UC San Diego
UC San Diego Electronic Theses and Dissertations

Title

Identification and characterization of a conserved isoform of the Drosophila Dorsal-related immunity factor, Dif

Permalink

https://escholarship.org/uc/item/1ks2b9qn

Author

Mayo, Jason David

Publication Date

2008

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Identification and Characterization of a Conserved Isoform of the
Drosophila Dorsal-Related Immunity Factor, Dif

A Thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Biology

by

Jason David Mayo

Committee in charge:

      Professor Steven A. Wasserman, Chair
      Professor Ethan Bier
      Professor William McGinnis

2008

The Thesis of Jason David Mayo is approved and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____
Chair

University of California, San Diego

2008

Table of Contents

List of Figures

<u>Acknowledgements</u>

ABSTRACT OF THE THESIS

Identification and Characterization of a Conserved Isoform of the
Drosophila Dorsal-Related Immunity Factor, Dif

by

Jason David Mayo

Master of Science in Biology

University of California, San Diego, 2008

Professor Steven A. Wasserman, Chair

The Dorsal-related immunity factor, Dif is a *Drosophilid* specific transcription factor that likely arose via gene duplication of Dorsal. An alternative splice variant was discovered for Dif based on bioinformatic analysis and conservation of Dorsal B homologs. This isoform, which I will term Dif B, is expressed in whole flies and S2 cells. The structure of Dif B has been characterized as a 987 amino acid protein that is identical to Dif up to the C-terminal end of the rel homology domain where it diverges completely. A hybrid Dif B cDNA was constructed for future use. Exon-specific RNAi suggests a possible inhibitory role for the Dif B isoform.

Chapter 1:

Introduction

Innate Immunity and the Toll Pathway

Pathogenic microbes are an ever-present threat to multicellular organisms. Prior to the evolutionary emergence of the adaptive immune system, the only line of defense for those organisms was the innate immune response. This ancient system, which is present in all multicellular life forms, carries out pathogen detection, upregulation of anti-microbial molecules and initiation of secondary immune responses such as inflammation.

In *Drosophila melanogaster*, cells respond to invasion by Gram (+) bacteria and fungi via the Toll pathway (Figure 1). This pathway is the counterpart of the mammalian Toll-like receptor (TLR) pathways, which mediate functions of the innate immune response in humans. In flies, recognition of Gram (+) bacteria or fungi leads to the cleavage of the protein Spätzle, which binds and activates the Toll receptor on the surface of the cell(Morisato and Anderson, 1994; Roth and Schüpbach, 1994). Activated Toll undergoes a conformational change, activating its cytoplasmic Toll/Interleukin-1 receptor (Ghiglione et al.) domain and initiating a signaling cascade within the cell(Bowie and O'Neill, 2000). The adaptor module MyD88 has an N-terminal death domain (DD) and a C-terminal TIR domain, which binds the TIR domain of Toll. The death domain of MyD88 forms a trimeric complex with the death domains of the proteins Tube and Pelle, continuing the signal cascade (Sun et al., 2002)

Downstream effectors of the Toll pathway are members the NF-κB (nuclear factor-kappa B) family of transcription factors. In response to innate immune pathway activity, these molecules bind to specific κB sites on DNA regions coding for anti-microbial peptide (AMP) genes. In mammals, the NF-κB members include p50 (NF-κB1), p52 (NF-κB2), p65 (Rel A), c Rel and Rel B. Each has a highly conserved N-terminal Rel Homology Domain (RHD) which has the capacity for DNA binding, nuclear localization, and protein dimerization (Ganchi et al., 1992). Prior to an immune response, the NF-κB proteins are held in an inactive state within the cytoplasm by the IκB family of inhibitory proteins. In response to signaling, the inhibitor is polyubiquitinated and concomitantly destroyed via proteosomal degradation. Free NF-κB translocates to the nucleus, where it upregulates genes necessary for the innate immune response.

In *Drosophila*, the NF-κB members of the Toll pathway are Dorsal and Dorsal-related immunity factor (Dif), each of which is kept inactive, under quiescent conditions, by the IκB homolog product encoded by the *cactus* gene. Dorsal was discovered and initially characterized as a maternal morphogen that regulates dorsoventral patterning in early embryos (Roth et al., 1989); it was later found to contribute to larval immunity(Reichhart et al., 1993). Dif functions only in immunity.

In response to Gram (+) bacterial or fungal infection, Dorsal and Dif bind to κB sites, directing expression of many AMPs as well as pathogen recognition proteins (Gross et al., 1996). One such AMP is the anti-fungal agent Drosomycin. Flies lacking both Dorsal and Dif are unable to produce Drosomycin (Rutschmann et al., 2000). Dorsal and Dif function similarly, since Dif can rescue Dorsal mutants in establishing the dorsoventral axis (Stein et al., 1998).

Discovery of an Alternate Dorsal Isoform

It was noticed in the early 1990's that following immune challenge, levels of a 4.4kb Dorsal transcript increased (Reichhart et al., 1993). This product was nearly double the size of the maternal Dorsal transcript. This larger transcript was later characterized and termed Dorsal-B (Gross et al., 1999). The protein product of Dorsal B contains 994 amino acid residues. As seen in figure 2, the A and B forms share the Rel homology domain, encoded by the first six exons, but diverge in their C-terminal domains. Dorsal B is an alternative splice variant in which a large intronic portion of Dorsal A is not spliced, retaining a coding region with a translational stop site near the 3' end. The Dorsal B transcript terminates at a polyadenylation within the unspliced exon. Gross and colleagues showed that Dorsal B is expressed only zygotically, differing from the maternally and zygotically expressed Dorsal. The researchers also explored the possibility of synergistic effects between Dorsal and Dorsal B in transcriptional activation. They co-transfected Dorsal and Dorsal B expression vectors into Schneider cells along with a luciferase reporter gene. The results indicated that when 1µg of both forms was transfected into the cells there was an increase in induction by 8-fold over transfection of 1µg Dorsal alone.

Dorsal B Isoforms are not specific to *Drosophila.* In 2005 Shin et al classified two isoforms of the *Aedes aegypti* Dorsal homolog Rel1. The Rel1-A specific splice product is an 844 residue peptide, while that of Rel1-B gives an 883 amino residue peptide that diverges from Rel1-A at amino acid 330. The B form was found to have a bipartite NLS, which can also be found in Dorsal B. The two forms were shown to exhibit synergistic activation of mosquito innate immunity defense molecule expression. In their functional

characterization, however, the researchers only designed dsRNA against REL1, which knocks out both A and B isoforms, leaving little basis for interpretation as to the specific function of each.

Exon-Specific RNAi in *Drosophila*

A potential method to approach function of alternative splice variants is to use RNAi to knock out each form separately. In order to use RNAi against isoforms of the same gene an exon-specific approach must be taken. There are, however, known systems of RNAi mechanisms that prevent exon-specific RNAi in some organisms. The transitive RNAi mechanism, which occurs in *C. elegans* (Sijen et al., 2001), results in knockdown of regions upstream and downstream of the targeted portion of the mRNA transcript. When Dicer cleaves the targeted mRNA, the secondary siRNA fragments are directed via an RNA-dependent RNA polymerase (RdRP) to other regions throughout the mRNA. Thus, if an isoform-specific exon is targeted, the transitive system detects and knocks down gene portions up or downstream that are common to all splice variants. This system, however, is not active in *Drosophila* cultured cells (Celotto and Graveley, 2002). Researchers have found that dsRNA-producing transgenes that were targeted to specific regions of alternative mRNA transcripts outside of the initial target allowed for isoform specific RNAi in *Drosophila* (Roignant et al., 2003). They hypothesize that the mechanisms that promote transient RNAi effects in other species are not conserved in *Drosophila*.

Thesis

In this work I show that Dif is *Drosophilid* specific and that it likely arose via duplication of the Dorsal gene. I then show that, like Dorsal, Dif exhibits an alternately spliced and alternately terminating isoform which I will call Dif B. I will demonstrate that this isoform is expressed *in vivo* and that it is a component of the Toll pathway. Through the use of bioinformatics, molecular genetics and exon-specific RNAi, I show that Dif B is an alternative isoform of the *Drosophilid* specific NF-κB transcription factor Dif and that it acts as an inhibitory component within the Toll pathway innate immune response.

Chapter 1 Figures
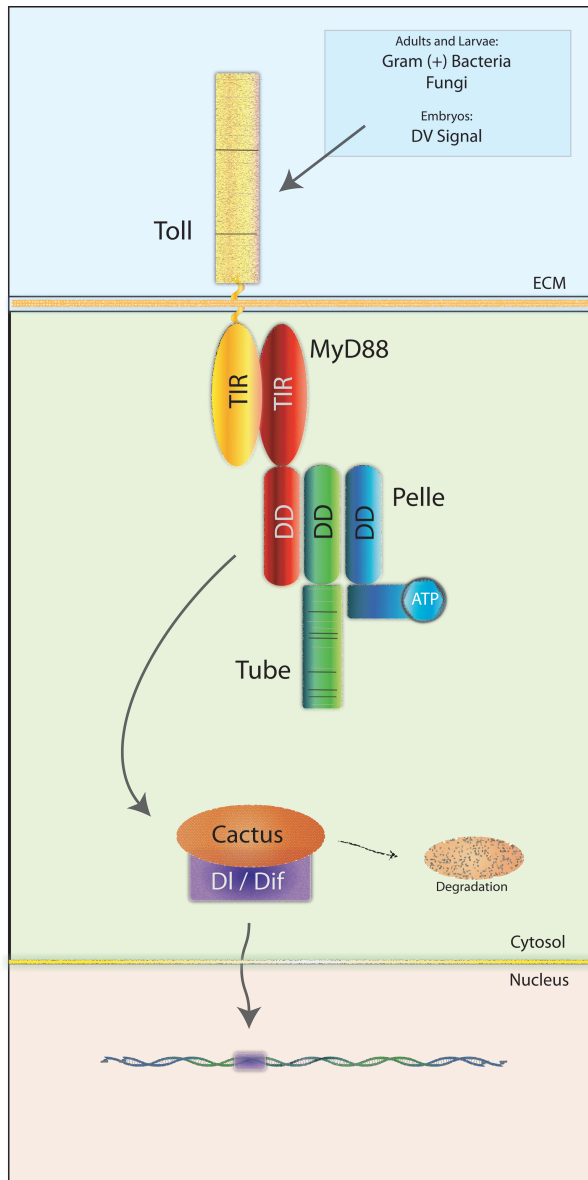
Figure 1. Drosophila Toll Pathway

The Toll pathway is activated in response to invasion from gram (+) bacteria of fungi in larvae and adult flies, or dorsoventral signaling in early embryos. A signal from the trimeric complex between MyD88, Tube and Pelle leads to ubiquitination and subsequent proteosomal degradation of Cactus, unmasking the NLS of Dif and Dorsal allowing them to translocate to the nucleus where they activate the transcription of innate immune genes.

