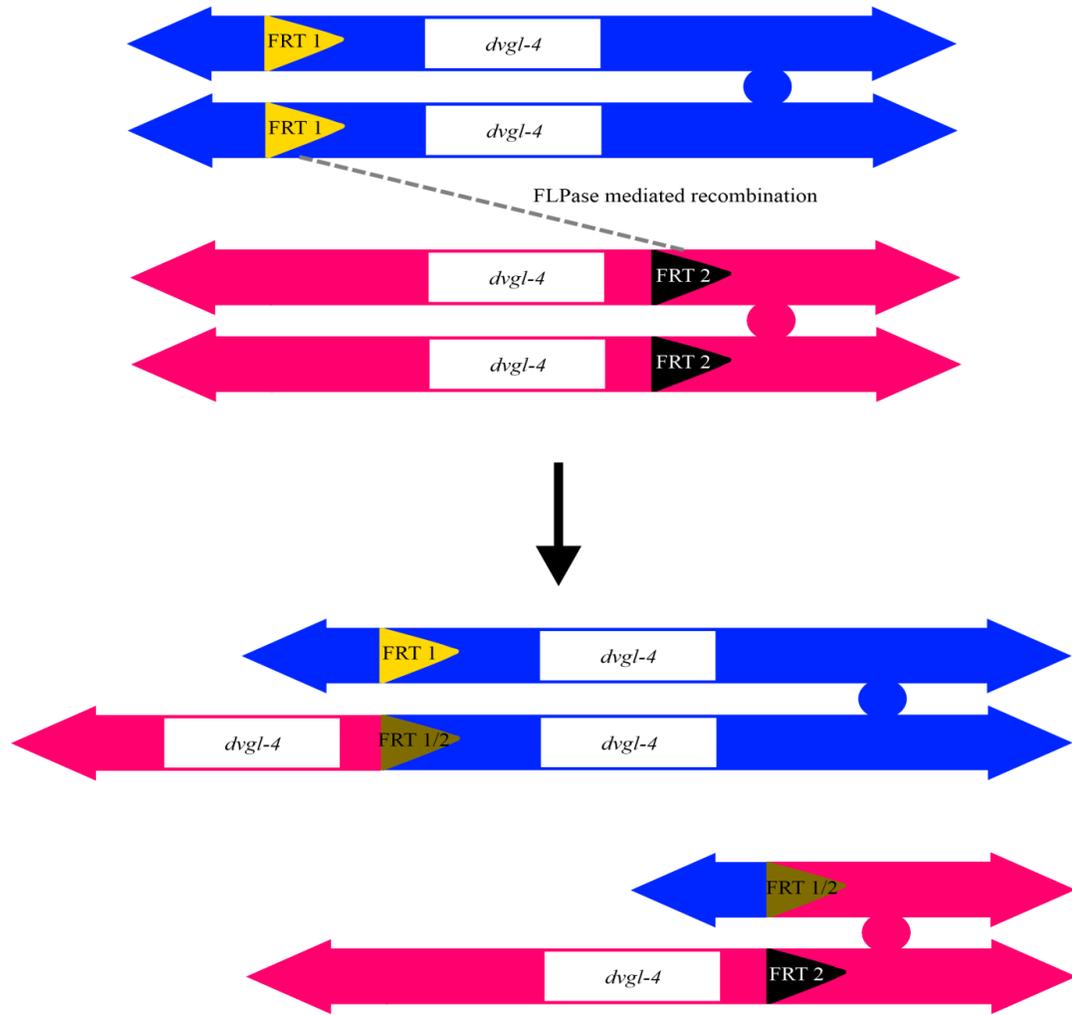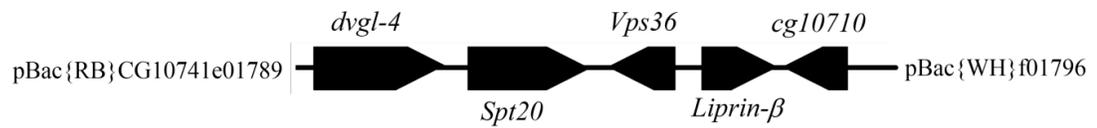
pBac{RB}CG10741e01789 *Spt20* *Liprin-β* pBac{WH}f01796

