University of Dayton

eCommons

Spring 4-2014

Identifying the DNA Sequence Requirements for a Synergistic Interaction Between Two Cis-Regulatory Elements

Kaitlyn R. Francis
University of Dayton

Follow this and additional works at: https://ecommons.udayton.edu/uhp_theses

Part of the Biology Commons

Identifying the DNA Sequence Requirements for a Synergistic Interaction Between Two Cis-Regulatory Elements

Honors Thesis

Kaitlyn R. Francis

Department: Biology

Advisor: Thomas M. Williams, Ph.D.

April 2014

Identifying the DNA Sequence Requirements for a Synergistic Interaction Between Two Cis-Regulatory Elements

Honors Thesis

Kaitlyn R. Francis

Department: Biology

Advisor: Thomas M. Williams, Ph.D.

April 2014

Abstract

My thesis research studies the genetic material that is the blue print to make animal life. In animals, a key type of genetic material is sequences collectively referred to as cis-regulatory elements (CREs). These sequences control the expression of genes; more specifically they instruct when to turn "ON" or "OFF" the production of a gene's functional product. My research investigates the interaction between the two CREs, the Anterior Element and the Dimorphic Element of the fruit fly species *Drosophila melanogaster*. These two CREs act synergistically to produce a pattern of expression for the bab1 and bab2 genes that differs between male and female flies. As synergistic CRE interactions have seldom been reported, my research has sought to identify the necessary sequences for this interaction. Learning more about CRE functions in fruit flies will facilitate a better understanding as to how CREs function in our own genetic material.

UNIVERSITY *of*

DAYTON

Table of Contents

CHAPTER I: INTRODUCTION

Gene Regulation in Animal Species:

Humans and all animal species are made of various traits such as one's hair color or eye color. The reason animal species possess such varying traits is because of their genetic makeup. Multicellular organisms begin as a single celled zygote which possesses the organism's complete set of genes. Through the process of development the zygote divides and cell differentiation occurs. The organism becomes a complex organism with thousands to upwards of trillions of cells that have various functions. Each cell in the organism's body possesses the entirety of their genes, yet only unique subsets of the genes are expressed in a cell (Davidson, 2006). For example, the genes turned "ON" in an individual's liver cell are different than the genes turned "ON" in an individual's skin cell. Therefore, in the process of development, the genes in the cells of the organism require genes to be capable of switching "ON" and "OFF."

Because genes can determine whether an individual has a genetic predisposition for ailments such as a myocardial infarction (Musunuru et al., 2010), investigating how genes are turned "ON" and "OFF" or regulated is vital for animal species. For instance, the *SORT1* gene is regulated by a sequence of DNA known as a *cis*-regulatory element (CRE) which is located over 120,000 base pairs away from the *SORT1* gene's promoter where its transcription is initiated. A one nucleotide mutation in a CRE controlling *SORT1* expression in the liver causes an increased risk of a myocardial infarction (Musunuru et al 2010). Thus, mutations in CREs have serious health implications and understanding the way genes are regulated and the sequences which regulate them is of vital importance.

Gene Regulation Mechanism:

Particular regions of DNA called *cis*-**regulatory elements** (**CREs**) direct gene expression; specifically when RNAs for genes are transcribed or simply put, whether genes are turned "ON" or "OFF". Each CRE can direct gene expression in a specific cell type or types and at a specific time point(s) during an organism's life span. Within a genome, the full repertoire of CREs can direct expressions of genes in diverse spatial areas, cell types, and at diverse time points (Davidson, 2006). CREs are typically sequences of non-coding DNA that are embedded in introns and flank protein coding regions (Arnone & Davidson, 1997; Carroll, 2005). They may reside near and on the same chromosome as the gene they regulate, far away from the gene they regulate (Bagga & D'Antonio, 2013; Lettice et al., 2003) or even on a different chromosome from the target of regulation (Lomvardas et al., 2006; Miele & Dekker, 2009; Spilianakis, Lalioti, Town, Lee, & Flavell, 2005; Xu, Tsai, & Lee, 2006). For instance, in the gene *Sonic hedgehog (Shh),* a CRE that controls *Shh* gene expression in the developing limb bud of the mouse embryo, lies roughly a million base pairs upstream of the first exon of the *Shh* gene and within the gene *Lmbr1*(Lettice et al., 2003; Rebeiz & Williams, 2011).

CREs possess short DNA sequence motifs that are binding sites for particular transcription factor proteins (Arnone & Davidson, 1997; Davidson, 2006) (Figure 1). Transcription factors possess a DNA-binding domain that interacts with specific DNA sequences that are usually 5-12 base pairs in length. When bound, transcription factors can promote the repression or activation of transcription. Therefore, the transcription regulatory activity of a specific CRE depends on the binding sites it possesses for transcription factors (Figure 1). Since genomes contain hundreds to over a thousand

transcription factors (Pfreundt et al., 2010), the combinations of binding sites and patterns

of expression are vast. In addition, CREs can function in many different contexts. The

signaling proteins, receptors, and most transcription factors whose activities are needed

for CRE function are themselves expressed in multiple cell types, tissues, and organs

(Carroll, 2005). Because signaling proteins, receptors and most transcription factors are

deployed in multiple body parts, the regulatory function and morphogenetic outcome of a

specific CRE is context dependent (Carroll, 2005, 2008).
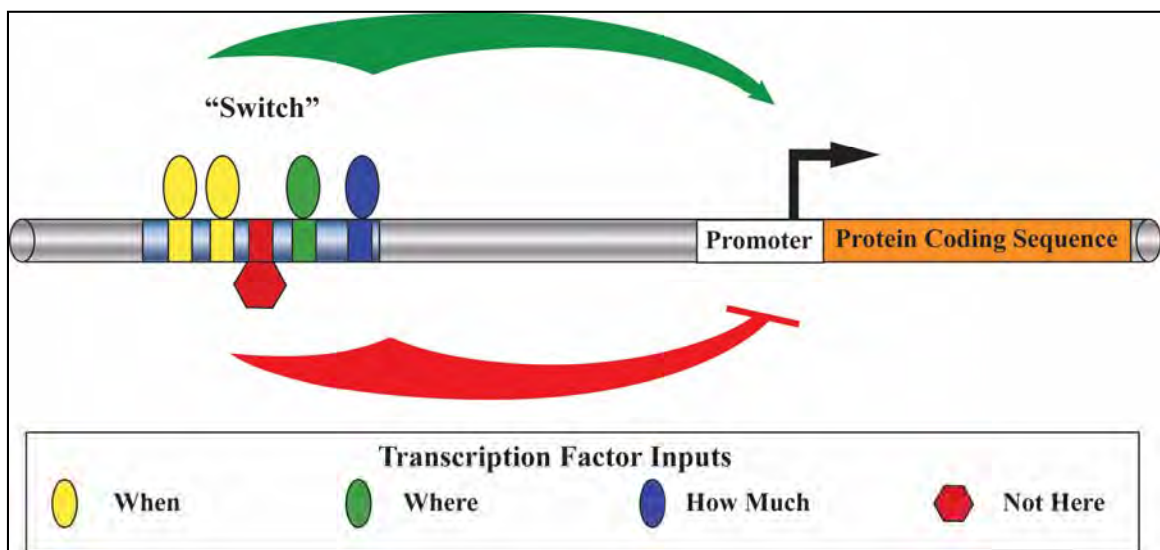
Figure 1. Representation as to how a *cis*-regulatory element controls gene expression.

The "Switch" or CRE possesses binding sites for particular transcription factor proteins.

The combination of transcription factors regulate the CRE which in turn regulates

transcriptional activation at a gene's promoter and thus controlling when, where, or how

much protein product is made.

Along with CREs and transcription factor proteins, promoters are also vital in the

gene expression regulation because promoters are regions of the DNA where RNA

polymerase attaches and initiates transcription of a gene (Griffiths, Wessler, Carroll, & Doebley, 2011). Promoters lie upstream a gene's first exon and are oriented in direction of transcription. Within the promoter of a gene is the transcription start point which is the nucleotide where RNA synthesis actually begins. The promoter of a gene typically extends several dozen base pairs upstream from the transcription start point and can extend downstream of the start site too (Kadonaga, 2012; Lagha, Bothma, & Levine, 2012). Within promoters are specific DNA sequences that provide binding sites for RNA polymerase and for transcription factor proteins which recruit RNA polymerase. In fact, in eukaryotes, RNA polymerase II cannot bind to the promoter until certain transcription factors have already attached to the promoter (Griffiths et al., 2011). RNA polymerase is the enzyme that links ribonucleotides into an RNA chain during transcription. Therefore, promoters provide a binding site for RNA polymerase so that transcription and thus gene expression can occur.

Cis-regulatory Elements Acting in Isolated (additive) Manner:

The expression of an individual animal gene is often regulated by the additive activity of an array of discrete CREs (Carroll, 2005, 2008; Rebeiz & Williams, 2011). For instance, the complex spatial pattern of *Eyeless* expression in *Drosophila*, which is a mosaically pleiotropic toolkit gene, is a composite of multiple CREs' regulatory actions (Carroll, 2008). In particular, *Eyeless* is regulated by the additive activity of six distinct CREs, averaging 1 kilobase in size. Each CRE drives a particular spatial pattern of *Eyeless* expression, such as in the eye, various lobes and cell types of the developing embryonic, larval and adult brain, and within the central nervous system. However,

whether CREs can interact in synergistic or non-additive ways has not received much consideration. In fact it is generally assumed that CREs act in additive manners though experiments seldom are designed to test for synergistic outcomes. For instance, consider the possible scenario that CRE A drives the expression of *gene A* in horizontal striped pattern, while CRE B drives a pattern of expression in vertical stripes (Figure 2). In an additive manner, gene A's expression would occur in a cross hatched pattern. However, a synergistic outcome is also possible, where gene A expression driven by CRE A and B occurs in more patterns than those directed by individual CREs (Figure 2). The idea of synergism has not received much consideration and is worth further investigation.
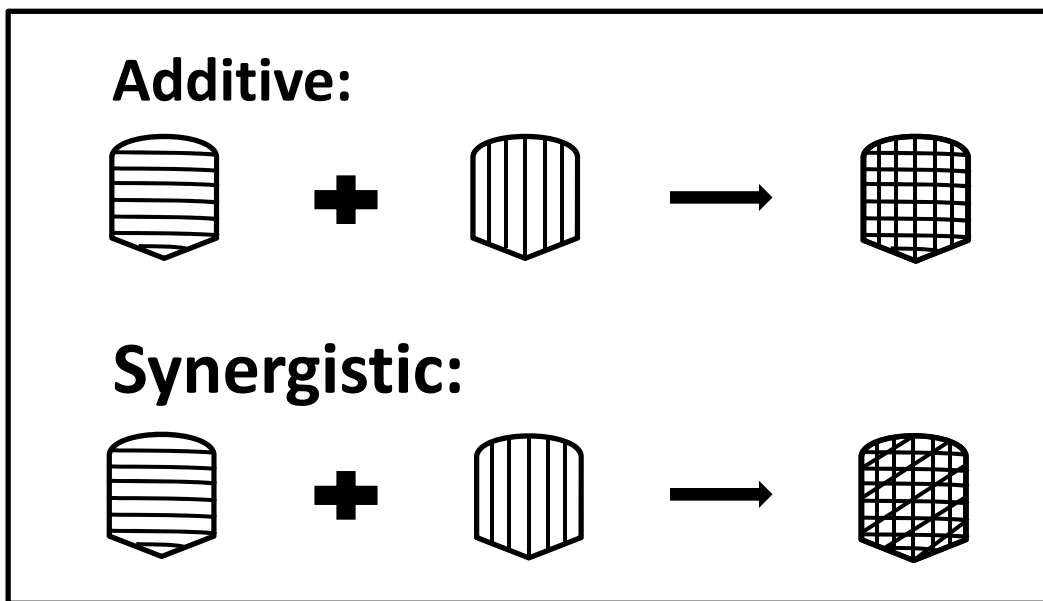
Figure 2. Simplified model for additive and synergistic gene expression outcomes. To the left of the arrow represents gene expression patterns for two separate CREs when studied in isolated reporter transgene assays. To the right of the arrow shows the endogenous pattern of expression driven by the two CREs. The additive outcome is where expression is the sum of the individual CRE activities, whereas a synergistic outcome is where the outcome is more or less than the sum of the individual activities.

In order to study *cis*-regulatory elements, reporter transgenes are used to show the regulatory actions of the *cis*-regulatory element. Reporter transgenes can be fused to the *cis*-regulatory element of interest so that the regulatory activity of the *cis*-regulatory element is exhibited through reporter transgene expression (Rebeiz & Williams, 2011) (Figure 3). A common reporter gene is the gene for *Enhanced Green Fluorescent Protein* (**EGFP**) (Figure 3). The GFP gene was originally sequestered from the jellyfish species *Aequorea Victoria* and then modified to produce the *EGFP* gene. The EGFP gene encodes the EGFP protein which when expressed is excited by light at a 488 nm. When EGFP protein is excited, it emits light at a wavelength of 509 nm and is thus easily recognized making EGFP a useful reporter gene (Rebeiz & Williams, 2011). 509 nm is in the green portion of the visible spectrum so where bright green is present represents the protein product of Enhanced Green Florescent Protein and consequently the regulatory actions of the CRE of interest (Figure 3).

Figure 3. Reporter transgenes include a CRE sequence placed next to a heterologous promoter (arrow) and the coding sequence for an easily monitored reporter protein, such as Enhanced Green Fluorescent Protein (EGFP).