Figure 2. Dorsal B Characterization

Adapted from Gross et al 1999. A) A comparison of the functional domains of Dorsal and Dorsal B. B) A comparison of the mRNA structures of Dorsal and Dorsal B and the generation of Dorsal B by alternative splicing.

Chapter 2:

Identification and Molecular Genetic Characterization of
an Alternate Dif Isoform

Here I show that Dif is a *Drosophila*-specific product that likely originated as a duplication of the Dorsal gene. I demonstrate the remarkable simil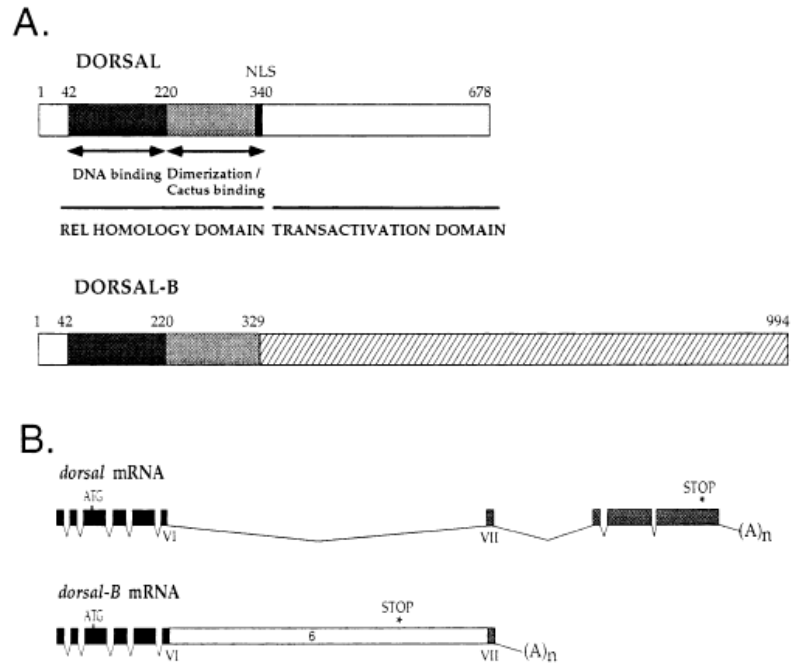arity between the Dorsal and Dif gene regions as a basis for the discovery of the alternate Dif transcript, Dif B. I then detail the steps taken to show the presence of the expressed Dif B protein *in vivo* and to further classify the sequence and translation of the product.

Dif is *Drosophilid* Specific

Using the amino acid sequence of *Drosophila melanogaster* Dif as a *tblastn* algorithm query, I searched for Dif orthologs in other insects such as *A. aegypti, A. mellifera* and *T. casteneum* and selected sequences with high statistical similarity.

I then used each of these sequences as a query in a *tblastn* search against *Drosophila*. In each case the search resulted in Dorsal protein as the top match statistically, with Dif having a less significant match (Figure 3). Due to its similarity to other genes on the phylogenetic tree, Dorsal appears to be the ancestral gene to Dif. Dif, it seems, is a Dorsal-related gene that requires further inquiry into its origin.

I next asked whether a Dif-specific protein could be found in other species, or if it is confined to a smaller branch on the evolutionary tree. I looked at what amino acid sequence motifs distinguish Dorsal proteins from Dif proteins as a basis for a search as to when the Dif gene might have arisen. Conserved sequences, specific to either form, flank

and include the RHD and allow for differentiation. Note that each of these sequences was also compared against Relish, another *Drosophila* NF-κB protein, to determine and validate specificity. Figure 3 shows the amino acid sequences specific to Dorsal-like proteins highlighted with blue. Each of the Dorsal specific sequences, CTLEINS and DKKAMSDL, lie within the RHD and a BLAST search reveals their presence across species, including Anopheles and Tribolium. Figure 3 also displays the amino acid motifs specific to Dif highlighted in red. The first, SLPVMPSHIPL, lies in the N-terminus just prior to the start of the RHD. The other sequences, PPEERRL and AKFDHKD, lie within the RHD, but are not found in Dorsal or any Dorsal homolog. These sequences are the basis for an evolutionary search as to when the Dif gene might have arisen. A protein BLAST search using the Dif specific motifs as a query did not reveal any homologs of Dif outside the *Drosophilid* branch of Brachycera flies. The top hit is always highly similar (near 100%) to the *Drosophila melanogaster* Dif product, showing very little divergence within the species. Outside of the species however, no variation of Dif, as opposed to Dorsal, can be found with any degree of divergence within the NCBI database. One theory states that Dif is specific to brachycera (Wiegmann et al., 2003), an idea that can be tested in the future with the continual additions of sequenced genomes.

Origin of Dif

Dif appears to have arisen by duplication of the Dorsal gene, based on evidence of similarity, proximity and splice structure.

Bioinformatic comparative analysis thoroughly establishes that Dif and Dorsal are closely related in both sequence and structure. As mentioned above, BLAST analysis

reveals the high degree of similarity between the two proteins. A direct alignment comparison between the two proteins shows regions of high conservation, especially within the N-terminal rel-homology domain. The protein Relish is another *Drosophila* immunity factor containing a RHD. Within this domain, Relish matches closely to both Dif and Dorsal. Outside of the RHD, however, the similarities between Dif and Dorsal are much more prominent than those of the two proteins with Relish. Even as far out in the evolutionary tree as humans, the Dorsal homolog p65 matches much more closely to Dif than the relish homolog p52. Within and across species, no protein exhibits a closer homology to Dorsal, including Dorsal isoforms, than Dif.

The splice patterns of the Dorsal and Dif transcripts offer further evidence of Dif's origin. Figure 4 shows the splice pattern exhibited by the "A" forms of the two genes. There is a striking similarity in the overall pattern and conservation of splice sites relative to protein coding regions. A 5'UTR is followed by a large intron region followed by exon 5 containing the rel homology domain coding sequence. This is followed by another large intron region with small exons interspersed and ending with a final exon and 3'UTR.

Finally, the Dorsal gene region is shortly followed at the 3' end by that of Dif along the 2L arm of the *Drosophila* genome. The two gene regions are separated by a 10kb span of DNA consisting of only two small protein products. Such a close localization on the chromosome between two closely related proteins is indicative of a gene duplication event that occurred approximately 70 million years ago around the time when the *Drosophilids* emerged (Wiegmann et al., 2003).

Alternate Isoform of Dif

Given that Dif appears to be an ancient derivative of Dorsal, and given the similarities in the structure and splice pattern of the two gene regions, I asked whether there was also an alternative isoform of Dif.

The first approach was to use bioinformatics to analyze the Dif gene and determine if expression of alternative sequences is possible. I looked for what, if any, conservation existed at the protein level. Open reading frame analysis revealed an approximately 3kb ORF 3' to the RHD encoding exon that, aside from the first 25 amino acids, is a non-coding portion of Dif A. The 25 amino acids at the start of the frame correspond to the last 25 amino acids of the RHD. The frame continues for 3kb from these residues until encountering a UAG stop signal. The open frame, not including the 25 RHD-encoding nucleotides, was translated and used as a BLAST query. The top hit was of Dorsal B in *Drosophila* and other Dorsal B isoforms from *Anopheles* and *Apis.*

I then wanted to ascertain how widespread this B isoform was within the evolutionary tree. I conducted an amino acid residue comparison across multiple species. In order to do this, the initial BLAST result was screened to find the largest and most conserved C-terminal motifs between the Dif sequence and Dorsal B in *Drosophila*. The N-terminal RHD conserved motifs were omitted due to their known conservation across species. The C-terminal amino acid sequences were then used for *tblastn* against nucleotide collections, non-human non-mouse ESTs and whole-genome shotgun reads. The resulting matches extended beyond the *Drosophilid* query species to include *Aedes, Apis* and *Culex.* Those sequences were then used to create an alignment of the amino acid

residues shown in Figure 5. The result was several highly conserved motifs across several species suggesting that the Dorsal B-like isoforms are indeed widespread.

Another objective was to evaluate the Dif transcript splice sites, particularly the 5' and 3' sites flanking the intron that is retained as an exon in an alternate splice variant. This query was performed by Sharon Torigoe, another student in the lab. The 5' splice site is located at the 3' end of exon 5 in the Dif A gene and has the sequence CAG|gtacag. The consensus sequence for the 5' end of a spliced intron is MAG|gtragt, (M= A or C; r= a or g). The Dif splice site completely diverges from the consensus for the last three nucleotides. All splice sites in Dif were analyzed and this site was the most divergent of the six splice sites. Note that nucleotides 3-5 of the 5' consensus, which are essential for splicing, are preserved in the site. The 3' Dif splice site carries the sequence tag|A, compared to the 3' consensus cag|G. The second and third nucleotides match the consensus while the first and last diverge. A Clustal W alignment reveals that the non-consensus sequences are perfectly preserved across the *Drosophilids* including *D. erectus, D. yakuba* and *D. simulans.*

 Expression and Sequencing of an Alternate Dif Isoform

The information gathered from the bioinformatic screen presented clear evidence that an alternate form of Dif could exist in Drosophila. This hypothesis was tested by screening whole fly and S2 cell culture cDNA using probes specific to the alternative isoform. Here I show that an alternate isoform of Dif does indeed exist *in vivo,* which I will refer to as Dif B.

Using reverse transcription, Sharon Torigoe demonstrated the existence of Dif B in adult flies. Whole male and female adults flies were ground with a pestle and their RNA isolated and purified. The RNA extract was then reverse transcribed via RT-PCR into cDNA which was probed to the region specific to Dif B. Primers were designed to amplify a region specific to a B isoform (Figure 6a). The forward primer was designed to sit near the 3' end of exon 3 within the RHD. The reverse primer was designed complementary to an intron region beyond exon 4. This region would, thus yield product from reverse transcription only if it exists as exon in an alternative isoform. Figure 6b shows the expression of Dif B in adult whole flies. Although some genomic DNA contamination is present, the 700bp Dif B signal from whole fly cDNA is robust. Thus, the region of interest is indeed expressed as coding sequence, although little concerning the structure of the alternate isoform could yet be discerned.

At this point we knew that an alternate isoform existed, but we only knew that it was present in whole flies. Knowledge of Dif B's expression in S2 cells was needed in order to establish that cell culture assays could be performed to explore Dif B function. Additionally, a matter of interest was to monitor induction of Dif B within the Toll pathway. In order to assay S2 cell expression, RNA was isolated from two samples of *Drosophila* S2 cells, one population induced, the other uninduced. Both populations were diluted 1:4 or 1:20. Figure 6c shows the Dif B cDNA expressed in S2 cells. Induced S2 cells show an approximately two-fold increase in Dif B expression over uninduced cells.

Characterization of Dif B

Knowing that Dif B is made in flies, we were next interested in the sequence and structure. We therefore sequenced the *in vivo* expressed Dif B cDNA using a rapid amplification of cDNA ends (RACE) procedure. Due to the similarity found in 5' UTR of the Dorsal B homologs to their Dorsal counterparts, we hypothesized that the 5' end of the Dif B transcript would not diverge from that of Dif A. On the other hand, the structure of the 3' end was difficult to predict without further analysis, aside from a probable similarity in location and structure to Dorsal B. It was hypothesized that, like Dorsal B, the Dif B mRNA transcript would contain an alternate transcriptional termination site.