Figure 3.8. Gene replacement using ends-out recombination. The 5' d*vgl-4* + 3' d*vgl-4* element (which contains an interior *w*⁺ mini-gene and is referred to in the text as d*vgl-4* AB) is excised and linearized by the actions of FLP recombinase (at the FRT sites) and I-*Sce*I endonuclease (at the I-sites), respectively. The 5' and 3' elements then recombine with their counterparts in the endogenous d*vgl-4* gene replacing the interior of that gene with the *w*⁺ mini-gene. Chr = chromosome. Modified from Gong and Golic, 2003.
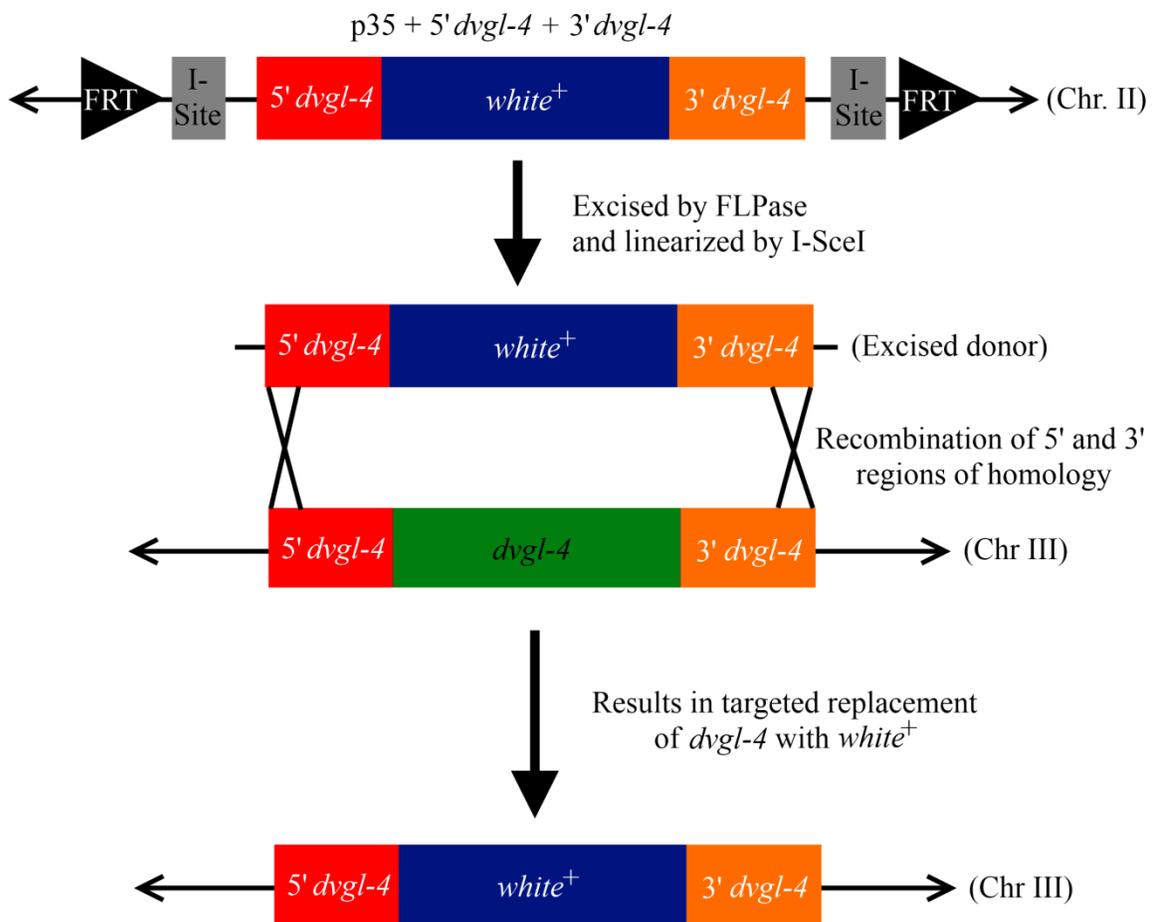