CRE Evolution Controls an Evolved Pattern of Gene Expression and Pigmentation:

Drosophila (*D.*) *melanogaster* fruit flies provide an evolutionary model for CREs because they are sexually dimorphic and their sexual dimorphism evolved from a

monomorphic state. *D. willistoni* is monomorphic because the *bric-à-brac (bab)* genes are fully expressed in A2 through A6 abdominal segments of both males and females (Kopp, Duncan, & Carroll, 2000; Williams et al., 2008). The *bab* locus contains genes that repress pigmentation in the abdomen of both males and females (Couderc et al., 2002); all *D. willistoni* flies are nearly identically pigmented on the abdomen (Salomone, Rogers, Rebeiz, & Williams, 2013) (Figure 4G and 4J). However, *D. melanogaster* is sexually dimorphic and has evolved from this monomorphic state (Figure 4A and 4D). The sexual dimorphism of *D. melanogaster* is exhibited by the abdominal pigmentation differences between males and females. Males have fully pigmented posterior-most two abdominal tergites whereas the female abdomen has alternating dark-light stripe pattern that is characteristic of all abdominal tergites (Couderc et al., 2002). The pigment is the black color on the abdomen.
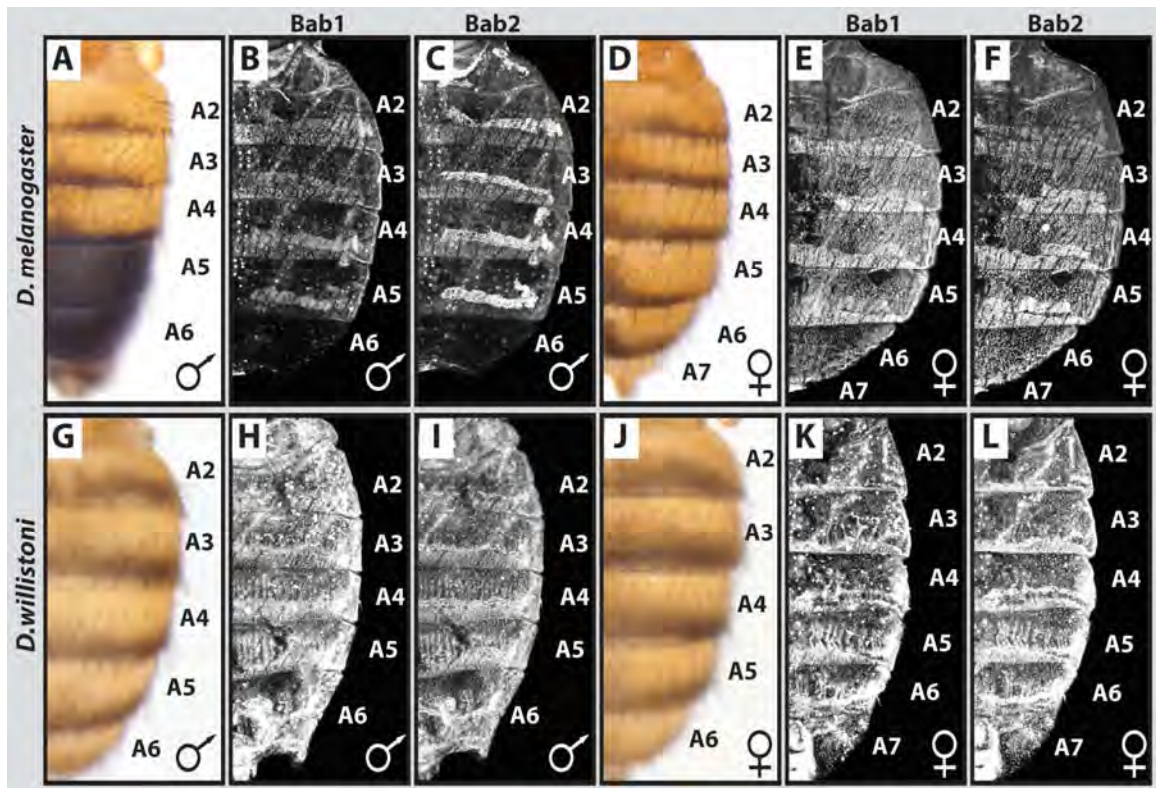
Figure 4. Dimorphic Bab expression and abdominal pigmentation evolved from a monomorphic state. *D. melanogaster* has a sexual dimorphic pigmentation pattern on the A5 and A6 abdominal segments (A and D). *D. willistoni* has a monomorphic pigmentation pattern on all abdominal segments (G and J). Bab1 (B, E, H, and K) and Bab2 (C, F, I and L) proteins are expressed in a dimorphic pattern in *D. melanogaster* but a monomorphic pattern in *D. willistoni*. Images are adapted from (Salomone et al., 2013).

The *bric-à-brac* genes in the *Drosophila melanogaster* fruit fly:

Fruit fly pigmentation is an excellent model to study the regulatory actions of CREs on gene expression because the *D. melanogaster* pigmentation pattern is due to a complex network involving the interaction of CREs. The sexually dimorphic pigmentation pattern seen in *D. melanogaster* is controlled by regulatory genes which encode transcription factors that control the expression of other structural genes, which encode enzymes that make up the biochemical pathways for pigment synthesis (Wittkopp, Carroll, & Kopp, 2003). The male-specific phenotype seen in *D. melanogaster* is the product of this type of complex network of transcription factor genes which suppress expression of enzymes required to make black melanin pigments. The *bric-à-brac* (*bab*) locus in *D. melanogaster* acts as a morphogenetic regulator in the *D. melanogaster* fruit fly's development of body parts, including the abdomen (Couderc et al., 2002). The suppression of pigmentation in the female abdomen is due to the *bab* locus on chromosome III. The *bab* locus, contains the tandem duplicate *bab1* and *bab2* genes, collectively referred to as *bab* (Williams 2008, Kopp 2000, Couderc 2002). The *bab* genes encode the transcription factors Bab1 and Bab2, which suppress the expression

of the *yellow* and *tan* genes that encode pigmentation enzymes (Couderc et al., 2002; Jeong et al., 2008; Jeong, Rokas, & Carroll, 2006). Thus, the *yellow* and *tan* genes are required to make black melanin pigment (Jeong 2008, Jeong 2006). Therefore, the *bab* genes encode transcription factors that repress the expression of pigmentation enzymes (Wittkopp 2003, Couderc 2002, Jeong 2006, Jeong 2008). The Bab proteins are expressed in the A2-A4 segments of males and in the A2-A6 segments of females (Figure 4B, 4C, 4E, and 4F). As a result, males exhibit pigmentation in A5 and A6 segments of their abdomen and female abdomens lack pigmentation (Figure 4A and 4D).

CREs in the *D. melanogaster* fruit fly:

   D. melanogaster sexual dimorphism results from the expression of Bab protein which is regulated by the activity of two CREs known as the anterior element (AE) and the dimorphic element (DE) located in the large first intron of the *bab1* gene (Figure 5A) (Williams et al., 2008). The AE regulates *bab* expression in segments A2-A4 of both sexes and the DE regulates *bab* expression in the posterior segments A5-A7 of females. The DE is regulated by ABD-B and DSX transcription factors. When the DE is bound by ABD-B and sex specific isoforms of the DSX protein, it acts as a genetic switch which directs female-specific activation and male-specific repression of *bab* in the posterior segments. As a result, *bab* in the A5 and A6 segments in females represses *yellow* and *tan* gene expression and consequently pigmentation, while male-specific repression of *bab* in the posterior segments allows for pigmentation (Kopp et al., 2000).
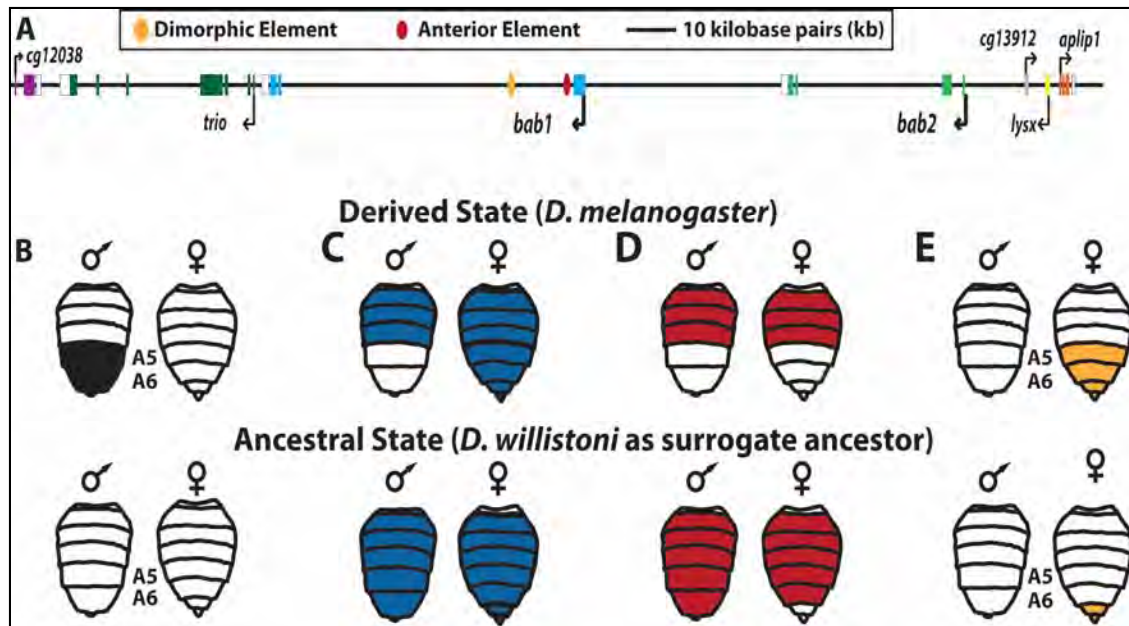
Figure 5. The *bab* locus and its role in pigmentation development and evolution. (A) To scale diagram of the *bab* locus and the locations of the dimorphic element and anterior element CREs. (B) Pigmentation patterns for *D. melanogaster* and *D. willistoni*. (C) Depiction of the endogenous pattern of Bab expression. Depiction of the Bab expression controlled by the (D) anterior element and (E) dimorphic element

Thesis Aim: Determine which DNA Sequences in the Dimorphic Element Contribute in the Synergistic Switch Combination Outcome.

The activities of the dimorphic element, anterior element, and the combination of the two CREs (called "Switch Combination") on driving expression of the Enhanced Green Fluorescent Protein (EGFP) gene were previously assessed by Eric Camino. It was found that the dimorphic element robustly activated EGFP protein expression in the A5 and A6 segments of females. The anterior element weakly activated EGFP protein expression in the A2-A4 segments in both sexes. Surprisingly, the Switch Combination

activated EGFP protein expression in the A2-A4 segments of males and females in a non-additive or synergistic way. The outcome of the EGFP protein expression was produced by neither CRE alone nor the sum of their independent activities.

In order to find the synergism encoding sequences of the dimorphic element for the synergistic Switch Combination outcome, 9 Switch Combination reporter transgenes were created that included a region(s) of mutant dimorphic element sequence. One of the 9 scanning mutants had mutated flank regions and a non-mutated core region. The other 8 scanning mutants had non-mutated flank regions and a specific part of the core region mutated (Figure 6). The mutated regions of the 8 core mutants partially overlapped, and collectively spanned the central region of the DRE. By integrating these reporter transgenes into *D. melanogaster* fruit flies, the effects of the mutations can be analyzed. The effects were analyzed by utilizing EGFP expression and comparing EGFP expression driven by non-mutant Switch Combination to EGFP expression driven by mutant Switch Combination. Synergism encoding sequences were identified by noting the mutants that resulted in a loss of the robust A2-A4 EGFP expression.

Surprisingly, results found that the dimorphic element possesses sequences distributed throughout the core region of the dimorphic element that impart additive, synergistic, activating, and repressive effects. Consequently, there is not one specific sequence of DNA in the dimorphic element that is solely responsible for the synergistic activity. Instead, sequences of the dimorphic element could be working together. So, in the future, an experiment mutating two different regions of DNA in the dimorphic element could be done to test for combinations of sequences in the dimorphic element causing the synergistic expression seen with the Switch Combination.

Thesis Aim: Determine which sequences in the Anterior Control Element are necessary for the robust EGFP expression

Previously Bab expression at different time points during pupal development was investigated. Bab expression in males is down regulated during late development, however male Bab expression during early pupal development is robust (Salomone et al., 2013). To investigate how Bab expression in males is down regulated during late development the anterior element and its neighboring sequences were evaluated and became known as the Anterior Control Element (ACE). The Anterior Element was expanded from 1400 bp to 2400 bp to make up the ACE. We compared the activity of the anterior element and the anterior control element to drive expression of the EGFP gene. When evaluating the anterior element alone, a weak expression of EGFP was seen. However, the anterior control element drives a pattern of EGFP expression that is robust and mimics male Bab expression during early pupal development. My thesis research aimed to identify the sequences in the anterior control element that are necessary for the robust expression of Bab during early pupal development. My research investigated for additional encodings or altogether separate CREs to explain how Bab expression in males is down regulated during late development.

To map sequences mediating AE activity I created 5 truncations of the ACE. The results showed that some additional sequences are necessary for robust AE activity because certain truncations of the ACE exhibited robust EGFP expression. Future works need to map which sequences are important for the pattern of expression driven by the AE and those sequences that work synergistically with the dimorphic element.

CHAPTER II: MATERIALS AND METHODS

Production of 9 DRE scanning mutants:

In order to identify sequences required for synergistic CRE activity, I designed nine scanning mutant versions of the **dimorphic regulatory element** (or **DRE**). The DRE name is used for the originally characterized dimorphic element (Williams et al., 2008) when it is flanked on each side by ~500 **base pairs** (**bp**) of endogenous *bab* locus sequence (Rogers et al., 2013). The mutated region in each scanning mutant was a series of non-complimentary nucleotide transversions (changing Adenine to Cytosine or Thymine to Guanine and vice-versa) located at every other base pair. These mutant sequences were created using the Perl script referred to as 'Scrambler' that was written by Mark Rebeiz (University of Pittsburgh, unpublished). With the 'Scrambler' the wild type sequence input was transformed to a scanning mutant output sequence. The first scanning mutant had the right and left flank regions mutated while the core region was left untouched (Figure 6 and Figure 7). The left and right flank regions were 500 base pairs and 400 base pairs, respectively. The number one scanning mutant (core mutant 1 or CM 1) consisted of the first ~80 base pairs of the core region of the DRE mutated, with every other area in the DRE untouched. Proceeding, core mutant 2 (CM 2) had the next ~ 80 base pairs of the DRE core region mutated, with their being 10 base pairs of overlap between the end of CM 1 and the beginning of CM 2. CM 3 through CM 8 followed the same pattern and the overlap for any two mutants was 10 base pairs (Figure 6 and Figure 7). In all of the core mutants, the sequences for the characterized 14 Abd-B binding sites and the 2 Dsx bindings sites were not mutated (Williams et al., 2008).
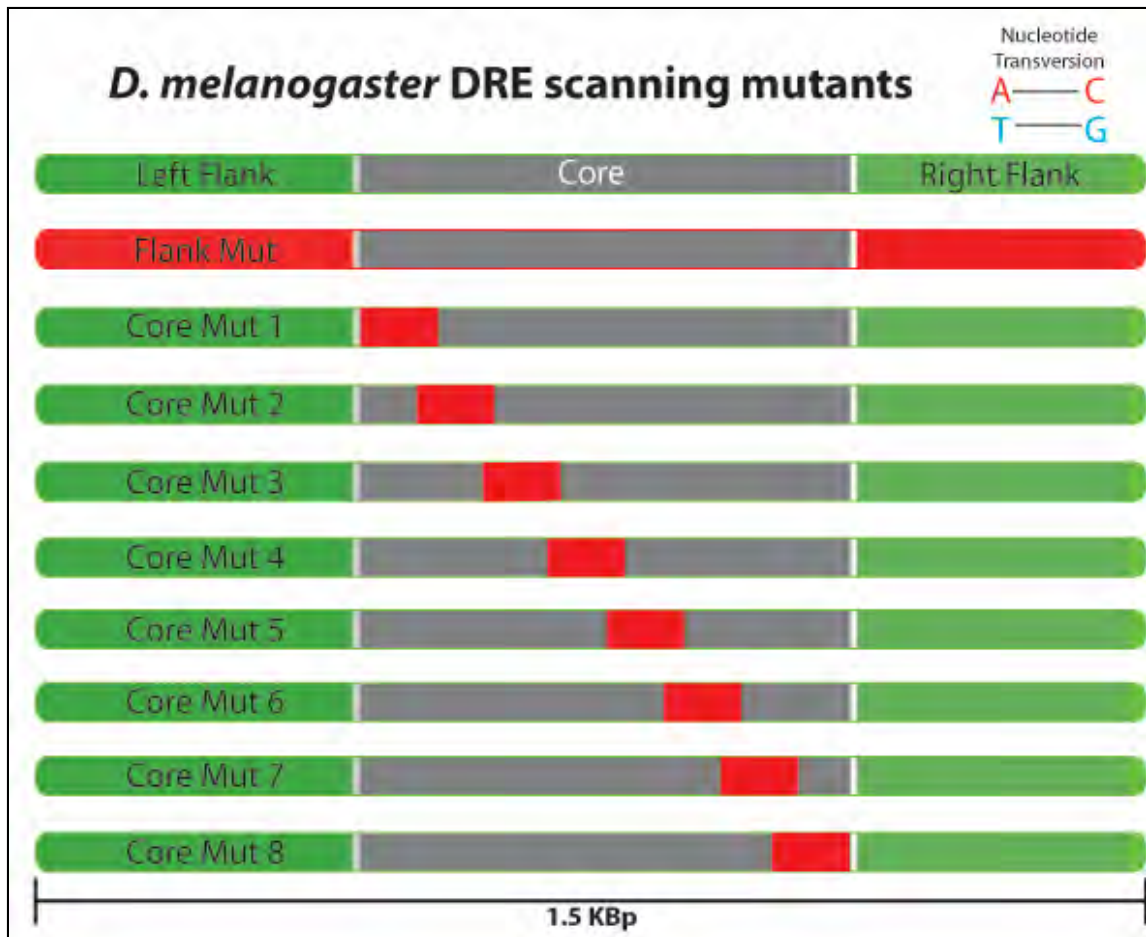
Figure 6. Schematic describing the design for the mutant dimorphic regulatory elements (DRE). 9 scanning mutants of the DRE were created; the red color blocks above represent the locations for the mutant sequence. Right and left flank mutant regions were ~400 bp and ~500 bp, respectively. Mutated core regions were ~80 bp and there was 10 bp of overlap between adjacent core mutant regions.
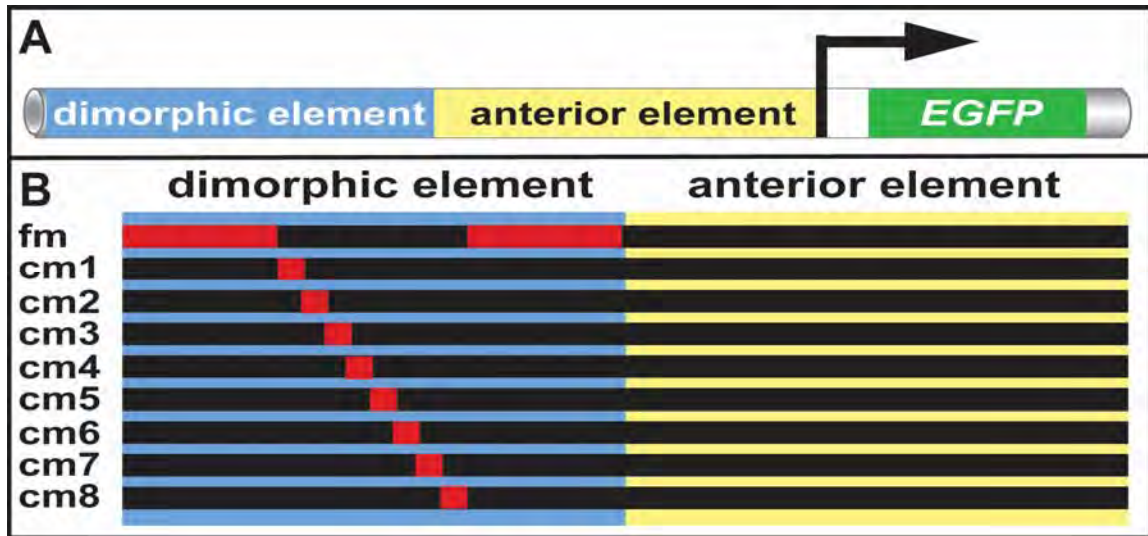
Figure 7. Schematic for the reporter transgenes that were used to test for synergistic CRE interactions. Mutant DRE sequences (listed above as dimorphic element) were placed adjacent to the anterior element. These CRE cassettes were placed upstream of the composite *D. melanogaster bab2* and *hsp70* promoters and the coding sequence for the *Enhanced Green Fluorescent Protein* gene (*EGFP*).

The non-mutant CRE sequence from the Switch Combination was replaced by a mutant DRE version in the reporter transgene vector called BPS3aG (Rogers et al., 2013). This vector has a *bab2* gene promoter next to the hsp70 promoter, coding sequence for enhanced green fluorescent protein, and an *attB* site for genomic integration (Groth, Fish, Nusse, & Calos, 2004). In order to replace the non-mutant DRE from the Switch Combination with the mutant DRE in the reporter transgene vector called BPS3aG, a series of protocols and experiments had to be performed. The mutated DRE had to be taken from the pUC57 vector and ligated into the BPS3aG vector.

Preparing the BPS3aG vector:

To prepare the BPS3aG vector that was used for the flank mutant and core mutants, a digestion was done to take out the non-mutated DRE from the BPS3aG vector (Kaitlyn Francis notebook: Dated 2/16/12), so that eventually a mutated DRE could be put in its place. The digest consisted of placing the following into a 1.5 ml tube at 37 degrees Celsius for 30 minutes: 10 ul DRE .04 vector, 10 ul 10x BSA, 10 ul 10x NEB 4, 2 ul AscI, 2 ul SbfI, and 54 ul Milli-Q, totaling 88 ul. The enzymes, in this case AscI (New England Biolabs or NEB, Catalog #R0558L) and SbfI (NEB, Catalog #R0642L), were always added last in digests. After the 30 minutes, 10 ul of Antarctic Phosphotase Buffer was added and then 2 ul of Shrimp Alkaline Phosphatase or SAP enzyme (NEB, Catalog #M0371S) was added, totaling 100 ul, and was kept at 37 degrees Celsius for 1 hour. The digest was run in a 0.7% gel (0.7 g agarose, 100 ml TAE) for 30 minutes using gel electrophoresis. The gel was prepared by first mixing 0.7 g agarose and 100 ml TAE in a beaker. Then the mix was heated in a microwave for 1 minute and 5 ul ethidium bromide (10 mg per ml stock solution) was added after being heated. The gel solidified in a gel casting tray for ~30 minutes and then was ready for use. The digest and 2-log ladder (NEB, Catalog #N3200L) were loaded into separate wells of the gel and then run using gel electrophoresis for 30 minutes with 130 V. From gel electrophoresis, the vector without the DRE and the DRE separated in the gel. The vector that no longer had the DRE in it, was cut out of the gel and purified using a large DNA gel purification (Appendix 3) (Kaitlyn Francis Notebook: Dated 2/24/12).

Preparing the flank mutant:

   To properly replace the non-mutated DRE with the mutated flank DRE, a digest was done first to cut the piece with the core and mutated flanks (entire DRE) out of the pUC57 vector originally in when received from

 Inc (Appendix 1). The digest consisted of placing the following into a 1.5 ml tube at 37 degrees Celsius for 30 minutes: 40 ul of the vector, 10 ul of 10x BSA, 10 ul 10x NEB4, 2 ul AscI, 2 ul SbfI, 36 ul Milli-Q, all totaling 100 ul. After digested, the mix was run in a 1% agarose gel. The agarose gel was prepared by mixing 1 g agarose with 100 ml TAE, microwaving that for 1 minute, then adding 5 ul ethidium bromide stock solution and pouring into the gel casting tray to solidify. The digested mix was then poured into one of the wells of the agarose gel and subject to gel electrophoresis. The DRE region consisting of the core and mutated flank regions of that vector were cut out and purified using the small DNA gel purification protocol (Appendix 2) (Kaitlyn Francis Notebook: Dated 2/14/12). The DRE is ~ 1.5 **kilobases** or **kb** in length so it is located using the 2-log DNA ladder in an adjacent gel well. A ligation was done to ligate the DRE flank mutant (entire DRE with mutated flanks) with the BPS3aG vector that does not have the DRE but that contains the anterior element. 6.5 ul of insert/ DRE Flank Mutant, 1 ul T4 DNA Ligase Buffer, 1 ul T4 DNA ligase (added last; NEB, Catalog #M0202S), and 1.5 ul vector BPS3aG (–DRE+anterior element) were mixed in a test tube and stored in a 4 degree Celsius fridge overnight. A control ligation was also done, using the same protocol as before, but substituting milli-Q in place of the DRE Flank Mutant. Transformations of the ligations were done (Appendix 4). Mini prep (Qiagen Inc., Catalog #27106) cultures were prepared from the transformations by labeling 6 large snap cap tubes, adding 3.5 ml Luria

Broth (LB) Broth + 100 ug/ml of Ampicllin (Amp) to each tube using a pipetman, then, with a glove on, using autoclaved toothpicks to pick a colony of bacteria for each tube. One toothpick with a colony on it was placed in each large snap cap tube. The snap cap tubes were stored in a 37 degrees Celsius incubator slowly spinning overnight for ~18 hours of growth. Plasmid DNA mini preps of the DRE flank mutants using the mini prep cultures were done (Appendix 5). A restriction digest was then completed consisting of mixing1 ul 10 x NEB4, 1 ul 10xBSA, 0.25 ul AscI, 0.25 ul SbfI, 2.5 ul Milli-Q and 5 ul mini prep DNA of DRE flank mutant in a tube at 37 degrees Celsius for 30 minutes (done for each mini prep). Then the restriction digests were run in a 0.7% test gel using gel electrophoresis. The test gel showed the DRE flank mutant was in there, so mini preps were sent to DNA Analysis LLC to be sequenced. The sequence results matched what we wanted the mutant to be and so the flank mutant mini preps were sent to Best Gene Inc. to be midi prepped and injected into the embryos of the *Drosophila melanogaster* fruit flies. These constructs were injected into embryos of Bloomington Drosophila Stock Center #8622. This stock contains the attP2 site on the 3rd chromosome for site specific integration mediated by a genomic source of PhiC integrase (Groth et al., 2004). The attP site is often referred to as "3-Calos" in the Williams lab.

Preparing the 8 DRE core mutants:

To prepare the pUC57 vector that was used for the 8 DRE core mutants, a restriction enzyme digest was done to take out the core of the DRE in the pUC57 vector. The digest contained the following: 10 ul DRE .04 (Sac II/ NheI) Flanked Core, 10 ul 10xNEB4, 10 ul 10xBSA, 2 ul Nhe1-HF enzyme (NEB, Catalog #R3131S), 2 ul Sac II

enzyme (NEB, Catalog #R0157S), and 54 ul Milli-Q totaling 88 ul in a test tube at 37

degrees Celsius for 30 minutes. After the 30 minutes, 10 ul Antarctic Phosphotase Buffer

was added and then 2 ul SAP enzyme was added. With these added, the mix totaled 100

ul and was kept in an incubator at 37 degrees Celsius for 60 minutes.

To prepare the 8 core mutants, the same set and order of protocols was done for

each one. Therefore, the same way core mutant 1 was prepared, is how all core mutants

were prepared just the appropriate mutated DRE Core DNA was substituted for each. To

prepare core mutant 1 (CM 1), first a restriction enzyme digest was done to cut out and

purify the mutated DRE core in the pUC57 vector provided by GenScript Inc (Appendix

1). The restriction digest contained the following (Kaitlyn Francis Notebook: Dated

3/26/12): 10 ul DRE scramble 1 DNA, 10 ul 10xBSA, 10 ul 10xNEB4, 2 ul NheI-HF

enzyme, 2 ul SacII enzyme, and 16 ul Milli-Q, totaling 50 ul in the test tube at 37 degrees

Celsius for 30 minutes. The digested mix was run in a 1% agarose gel using gel

electrophoresis. The mutated core- scramble 1 was cut out of the gel and purified using

the small DNA gel purification protocol (Appendix 2). A ligation was done to ligate the

mutated core scramble 1 into the DRE –core pUC57 vector described above. The ligation

contained the following in a 1.5 ml test tube: 6.5 ul insert (mutated core scramble 1), 1 ul

T4 DNA Ligase Buffer, 1 ul T4 DNA ligase enzyme added last, and 1.5 ul DRE-core

pUC57 vector. A control ligation was also done using the same protocol except for using

6.5 ul Milli-Q in place of 6.5 ul mutated core scramble 1. The mix was stored in a 4

degrees Celsius fridge overnight. A transformation was done of the ligation (Appendix

4). Mini prep cultures from the transformations were prepared. Using the mini prep

cultures 4 mini preps were done of the mutated DRE core scramble 1 in the pUC57

vector. Using the mini prep DNA, a restriction digest was done to test for the mutated

core plus flanks (entire DRE) in the mini preps (Kaitlyn Francis Notebook: Dated

4/10/12). 1 ul 10xNEB4, 1 ul 10xBSA, 12.5 ul Milli-Q, 0.25 ul AscI enzyme, 0.25 ul SbfI

enzyme and 5 ul miniprep DNA was added to each test tube totaling 10 ul, and was

placed at 37 degrees Celsius for 30 minutes. The restriction digest was run in a 0.7% test

gel using gel electrophoresis to confirm the mutated DRE was in the pUC57 vector. The

test gel showed it was and so the process continued.