The result of the RACE (not shown) reveals that the hypotheses concerning the structure of the isoform held. Dif B exhibits an identical 5' UTR sequence to that of Dif A in all aspects. The 3' RACE result shows that the Dif B UTR terminates prior to the start of the coding sequence for the final exon of Dif. The resulting 3' UTR closely mimics that of Dorsal B in sequence structure and relative location within the gene.

Further characterization of the Dif B sequence provided further evidence for Dorsal gene duplication as the origin of Dif. The Dif B transcript, like Dorsal B is identical to the A isoform up to the final amino acid of the RHD (Figure 7). The remainder of the transcript is B-form specific, completely divergent aside from exon 5 of Dif A that is incorporated, out of frame, into the Dif B 3' UTR. Figure 8 shows the nucleotide sequence of Dif B and the corresponding amino acid translation. The transcript is 4.3kb nucleotides with an ORF encoding a 987 residue protein. The rel homology domain is highlighted in boldface.

Construction of Dif B cDNA

Studies of Dif B by GFP tagging for localization analysis, co-transfection assays and antibody synthesis would require expression of the gene in bacteria or flies. As a basis for these, and other future experiments I constructed a cDNA specific to Dif B. The strategy in developing the cDNA was to make a hybrid construct from existing Dif cDNA and a PCR product specific to Dif B.

Due to the fact that Dif A and Dif B share an exact N-terminal sequence up to the end of the RHD, I planned to excise the C-terminal Dif A-specific fragment from the existing Dif cDNA and insert a Dif B specific construct. A transcript unique DraIII restriction cleavage site is approximately 8 nucleotides from the 3' end of the transcript common region. A NotI restriction site was encoded within the pNB40 vector just downstream of the 3' polyadenylation within the existing Dif cDNA.

Figure 9 demonstrates the strategy encompassing the construction of the Dif B cDNA. A nested PCR approach was taken to generate a smaller fragment of DNA from genomic template, which could then be used to construct the Dif B-specific sequence. Primers were used to amplify an approximately 3kb length of DNA corresponding to several nucleotides upstream of the C-terminal encoding region of the RHD and several nucleotides downstream of the 3' end of the hypothetical Dif B transcript. I then designed primers that would generate a 2.7kb sequence corresponding to Dif B. The forward primer included the DraIII restriction site followed by the remaining nucleotides of the common Dif transcript sequence; this was immediately followed by the first nucleotides

of the Dif B specific coding sequence. The reverse primer consisted of the final nucleotides of the Dif B specific 3'UTR fused to sequences within the vector up to and including the NotI restriction enzyme cleavage site.

The existing Dif cDNA was then cut using the DraIII and NotI restriction enzymes and loaded to an agarose gel to excise the larger vector fragment. This piece was then ligated to the engineered Dif B sequence and transformed into competent cells. The cDNA was then examined by restriction enzyme digestion to ensure fidelity.

At the time of completion of this work a Dif B cDNA was created. However, the final product was contaminated with large amounts of the original Dif cDNA that had re-ligated within the reaction. Due to the expiration of my time with the project, further purification will be delegated to a future lab member.
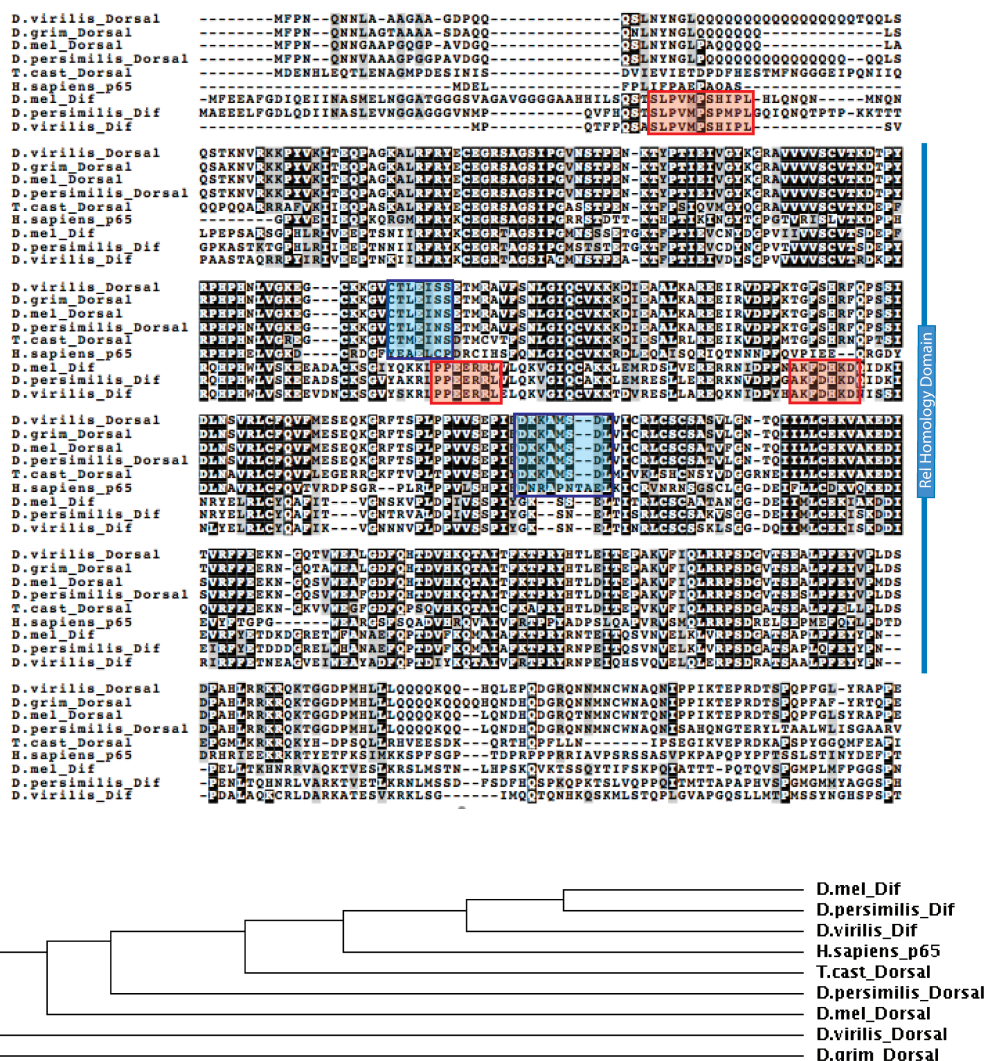
Chapter 2 Figures

Figure 3. Alignment of Dif and Dorsal Rel Homology Domains Across Species

A) An alignment of the N-terminal portion of Dif in *Drosophilids* and Dorsal in *Drosophilid, Tribolium* and *H. Sapiens* p65. This alignment does not include the remainder of the C-terminal transactivation domain. Dorsal-specific sequences are highlighted in blue while Dif-specific sequences are highlighted in red. Below the alignment is a cladogram tree generated based on the statistical conservation of the aligned residues.
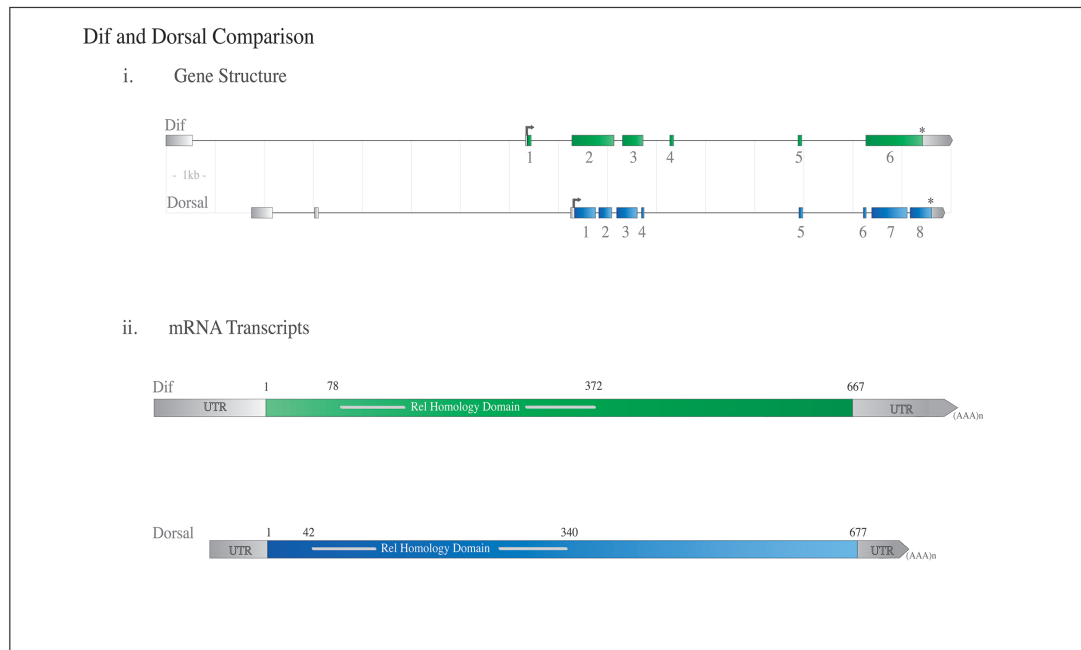
Figure 4. Structure of Dorsal and Dif Reflect Gene Duplication

i) The gene regions for Dorsal and Dif are illustrated. On the chromosome, only 10kb separates the 5' end of Dif from the 3' end of Dorsal. Note the similarity in splice structure between the two genes. ii) A comparison of the mRNA transcripts of Dif and Dorsal. An approximately 300 residue rel homology domain is encoded at the N-terminal region. Dif mRNA yields a 667 residue product while Dorsal has 677 amino acids.

Figure 5. Broad Conservation of C-Terminal Domains of B Isoforms

A comparison of Dorsal B and Dif B in *Drosophilids, Apis Mellifera (honeybee), Aedes aegypti(mosquito) and Culex pipiens (mosquito)*. The comparison is of the C-terminal domains specific to the B isoforms. The N-terminal RHD is not shown.