Figure 3.9. Phenotypes of *3xFLAG*-d*vgl-4* overexpression animals. A-C) *UAS-3xFLAG*-d*Vgl-4* was expressed using: *sd*-GAL4 (A), *act*-GAL4 (B) or *pnr*-GAL4 (C). In (A), the filled and open arrowheads mark the sites of the missing wing and haltere tissues, respectively.   The wings were removed from (C) in order to aid in the visualization of the notum.
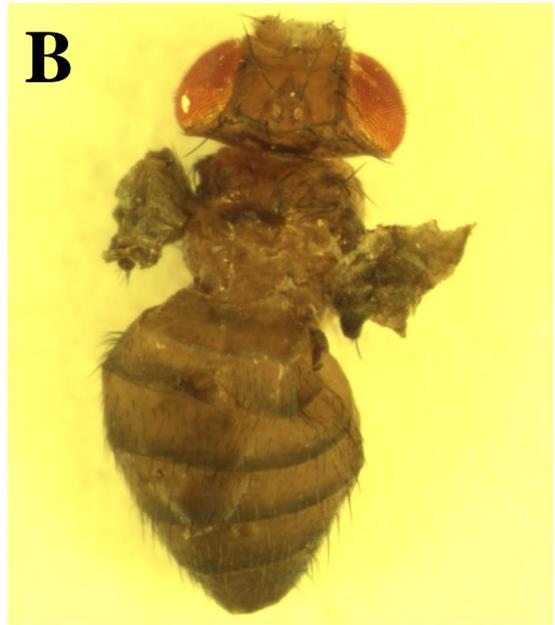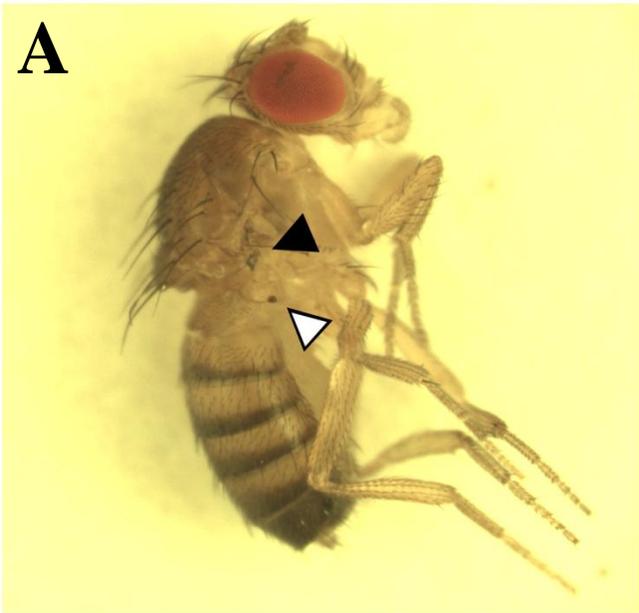
Figure 3.10. Schematic of P-element insertion into d*vgl-4* locus. The d*vgl-4*^{e01789} allele contains a P-element inserted at the 3' end of the first exon of the RB isoform of d*vgl-4* as indicated by the orange arrow. This insertion is not present in the RAisoform, which is also shown. Modified from Tweedie et al., 2009.