At this point, the mutated DRE core scramble 1 and non-mutated flanks (entire

DRE) was in the pUC57 vector. Now, the goal was to place the mutated DRE from the

pUC57 vector into the BPS3aG vector prepared above. A restriction digest was done of

the mutated DRE core scramble 1 mini prep DNA to cut out the DRE from the pUC57

vector. The restriction digest contained the following: 20 ul miniprep DNA, 10 ul BSA,

10 ul NEB4, 2 ul AscI (added last), 2 ul SbfI (added last), and 56 ul Milli-Q, totaling 100

ul in a 1.5 ml test tube. The mix was placed in a 37 degrees Celsius incubator for 45

minutes. After 45 minutes the digest was run in a 1% agarose gel using gel

electrophoresis for 45 minutes. The entire DRE that had mutated core scramble 1 was cut

out of the gel and purified using the small DNA gel purification protocol (Appendix 2). A

ligation was done to ligate the purified DRE mutated core scramble 1 into the BPS3aG

vector that does not have the DRE in it (prepared above) but that contains the anterior

element. The ligation contained the following: 6.5 ul DRE mutated core scramble 1, 1 ul

T4 DNA ligase Buffer, 1 ul T4 DNA ligase enzyme, 1.5 ul BPS3aG vector (–

DRE+anterior element). A control was also done using Milli-Q in place of the DRE

mutated core scramble 1 insert. The ligation was stored in a 4 degrees Celsius fridge

overnight. The ligation was transformed the following day (Appendix 4). Mini prep

cultures of the transformation were prepared. Six mini preps of the mutated DRE core

scramble 1 in the BPS3aG vector were completed. A restriction digest of the six mini

preps was done to test for the mutated core + flanks (entire DRE), Anterior Element and

BPS3aG vector. To do this, a master mix was made containing the following: 7 ul

10xNEB4, 7 ul 10xBSA, 1.75 ul AscI enzyme, 1.75 ul NheI enzyme, and 17.5 Milli-Q

totaling 35 ul. 5 ul master mix and 5 ul mini prep DNA were put in each tube, totaling 10

ul in each one of the 6 tubes at 37 degrees Celsius for 30 minutes. After the 30 minutes,

the digests were run in a 0.7% test gel using gel electrophoresis. The gel electrophoresis

showed that 5 of the mini preps worked exhibiting a band for the entire DRE, a band for

the AE, and a band for the BPS3aG vector. The successful mini preps were mini preps 2-

6 (Kaitlyn Francis Notebook: Dated 6/15/12). The successful mini preps of the DRE

mutated core scramble 1 in the BPS3aG vector were sent out for sequencing to DNA

Analysis LLC. The sequenced results exhibited 100% accuracy with what we wanted

(Kaitlyn Francis Notebook: Dated 7/11/12). Thus, the successful mini preps were sent out

to Best Gene Inc. to be midi prepped and further injected into fertilized eggs of

Drosophila melanogaster fruit flies. Once again the transgenes were inserted into the

attP2 site (or 3 Calos).

Production of 5 ACE truncations:

The Anterior Control Element (ACE) was broken into 5 truncations so that each

segment of DNA could be analyzed to find which is responsible for ACE expression

(Figure 8). To start, *Drosophila melanogaster* genomic DNA (gDNA) was prepped using

the Qiagen "DNeasy" kit (Qiagen Inc., Catalog #69514) Genomic DNA Prep, so that the

gDNA could be used for a PCR (Appendix 6). Each ACE sub construct went through the

same set and order of protocols. So, the same way ACE sub 1 was prepared, is the same

way ACE sub 2-5 were also prepared. A Polymerase Chain Reaction (PCR) was done

with the Phusion High-Fidelity DNA Polymerase enzyme (NEB, Catalog #M0530L) to

replicate the ACE sub 1 DNA. The Phusion PCR was done by mixing 10 ul 5x Phusion

HF buffer, 4 ul 2.5 mM dNTPs, 2.5 ul Forward Primer, 2.5 ul Reverse Primer, 1 ul

gDNA, 29.5 ul sterile Milli-Q, and 0.5 ul Phusion High-Fidelity DNA Polymerase

enzyme in a tube and then run in the PCR machine. Table 1 provides the forward and

reverse primers used for each ACE sub construct. Once the PCR was completed, the PCR

product was run in a 1% agarose gel using gel electrophoresis to separate the ACE sub 1

DNA segment from the rest of the gDNA. Using the 2-log DNA ladder as comparison in

the gel, the ACE sub 1 PCR band was located. The ACE sub 1 band of DNA was cut out

of the gel and purified using the small DNA gel purification protocol (Appendix 2??). A

DNA digest of the purified ACE sub 1 PCR product was completed by mixing in a 1.5 ml

tube the following: 52 ul PCR clean ACE sub 1 DNA, 7 ul 10x NEB4, 7 ul 10xBSA, 2 ul

AscI enzyme, and 2 ul SbfI enzyme, totaling 70 ul in the tube. Enzymes are always added

last. The mix was incubated at 37 degrees Celsius for 30 minutes. Enzyme removal of the

digest was completed using the QIAquick PCR purification kit (Qiagen Inc., Catalog

#28106) (Appendix 7). To test whether the ACE sub 1 band was present and intact at this

point, the ACE sub 1 clean PCR product was run in a 0.7% test gel using gel

electrophoresis. Gel electrophoresis exhibited the proper band representing the ACE sub

1 sequence of DNA, so the process continued. Table 2 provides the length of each ACE
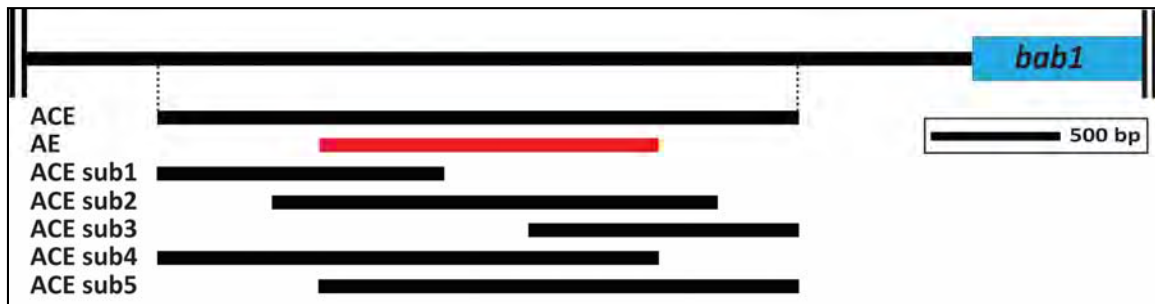
sub construct.

Figure 8. Schematic of the relationship between the ACE truncations. This figure

details the sequences included within each ACE truncation with respect to the other

truncations, the larger ACE, and the more modest anterior element (AE). To the right is

the end of the 1st exon of the *bab1* gene.

Cloning PCR-Amplified Sequences into the Reporter Transgene Vector:

The ACE sub 1 AscI/SbfI digested and enzyme purified DNA was ligated into the

BPS3aG vector. The ligation consisted of mixing the following: 1.5 ul BPS3aG vector (-

DRE and –anterior element), 1ul 10x T4 ligase buffer, 0.75 ul NEB T4 DNA ligase, and

6.75 ul insert DNA, totaling 10 ul in a 1.5 ml tube. The insert DNA in this case, was 1:10

dilutions and 1:50 dilutions of the ACE sub 1 clean DNA. A control ligation was also

done, using the same protocol, except Milli-Q in place of the insert DNA.

Transformations of the ligations were done (Appendix 4). Mini prep cultures of the ACE

sub 1 transformations were prepared and stored at 37 degrees Celsius slowly spinning for

~18 hours. After 18 hours of growth mini preps of the mini prep cultures were completed

(Appendix 5). A restriction digest of the mini prep DNA was completed to test for the

ACE sub 1 segment of DNA. The restriction digest contained: 1 ul 10xNEB4, 1 ul

10xBSA, 0.25 ul AscI, 0.25 ul SbfI, 2.5 ul Milli-Q and 5 ul mini prep DNA, totaling 10

ul in each tube. The mix was placed in an incubator at 37 degrees Celsius for 30 minutes.

The digest of the ACE sub 1 mini prep DNA was run in a 0.7% test agarose gel using gel

electrophoresis for 30 minutes. The test gel showed the correct size of bands to signify

the ACE sub 1 DNA segment present in the mini prep ACE sub 1 DNA. The mini prep

DNA was tested for the concentration level of DNA and A260 in the sample. The

concentrations of the DNA samples were manipulated to be 75 ng/ul. Once 75 ng/ul the

samples were sent to DNA Analysis LLC to be sequenced. The sequenced results

exhibited 100% accuracy with what we wanted. So, the mini preps were sent out to Best

Gene Inc. to be midi prepped and further injected into fertilized eggs of *Drosophila*

melanogaster fruit flies. Trasgenes were integrated into the attP2 site (3 Calos).

Table 1. The forward and reverse primers used in the PCRs for the ACE truncations.

Construct	Forward Primer Name Sequence (5'-3')	Reverse Primer Name Sequence (5'-3')
ACE Sub 1	2.2 Ortho F2 TTCCGggcgcgccGAGTGACGAGTGTAA TGACTTTTGTCG	ACE left Rvs TTGCCcctgcaggCGCCACAGATACACC CG
ACE Sub 2	ACE mid Fwd AscI TTCCGggcgcgccGCGGTTTGGCCCTCAT CTGTGC	ACE mid Rvs SbfI TTGCCcctgcaggGTCCTAATGCGGCTT TGCTTC
ACE Sub 3	ACE rt. Fwd. AscI TTCCGggcgcgccGCCGGCAACACTAAT ACCCAGGC	2.2 Ortho R2 SbfI TTGCCcctgcaggGCATTTTGTGAC TTGTTGTGGTTCCC
ACE Sub 4	2.2 Ortho F2 TTCCGggcgcgccGAGTGACGAGTGTAA TGACTTTTGTCG	AEmel Rvs TTCCGcctgcaggACTGCGACTGCATTA GCACCGA
ACE Sub 5	AE conserved Fwd TTCCGggcgcgccCCTAATAAAAACTTCC TCTGCGCC	2.2 Ortho R2 SbfI TTGCCcctgcaggGCATTTTGTGAC TTGTTGTGGTTCCC

 Table 2. Size of ACE sub truncations. This table shows the size of each ACE sub truncation.

Construct	Size of Truncation (base pairs)
ACE Sub 1	1200
ACE Sub 2	1700
ACE Sub 3	1000
ACE Sub 4	2000
ACE Sub 5	1800

Transgenic fruit fly lines:

To generate transgenic lines, the transgenes were sent to Best Gene Inc. (Chino Hills, CA) to be injected into *D. melanogaster* embryos. All mutant Switch Combination reporter transgenes and all ACE sub reporter transgenes were integrated into the attP2 genomic landing site on the third chromosome referred to as 3-Calos (Best Gene, Inc.) (Groth et al., 2004). Inserting transgenes into the same genomic landing site eliminates variation due to transgene insertion site (Williams 2008, Rogers 2011). When transgenic lines were received, transgenes were crossed so that they would be homozygous for the transgene and thus have two copies of the transgene. The transgenic lines were made homozygous by crossing virgin male and female flies that have a dark red eye color phenotype. The red eye color indicates flies are homozygous for the transgene, while orange eye color indicates flies are heterozygous for the transgene (Rogers & Williams 2011). Multiple lines were analyzed to eliminate variation in reporter gene activity due to factors other than the transgene's sequence.

Visualizing Reporter Transgene activity:

By use of forceps and specimen magnification via a stereomicroscope (Olympus), transgenic pupae were removed from their puparium at ~65-70 hours after puparium formation (hAPF). Pupation runs for ~100 hours for *Drosophila melanogaster* raised at 25 degrees Celsius. Specimens were fixed in halocarbon oil on a microscope glass cover slip with their dorsal abdomen facing downwards. EGFP expression was visualized in the dorsal abdominal epidermis using a confocal microscope (Olympus FV1000) (Rogers & Williams, 2011). With a confocal microscope (Fluoview 1000) with the 10X objective, a series of images were taken of each specimen along the z-axis at increments of 10 microns. These series of images were taken at the settings in Table 3 and at a Kalman line average setting of 3.

Table 3. Confocal Microscope Settings. These are the settings the confocal microscope was at when viewing the reporter transgenes.

Construct	Laser	HV	GAIN	OFFSET	Aperture
Tests for Synergism	10%	700	1	1	200
DRE scan mutants/AE	10%	700	1	1	200
ACE truncations	10%	700	1	1	200

A projection image using Fluoview software was exported in Tagged Image File Format (TIFF) after the series of images were completed. For each reporter transgene multiple specimens of both males and females were used to study the effects of the transgene. Images of multiple specimens for each reporter transgene ensure that a

representative image is included in figures. All images included in figures were processed in the same way through Adobe Photoshop CS3.

<u>CHAPTER III: RESULTS</u>

<u>*The Original Test of Synergism vs. Additivity:*</u>

In reporter transgene assays, the dimorphic element activates EGFP expression in the dorsal abdominal epidermis of the A5, A6, and A7 segments of females but not males (Figure 9G). In contrast, the anterior element activates EGFP expression in the dorsal abdominal epidermis of the A2, A3, and A4 abdominal segments of both males and females (Figure 9H). This low level of anterior element activity occurs when this CRE bearing transgene is inserted in the attP2 (referred to in the Williams Lab as "3-Calos") genomic landing site (Groth et al., 2004) and contrasts from a more robust pattern of EGFP expression that was observed in the original publication describing the anterior element (Williams et al., 2008). Thus it appears the anterior element is more sensitive to the chromatin environment of transgene placement than the dimorphic element.