Figure 6. Dif B Expression

A) Highlighted in blue is the target region of the Dif gene to which primers were designed to detect Dif B expression. Note that the reverse primer is complementary to intron sequence in Dif A. B) Expression of Dif B in whole flies. RNA from male and female flies was isolated and reverse transcribed. Primers specific to the region in 5a were then used to detect the ~0.7kb Dif B-specific fragment. Lane 1 is the amplified genomic product at ~1kb. Some genomic contamination is present in the whole fly cDNA samples. Dorsal and Dif cDNA were used as controls. C) Detection of expressed Dif B in S2 cells. RNA from S2 was isolated and reverse transcribed. Primers specific to the region described above were then used to detect the ~0.7kb Dif B-specific fragment. The Toll pathway in the S2 cells was either induced or uninduced prior to RNA isolation

A. Primer Target
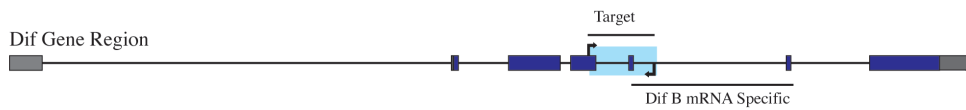
B. Dif B Expression in Whole Flies
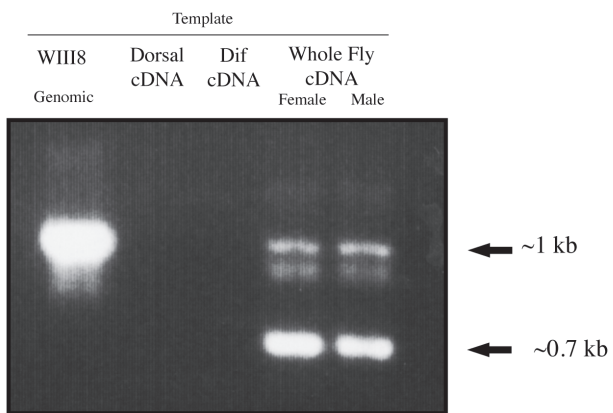
C. Dif B Expression in S2 Cells
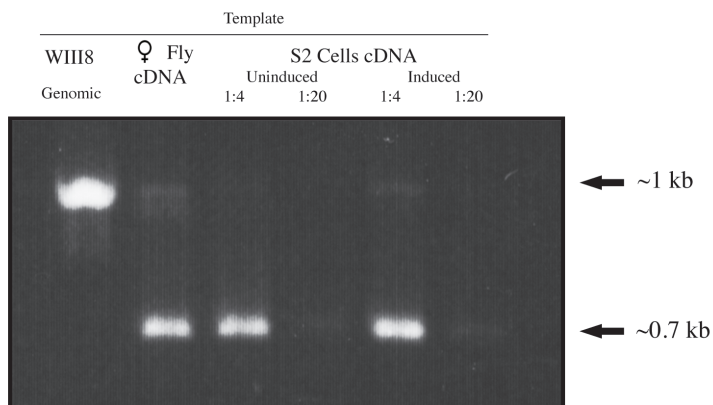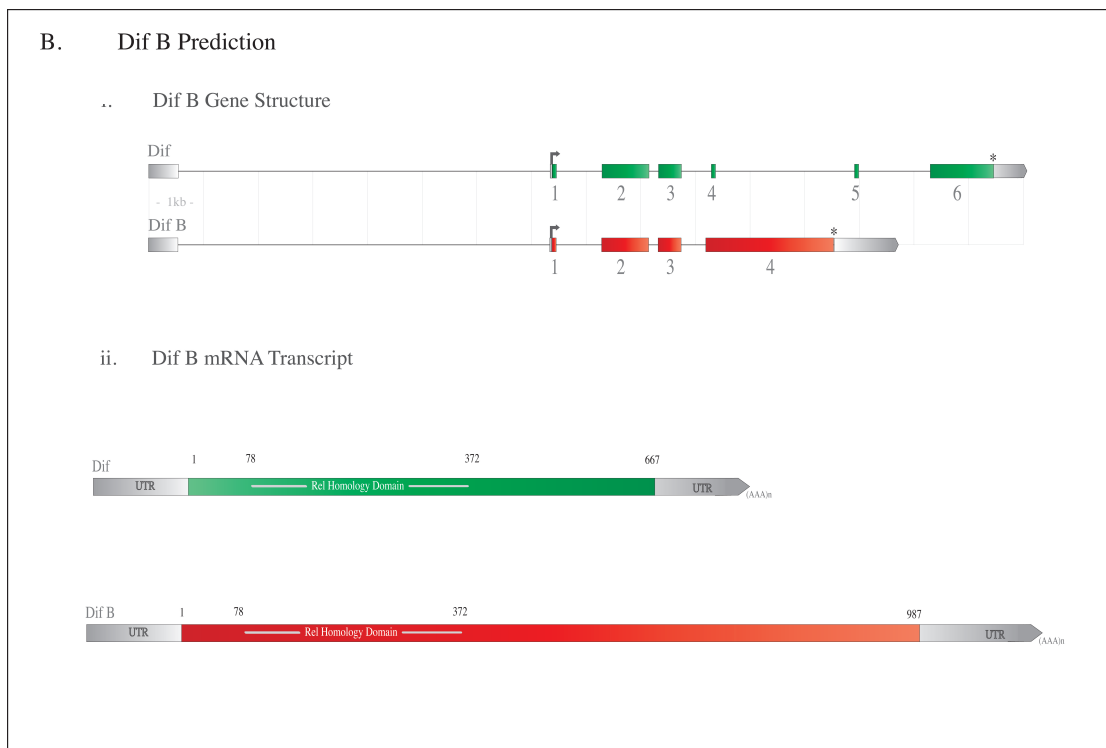
Figure 7. Dif B Characterization