dvgl-4-RB

P-element insertion site

dvgl-4-RA

References

Ayoubi, T., and Van De Ven, W. (1996). Regulation of gene expression by alternative promoters. The FASEB Journal *10*, 453 -460.

Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Bischof, J., Maeda, R. K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific φC31 integrases. Proc Natl Acad Sci U S A *104*, 3312-3317.

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.

Campbell, S., Inamdar, M., Rodrigues, V., Raghavan, V., Palazzolo, M., and Chovnick, A. (1992). The scalloped gene encodes a novel, evolutionarily conserved transcription factor required for sensory organ differentiation in *Drosophila*. Genes & development *6*, 367-379.

Carthew, R. W., and Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell *136*, 642-655.

Chen, H. H., Mullett, S. J., and Stewart, A. F. (2004). Vgl-4, a novel member of the vestigial-like family of transcription cofactors, regulates alpha1-adrenergic activation of gene expression in cardiac myocytes. Journal of Biological Chemistry *279*, 30800-30806.

Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004). WebLogo: A Sequence Logo Generator. Genome Res *14*, 1188-1190.

Dietzl, G., Chen, D., Schnorrer, F., Su, K.-C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S., et al. (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. Nature *448*, 151-156.

Faucheux, C., Naye, F., Treguer, K., Fedou, S., Thiebaud, P., and Theze, N. (2010). Vestigial like gene family expression in Xenopus: common and divergent features with other vertebrates. Int. J. Dev. Biol. *54*, 1375-1382.

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

Groth, A. C., Fish, M., Nusse, R., and Calos, M. P. (2004). Construction of transgenic *Drosophila* by using the site-specific integrase from phage phiC31. Genetics *166*, 1775-1782.

Hammond, S. M., Bernstein, E., Beach, D., and Hannon, G. J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. Nature *404*, 293-296.

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.

King-Jones, K., Horner, M. A., Lam, G., and Thummel, C. S. (2006). The DHR96 nuclear receptor regulates xenobiotic responses in *Drosophila*. Cell Metab *4*, 37-48.

Lee, Y. S., and Carthew, R. W. (2003). Making a better RNAi vector for *Drosophila*: use of intron spacers. Methods *30*, 322-329.

Liu, Q., and Paroo, Z. (2010). Biochemical Principles of Small RNA Pathways. Annu. Rev. Biochem. *79*, 295-319.

Maeda, T., Chapman, D. L., and Stewart, A. F. (2002). Mammalian vestigial-like 2, a cofactor of TEF-1 and MEF2 transcription factors that promotes skeletal muscle differentiation. Journal of Biological Chemistry *277*, 48889-48898.

Mummery-Widmer, J. L., Yamazaki, M., Stoeger, T., Novatchkova, M., Bhalerao, S., Chen, D., Dietzl, G., Dickson, B. J., and Knoblich, J. A. (2009). Genome-wide analysis of Notch signalling in *Drosophila* by transgenic RNAi. Nature *458*, 987-992.

Parks, A. L., Cook, K. R., Belvin, M., Dompe, N. A., Fawcett, R., Huppert, K., Tan, L. R., Winter, C. G., Bogart, K. P., Deal, J. E., et al. (2004). Systematic generation of high-resolution deletion coverage of the *Drosophila melanogaster* genome. Nat Genet *36*, 288-292.

Pham, J. W., Pellino, J. L., Lee, Y. S., Carthew, R. W., and Sontheimer, E. J. (2004). A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in *Drosophila*. Cell *117*, 83-94.

Rousset, R., Bono-Lauriol, S., Gettings, M., Suzanne, M., Spéder, P., and Noselli, S. (2010). The *Drosophila* serine protease homologue Scarface regulates JNK signalling in a negative-feedback loop during epithelial morphogenesis. Development *137*, 2177 -2186.

Schneider, T. D., and Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. Nucleic Acids Res *18*, 6097-6100.

Schwarz, D. S., Tomari, Y., and Zamore, P. D. (2004). The RNA-induced silencing complex is a Mg2+-dependent endonuclease. Curr. Biol *14*, 787-791.

Thibault, S. T., Singer, M. A., Miyazaki, W. Y., Milash, B., Dompe, N. A., Singh, C. M., Buchholz, R., Demsky, M., Fawcett, R., Francis-Lang, H. L., et al. (2004). A complementary transposon tool kit for *Drosophila melanogaster* using P and piggyBac. Nat. Genet *36*, 283-287.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S. E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S. E., et al. (2002). Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. Genome Biol *3*, 1-88.

Tweedie, S. et al., 2009. FlyBase: enhancing Drosophila Gene Ontology annotations. *Nucleic Acids Research*, 37(Database), p.D555-D559

Vaudin, P., Delanoue, R., Davidson, I., Silber, J., and Zider, A. (1999). TONDU (TDU), a novel human protein related to the product of vestigial (vg) gene of *Drosophila melanogaster* interacts with vertebrate TEF factors and substitutes for Vg function in wing formation. Development *126*, 4807-4816.