We were curious as to what the regulatory activities of the dimorphic element and anterior element would be when they occur in a single transgene regulating EGFP expression (Figure 9F, Switch Combination). One possible outcome is an additive pattern of expression, where EGFP expression is seen in a pattern that is merely the sum of the individual patterns directed by these two CREs. The other possible outcome is a synergistic pattern of expression, where EGFP expression occurs in either a less elaborate or more elaborate pattern than that directed by the CREs in isolation. We found that the synergistic outcome occurred, where a non-additive pattern of EGFP was observed (Figure 9I). Notably, EGFP expression in the A2-A4 segments was significantly increased (Figure 9I, yellow arrowheads) and more similar to the endogenous pattern of Bab1 expression in females (Figure 9D).
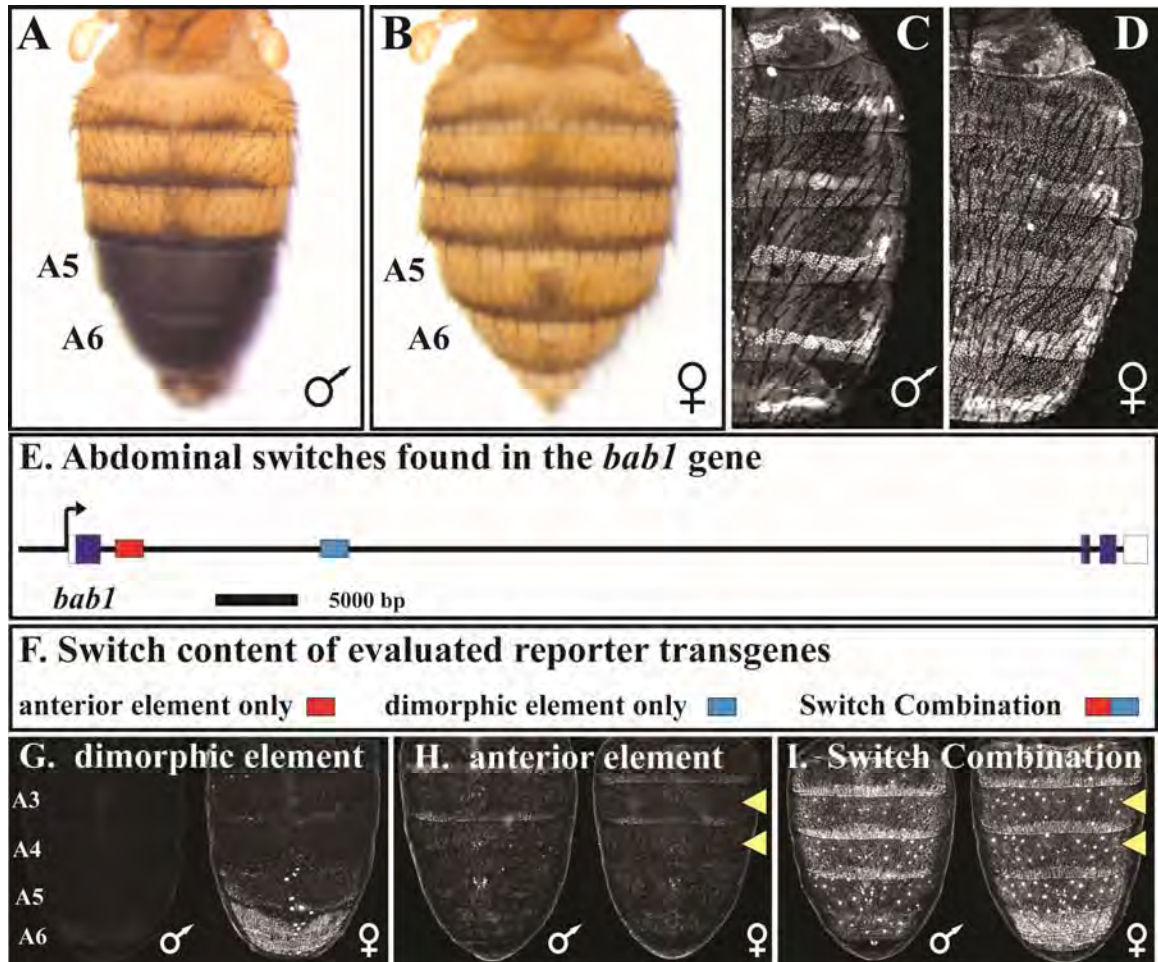
Figure 9. Comparing *bab* locus CRE activities to the endogenous pattern of Bab1 expression and abdominal pigmentation. Abdominal pigmentation pattern for a *Drosophila melanogaster* (A) male and (B) female abdomen. (C and D) Nucleus-localized pattern of Bab1 expression as revealed through the uses of an antibody specific for the Bab1 protein (images kindly provided by David Tacy). Notice that Bab1 expression is robust throughout the female abdomen, including the A2-A4 segments, whereas male expression is weak at this late pupal development time point. (E) Schematic of the *bab* locus showing the location of the dimorphic element and the anterior element in the 1ˢᵗ intron of *bab1*. (F) Symbolism for CREs. EGFP expression

driven by reporter transgenes possessing the (G) dimorphic element, the (H) anterior element, and (I) the dimorphic element and anterior element.

The effect of Dimorphic Element scanning mutations on synergism:

The generally accepted model for the function of CREs is where each CRE is modular and imparts a pattern of regulation upon a gene promoter. The synergistic outcome observed here is an outlier, where the dimorphic element and anterior element when in close proximity to each other results in a synergistic outcome. I hypothesized that this synergistic pattern of EGFP expression depended upon DNA sequence features within the dimorphic element. Previously, Eric Camino tested whether the important sequence features were the binding sites for the transcription factors Abd-B and Dsx that were found to reside in the central "Core" region of the dimorphic element. However, mutating these sequences did not alter the synergistic pattern of expression in the A2-A4 segments. To test the hypothesis that some other sequence or sequences direct this synergistic outcome, I designed nine scanning mutant versions of the dimorphic element to test for sequences necessary for the non-additive regulatory activity (Figure 6).

Of the nine dimorphic element scanning mutants, reductions were seen for several in the A5 and A6 segments, the domain of expression directed by the dimorphic element (Figure 10). In particular the scanning mutants DE(CM1)-AE, DE(CM2)-AE, and DE(CM6)-AE resulted in reduced female specific A5 and A6 expression. Thus, these mutated regions of the dimorphic element revealed novel sequences that are necessary for the full functionality of the dimorphic element in these posterior abdominal segments. Interestingly, several scanning mutants resulted in reduced EGFP expression in the A2-

A4 segments, suggesting that the mutated sequences are necessary for the synergistic outcome and that multiple sequences contribute to synergism. These mutants include DE(CM1)-AE, DE(CM2)-AE, DE(CM6)-AE, and DE(CM7)-AE in the switch combination transgene. One additional and unexpected finding was the pattern of EGFP expression driven by the DE(CM5)-AE switch combination transgene. In this mutant, increased EGFP expression was seen in segments A2-A4. This outcome suggests that this mutated sequence disrupted encodings that normally act to reduce anterior element or perhaps even dimorphic element activity in the A2-A4 abdominal segments. Outside of the aforementioned binding sites for the transcription factors Abd-B and Dsx, no other functional sequences have been reported for the dimorphic element. Thus, these scanning mutants have identified several additional CRE regions that impact dimorphic element activity. Lastly, while mutations in the core region affected the pattern of EGFP expression, no alteration in expression was seen when the entire flanking sequences were mutated. Thus, it appears from this outcome that the dimorphic element's regulatory activity is confined within the core region and does not sprawl into either of the adjacent flank sequences.
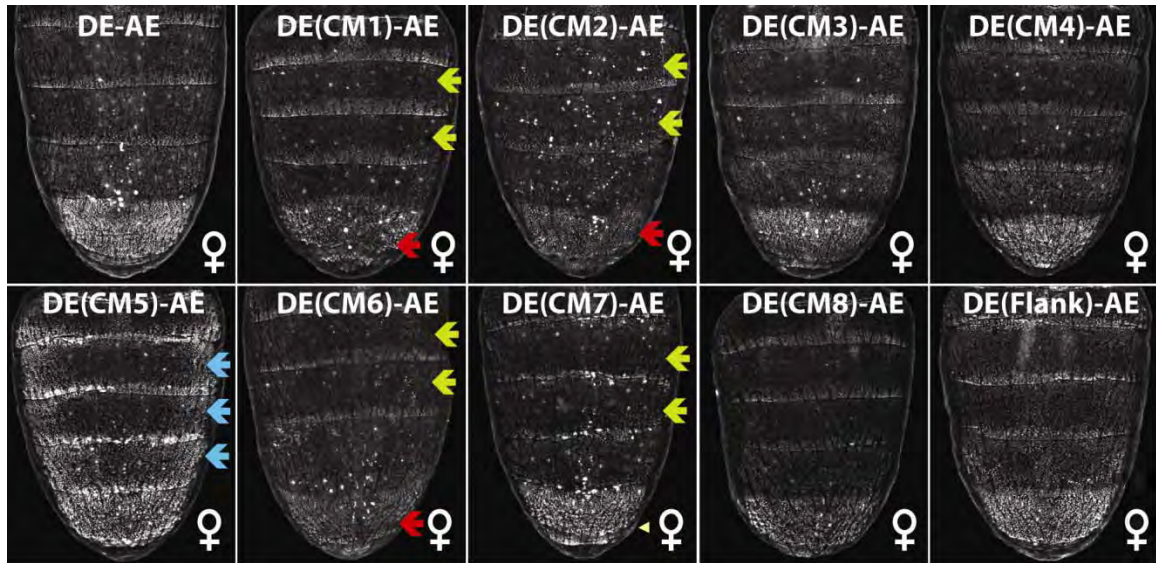
Figure 10. Effects of scanning mutations on the Switch Combination's regulatory activity. This figure shows the EGFP expression in the abdomens of transgenic female pupae at ~65-70 hours after puparium formation. These transgenes include the dimorphic element (DE) adjacent to the anterior element (AE). Scanning mutations in the core region of the dimorphic element are indicated as CM1-CM8. Scanning mutations in the entire flanking regions to the core is referred to as "Flank". Red arrowheads indicate reduced EGFP expression in the A6 segment of females compared to the wild type control (DE-AE). Yellow arrowheads indicate reduction in synergistic EGFP expression in the A3 and A4 segments. Blue arrowheads indicate segments in which a scanning mutation resulted in increased EGFP expression.

The Results of Anterior Control Element Truncations:

The synergistic EGFP expression pattern driven by the Switch Combination requires DNA sequence encodings mapped in the above scanning mutation reporter transgenes. Presumably synergism requires sequence encodings within the anterior

element. For instance the anterior element contains the information to activate Bab expression in the A2-A4 abdominal segments. Adding a further complication to an understanding as to how Bab expression is regulated during development was the finding that expression was dramatically reduced in the male abdomen during the late stages of pupal development (Salomone et al., 2013). This reduced pattern of expression looks similar to EGFP expression driven solely by the anterior element (Figure 11). We felt it was important to further investigate the anterior element to see whether there exist additional regulatory information that influences this CRE's activity.

In order to better identify the sufficient sequences responsible for the anterior element's individual regulatory activity, I created a series of additional reporter transgenes that include one possessing a larger sequence which includes the anterior element called "ACE", and a series of truncated versions of ACE called sub1-sub5 (Figure 11). These five truncations partially overlapped each other, so that they collectively spanned the entire ACE. The reporter transgenes were integrated into *D. melanogaster* attP2 genomic landing site, and the pattern of EGFP expression driven by the reporter transgenes was compared to EGFP expression when driven by the AE and the ACE. First, I found that the ACE sequence drove a more robust pattern of expression than the anterior element (Figure 11, AE). The sequences sufficient for robust EGFP expression driven by the ACE were located by identifying the truncation which mimics ACE driven EGFP expression. ACE sub4 behaved similarly to the AE, showing EGFP expression similar to that seen when driven by the AE alone (Figure 11). ACE sub1 showed light EGFP expression, similar to that when driven by the AE alone. These results of ACE sub1 and ACE sub4 suggest the AE right flank possesses additional

encodings or a separate CRE that directs early Bab expression (Figure 8). ACE sub2 exhibited robust EGFP expression similar to that seen when driven by the ACE (figure 11). Likewise, ACE sub3 and sub5 also exhibited robust EGFP expression similar to that seen when driven by the ACE (Figure 11). The sub2, sub3, and sub5 sequences behave similarly to the ACE and are similar DNA sequence content in that they each possess the right side of the anterior element and some flanking sequence (Figure 8). This suggests that the sufficient sequences for ACE driven EGFP expression are found in the shared sequence for the three truncations. Thus, some additional encodings to the right of the anterior element appear to augment the regulatory activity of the anterior element or may even be an entirely separate though closely associated CRE. Future work is needed to distinguish the specific regulatory mechanism that is occurring.
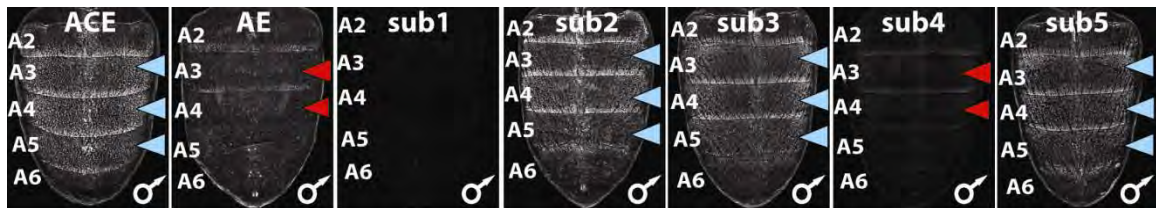

Figure 61 EGFP expressions patterns driven by the Anterior Control Element – ACE - truncation reporter transgenes. This figure compares EGFP expression patterns when driven by Anterior Control Element (ACE), the anterior element (AE), and five truncated CRE forms (sub1-sub5). The DNA sequence included in each CRE form is depicted in Figure 8.

<u>CHAPTER IV: DISCUSSION</u>

My thesis focused on studying how particular regions of DNA called ***cis-regulatory elements (CREs)*** control the manner in which gene expression is regulated. While the majority of studies on CRE function have focused on their modular effects, my thesis explored an occurrence where CREs can work in a non-modular manner to yield a synergistic pattern of gene expression.

CREs are critically important to study because mutations in CREs have serious health implications. For example, when a one nucleotide mutation is present in a CRE directing *SORT1* gene expression in the liver, an individual has an increased risk of a myocardial infarction (Musunuru et al., 2010). However, it remains mysterious why this one particular mutation impacts *SORT1* expression, as there are many other mutations that distinguish *SORT1* loci between a population of individuals, yet apparently have no health consequences. Several recent studies have suggested that variation in CRE sequences are likely to be the predominant reservoir of mutations for human phenotypic variation (Sethupathy & Collins, 2008; Visel, Rubin, & Pennacchio, 2009). Unfortunately, CREs remain one of the least well understood components of genomes and this shortcoming is exacerbated by the fact that CREs outnumber the genes in a genome (Encode & Consortium, 2011; Pennisi, 2012; The ENCODE Project Consortium, 2012). In order to better understand the role of CREs in shaping the human condition, basic questions need to be addressed about CRE function which can be expedited by studies in model organisms.

Drosophila melanogaster fruit flies provide an excellent model to study the regulatory actions of CREs on gene expression because the *D. melanogaster*

pigmentation pattern is due to a complex network of genes that are connected by the activities of CREs (Jeong et al., 2008, 2006; Rebeiz, Pool, Kassner, Aquadro, & Carroll, 2009; Rebeiz, Ramos-Womack, et al., 2009; Williams et al., 2008; Wittkopp et al., 2003).

My research focused on two aims. The first aim was to investigate a synergistic interaction between the dimorphic element and anterior element in *D. melanogaster*. The second aim was to identify sequences in a CRE called the Anterior Control Element (ACE) that are sufficient for robust Bab expression during early *Drosophila melanogaster* pupal development. My research investigating the synergistic interaction uncovered repressive and activating sequences in the dimorphic element, although, a single synergism motif was not found. Thus, multiple sequence motifs are necessary to produce the synergistic EGFP expression driven by the Switch Combination. The results from examining the ACE showed that additional sequences are necessary for robust anterior element activity because certain truncations of the ACE exhibited robust EGFP expression. Therefore, additional encodings or a separate CRE that directs early Bab expression may be located in the anterior element right flank (Figure 8 and Figure 11).