i) A comparison of the Dif gene structure to that of Dif B. The Dif B gene is identical to Dif up the 3' end of Dif exon 4 where an intron splicing does not occur and the large Dif B exon 4 is retained, creating the alternative splice variant. Note the early termination of the Dif B transcript relative to Dif. ii) The mRNA transcript of Dif B encodes a 987 amino acid protein product. The transcript is identical to Dif up to the 3' end of the RHD at residue 372.

```
AGTTTCAATCTGAACTTGGCCGCCGACGGACGTGAAGCGCCGACTTGTTCACACATAGATATCCATACATATATATATACATATAGATGTACATATAACG   100
CATATTGTGCAAGTCGCACGCGACTATATAGCTGCCAGTGTATTGGTTAATTATACGAAGCACGACAACTACAACGTAAACTACAGATGTGTGTATAAAT   200
CAATGTCCGCTCTAAACAGGCGGCTGCAGTGTCAATAGCAAAAATTAAACAAAACTCAAAAGTAAAATACATGAGCGGAGCAACAACGTCGCTTGTGGGA   300
GCAACTTTTGTTTTTGTTTTCAATTTCTTCATAATTTTTTTGGCGCGCCTTTTGTGAGTCGCGCGCGATAAGTCCGACACATAAACCGAAAACCAGTCAA   400
TCATAAAAGACAAATTTAATCAATTACACGTGTCAACACAACAAAGAGCTAATTCGATTTCTGTCTGACCCAGTGCATTCGGATCGAGAATCCAAATTTA   500
ATCAAGCATTTAAAGCAGCCTACCTACTACACCATAAAAGGCAATCAACTGTGTAATCCAATGTTTGAGGAGGCTTTCGGCGGATATACAGGAAATTATCA   600
                                                        M  F  E  E  A  F  G  D  I  Q  E  I  I  N
ATGCCAGCATGGAGTTGAATGGTGGCGCCACCGGCGGAGGCAGTGTGGCAGGAGCTGTTGGCGGAGGCGGTGCTGCACATCACATATTATCGCAGTCGAC   700
  A  S  M  E  L  N  G  G  A  T  G  G  G  S  V  A  G  A  V  G  G  G  A  A  H  H  I  L  S  Q  S  T
TTCCCTGCCGGTAATGCCGTCGCACATTCCGCTCCACCTGCAGAATCAGAATATGAATCAGAATCTGCCCGAGCCAAGTGCAAGAAGTGGTCCCCACCTG   800
  S  L  P  V  M  P  S  H  I  P  L  H  L  Q  N  Q  N  M  N  Q  N  L  P  E  P  S  A  R  S  G  P  H  L
CGTATCGTGGAGGAGCCGACAAGCAATATAATCCGCTTTCGCTACAAATGCGAGGGTCGCACCGCCGGTTCGATTCCGGGCATGAACTCCAGCTCGGAAA   900
  R  I  V  E  E  P  T  S  N  I  I  R  F  R  Y  K  C  E  G  R  T  A  G  S  I  P  G  M  N  S  S  S  E  T
CGGGCAAGACCTTTCCCACCATCGAAGTGTGCAACTACGATGGACCCGTCATCATCGTGGTCTCCTGTGTGACAGCGACGGCCGCTGGTGCTCCAAAAGTG   1000
  G  K  T  F  P  T  I  E  V  C  N  Y  D  G  P  V  I  I  V  V  S  C  V  T  S  D  E  P  F  R  Q  H  P
CCACTGGCTGGTCAGCAAGGAGGAGGCGGATGCCTGCAAGTCGGGCATCTACCAAAAGAAATTGCCGCCAGAGGAGCGGCGGCTGGTGCTCCAAAAAGTG   1100
  H  W  L  V  S  K  E  E  A  D  A  C  K  S  G  I  Y  Q  K  K  L  P  P  E  E  R  R  L  V  L  Q  K  V
GGCATACAGTGCGCCAAGAAGCTGGAGATGCGCGACTCGCTGGTGGTGGAGAGGGAAAGGAGAAACATCGATCCCTTCAATGCCAAATTCGATCACAAGGACC   1200
  G  I  Q  C  A  K  K  L  E  M  R  D  S  L  V  E  R  E  R  R  N  I  D  P  F  N  A  K  F  D  H  K  D  Q
AGATCGACAAGATCAATCGGTATGAGTTGCGCCTCTGCTACCAGGCCTTCATCACAGTGGGCAACTCAAAAGTGCCCCTGGATCCCATCGTATCCTCACC   1300
  I  D  K  I  N  R  Y  E  L  R  L  C  Y  Q  A  F  I  T  V  G  N  S  K  V  P  L  D  P  I  V  S  S  P
GATTTACGGCAAGAGCAGCGAACTGACCATCACCCGGCTGTGCAGTTGCGCGGCCACTGCGAACGGCGGCGACGAGATCATCATGCTGTGCGAGAAGATT   1400
  I  Y  G  K  S  S  E  L  T  I  T  R  L  C  S  C  A  A  T  A  N  G  G  D  E  I  I  M  L  C  E  K  I
GCCAAGGACGACATCGAGGTGCGATTCTATGAGACGGACAAGGATGGACGGGAGACGTGGTTCGCCAACGCAGAGTTCCAACCCACGGACGTGTTCAAAC   1500
  A  K  D  D  I  E  V  R  F  Y  E  T  D  K  D  G  R  E  T  W  F  A  N  A  E  F  Q  P  T  D  V  F  K  Q
AGATGGCCATAGCTTTTAAGACGCCGCGCTACAGGAACACCGAGATCACACAAAGTGTTAATGTGGAGCTGAAACTAGTGAGACCCTCGGATGGAGCGAC   1600
  M  A  I  A  F  K  T  P  R  Y  R  N  T  E  I  T  Q  S  V  N  V  E  L  K  L  V  R  P  S  D  G  A  T
GAGTGCCCCACTGCCGTTCGAGTACTACCCGAATCCAGGTACAGTAACCTTTGCCCGGCTACAGCGAAAGTTGAAACGCCGCCAGGAACTGGACGTGTTC   1700
  S  A  P  L  P  F  E  Y  Y  P  N  P  G  T  V  T  F  A  R  L  Q  R  K  L  K  R  R  Q  E  L  D  V  F
CAGCAGATCCTTTCCCTAGACAGCGAAACGACGGCCACGAAATACTCGTGCCCACCGATGGATCTGGAGGAGGATAATAACACGGTTTCATCCGATGACC   1800
  Q  Q  I  L  S  L  D  S  E  T  T  A  T  K  Y  S  C  P  P  M  D  L  E  E  D  N  N  T  V  S  S  D  D  Q
AACAGAGTTCGGGAACTCTCAACGAATTCGAGATAGCCGTCAAGAAAATGCAATGCGTCAGAGGGGTGAATCGCACGAAATGGATGGTGGCACTGTCAC   1900
  Q  S  S  G  T  L  N  E  F  E  I  A  V  K  K  M  Q  M  R  Q  R  G  E  S  H  E  M  D  G  G  T  V  T
CGAGTACACGGATAACGATAACGAACCAGATATGGAAATAGAGCTGGAGGTTCCGCAATTGTTGCCCCAACTGGCCGATGACTGCACCCAGACGTCCACG   2000
  E  Y  T  D  N  D  N  E  P  D  M  E  I  E  L  E  V  P  Q  L  L  P  Q  L  A  D  D  C  T  Q  T  S  T
CCGATGGAGGAGGATTCTGCAGCATCCGCAGCTGCAATCGCGACCCCTGTCCAATGTGGAAACAATCACCGAGTGGATGAAGAGCAGCGAGTTCGAGAGAA   2100
  P  M  E  E  I  L  Q  H  P  Q  L  Q  S  R  P  L  S  N  V  E  T  I  T  E  W  M  K  S  S  E  F  E  R  T
CGGATAGCCTGACGGTGGAGAACGGGGACACGCAATCCCTATCGAATAGCACGGCCCAAACAGCATTTACCAATGTCAGCAGCTTGACAGCCAATACGAC   2200
  D  S  L  T  V  E  N  G  D  T  Q  S  L  S  N  S  T  A  Q  T  A  F  T  N  V  S  S  L  T  A  N  T  T
CATAACGAACCAGTTGGACGAGGAAGCCCACGGTCTTTTGGCCAATGGTTCTCGCAGGGATACCCTCGAGAATCCGGCCGTGAAAGAGCTACCGGAAAAT   2300
  I  T  N  Q  L  D  E  E  A  H  G  L  L  A  N  G  S  R  R  D  T  L  E  N  P  A  V  K  E  L  P  E  N
GAAAAAGAACCGGAAATGGTTTCAGCCAGTGCCCAGGCCTCTATGGATTTAGATGAAAACTTCGATGAGACAGCCATATACGAGCCTACAAATAGCCT   2400
  E  K  E  P  E  M  V  S  A  S  A  Q  A  S  M  D  L  D  E  N  F  D  E  T  A  T  Y  T  S  L  Q  I  A  F
TTAATAATCCCATAGACATAGGATTGCCCAAAGAAAGCAAACAAGCAGCCCCCATTTATCCACCCACCAATCCGTTTGGTCAGTAGATGATGACGAACT   2500
  N  N  P  I  D  I  G  L  P  K  E  S  K  Q  A  A  P  I  Y  P  P  T  N  P  F  G  Q  L  D  D  D  E  L
GGTGGTGCTGACAAATCCACCAGCTCCCAGGATCAAAGTGAACGCTGCTCAGACACCAGCCACGCCCCCAGCCTCACCGCCCCTGCCATTGCCACCTCGA   2600
  V  V  L  T  N  P  P  A  P  R  I  K  V  N  A  A  Q  T  P  A  T  P  P  A  S  P  L  P  L  P  P  R
ACTCCTTCCCCAGTTTCAGATCAGAGATTACCTCCCTTACCACCGAAACCGCAGAGAAACTCCCAGGATCTGAGTGCTGTCAGTGATTCAATTTCTGTGA   2700
  T  P  S  P  V  S  D  Q  R  L  P  P  L  P  P  K  P  Q  R  N  S  Q  D  L  S  A  V  S  D  S  I  S  V  S
GCCGCTCGCGAAACGGCAGCCTAAGCTCCAATAGAACTACACCTTCGCCTATAATTATAATGCGAACCCCTGATCAAAGTCCCACCAAAAGGCAACCGAC   2800
  R  S  R  N  G  S  L  S  S  N  R  T  T  P  S  P  I  I  I  M  R  T  P  D  Q  S  P  T  K  R  Q  P  T
GCCCAGTGCTTCGCCAAAGAAGCGTCAGGGATTCTTCTCGAGATTCTTTTCTCGCCGAAGGAGCAAGACGGATGACGAAGAATCTCTGACACCTGGTGGC   2900
  P  S  A  S  P  K  K  R  Q  G  F  F  S  R  F  F  S  R  R  R  S  K  T  D  D  E  E  S  L  T  P  G  G
AGCCCCAATAAGCAACTACCAGGGGGATCGAAAGCCACAACACCCACCGGCAGTAGGGAACCAAGCGTTGCGCACTTTTCCCTGGGAGATCCCAATCGCA   3000
  S  P  N  K  Q  L  P  G  G  S  K  A  T  T  P  T  G  S  R  E  P  S  V  A  H  F  S  L  G  D  P  N  R  S
GCTCCACAAGATCGATGCAGCCACTGGGAAAAAGTGAGGGTCGAAATGGCAAGCCAGTCGGTCGTAGTTCGAGCAGTGTATCGGGTAAAAGACCTGCCCA   3100
  S  T  R  S  M  Q  P  L  G  K  S  E  G  R  N  G  K  P  V  G  R  S  S  S  S  V  S  G  K  R  P  A  H
CCTGGATGCGGATGTTATTCACATACCATTGAAGGGAGGCGATAGCGAGAACAGTTTGCTACGACCGGAGAGCTACTCGAATGCCAGCACCCTTAGTTAC   3200
  L  D  A  D  V  I  H  I  P  L  K  G  G  D  S  E  N  S  L  L  R  P  E  S  Y  S  N  A  S  T  L  S  Y
GGTCGACCGCTGGATAGGAAGACCCTGAGCACTCTGCAACTGGCGGATATTCCCATAAGTGACGGCAATATGGAATTGATTGCCATCGCCGACAGGCAGA   3300
  G  R  P  L  D  R  K  T  L  S  T  L  Q  L  A  D  I  P  I  S  D  G  N  M  E  L  I  A  I  A  D  R  Q  S
GCATCAAGAATCTGTGCGAGGGAGCCTATGGTGTTGTGCTGGATCCCAGTGTGGATCTCTCCGAGGCGGAACACTTTGCTTTGTACACCAGCAAAAGTCC   3400
  I  K  N  L  C  E  G  A  Y  G  V  V  L  D  P  S  V  D  L  S  E  A  E  H  F  A  L  Y  T  S  K  S  P
GGAACCACCATTTGGCGGGGAGTGCGGCGAAGGTGGAGCAGCAACTGGTGGAGTGGATACGACCGATTTTTTGAGTGCGGATGAAATAGCACGGCGGCTG   3500
  E  P  P  F  G  G  E  C  G  E  G  G  A  A  T  G  V  D  T  T  D  F  L  S  A  D  E  I  A  R  R  L
GCGGAGGCAAATGGCTTGGAGTAGATGAATTGGGTTACGGATAACTGATTTGAAGACAAAATGGGATAACGAGATATTTTCTTAAATGTGCGGATTTATA   3600
  A  E  A  N  G  L  E  *
ATCGAATCAGTTAACCAAAACCCTAGCATTAGTTAGTGCTTAACAGCAAAACACACAACTGCACCAACAAATTAAACTTTTCTAGTCGTCTTTTTGTTTC   3700
AGTATTAATTATAATGATTACACAAATTTCTAGCTAAGAACCCCTGAGATCGATCGCTCACCAAATTCAATTGTATACGTATCTTTTTTTAATCGACATA   3800
AAATTTAAATATCTTTGTAGAACTCCTAACCAAACACAATCGCCGTGTTGCCCAAAAAACAGTTGAAAGCCTAAAGCGCAGTCTCATGAGCACCAATTTA   3900
CATCCATCCAAGCAGGTTAAAACCTCCAGTAAGTTTTCACGAATTTTCCGCATTTATACTGGTGCCGGGACTGGGTTTCTTTATAGAAGCATTTCATCTG   4000
TGCAGTTCCTAGAAAAACTCATTAACCACACAAAAATACACAAACTCATCACTCGATGGTGTTATTTATATTGACTAGCAAAAATGTTTGCTGAAAATTC   4100
ACTATTGATAGTACAGGATTGAATACCTTGTGGACCAATGGATAAAGTTGTTGTCCGTCCCAGAATTGTTGGGAAATGTAACGTTGTAGTCGAAGAACCC   4200
CATTAATGAATATATAGACCGTTCGAATAGAAAAATGTTTGTTTCACCCGTTCTTATTGTATTCAATATTAATG**AATAAA**AGATTTTCGTAAAAGTCGTG   4300
GCATTGAA                                                                                           4400
```

Figure 8. Complete Nucleotide Sequence of Dif B.

A translation of the 987 amino acid open reading frame is shown below the respective nucleotides. The rel homology domain is shown in boldface and the polyadenylation signal is shown in underlined bold.

Figure 9. Dif B Hybrid cDNA Construction

A diagram of the steps taken to construct the Dif B cDNA. Nested PCR from a genomic portion of Dif surrounding the region specific to Dif B yielded a fragment to which Dif B specific primers were targeted. The primers were tagged with excess sequence needed to both ligate the insert back into the vector and to insert the added nucleotides necessary for a complete, functional cDNA.