Chapter Four: General conclusions and future directions

Nuclear localization of Sd

In Chapter two, evidence was provided that shows quite clearly that Sd contains a functional nuclear localization signal (NLS). As discussed in that chapter, this is the first time an NLS has been proven to be functional in a member of the TEAD family of proteins. However, the available evidence also indicates that this NLS, while necessary for the proper nuclear of localization of Sd, is not sufficient. Indeed, it is clear that at least one other signal – present within the C-terminal domain of Sd – is required as well. More intriguingly, evidence that Sd also contains a nuclear export signal (NES) was presented as well. If Sd does indeed contain an NES, then that is suggestive that Sd may cycle bidirectionally between the cytoplasm and nucleus, rather than unidirectionally into the nucleus. If so, understanding the regulation of this process would thus provide insight into the regulation of Sd function vis-à-vis transcription.

A major problem with analyzing Sd localization, and therefore understanding the regulation thereof, has been the lack of working antibody. For instance, while a working antibody against TEF-1 has been developed, it is not compatible with Sd. Furthermore, two attempts to make a Sd-specific antibody by our laboratory have met with limited success. The first was unable to detect Sd under any condition, while the second (made in a similar fashion to the dVgl-4 antibody discussed in Chapter three), could weakly detect (via immunostaining or Western blot) exogenously expressed Sd, but not the endogenous protein (data not shown). This has made it difficult to study the endogenous nuclear localization of Sd, and moreover, to test the hypothesis that Sd may shuttle between the nucleus and the cytoplasm under certain conditions. Although it is possible that the correct conditions for immunostaining with Sd antibodies has not been found, it is also possible that Sd is expressed in low quantities and thus is difficult to detect, and of course these possibilities are not mutually exclusive. If

the latter is true, one way around this problem would be to employ the tyramide signal amplification system (PerkinElmer), which can greatly amplify the sensitivity of immunostaining. Further, there are a many different fixation techniques that could be tried in order to improve accessibility of the antibody to the antigen. If the antibody could be reliably detected, a host of experiments could be done. One would be to examine the *in vivo* localization behaviour of the various Sd mutants discussed in Chapter two (by making and expressing transgenic *UAS*-mutant sd lines). It would also be worthwhile to examine the localization of Sd in *sd* mutants (of particular interest would be sd^{68L}, which shows the curious *vg* mislocalization phenotype).   Lastly, the spatial, temporal and sub-cellular localization could be studied at various time points, ranging from embryonic to pupal development, in order to confirm the in situ and enhancer trap localization patterns previously described as well as to determine if there are times/tissues in which Sd shows enrichment in the cytoplasm – which would be more evidence that Sd may be regulated by nuclear-cytoplasmic shuttling. However, barring an improvement in the antibody efficacy, other experiments could also be attempted. The first would be to conduct an RNAi screen, using the localization of several eGFP-tagged isoforms of Sd (such as Sd itself, Sd mNLS^{N+C} and SdΔ392-440; see Chapter two) as a read-out. The rationale behind this experiment would be two-fold. First, it could act to confirm that Imp-α3 is involved in the nuclear translocation of Sd as well as determining whether either of the other two Imp-α proteins are also involved.   Secondly, the screen would hopefully identify other genes which could also modulate Sd translocation – ideally identifying proteins which bind to the C-terminus of Sd in order to mediate that function. It would also strengthen the argument that Sd contains an NES if it could be shown that RNAi against the exportin *Crm1* differentially rescues the localization of Sd mutants containing the putative NES signal. If one or more of the *Imp-α*'s are identified in this approach, but no additional genes, it would consistent with the idea that there is an additional NLS signal in the C-terminal domain of Sd.