Future experimentation is needed to better locate the necessary sequences in the Switch Combination and specifically the anterior element to understand how two CREs can contribute to a pattern of expression that is greater than the sum of the parts. Specifically, the transcription factors need to be identified that interact with the novel dimorphic element sequences that I uncovered by my scanning mutagenesis approach (Figure 10). For the anterior element and the even more expansive ACE CRE, no transcription factors have been identified to date that shape the encoded regulatory activity. These factors and their binding site sequences need to be identified in order to

fully understand this case of synergism and the temporally changing and sex specific pattern of *bab* expression.

Previous research identified the additive regulatory activity of discrete CREs on individual genes (Carroll, 2005, 2008; Rebeiz & Williams, 2011). For example, the spatial pattern of *Eyeless* in *Drosophila*, is a composite of the additive activity of six distinct CREs (Carroll, 2008). However, not only can CREs act in additive ways, but research done by Eric Camino of the Williams led to a finding for *bab*, that CREs can also interact in synergistic or non-additive ways. Unfortunately, most studies seeking CREs and characterizing CREs do not test whether these CREs may possess additional synergistic functions that occur in the presence of a second or more CRE. One important future research direction is to perform similar tests to that done here for the anterior element and dimorphic element of the *bab* locus.

Since the *Drosophila melanogaster* pigmentation pattern is due to a complex network involving the interaction of CREs, *Drosophila melanogaster* fruit flies provide an excellent model to study the regulatory actions of CREs on gene expression. Specifically, my research focused on investigating a synergistic interaction between the dimorphic element and anterior element in *Drosophila melanogaster.* Eric Camino had previously assessed the activities of the dimorphic element, anterior element, and the combination of the two CREs (called "Switch Combination") on driving expression of the Enhanced Green Fluorescent Protein (EGFP) gene. The dimorphic element robustly activated EGFP protein expression in A5 and A6 abdominal segments of females while the anterior element weakly activated EGFP protein expression in the A2-A4 segments in both sexes. Surprisingly, the Switch Combination activated EGFP protein expression in

the A2-A4 segments of males and females in a non-additive or synergistic way. The

EGFP protein expression driven by the Switch Combination was produced by neither

CRE alone nor the sum of their independent activities.

My research sought to identify the synergism encoding sequences of the

dimorphic element for the synergistic Switch Combination outcome. To identify these

sequences, 9 Switch Combination reporter transgenes were created that included a

region(s) of mutant dimorphic element sequence. One of the mutants had mutated flank

regions and a non-mutated core region. The other 8 scanning mutants had non-mutated

flank regions but a mutated core region. The mutated regions of the 8 core mutants

partially overlapped, and collectively spanned the entire dimorphic element core. The

reporter transgenes were integrated into *Drosophila melanogaster* fruit flies to study the

effects of the mutations and ultimately identify synergistic encoding sequences. The

effects were analyzed using EGFP expression and comparing EGFP expression driven by

non-mutant Switch Combination to EGFP expression driven by mutant Switch

Combination. Synergism encoding sequences were identified by noting the mutants that

resulted in a loss of robust A2-A4 EGFP expression.

Surprisingly, results found that the dimorphic element possesses sequences

throughout the core region that impart additive, synergistic, activating and repressive

effects. Thus, repressive and activating sequences in the dimorphic element were

uncovered, but a single synergism motif was not found. Both DE(CM6)-AE and

DE(CM7)-AE exhibited reductions in the synergistic expression. Consequently, the

mutated sequences of DE(CM6)-AE and DE(CM7)-AE, when not mutated, could be

working together to create EGFP expression driven by the Switch Combination. Multiple

other sequences may be working together to produce the synergistic expression as well, yet the effects of knocking out one sequence motif did not have a strong enough impact to produce a significant reduction in EGFP expression. In conclusion, multiple sequences in the dimorphic element are necessary for the synergistic interaction between the dimorphic element and the anterior element.

Again, *Drosophila melanogaster* fruit fly pigmentation was utilized to study the regulatory action of CREs on gene expression. However, this part of my thesis research specifically focused on investigating the CRE called the anterior element. Previously, Bab expression at different time points during pupal development was examined. Bab expression in males is down regulated during late development, yet male Bab expression during early pupal development is robust (Salomone et al., 2013).

To explore how Bab expression in males is down regulated during late development the anterior element and its neighboring sequences were evaluated and became known as the Anterior Control Element (ACE). The Anterior Element was expanded from 1400 bp to 2400 bp to make up the ACE. The activity of the anterior element and the anterior control element to drive expression of the EGFP gene were compared. When evaluating the anterior element alone, a weak expression of EGFP was seen. However, when evaluating the ACE, a robust EGFP expression pattern that mimics male Bab expression during early pupal development was seen. My research sought to identify the sufficient sequences in the anterior control element for robust Bab expression during early pupal development. My research investigated for additional encodings or even completely separate CREs to explain how Bab expression is down regulated during

late development. Five truncations of the ACE were created to map sequences mediating AE activity.

The results showed that some additional sequences are necessary for robust anterior element activity because certain truncations of the ACE exhibited robust EGFP expression. For example, ACE sub2, ACE sub3, and sub5 exhibited robust EGFP expression similar to that seen when driven by the ACE (Figure 11). Therefore, additional encodings or a separate CRE that directs early Bab expression may be located in the common sequences for these truncations (Figure 8). In agreement with this is that ACE sub4 and ACE sub1 showed light EGFP expression, similar to that seen when driven by the anterior element (AE) alone (Figure 11). These results also suggest the AE right flank possesses additional encodings or a separate CRE that directs early Bab expression. ACE sub2 and ACE sub3 possess sequences of DNA that lie in the right flank of the AE thus also supporting this assertion. On the contrary though, because ACE sub1 and ACE sub4 result in light EGFP expression mimicking that when driven by the AE alone, a repressor may be located in the AE left flank. In addition, the robust expression driven by ACE sub2 and ACE sub3 confirm this assertion because they do not possess all the sequences in the AE left flank; therefore, ACE sub2 and ACE sub3 sequences do not include the hypothesized repressor. The activity of ACE sub5 was robust like ACE even though it included the anterior element sequence in which a repressor activity was not previously ruled out. Since robust expression occurs (Figure 11), it can be concluded that such a repressor input does not exist, but rather the right flank is adding an input or inputs that impart a more robust activity upon the anterior element.

Future research is needed to resolve with greater precision the necessary synergistic encoding sequences in the dimorphic element. To better identify the necessary synergistic encoding sequences in the dimorphic element, future research studies should examine the effects of simultaneously mutating multiple regions of DNA in the dimorphic element to test for sequences working together to produce the synergistic EGFP expression. In addition, further research is necessary to identify what binds to the interesting sequences of the dimorphic element; thus additional examination can be done potentially to identify transcription factors that regulate the Switch Combination.

Future works is needed in order to map which sequences in the anterior element/ACE work synergistically with the dimorphic element. To do this, a similar approach can be taken, as I used for the dimorphic element, with the anterior element/ACE (Figure 12). I would propose to use ten scanning mutants of the anterior element in the Switch Combination reporter transgene configuration. The mutated regions would partially overlap and collectively span the entire anterior element (Figure 12). The effects of the mutations could be analyzed by the outcome on EGFP reporter protein expression. By noting which mutant sequences knock down or increase synergistic EGFP expression driven by the Switch Combination, one can identify the necessary sequences for the synergistic expression driven by the Switch Combination. The goal would be to identify the dimorphic element's and anterior element's necessary sequences for the synergistic Switch Combination expression. From there, the aim would be to investigate whether these necessary sequences are transcription factor binding sites and, if so, to identify the transcription factors regulating the Switch Combination.
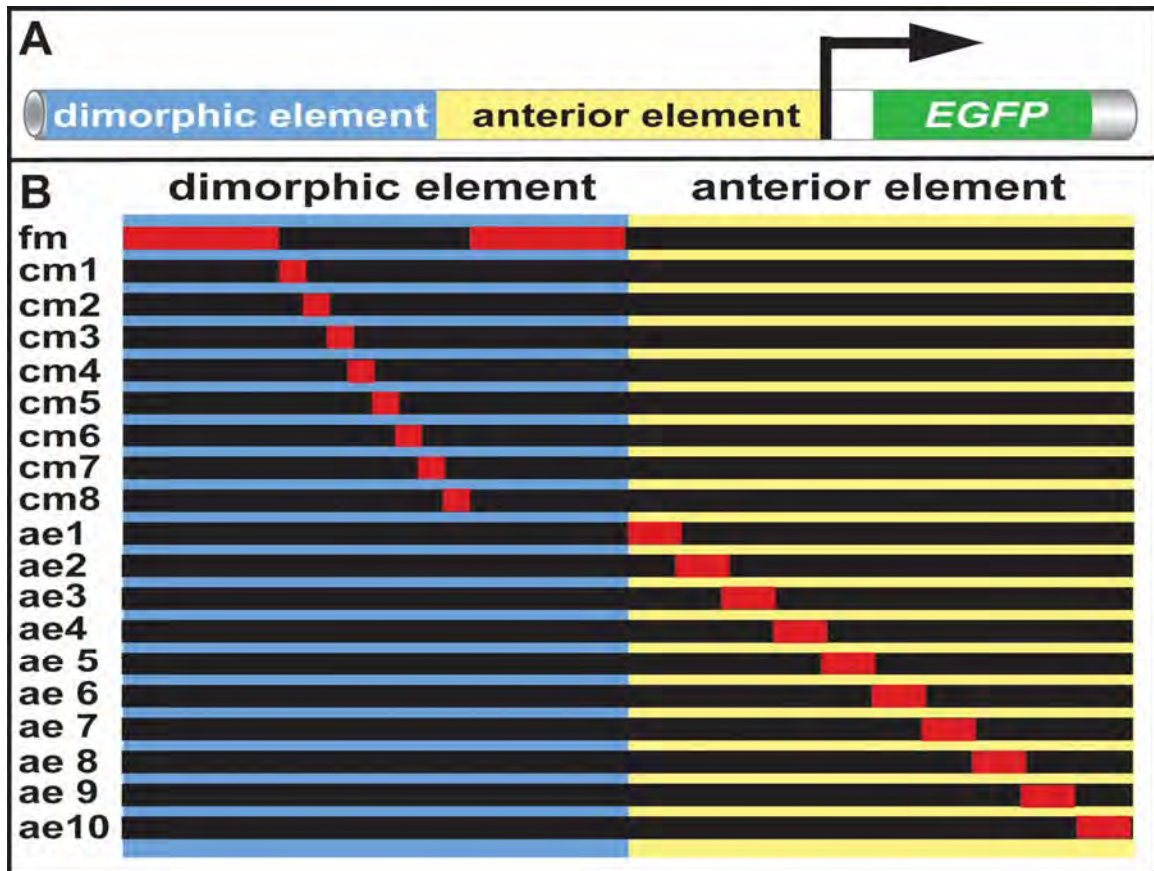
Figure 72. An approach to identify synergistic CRE encodings. (A) The configuration of the Switch Combination reporter transgene. The dimorphic element and anterior element are placed right next to each other and are linked to EGFP, so that EGFP expression exhibits the regulatory activity of the Switch Combination. (B) A schematic for scanning mutants of the dimorphic element and proposed for the anterior element. The dimorphic element scanning mutants were completed and analyzed in my thesis. Future works is needed to carry out the anterior element scheme. The red regions represent mutated sequences.

One puzzling question that my thesis work did not answer, is how Bab expression becomes reduced in the male abdomen during late pupal development (Salomone et al., 2013). The switch combination transgenes had EGFP expression results that were similar

in both males and females. One possible explanation for this might be that the EGFP

reporter protein persists in the male abdomen long after CRE activity has ceased. *In vivo*,

Bab protein may have a reduced lifespan in the male abdomen. Thus, reporter transgene

tools are needed that can report temporal changes in reporter expression. A second

possible explanation is that there exists a third or more CRE that synergistically interacts

with the anterior element and the dimorphic element. A possibility that requires more

elaborate combinations of CRE sequences in reporter transgenes or an altogether

different experimental approach.

CHAPTER V: APPENDICES

1. **Listed below are the DRE scanning mutant sequences that were designed and synthesized by GenScript Inc.:**

>**DREScramble1**
CCGCGGATATGTATATGTTCAAGTGTCAATTTTATGAATATGACTATAAACAC
GATTGATCGCTTCTGTCAATCGGTAAACAGGAAAGAggtcATAAAAAGTTGCA
GGAGGCATGTTGCCAGTTGCCTGCAACCGGCAACATTCGCAGAACAGCAGCA
ACATCGTAAAATAACTTCTTGCTCTGCGGTCTGAGTTTGGCCGCAACAATGTT
GCTGCATTTATTCGTATTATTATTACATTTTAATGAATAATTCTAATTATATGC
AACTTGAATAAGCCCGCCGATGCCAATAAAAAGCGGCGTGGCAAAGTGGAG
TGGACTGGGTTTGTGTGGCGCCCCTGCTAGTGGCACATAAAAATTGGCGCAA
GTTAATTGTGGTAGTTATTTGCTGTTTTGCCATTTGGTCATTTTACAATTTTAC
CATTTCAGCCACAACTTTTCGCACTGCTCCCCCCCTTTCCCAGCACAACAATG
TTGCGGCATTCTCGCACTTTACGAGGCGTTTTTTTTTATATCACTTACTTTAC
TTAGTTGATTAAGGGCGTGGCCGATGGGCCAGATACATGCTTAGATTTGCTCC
AGCAGTGGGCTGCATTTTACGACCCTCAAAACCCGATCCAAATGGAAAATAT
GAAAATACGGCTAATCCGCTTATGAGCACAACAAATTGGTTCACACACGCAG
C

>**DREScramble2**
CCGCGGCTCTTTCTCTTTGCCATTTTAACTTTTATTACTCTTAATATAAAAAG
CTGGCTAGATGCGGGCCAGCTGTAAACAGGAAAGAGTTAATAAACATTGGAA
TGCGTCCTTTGGACCGGTTCATTCCAACTGAACCCTGCTCCGCAAATCCGAAC
CCTAGTAAAATAACTTCTTGCTCTGCGGTCTGAGTTTGGCCGCAACAATGTTG
CTGCATTTATTCGTATTATTATTACATTTTAATGAATAATTCTAATTATATGCA
ACTTGAATAAGCCCGCCGATGCCAATAAAAAGCGGCGTGGCAAAGTGGAGT
GGACTGGGTTTGTGTGGCGCCCCTGCTAGTGGCACATAAAAATTGGCGCAAG
TTAATTGTGGTAGTTATTTGCTGTTTTGCCATTTGGTCATTTTACAATTTTACC
ATTTCAGCCACAACTTTTCGCACTGCTCCCCCCCTTTCCCAGCACAACAATGT
TGCGGCATTCTCGCACTTTACGAGGCGTTTTTTTTTATATCACTTACTTTACT