Target PCR Region

Dif Genomic Region

Dif B Specific

Nested PCR

Forward Primer with Tag

Dif B Specific

Reverse Primer with Tag

PCR

DraIII

Common Dif

Dif B Specific

End of Transcript

NotI

Tag

Tag

Ligation

DraIII

Common Dif

pNB40 Vector

Dif B cDNA

RE Digest:
DraIII
NotI

Current Dif cDNA

DraIII

Common Dif cDNA

Dif A Specific

NotI

pNB40 Vector
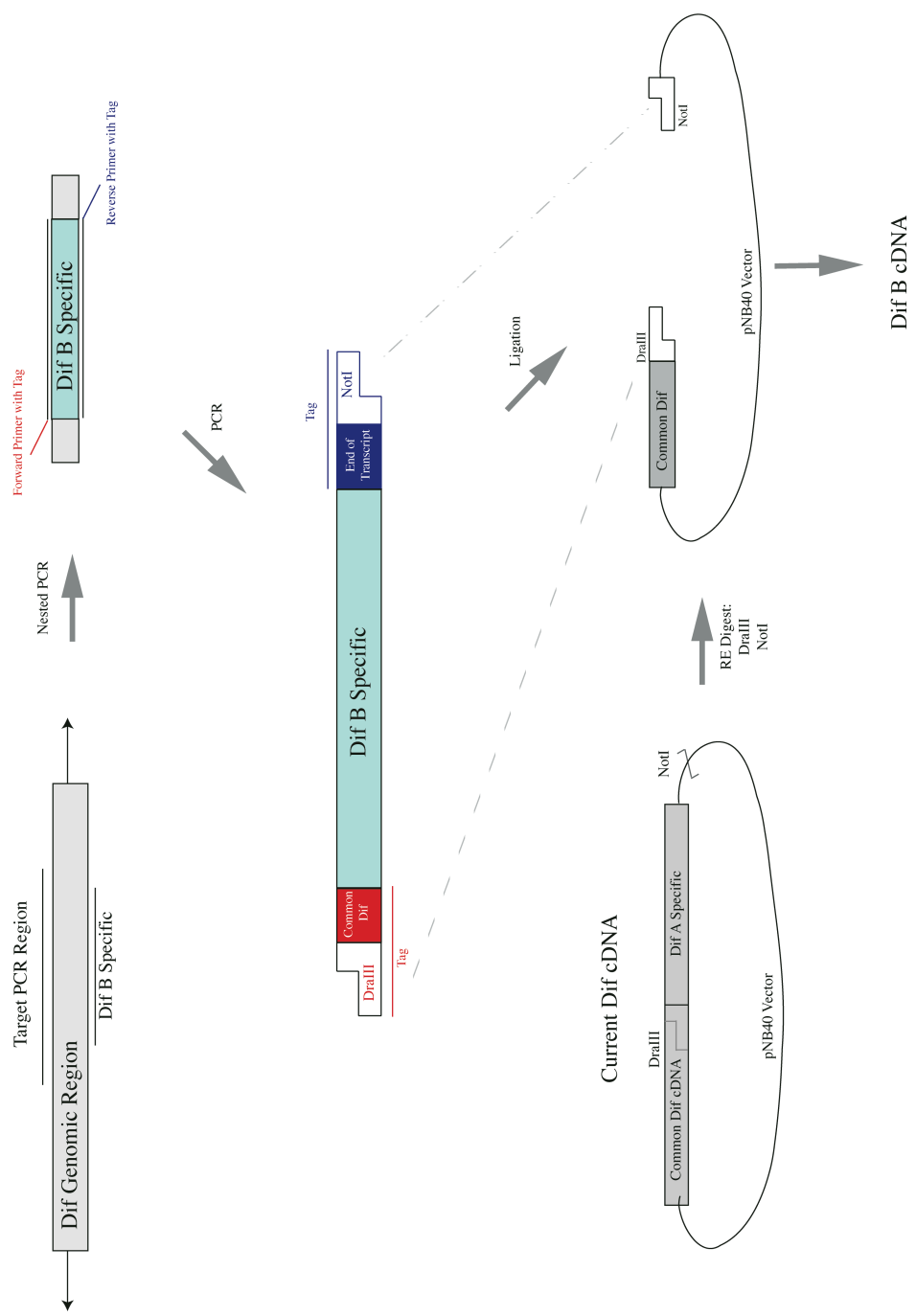
Discussion:
Discovery Analysis

In this chapter I have demonstrated that Dif is an NF-κB transcription factor that likely arose via gene duplication of Dorsal. This event occurred some 70 million years ago around the time of the emergence of *Drosophilid* flies, whereby the Dif protein has not been found outside the species. I have shown that Dif, like Dorsal, exhibits an alternative isoform, which I have termed Dif B and that said isoform exists in whole flies as well as S2 cell culture. I then described the steps taken to engineer a Dif B cDNA for use in future experiments.

Evolutionary Emergence of Dif

At present, my assertion that Dif is *Drosophilid* specific appears to be reasonable. I concede, however, that it is most likely a provisional concept, as so few genomes directly surrounding the *Drosophilid* branch have been mapped. In fact, Dif is similar to Dorsal, but has diverged enough from the ancestral gene to establish its own identity and function. Dif, however, among *Drosophilids* is highly conserved, although most proteins show high conservation among species. In light of these observations the most likely scenario is that the Dorsal gene was duplicated at some point within the *Brachyceran* flies and eventually diverged to become the Dif protein among the *Drosophilds*. In the future as genome mapping becomes a routine task, the evolutionary tree will vastly increase in resolution allowing researchers to pinpoints moments of divergence across

and throughout species. For now, I have identified Dif-specific sequences that can serve as an initial query for a search of Dif homologs.

Analysis of Dif B

I have established that Dif B is an alternative splice variant of Dif encoding a 667 amino acid product that is identical at the N-terminus, including the DNA binding, dimerization and rel-homology domains. Like Dorsal B, Dif B does not contain the NLS of the A isoform. However, analysis of the amino acid residues reveals alternative candidate NLS motifs in two separate locations. Evidence of a bipartite NLS is also found in Dorsal B. This implies that Dif B, like Dif A, could possess the ability to bind DNA and dimerize as well as translocate into the nucleus.

Dif A and B diverge C-terminal to the RHD. I have, as yet, not conducted any experiments to examine the function of the C-terminus, aside from the observation of the bipartite NLS. The initial evaluation of Dorsal B was interpreted as indicating that the C-terminal domain contains scattered negatively-charged acidic motifs that serve as trans-activating domains, similar to Dorsal A and Dif A. Such analysis for Dif B has yet to be accomplished.

Given that Dif A and Dif B share identical protein dimerization domains, the proteins may form heterodimers with each other. This is an interesting topic for further study: Experiments should be done to demonstrate whether or not they actually form homodimers and whether or not they can form heterodimers with Dif A. If true, functional studies could reveal interesting data concerning the three combinations of the proteins, AA, AB or BB. Does each hold a specific function? Along the same lines of

discussion is the question of the ability of Dif B to interact with the IκB homolog, Cactus. The amino acid residues for Cactus binding in Dorsal and Dif occur within the sequences common to all isoforms. So an interaction of Dif B with Cactus is likely.

These additions to the established model for Dorsal and Dif function could yield a more intricate and regulated system of immune response to bacteria and fungi than previously thought. Future study is necessary to further define this portion of the innate immune response and establish a higher resolution model for the interaction of these and other NF-κB homologs.

Potential Genetic Studies with Dif B

The discovery of Dif B and the novel idea that the Dif gene encodes more than one transcript opens the door for genetic studies. Currently there are two established Dif mutants. Dif^1 is a point mutation replacing the phenylalanine at residue 181 with a glycine. Dif^2, also a point mutation, changes phenylalanine at position 245 to a serine. Both mutants map to the RHD of the proteins, meaning that their phenotypic outcome is a genetic response to a loss in both Dif A and Dif B. More specific genetic studies would need to incorporate mutations specific to either isoform. This would require mutations in either exon 6 of Dif A or exon 4 of Dif B. Each of these regions specifically maps to amino acids found within the primary structure of one isoform, but not the other.

Experimental Outlook with Dif B cDNA

Finally, utilization of the Dif B cDNA will provide valuable insight into the functions and properties of Dif B. By adding a GFP tag, for example, one could track the

expression and localization of Dif B in response to immune challenge. Dorsal B is thought to be upregulated and translocated into the nucleus, although not in its native form (Gross et al.). Nevertheless, a similar outcome could be possible for Dif B. One could also use multiple GFP tags to monitor, for example, Dif A and Dif B, allowing one to map the similarities and variations and possibly see interaction or dimerization. Furthermore, co-transfection experiments, similar to those of Dorsal B and *Aedes* Rel, can be performed with updated techniques.

A Dif B antibody would be a desirable reagent. Using the Dif B cDNA, one could express the protein in bacteria and immunize animals to produce the antibody. This could then be used in a multitude of ways to enhance our knowledge of Dif B and other Dorsal B homologs. Western blotting, for example, could reveal roles of Dif B in Toll pathway dynamics and Chip analysis could determine when and if Dif B binds to DNA.

Chapter 3:

Functional Characterization of Dif B

I next wanted to explore possible functions of Dif B within the Toll pathway. Mutational genetics in whole flies presented a problem because the only known chromosomal mutations to Dif, Dif^1 and Dif^2, affect residues that map to the RHD and thus affect both isoforms of Dif. The desired studies, however, would require that only one isoform be manipulated at a time with the other being left unchanged. The solution was to use exon-specific RNAi in *Drosophila* S2 cells that would effectively knockdown either or both forms as needed. In our lab we have developed S2 cells that possess a stably transfected chimeric Toll receptor with a cytoplasmic Toll TIR domain fused to an extracellular epidermal growth factor receptor. The reliability of this modified Toll protein has been proven in past literature (Sun et al.) and allows me to stimulate the Toll pathway as needed with the addition of murine EGF. Transcription of the gene for the anti-fungal agent Drosomycin is activated by Dorsal and Dif following immune challenge. The assay was designed to monitor Drosomycin induction in S2 cells, with the reporter fused to luciferase. A loss of signal efficiency within the Toll pathway corresponds to a decrease in Drosomycin induction and, therefore, a decrease in luciferase light output as measured by a luminometer.

Adaptor protein Tube as RNAi Control

I designed several dsRNA products to target the Toll pathway adaptor Tube to serve as controls. In our lab, we routinely use dsRNA products of ≥500bp targeted to a coding region of a protein. The Tube products were used to control for overall RNAi integrity, dsRNA size variations, essential vs. non-essential region targeting within proteins and UTR RNAi.

The death domain of Tube is essential in maintaining integrity in the Toll pathway. Without it, the signal dies and there is no downstream induction (Grosshans et al., 1994). The results of RNAi of the Tube controls are shown in Figure 10. A dsRNA product was targeted to a Tube death domain encoding exon to serve as an overall control as to the efficacy of RNAi. RNAi for this product reduced induction by >90%. I designed another Tube product targeted to the non-essential repeat region of Tube. RNAi of this product also knocked the signal below 90%. This control informed me that dsRNA products might not need to be designed to target essential regions within the protein of interest. However, studies have shown that the death domain alone is unstable (Sun et al.). As mentioned in the introduction, transient RNAi mechanisms are not active in *Drosophila*. This non-essential region targeting is effective because the entire mRNA transcript containing those target amino acids is knocked down. With this knowledge I could confidently create dsRNA products that target regions outside of the RHD in Dif and Dorsal.

I also asked whether targeting the UTR of Tube would have an inactivating effect. The results showed that UTR targeting is highly effective. 5' UTR RNAi gave an

induction of only 17±4% and the 3'UTR RNAi gave an induction of only 9±0.4%. The combination of the 5' and 3' UTR dsRNA knocked induction down to 4±0.8%.

Finally, the Tube coding region was divided into approximately 250, 500, 1000 and 2000 base pair regions to analyze the effect of smaller versus larger dsRNAs. Targets A thru E represent approximately 250bp dsRNA fragments targeted to the region shown in the inset map. The results showed that size had no effect, with each product effectively eliminating the signal to <90%.