A small pilot-screen has already been done, expressing eGFP-Sd together in S2 cells treated with RNAi against *Imp-α1-3*, *vg*, *dvgl-4* and *sd*. The first three RNAi treatments (alone or in combination) resulted in almost complete abolishment of transfection in the treated cells (data not shown). The ability of RNAi treatment to interfere with transfection seems to be a general complication that has been observed by others (D. Bond, personal communication), although this effect may also have been worsened as a consequence of interfering with the nuclear import machinery, since the RNAi against the latter three genes mentioned above did not interfere with transfection, but nor did they disrupt eGFP-Sd localization. In order to overcome the difficulties of transiently transfecting RNAi treated cells, stable lines expressing the various constructs could be generated and used for the screen.   As far as the lack of phenotype when treating with *vg*, *sd* and *dvgl-4* RNAi, there were two limitations to the study: First, no measurements were made comparing transcript levels between control and RNAi treated cells, so it is possible the RNAi was not efficient in reducing the transcript levels of one or more of the target genes. Second, only full-length Sd was tested. If mutant forms of Sd were also tested, phenotypes may have been observed. Particularly in the case of Sd mNLS^{N+C}, one could imagine that binding to endogenous Sd – which has an intact NLS – could be responsible for some or all of the observed translocation seen in this mutant. However, that would not explain why the C-terminal domain is required for proper localization even when the NLS is intact (e.g. in the case of SdΔ392-440). Ideally, this experiment would provide a gateway to revealing new Sd interacting proteins, and possibly give insight into the regulation of Sd function via control of nuclear import and export – based on the assumption that some of these new interactors would be involved in that regulation.

dVgl-4

It is difficult to say what role, if any, dVgl-4 has in *Drosophila* development. Certainly, the fact

that it is able to interact with Sd, and that there is clearly conservation between the sequence and

arrangement of the dVgl-4 and hVgl-4 TDU domains is suggestive that the protein is needed for some

undefined biological process. It is also clear that the protein is capable of biological activity, based on

the over-expression phenotypes. However, a reasonable assumption that could be made is that the

dVgl-4 over-expression phenotypes are simply due to binding to and interfering with endogenous Sd,

whether or not the physical interaction demonstrated is biologically relevant *in vivo*. Indeed, one could

imagine that by the erroneous (at least in those tissues) binding of dVgl-4 to Sd, the correct TIFs of Sd

would have their own ability to bind Sd reduced (i.e. there would be competition by dVgl-4 for Sd

binding). That said, while an enhancer trap detects *sd* expression in the presumptive scutellum (which

gives rise to the posterior end of the dorsal thorax) and the mesopleura (which gives rise to the lateral

portion of the thorax) of the wing disc, no defects are seen when either of two *UAS-sd* RNAi constructs is

driven by *pnr*-GAL4, although the transgene is able to interfere with Sd function in the wings and eyes of

the fly using drivers specific to those tissues (Zhang et al., 2008; and data not shown). Additionally, the

cuticle and bristles of the notum show a great deal of disorganization in the *pnr*>dvgl-4 animals, and

there is no evidence that sd is expressed in the presumptive notum of the wing disc.   Thus, *pnr*-GAL4 is

the only driver tested which causes defects when driving *UAS-3xFLAG*-d*vgl-4* that are not easily

explained by a possible interaction with Sd. Indeed, the *act*-GAL4 driver gives results consistent with

this idea. Most of the phenotypes are possibly explained by Sd interactions; the lethality seen could be

due to defects in nervous system or cardiac muscle development, while the wing defects could also

reflect a dominant-negative interaction with Sd during the development of that tissue.   However, the

only other phenotype observed in adults is a thoracic phenotype similar to that seen when using the

pnr-GAL4 driver. Altogether, this could indicate that dVgl-4 is capable of a biological function beyond

interacting with Sd. Elucidating this function may be difficult however, especially if dVgl-4 has functional

redundancy with one or more other proteins. Generating a null or a knockdown and identifying

phenotypes for either or both must be the first priority. Possible ways to improve the efficiency of the

RNAi were already given in Chapter Three. By using in situ hybridization it may be possible to identify

the tissue(s) in which *dvgl-4* is expressed, and thus perhaps lead to the selection of a more specific

driver to use for the *UAS*-RNAi lines (although act-GAL4 and HS-GAL4 gave no phenotype with the RNAi

lines so there is no guarantee a more-specific driver would yield more informative results). As far as

generating a null, there have been improvements made to the ends-out replacement scheme (Huang et

al., 2008) and those improvements could be used in a second attempt to generate a null allele.