TAGTTGATTAAGGGCGTGGCCGATGGGCCAGATACATGCTTAGATTTGCTCC
AGCAGTGGGCTGCATTTTACGACCCTCAAAACCCGATCCAAATGGAAAATAT
GAAAATACGGCTAATCCGCTTATGAGCACAACAAATTGGTTCACACACGCTA
GC

>**DREScramble3**

CCGCGGCTCTTTCTCTTTGCCATTTTAACTTTTATTACTCTTAATATAAAAAG
CTGGCTAGATGCGGGCCAGCTGTAAAAATGCACGCGGTCATAAAAAGTTGCA
GGAGGCATGTTGCCAGTTGCCTGCAACCGGCAACATTCGCAGAACAGCAGCA
AAAGCGTAAACTCAATGCGTTCGCGGAGTTATTATTGTTGACGCAACAATGTT
GCGGAATTTATGCTTCTGAGTCTGAAAGTGTCAGGCAGAATTCTAATTATATG
CAACTTGAATAAGCCCGCCGATGCCAATAAAAAGCGGCGTGGCAAAGTGGA
GTGGACTGGGTTTGTGTGGCGCCCCTGCTAGTGGCACATAAAAATTGGCGCA
AGTTAATTGTGGTAGTTATTTGCTGTTTTGCCATTTGGTCATTTTACAATTTTA
CCATTTCAGCCACAACTTTTCGCACTGCTCCCCCCCTTTCCCAGCACAACAAT
GTTGCGGCATTCTCGCACTTTACGAGGCGTTTTTTTTTATATCACTTACTTTA
CTTAGTTGATTAAGGGCGTGGCCGATGGGCCAGATACATGCTTAGATTTGCTC
CAGCAGTGGGCTGCATTTTACGACCCTCAAAACCCGATCCAAATGGAAAATA
TGAAAATACGGCTAATCCGCTTATGAGCACAACAAATTGGTTCACACACGCT
AGC

>**DREScramble4**

CCGCGGCTCTTTCTCTTTGCCATTTTAACTTTTATTACTCTTAATATAAAAAG
CTGGCTAGATGCGGGCCAGCTGTAAAAATGCACGCGGTCATAAAAAGTTGCA
GGAGGCATGTTGCCAGTTGCCTGCAACCGGCAACATTCGCAGAACAGCAGCA
ACATCGTAAAATAACTTCTTGCTCTGCGGTCTGAGTTTGGCCGCAACAATGTT
GCTGCATTTATTCGTATTATTATTACATGTGACTTACTCAGTATCAGTCTCTTC
CAATGGCAGACGACAGACTAGGACCATAAACATCTGAGGGTCCACGGGTATT
TGCCGGTGGTGGTGTGGCGCCCCTGCTAGTGGCACATAAAAATTGGCGCAAG
TTAATTGTGGTAGTTATTTGCTGTTTTGCCATTTGGTCATTTTACAATTTTACC
ATTTCAGCCACAACTTTTCGCACTGCTCCCCCCCTTTCCCAGCACAACAATGT

TGCGGCATTCTCGCACTTTACGAGGCGTTTTTTTTTATATCACTTACTTTACT
TAGTTGATTAAGGGCGTGGCCGATGGGCCAGATACATGCTTAGATTTGCTCC
AGCAGTGGGCTGCATTTTACGACCCTCAAAACCCGATCCAAATGGAAAATAT
GAAAATACGGCTAATCCGCTTATGAGCACAACAAATTGGTTCACACACGCTA
GC

>**DREScramble5**

CCGCGGCTCTTTCTCTTTGCCATTTTAACTTTTATTACTCTTAATATAAAAAAG
CTGGCTAGATGCGGGCCAGCTGTAAAAATGCACGCGGTCATAAAAAGTTGCA
GGAGGCATGTTGCCAGTTGCCTGCAACCGGCAACATTCGCAGAACAGCAGCA
ACATCGTAAAATAACTTCTTGCTCTGCGGTCTGAGTTTGGCCGCAACAATGTT
GCTGCATTTATTCGTATTATTATTACATTTTAATGAATAATTCTAATTATATGC
AACTTGAATAAGCCCGCCGATGCCAATAAAAAGCGGCGTGGCAAAGTGGAG
TGGCCGGTGGTGGGGGGTCTCACATTCGATTTGAAAATAAACAGTTGAGAAC
GGTCAGTTTTGGATTGAGTGGATTTGTGGACCTGTTGGCCTTTTACAATTTTA
CCATTTCAGCCACAACTTTTCGCACTGCTCCCCCCCTTTCCCAGCACAACAAT
GTTGCGGCATTCTCGCACTTTACGAGGCGTTTTTTTTTATATCACTTACTTTA
CTTAGTTGATTAAGGGCGTGGCCGATGGGCCAGATACATGCTTAGATTTGCTC
CAGCAGTGGGCTGCATTTTACGACCCTCAAAACCCGATCCAAATGGAAAATA
TGAAAATACGGCTAATCCGCTTATGAGCACAACAAATTGGTTCACACACGCT
AGC

>**DREScramble6**

CCGCGGCTCTTTCTCTTTGCCATTTTAACTTTTATTACTCTTAATATAAAAAAG
CTGGCTAGATGCGGGCCAGCTGTAAAAATGCACGCGGTCATAAAAAGTTGCA
GGAGGCATGTTGCCAGTTGCCTGCAACCGGCAACATTCGCAGAACAGCAGCA
ACATCGTAAAATAACTTCTTGCTCTGCGGTCTGAGTTTGGCCGCAACAATGTT
GCTGCATTTATTCGTATTATTATTACATTTTAATGAATAATTCTAATTATATGC
AACTTGAATAAGCCCGCCGATGCCAATAAAAAGCGGCGTGGCAAAGTGGAG

TGGACTGGGTTTGTGTGGCGCCCCTGCTAGTGGCACATAAAAATTGGCGCAA
GTTAATTGTGGTAGTTATTTGCTGTTTTGCCCTGTTGGCCTTTTACCAGTTTAC
AAGTGCCGACCCCAATGTGCTCCCGGATACACACATGTACAATCACAACAAT
GTTGCTGAAGTATAGCACTTTACGAGGCGTTTTTTTTTATATCACTTACTTTA
CTTAGTTGATTAAGGGCGTGGCCGATGGGCCAGATACATGCTTAGATTTGCTC
CAGCAGTGGGCTGCATTTTACGACCCTCAAAACCCGATCCAAATGGAAAATA
TGAAAATACGGCTAATCCGCTTATGAGCACAACAAATTGGTTCACACACGCT
AGC

>**DREScramble7**

CCGCGGCTCTTTCTCTTTGCCATTTTAACTTTTATTACTCTTAATATAAAAAG
CTGGCTAGATGCGGGCCAGCTGTAAAAATGCACGCGGTCATAAAAAGTTGCA
GGAGGCATGTTGCCAGTTGCCTGCAACCGGCAACATTCGCAGAACAGCAGCA
ACATCGTAAAATAACTTCTTGCTCTGCGGTCTGAGTTTGGCCGCAACAATGTT
GCTGCATTTATTCGTATTATTATTACATTTTAATGAATAATTCTAATTATATGC
AACTTGAATAAGCCCGCCGATGCCAATAAAAAGCGGCGTGGCAAAGTGGAG
TGGACTGGGTTTGTGTGGCGCCCCTGCTAGTGGCACATAAAAATTGGCGCAA
GTTAATTGTGGTAGTTATTTGCTGTTTTGCCATTTGGTCATTTTACAATTTTAC
CATTTCAGCCACAACTTTTCGCACTGCTCCCCCCCTTTCCCAGCACAACAATG
TTGCTGAAGTATAGAAATTTACTATGAGGTGTGTGTTTATCTAAATGAATTTA
CGTCGGTTAGTCATGTCTTTGACTAGGTGACCGCTCCCTTCTTAGATTTGCTCC
AGCAGTGGGCTGCATTTTACGACCCTCAAAACCCGATCCAAATGGAAAATAT
GAAAATACGGCTAATCCGCTTATGAGCACAACAAATTGGTTCACACACGCTA
GC

>**DREScramble8**

CCGCGGCTCTTTCTCTTTGCCATTTTAACTTTTATTACTCTTAATATAAAAAG
CTGGCTAGATGCGGGCCAGCTGTAAAAATGCACGCGGTCATAAAAAGTTGCA
GGAGGCATGTTGCCAGTTGCCTGCAACCGGCAACATTCGCAGAACAGCAGCA
ACATCGTAAAATAACTTCTTGCTCTGCGGTCTGAGTTTGGCCGCAACAATGTT
GCTGCATTTATTCGTATTATTATTACATTTTAATGAATAATTCTAATTATATGC

AACTTGAATAAGCCCGCCGATGCCAATAAAAAGCGGCGTGGCAAAGTGGAG
TGGACTGGGTTTGTGTGGCGCCCCTGCTAGTGGCACATAAAAATTGGCGCAA
GTTAATTGTGGTAGTTATTTGCTGTTTTGCCATTTGGTCATTTTACAATTTTAC
CATTTCAGCCACAACTTTTCGCACTGCTCCCCCCCTTTCCCAGCACAACAATG
TTGCGGCATTCTCGCACTTTACGAGGCGTTTTTTTTTATATCACTTACTTTAC
TTAGTTGATTAAGGGCGTGGCCGATGGGCCCGCTCCCTTCGTCGCTGTTCGCA
ATCCGGGTGATTCCTTTTACTAACATAACACCACTAGCAACAGGTACACTCTT
ACACTCCTGATCAGCAGATGAGGCGAAAACCCACTGGTTGCCCCCCCGCTAG
C

>**DRE.04(SacII/NheI flanked core)**
GGCGCGCCCACATAAAAATCAGCAACAAAGTTGCTCTGGCCCCATAAAGAT
TGCAAACAAAAACAGAACAACAGAATGGCATGGAATAAAATTTATATGAAT
AACAAAAAGCAGCTAAAGCAAGCAGCAACAACAATAGTTTACTGCCCCGGC
TCAGCGGTACACTGTGCAAAACGTTGTACTCCTCCTCATAATAATATGAGTAT
ATAGAGTATATAATATACTATATATCTCCATTGATAATTTCGATCATTTTCAC
CTTTTAACTAATTTATGCCCAATGTAGTTGCATTTCTCTGAGTGTGCAGTAAG
TGCCCCAGAATGCGAATGCATCTCGGGTTCATCGGCGGGTCGAGTTTGTTGCA
ACAACCGAAGAACGAAGAAGTTGCAGCGTGCGTTCGGCATTAAAATTGTGTT
TATGCGTGTTCGGTAATTTTATAAAAGTTAAATTAGTTTTAAGACCCTAAATT
CAGCTCACTCTCTCTccgcggCTCTTTCTCTTTGCCATTTTAACTTTTATTACTCTT
AATATAAAAAAGCTGGCTAGATGCGGGCCAGCTGTAAAAATGCACGCGGTCA
TAAAAAGTTGCAGGAGGCATGTTGCCAGTTGCCTGCAACCGGCAACATTCGC
AGAACAGCAGCAACATCGTAAAATAACTTCTTGCTCTGCGGTCTGAGTTTGG
CCGCAACAATGTTGCTGCATTTATTCGTATTATTATTACATTTTAATGAATAAT
TCTAATTATATGCAACTTGAATAAGCCCGCCGATGCCAATAAAAAGCGGCGT
GGCAAAGTGGAGTGGACTGGGTTTGTGTGGCGCCCCTGCTAGTGGCACATAA
AAATTGGCGCAAGTTAATTGTGGTAGTTATTTGCTGTTTTGCCATTTGGTCATT
TTACAATTTTACCATTTCAGCCACAACTTTTCGCACTGCTCCCCCCCTTTCCCA
GCACAACAATGTTGCGGCATTCTCGCACTTTACGAGGCGTTTTTTTTTATATC
ACTTACTTTACTTAGTTGATTAAGGGCGTGGCCGATGGGCCAGATACATGCTT

AGATTTGCTCCAGCAGTGGGCTGCATTTTACGACCCTCAAAACCCGATCCAA
ATGGAAAATATGAAAATACGGCTAATCCGCTTATGAGCACAACAAATTGGTT
CACACACgctagcCGAAATTACTTGCGATCGCCATTTGATTGGTTTCAATGTATT
GCTTTAACTGGCAGGTGAACACTTTGTTTTTTATCTAACGATTCTTACTATTTA
ATATCCTAGTCAATTAATGTATTTTCCACTACTTCCATCGATATCACAGAGTT
CCCATTTCGCAAAGTCACATATTTGTTCTTTTATAACATGAACGCGTACCGCG
AAGGCCCCATAAAGTGTTCGCAATAAAATATATTGTGCAATAGTTATACAGC
CACTCATATACATTATATACAATATATATATGTGGATGTGTATGTGCACAA
CCATATAGATGTGTTGTATATAAATTGCCATCCCATTGCTTATCATCGCCTTTA
TAGGTAGAATGTAATTTCTTTTTATGTGCAGTTTTGCCTGCAGG

>**DRE.04(flanks mutant)**

GGCGCGCCaAaAgAcAcAgCcGaAcCcAcGgTtCgCgGtCaCaAgAcAcGcTgGaAcAaAc
AcAaAtAcCcAaAtAcTtGaAgGtAcTcAcAgTgAgAgGcAgAcCcAcAcGaAtCgAcAtCcAt
CcGaAcCcAaAcTcGgTgAaTtCaCaGtCgCcGaGtTcCcCgGgGaAcAcCtTgGgAaTaCgCa
TaAgAcTcAgAgGcGgAgAgAtAtTcTcTcAgAgAaTcTcTcTaTaCcTgGcTcAgTgatAgCc
TgTgCcCaTgTgAcCgAcTgTcTtCaCcAgGgAtTgGaAgTgCgCgGcGgGgGaAtTcAtTtCa
CaAtAcTtCtAcTtCcTaTaGtGgTaAgCtGaGtGgCtAtTgTtTgGaAcCcAaCtAcGcAaGcAt
AcGgTtCcGaGgGaGgTaGtCcTgAcAcTgGgGgTgAgGaGgGgTaGtCaGtGtCtCaCAtT
gAcAgTcGgTgTcAtAaCaTcAcTgCcGaTaAaTaTaTaTCCGCGGCTCTTTCTCTTTGCC
ATTTTAACTTTTATTACTCTTAATATAAAAAAGCTGGCTAGATGCGGGCCAGC
TGTAAAAATGCACGCGGTCATAAAAAGTTGCAGGAGGCATGTTGCCAGTTGC
CTGCAACCGGCAACATTCGCAGAACAGCAGCAACATCGTAAAATAACTTCTT
GCTCTGCGGTCTGAGTTTGGCCGCAACAATGTTGCTGCATTTATTCGTATTAT
TATTACATTTTAATGAATAATTCTAATTATATGCAACTTGAATAAGCCCGCCG
ATGCCAATAAAAAGCGGCGTGGCAAAGTGGAGTGGACTGGGTTTGTGTGGCG
CCCCTGCTAGTGGCACATAAAAATTGGCGCAAGTTAATTGTGGTAGTTATTTG
CTGTTTTGCCATTTGGTCATTTTACAATTTTACCATTTCAGCCACAACTTTTCG
CACTGCTCCCCCCCTTTCCCAGCACAACAATGTTGCGGCATTCTCGCACTTTA
CGAGGCGTTTTTTTTTTATATCACTTACTTTACTTAGTTGATTAAGGGCGTGGC
CGATGGGCCAGATACATGCTTAGATTTGCTCCAGCAGTGGGCTGCATTTTACG