Design of dsRNA to specific Dif Isoforms

To test the function of Dif B in the immune response, I designed dsRNA products to target the RHD domain common to the A and B form of either Dif and Dorsal as well as dsRNAs specific to the A or B forms. The Dif B specific dsRNA was targeted to a 1kb portion of the B form-specific exon 4. This region is an intron in Dif A and therefore not present in its RNAi-targeted mRNA transcript. The Dif A dsRNA was targeted to the sixth exon of Dif A which lies outside of the Dif B gene and includes both coding sequence and UTR. Additionally, control dsRNA products were designed to target introns in both Dorsal and Dif. A map of primers used to design dsRNA products is given in Figure 11a. These controls assured me that the Dif B dsRNA would achieve a specific effect and would not have an effect on the Dif A dsRNA target.

Dif B acts as an Inhibitor of Toll Signaling

To assess the function of Dif B, the assay incorporated all combinations of Dif A and Dif B. The strategy was to first knockdown all Dif isoforms and record the effects. I

would then use exon-specific RNAi to individually target Dif A or Dif B and examine the change relative to the common transcripts. Alternatively the Dif A and Dif B variations were combined with a Dorsal knockdown. This served to minimize the redundancy found in the immune function of both proteins.

The first experiment was to target the RHD, knocking down all forms of Dorsal and Dif. The result was an effective elimination of the Toll signaling response. The induced control was given a value of 1 (100%) and had a standard deviation of ±36%. In all trials, RNAi of the Dorsal and Dif RHD resulted in a knock down of signal to <5% of the induced control (Figure 11b). RNAi of Dorsal alone yielded an average induction of 62±3.5%, and RNAi of Dif gave 67±18%. This is evidence of redundant function between Dorsal and Dif as reported by Busse et al.

The next experiment was a specific knockdown of Dif A, Dif B or both. A Dif A knockdown gave a 163±5.6% induction. A knockout of Dif B alone gave the highest induction of any samples at 314±150% induction. The large standard deviation is indicative of the values obtained from three separate Dif B dsRNA experiments. The individual numbers are 330%, 450% and 160%, all well above the positive control. A knock down of Dif A and Dif B gave a 102±56% induction. This value, theoretically, should mimic the induction of the Dif RHD RNAi due to the fact that both forms are knocked down. Although the average is higher, it does fall within error, not invalidating the hypothesis.

Another experiment run in parallel mimicked the first but included an equal amount of Dorsal dsRNA product along with the Dif A and B combinations. A combination of Dorsal and Dif A RNAi gave an induction of only 15.8±7.3%. This low

value correlates with the Dorsal/Dif RHD knockdown induction. The Dorsal combined with both Dif A and Dif B should, in theory, yield a very low induction as the RNAi knocks down all forms of Dif and Dorsal. The induction however is at 28.6±2.7%, a higher value than expected. When Dorsal is combined with Dif B the induction rises to 152±72%. Again a knockdown of Dif B results in a very high induction, although within error, the number does match the control. Such variation necessitates repeated experiments as this particular combination was only conducted twice. Note that the combination of the Dorsal and Dif B targets gives a higher induction than for the Dorsal knockdown alone, even within error.

The right side of Figure 11 shows induction following RNAi of intron controls. These controls were necessary to demonstrate that exon-specific targeting of Dif B (which is an intron sequence in Dif A) will not have any effect other than that of the knockdown of the Dif B transcript. Two dsRNA products were designed to target the large 5' intron of Dif and Dorsal (See Figure 11a). In theory, targeting of an intron should have no effect on RNAi since splicing occurs within the nucleus while RNAi occurs within the cytoplasm. The results of the single intron RNAi reflect this theory. The Dorsal intron RNAi gave an induction of 116±27.9% while the Dif intron gave 128±7.3% induction. The combination of the two introns, however, shows a markedly lower induction of 72±17.4% . This value is lower than expected if intron-targeted RNAi does not hold any effect. There may be a level of toxicity to the cells involved in this type of RNAi targeting. Interestingly, we have found that Dorsal B is not expressed in S2 cells (data not shown). However, the slight average increase over the control between the Dif and Dorsal introns is the same. This means that if there is an unknown effect of intron

targeted RNAi it has no relevance to the presence or absence of Dif B. Otherwise, there

would be a clear difference between the Dorsal and Dif intron control.

Chapter 3 Figures
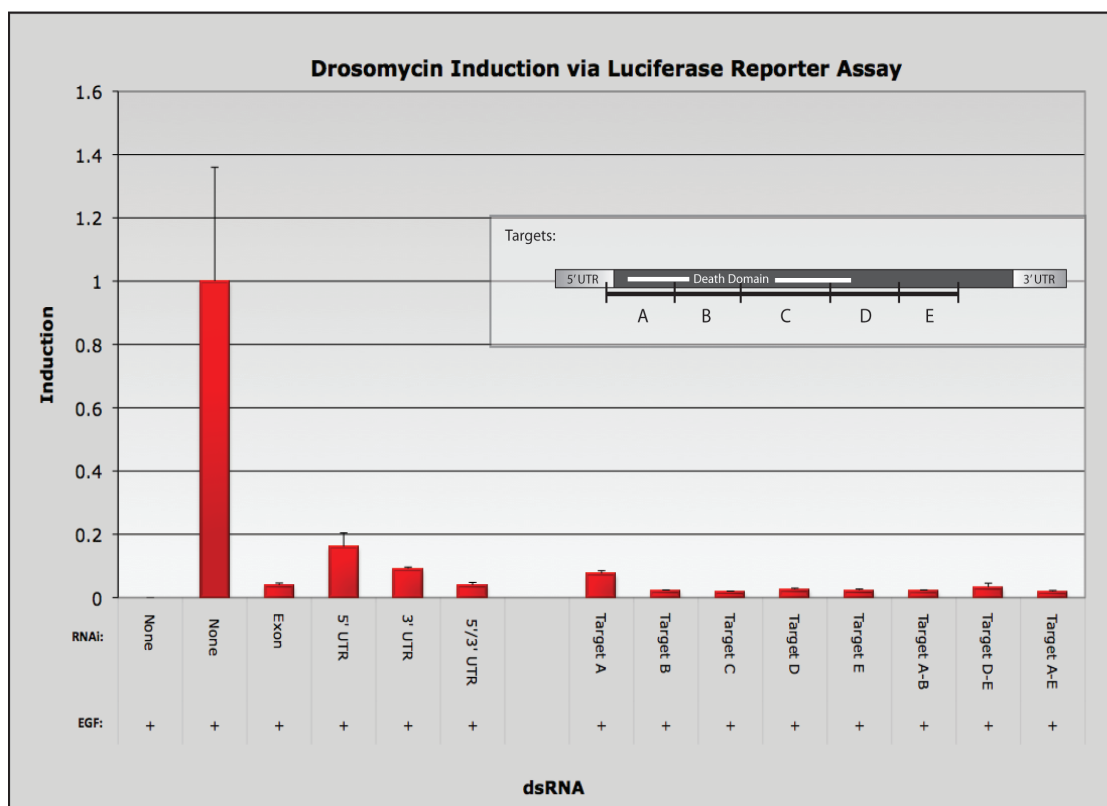
Figure 10. Tube as RNAi Control

Sample 3, labeled "exon", is the default dsRNA product used to RNAi tube and was used as a variable in each RNAi experiment in order to evaluate overall RNAi efficacy. It roughly maps to targets BC on the insert. RNAi of Tube 5' and 3' UTR are sample 4-6. RNAi of varying lengths of dsRNA are in samples 7-14. The insert shows a map of the Tube transcript with corresponding dsRNA targets, A-E. Note that target E lies outside of the death domain. Also note that Target A contains a small portion of 5'UTR sequence.

A. dsRNA Primer Map
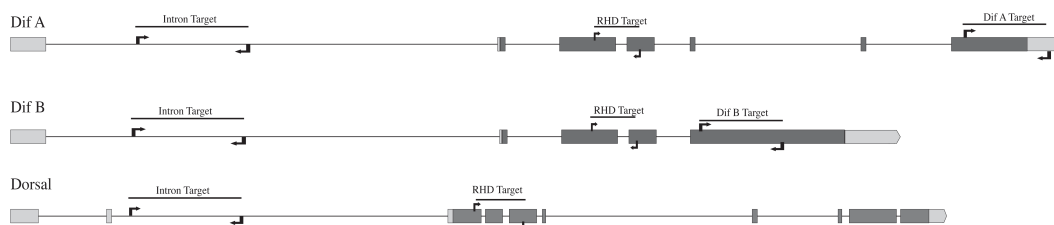
B. Exon-Specific RNAi of Dorsal and Dif
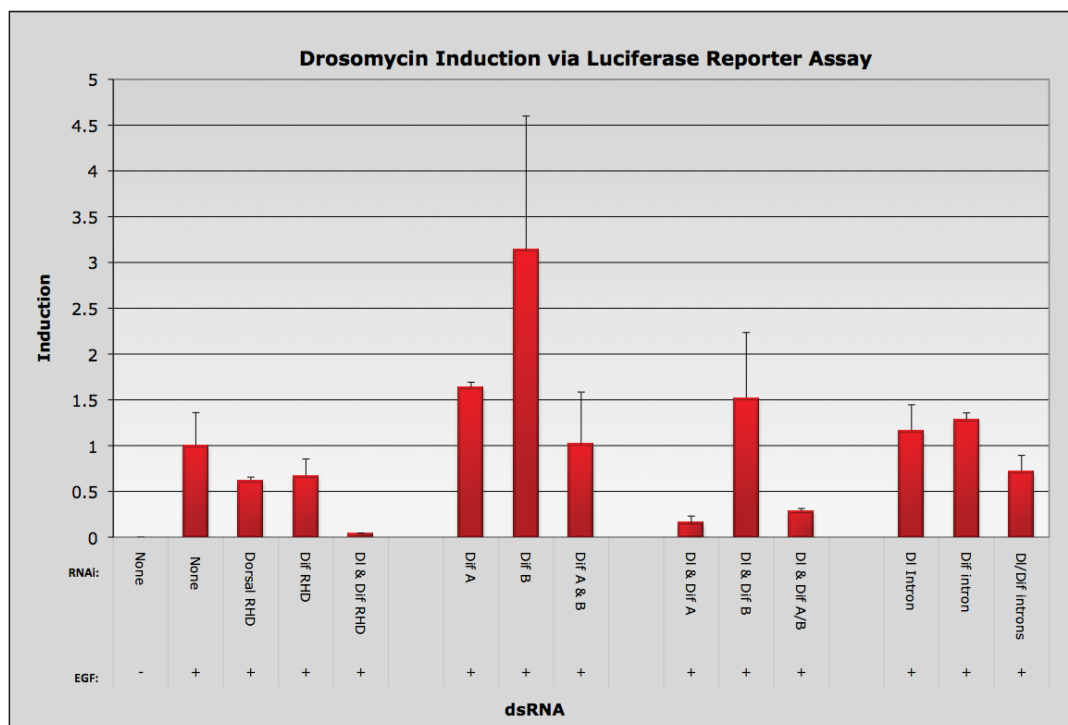
Figure 11. Exon-specific RNAi of Dif and Dorsal

A) A map of dsRNA targets for the Dif and Dorsal genes corresponding to the RNAi results in B. B) RNAi results for knockdown of Dorsal and Dif isoforms, Dif A and Dif B. Those labeled RHD (3-5) target all isoforms of that species. Samples 9-11 represent exon specific RNAi of Dif A and Dif B along with an equal amount of Dorsal RHD dsRNA. Samples 12-14 represent intron targeted RNAi.