Additionally, nearby P-elements could be used to generate partial or complete deletions in a variety of

ways (e.g. imprecise excision). See Hummel and Klämbt, (2008) for a review of this and other

techniques. Once a null has been generated, if a phenotype is seen, it will be evidence that the PB

isoform of dVgl-4 is sufficient for development in the absence of the PA isoform, but that the loss of

both is deleterious. If not, it will be further evidence that supporting the idea that there is functional

redundancy provided by the product of another gene. If this is the case, a screen would need to be

done in order to identify the compensatory gene. This could be done in two ways. First, flies

homozygous null for d*vgl-4* could be treated with a mutagen in order to generate second-site mutations

which show phenotypes in this background. Unfortunately, this procedure would generate many

mutations in genes that are not related in any way to d*vgl-4* and so the mutations generated would

need to be crossed back into a wildtype background to ensure the phenotype is dependent on both the

new mutation and the d*vgl-4* null. The second way would be to use large scale deficiencies which in

aggregate uncover the whole of the *Drosophila* genome, to generate a set of flies homozygous (or

hemizygous) for the dvgl-4 null which also contain the deficiency. This would generate the d*vgl-4* null,

plus hemizygotes for the genes uncovered by the deficiency. Hopefully, losing both copies of d*vgl-4* and

a copy of whatever gene is providing the compensatory function would result in an observable

phenotype.   Unfortunately, both techniques are labour intensive and neither is guaranteed to be

successful.

Suppressors of *sd*^{58d} mutants

Not discussed in this thesis was work done to characterize mutations which suppress the

extreme wing phenotypes of *sd*^{58d} mutants. These mutants are comprised of four independent lines

that were previously generated by feeding *sd*^{58d} flies the mutagen ethyl methanesulfonate (EMS; which

induces G:C to A:T transitions) and screening the progeny for suppressors of *sd*^{58d} as characterized by the

restoration of wing tissue relative to the original mutants.   All four of these mutations have been

successfully mapped to the right arm of the second chromosome (T. Kelly, unpublished results).

Moreover, they are all dominant mutations which either effect viability in the heterozygous state, have

a reduced penetrance, or both (data not shown).   Attempts to balance them have been largely

unsuccessful since the viability of the strains (already quite poor) is reduced even further when

maintained over balancer chromosomes. Due to the difficulties in generating balanced lines, it has not

been possible to determine how many complementation groups the mutations fall into or estimate the

penetrance of the phenotype, but given that they show a similar range of wing phenotypes in the *sd*^{58d}

background and that they all map to the right arm of chromosome two, it is possible that they are all

alleles of the same gene. While deficiency mapping was determined to be the most suitable method to

determine the gene(s) mutated in these lines, time constraints prevented the experiments from being

done. However, this information would be very useful, since one or more genes that interact - either

genetically or physically - with Sd in the context of wing development would be identified. This would

not only potentially lead to a better understanding of the function of Sd during wing development, but would also identify potential candidates which may regulate the sub-cellular localization of Sd.

Conclusion

In the course of this thesis a bipartite classical nuclear localization signal was confirmed to be present and functional within Sd. Additionally, hints that there may be complex regulation of the sub-cellular localization of the protein were also obtained. Indeed, there is evidence that Sd also contains an NES, is subject to post-translational modification and requires another NLS and/or binding to unknown protein(s)in order to be properly translocated to the nucleus. Additionally, a potential TIF of Sd has been identified in d*vgl-4*, and means to identify one or more additional TIFs (by characterizing the *sd*^{58d} suppressors and determining if there are additional proteins that modify the localization of Sd) have been provided. Identifying new TIFs of Sd could provide insight into further functional roles of Sd, ideally outside the tissue in which the function of Sd has been best-characterized – the wing imaginal disc.

References

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.

Hummel, T., and Klämbt, C. (2008). P-element mutagenesis. Methods Mol. Biol *420*, 97-117.

Zhang, L., Ren, F., Zhang, Q., Chen, Y., Wang, B., and Jiang, J. (2008). The TEAD/TEF Family of Transcription Factor Scalloped Mediates Hippo Signalling in Organ Size Control. Developmental Cell *14*, 377-387.