Discussion:

Analysis of Dif B Function

Observations from controls

The Tube RNAi controls were of great benefit to my analysis: The UTR RNAi succeeded in a knockdown of induction, target E saw effective knockdown even though it was targeted outside the death domain, and all dsRNA size variations equally brought down Drosomycin induction. The results, however, hold interesting points of discussion.

First, I showed that targeting the UTR of a transcript can effectively knock down target protein levels, especially when targeting both the 5' and 3' regions. This is an important finding because the UTR, by definition, contains no coding sequence. One could, therefore, create a synthetic transcript with a modified coding sequence and replace the UTR to differ from wild type. That modified transcript can then be transfected into cells . If one then targets the UTR of the native protein via RNAi, the native protein will be knocked down, while the protein from the transfected construct will be left untouched. This would introduce several levels of possible experimentation: One could completely delete a region of interest from a transcript and monitor its effects, or one could simply introduce point mutations into a protein as a test. The effects seen would all be as a result of the transfected protein only since the native protein has been knocked down via RNAi.

Second, I showed that dsRNA length and position seem to have no effect on RNAi efficacy. The several dsRNA products constructed varied in size and all of them

knocked down induction below 10%. This knowledge is important to a researcher if a protein of interest has only a narrow region to which dsRNA can be targeted. Equally important is the Tube control target E, which lies outside of the death domain yet still knocks down signal below 10%. In the future, myself and others in the lab can have greater confidence to target transcript regions that are considered non-essential at the protein level, although further study is needed. Again, the importance of this comes into play at times when there is little leeway in RNAi targeting.

A Possible Inhibitory Role for Dif B

The RNAi results suggest that Dif B has an inhibitory role in the Toll pathway, although a mechanism of action is difficult to predict. Knockdown of the Dif B product alone enhances the Toll signal, as does the combination of Dorsal and Dif B RNAi. RNAi of Dorsal leaves only Dif A and Dif B active to work downstream and gives an induction of 62%. The combination of Dorsal and Dif B dsRNA gives an induction of 152%, a 2.5 fold increase over Dorsal alone. The difference between the two being a loss of Dif B, resulting in higher induction.

A negative Dif B role could be a part of a negative feedback system on the Toll pathway following immune response. In mammals there are several levels of feedback systems including the heterodimerization of transcriptional activators with the repressive transcription factors p50 and p52 (Blank et al., 1991). The ability of Dif B to heterodimerize with the Dif A isoform is possible due to the fact that the dimerization domain is identical in both forms. This is only an assumption, however, and further experimentation must be done.

Gross and colleagues suggest a role of Dorsal B to be synergistic activation with Dorsal. Their claim is based on co-transfection experiments with Dorsal and Dorsal-B in which the combination of the two forms leads to a much higher induction than transfection of Dorsal alone. The experiment, however, holds less meaning than RNAi. With RNAi a clear phenotype if given for Dif B knock down: increased induction.

A simple test would be to RNAi Dorsal B and monitor the effect. If it is an activator, induction should drop. If it is inhibitory, the induction should increase. However, we have found (data not shown) that Dorsal B is not expressed in S2 cells. For now we can rely only on RNAi data for Dif B, a close homolog to Dorsal B.

A critical variable that is absent in the Dorsal B co-transfection experiments is induction following Toll activation. In our studies, RNAi results follow EGF stimulus. It is possible that the inhibitory effects of the B isoforms are not active prior to Toll signaling. Other variables that further complicate the model are those of dimerization, DNA binding and nuclear translocation. Although all of these occur in Dif A, there is insufficient data to draw conclusions for Dif B.

Methods

S2*/ERTL Cell Culture

The S2*/ERTL cell line was transfected as described previously (Sun et al., 2004). The cells were grown in Schneider's Drosophila medium (Invitrogen) as described by Sun *et al*, 2002 and further supplemented with 1X Penn/Strep and 1X Fungizone.

RNAi

The RNA interference protocol adheres to the methods of Sun *et al*, 2002. Cells were diluted to $1x10^6$ cells/ml in Serum-Free media and 1ml was added to each well of a six-well plate. Fifteen µg of dsDNA was immediately added to the appropriate well followed by 15 second of vigorous shaking. Two wells were left as controls and contained no dsRNA. The cells were incubated in this state for 45 minutes and then 2 ml of complete media was added. Following 4 days of incubation, transient transfections were performed as described by the *Drosophila* Expression System (Invitrogen), using 100 ng of Drosomycin-Luciferase Reporter and pAc-LacZ reporter plasmids.

Toll Stimulation and Harvest

Approximately 24 hours following transfection, the wells were aspirated and murine EGF was added at 0.1µg/ml. One of the two control wells received no stimulation, but was aspirated and supplemented with new media. After 4 hours the cells were then harvested and lysed as described previously (Busse et al., 2007)

Luciferase Assay

Twenty μl of lysate per sample was assayed for both Luciferase and ß-galactosidase activity via the Luciferase Assay System (Goueli and Jarvis) and the Galacto-Light Plus system (Tropix), respectively. The ß-gal activity was used to normalize the luciferase activity as a means of quantifying the efficiency of transfection. Induction was calculated as the reported normalized value divided by the baseline control of unstimulated sample, assigned an induction of zero.

Bioinformatics

The NCBI database (http://blast.ncbi.nlm.nih.gov/Blast) was utilized for multiple searches with the *tblastn*, *megablast* and *blastp* algorithms. Sequence alignments were performed using the Clustal W program from the EMBI Nucleotide Sequence Database (http://www.ebi.ac.uk/Tools/clustalw/). This tool also accounted for the phylogenetic tree map.

References

Blank, V., Kourilsky, P. and Israel, A. (1991) Cytoplasmic retention, DNA binding and processing of the NF-kappa B p50 precursor are controlled by a small region in its C-terminus. *Embo J*, **10**, 4159-4167.

Bowie, A. and O'Neill, L.A. (2000) The interleukin-1 receptor/Toll-like receptor superfamily: signal generators for pro-inflammatory interleukins and microbial products. *J Leukoc Biol*, **67**, 508-514.

Busse, M.S., Arnold, C.P., Towb, P., Katrivesis, J. and Wasserman, S.A. (2007) A kappaB sequence code for pathway-specific innate immune responses. *Embo J*, **26**, 3826-3835.

Celotto, A.M. and Graveley, B.R. (2002) Exon-specific RNAi: a tool for dissecting the functional relevance of alternative splicing. *Rna*, **8**, 718-724.

Ganchi, P.A., Sun, S.C., Greene, W.C. and Ballard, D.W. (1992) I kappa B/MAD-3 masks the nuclear localization signal of NF-kappa B p65 and requires the transactivation domain to inhibit NF-kappa B p65 DNA binding. *Mol Biol Cell*, **3**, 1339-1352.

Ghiglione, C., Carraway, K.L., 3rd, Amundadottir, L.T., Boswell, R.E., Perrimon, N. and Duffy, J.B. (1999) The transmembrane molecule kekkon 1 acts in a feedback loop to negatively regulate the activity of the Drosophila EGF receptor during oogenesis. *Cell*, **96**, 847-856.

Goueli, S.A. and Jarvis, B.W. (2001) Phospho-specific mitogen-activated protein kinase antibodies for ERK, JNK, and p38 activation. *Methods Enzymol*, **332**, 337-343.

Gross, I., Georgel, P., Kappler, C., Reichhart, J.M. and Hoffmann, J.A. (1996) Drosophila immunity: a comparative analysis of the Rel proteins dorsal and Dif in the induction of the genes encoding diptericin and cecropin. *Nucleic Acids Res*, **24**, 1238-1245.

Gross, I., Georgel, P., Oertel-Buchheit, P., Schnarr, M. and Reichhart, J.M. (1999) Dorsal-B, a splice variant of the Drosophila factor Dorsal, is a novel Rel/NF-kappaB transcriptional activator. *Gene*, **228**, 233-242.

Grosshans, J., Bergmann, A., Haffter, P. and Nusslein-Volhard, C. (1994) Activation of the kinase Pelle by Tube in the dorsoventral signal transduction pathway of Drosophila embryo. *Nature*, **372**, 563-566.

Morisato, D. and Anderson, K.V. (1994) The spatzle gene encodes a component of the extracellular signaling pathway establishing the dorsal-ventral pattern of the Drosophila embryo. *Cell*, **76**, 677-688.

Reichhart, J.M., Georgel, P., Meister, M., Lemaitre, B., Kappler, C. and Hoffmann, J.A. (1993) Expression and nuclear translocation of the rel/NF-kappa B-related morphogen dorsal during the immune response of Drosophila. *C R Acad Sci III*, **316**, 1218-1224.

Roignant, J.Y., Carre, C., Mugat, B., Szymczak, D., Lepesant, J.A. and Antoniewski, C. (2003) Absence of transitive and systemic pathways allows cell-specific and isoform-specific RNAi in Drosophila. *Rna*, **9**, 299-308.

Roth, S. and Schüpbach, T. (1994) The relationship between ovarian and embryonic dorsoventral patterning in Drosophila. *Development*, **120**, 2245-2257.

Roth, S., Stein, D. and Nusslein-Volhard, C. (1989) A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the Drosophila embryo. *Cell*, **59**, 1189-1202.

Rutschmann, S., Jung, A.C., Hetru, C., Reichhart, J.M., Hoffmann, J.A. and Ferrandon, D. (2000) The Rel protein DIF mediates the antifungal but not the antibacterial host defense in Drosophila. *Immunity*, **12**, 569-580.

Sijen, T., Fleenor, J., Simmer, F., Thijssen, K.L., Parrish, S., Timmons, L., Plasterk, R.H. and Fire, A. (2001) On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell*, **107**, 465-476.

Stein, D., Goltz, J.S., Jurcsak, J. and Stevens, L. (1998) The Dorsal-related immunity factor (Dif) can define the dorsal-ventral axis of polarity in the Drosophila embryo. *Development*, **125**, 2159-2169.

Sun, H., Bristow, B.N., Qu, G. and Wasserman, S.A. (2002) A heterotrimeric death domain complex in Toll signaling. *Proc Natl Acad Sci USA*, **99**, 12871-12876.

Sun, H., Towb, P., Chiem, D.N., Foster, B.A. and Wasserman, S.A. (2004) Regulated assembly of the Toll signaling complex drives Drosophila dorsoventral patterning. *Embo J*, **23**, 100-110.

Wiegmann, B.M., Yeates, D.K., Thorne, J.L. and Kishino, H. (2003) Time flies, a new molecular time-scale for brachyceran fly evolution without a clock. *Syst Biol*, **52**, 745-756.