ACCCTCAAAACCCGATCCAAATGGAAAATATGAAAATACGGCTAATCCGCTT
ATGAGCACAACAAATTGGTTCACACACGCTAGCaGcAcTgAaTgGaGcTaGaCcTg
TtAgTtGgTgCcAgGgAgTtCgTgAcCgGtCcGtTtAcCcCgTgGgTgTgTcTaTcAaGcTgCgT
cCgAgTgAcTcTaCgAtTaAcTgAcTtTcTgTgCaAaTcCgTaCcTaGcTcTaAaAtAtTgCaCc
TgTaGaAcAtTaAaAgAgTgGgTaTgTgAgAcCcTtAcCtCtTcCaGaGcAtGaCaCcTcAcGg
GgTaGaAcTcAcAgAgAgTtTtCcAgAtTgAgAaAtCaAaTaAgAgAaAgTcTcTcCcAgAgA
gAgAgAgGgGtAgGgGgAgGgGaAaAcCaAgAgAtAgGgGgTtTcTcTcAcTgGaCcTaCaA
gTtCgTcTaAgCtCaTgTcTcGtTcGcAgGgAcTgTaTgTgTcTtTtCcGgTgTtCCTGCAGG

2. Small DNA Gel Purification Protocol:

1. Pipet the respective amount of Buffer QG into the appropriate tubes.

   a. Weight of gel x 3000 = amount of Buffer QG need to add (microliters)

2. Place tubes in 50° C water bath for 10 minutes. Flick tubes every ~ 5 minutes.

3. Pipet 750 ul (or pour) liquid into purple spin column and spin down for 2 minutes and dump liquid

   a. Repeat until all original is gone.

4. Wash with PE Buffer- 750 ul

5. Spin down- 2 minutes and dump liquid

6. Spin down- 2 minutes and dump liquid

7. Transfer spin column to 1.5 ml tube

8. Pipet 58 ul EB Buffer

9. Spin down- 2 minutes- Discard spin column; place 1.5 ml tube with supernatant (DNA) in 20° C freezer for storage.

3. Purifying Large DNA Fragments (76 kb) Using the Qiagen QX1Kit

1. Prepare a 0.5-0.8% TAE-agarose gel with the comb possessing the very wide teeth.

2. Load ~15-20 ul of 2 log ladder in the narrow well. Then load DNA samples in the wide wells and run until the DNA bands are sufficiently separated (~ 30 min. can vary)

3. Record an image of the gel on the gel documentation system.

4. Use razor blade on UV light box to cut out the desired DNA fragments. Try to remove agarose without desired DNA.

5. Put each agarose slab with desired DNA fragment into a separate 15 ml snap cap tube and weigh the tube on the scale. Typical snap cap tube empty weighs 5 g. Tube with agarose slab, weight of 5.5 g means agarose weighs 0.5 g

6. Add 3 volumes of QX1 buffer (not QG buffer) to snap cap tube with agarose slab. 1 volume equals 100 ul per 0.1 g of agarose so a 0.5 g agarose slab one volume QX1 = 500 ul and therefore must add 1500 ul (1.5 ml) of QX1 to snap cap tube to achieve 3 volumes.

7. Add 2 volumes of milli-Q to appropriate snap cap tube with agarose slab.

8. Add 35 ul of QIAExII Suspension (slime) to each snap cap tube, snap tube closed tight and vortex for a few seconds to mix components.

9. Incubate sample at 65° C till agarose dissolves, giving brief vortex every 5 minutes. Dissolving usually take 15-20 minutes to dissolve completely.

10. Add 18 ml of dissolved gel solution to 2 ml tube and spin sample in microfuge for 2 minutes at high speed. Gently dump off supernatant into trash or sink. Repeat (adding more of same dissolved gel solution to same 2 ml tube). Until nearly all of solution from snap cap tube has been spun in a 2 ml microfuge tube.

11. After dumping off the last of supernatant from 2 ml, add 0.75 ml of fresh QX1 buffer to the tube, close and vortex until pellet has been resuspended. Then microfuge samples for 2 minutes at high speed in microfuge and dump off supernatant in trash.

12. Wash Step. Add 0.5 ml of Qiagen Buffer PE to each sample, close tube and briefly vortex to resuspend the pellet. Then spin samples in microfuge for 2 minutes at high speed. After spin pour off supernatant. Repeat Wash Step.

13. Aspirate supernatant from microfuge tube, trying to remove most of liquid. Do not touch pellet as it contains immobilized DNA fragments.

14. Place tubes in heat block set to 50°C with lids open, let samples dry for 5-10 minutes. Stop drying sample before the pellet has completely dried (white).

15. Elution of DNA from QIAEXII Suspension. Add Qiagen Buffer EB to microfuge tubes, typically 50-75 ul. Close tubes and vortex to resuspend pellet.

16. Float tubes in water bath or incubator set to 65°C for about 10-30 minutes, flicking tube every five minutes.

17. Spin the tubes in the microfuge for 3 minutes at high speed.

18. Remove supernatant by pipetting and transfer it to a fresh 1.5 ml microfuge tube. This tube contains your purified DNA. Label tube—name and date, store at -20°C.

4. Transformation

1. Set water bath to 42°C

2. Remove ligation reaction from 4°C fridge, place on ice

3. Get competent cells from freezer, place on ice

4. Pipette 2 ul of ligation reaction into snap cap tubes

5. Pipette 50 ul of competent cells (if lucigen cells pipet 40 ul instead of 50 ul) into snap cap tubes

6. Place in ice for 30 minutes

7. Water bath at 42°C for 45 seconds.

8. Place back in ice for 1 minute.

9. Remove from ice and pipette 70 ul SOB

10. Incubate at 37°C for 1 hour (in wheel). Put LB + Amp plates in incubator too.

11. Spread onto LB with Amp plates

12. Store overnight at 37°C. Allow ~18 hours of growth

5. Plasmid DNA mini prep

1. Transfer DNA from snap tubes to 2 ml flat tubes

2. Spin down for 10 minutes at 13,000 rpm—discard supernatant. (Do this step twice)

3. Re-suspend pellet using 250 ul Buffer P1 to each. Rest 3 minutes.

4. Vortex for 30 seconds

5. Add 250 ul P2 and let tubes rest for 5 minutes (will be blue)

6. Add 350 ul N3 and invert 4-6 times (until all blue is gone from tubes; neutralized)

7. Place at -20° C for 15 minutes.

8. Spin down for 10 minutes

9. Transfer supernatant to blue spin columns and spin down for 2 minutes

10. Discard supernatant and add 750 ul wash buffer PE

11. Spin down for 2 minutes—then transfer spin columns to 1.5 ul tubes

12. 60 ul Elution Buffer (EB) to each tube—let rest 1 minute.

13. Spin down for 2 minutes—throw out spin columns

14. Store at 4°C (if only a few days) or at -20°C (longer durations)

6. Qiagen "DNeasy" Kit Genomic DNA Prep

1. Put on gloves and spread ethanol on gloves. Add 180 ul 1 x PBS to 1.5 ml tube and homogenize anesthetized flies.

2. Add 20 ul Qiagen Proteinase K

3. Add 200 ul of Buffer AL→ vortex then heat at 56° C for 60 minutes

4. Pellet debris by centrifugation at 12,000 RPM—3 minutes. Transfer supernatant to a fresh tube

5. Add 200 ul of 100% Ethanol to sample and vortex

6. Transfer supernatant to a DNeasy mini column and centrifuge for 1 minute at 6000 xg

7. Transfer column to fresh collection tube, add 500 ul of Buffer AW1 and centrifuge for 1 minute at 6000 xg.

8. Transfer column to fresh collection tube. Add 500 ul of Buffer AW2 and centrifuge for 3 minutes at 20,000 xg.

9. Dump collection tube and centrifuge column for 1 more minute

10. Place DNeasy column into a fresh 1.5 ml tube. Add 200 ul Buffer AE to the column, let column stand for 1 minute.

11. Centrifuge column for 1 minute at 6000 xg to elute, dispose of column and label lids of 1.5 ml tubes.

7. **Enzyme Removal- Qiagen PCR clean-up**

    1. PB buffer- 5 x amount of DNA digest- PB buffer 350 ul (70 x 5) pipetted into spin column with DNA digest

    2. Spin down for 2 minutes at max, remove supernatant

    3. PE buffer 750 ul → spin down for 2 minutes at max → repeat spin down

    4. Transfer spin column to 1.5 ml tube

    5. EB Buffer 35 ul, rest for 1 minute, spin down for 2 minutes, discard spin column

CHAPTER V1: REFERENCES

Arnone, M. I., & Davidson, E. H. (1997). The hardwiring of development: organization and function of genomic regulatory systems. *Development (Cambridge, England)*, *124*(10), 1851–64.

Bagga, J. S., & D'Antonio, L. A. (2013). Role of conserved cis-regulatory elements in the post-transcriptional regulation of the human MECP2 gene involved in autism. *Human Genomics*, *7*, 19. doi:10.1186/1479-7364-7-19

Carroll, S. B. (2005). Evolution at two levels: on genes and form. *PLoS Biology*, *3*(7), e245. doi:10.1371/journal.pbio.0030245

Carroll, S. B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*, *134*(1), 25–36. doi:10.1016/j.cell.2008.06.030

Couderc, J.-L., Godt, D., Zollman, S., Chen, J., Li, M., Tiong, S., … Laski, F. A. (2002). The bric a brac locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in Drosophila. *Development*, *129*(10), 2419–2433. Retrieved from http://dev.biologists.org/content/129/10/2419.short

Davidson, E. H. (2006). *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution*. Burlington, MA: Elsevier Inc.

Encode, T., & Consortium, P. (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biology*, *9*(4), e1001046. doi:10.1371/journal.pbio.1001046

Griffiths, A., Wessler, S., Carroll, S. B., & Doebley, J. (2011). *Introduction to Genetic Analysis* (10th ed.). New York: W. H. Freeman and Company.

Groth, A. C., Fish, M., Nusse, R., & Calos, M. P. (2004). Construction of Transgenic Drosophila by Using the Site-Specific Integrase From Phage phiC31. *Genetics*, *166*(April), 1775–1782.

Jeong, S., Rebeiz, M., Andolfatto, P., Werner, T., True, J., & Carroll, S. B. (2008). The evolution of gene regulation underlies a morphological difference between two Drosophila sister species. *Cell*, *132*(5), 783–93. doi:10.1016/j.cell.2008.01.014

Jeong, S., Rokas, A., & Carroll, S. B. (2006). Regulation of Body Pigmentation by the Abdominal-B Hox Protein and Its Gain and Loss in Drosophila Evolution. *Cell*, *125*(7), 1387–1399. doi:10.1016/j.cell.2006.04.043

Kadonaga, J. T. (2012). Perspectives on the RNA polymerase II core promoter. *Wiley Interdisciplinary Reviews. Developmental Biology*, *1*(1), 40–51. doi:10.1002/wdev.21

Kopp, A., Duncan, I., & Carroll, S. B. (2000). Genetic control and evolution of sexually dimorphic characters in Drosophila. *Nature*, *408*(6812), 553–9. doi:10.1038/35046017

Lagha, M., Bothma, J. P., & Levine, M. (2012). Mechanisms of transcriptional precision in animal development. *Trends in Genetics : TIG*, 1–8. doi:10.1016/j.tig.2012.03.006

Lettice, L. A., Heaney, S. J. H., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., … de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, *12*(14), 1725–1735. doi:10.1093/hmg/ddg180

Lomvardas, S., Barnea, G., Pisapia, D. J., Mendelsohn, M., Kirkland, J., & Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. *Cell*, *126*(2), 403–13. doi:10.1016/j.cell.2006.06.035

Miele, A., & Dekker, J. (2009). Mapping cis- and trans- chromatin interaction networks using chromosome conformation capture (3C). *Methods in Molecular Biology (Clifton, N.J.)*, *464*, 105–21. doi:10.1007/978-1-60327-461-6_7

Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N. E., Ahfeldt, T., Sachs, K. V, … Rader, D. J. (2010). From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature*, *466*(7307), 714–9. doi:10.1038/nature09266

Pennisi, E. (2012). ENCODE Project Writes Eulogy For Junk DNA. *Science*, *337*(7 September), 1159–1160.

Pfreundt, U., James, D. P., Tweedie, S., Wilson, D., Teichmann, S. A., & Adryan, B. (2010). FlyTF: improved annotation and enhanced functionality of the Drosophila transcription factor database. *Nucleic Acids Research*, *38*(Database issue), D443–7. doi:10.1093/nar/gkp910

Rebeiz, M., Pool, J. E., Kassner, V. A., Aquadro, C. F., & Carroll, S. B. (2009). Stepwise modification of a modular enhancer underlies adaptation in a Drosophila population. *Science (New York, N.Y.)*, *326*(5960), 1663–7. doi:10.1126/science.1178357

Rebeiz, M., Ramos-Womack, M., Jeong, S., Andolfatto, P., Werner, T., True, J., … Carroll, S. B. (2009). Evolution of the tan locus contributed to pigment loss in Drosophila santomea: a response to Matute et al. *Cell*, *139*(6), 1189–96. doi:10.1016/j.cell.2009.11.004

Rebeiz, M., & Williams, T. M. (2011). Experimental Approaches to Evaluate the Contributions of Candidate Cis- regulatory Mutations to Phenotypic Evolution. In V. Orgogozo & M. V. Rockman (Eds.), *Methods in Molecular Biology* (Vol. 772, pp. 351–375). Totowa, NJ: Humana Press. doi:10.1007/978-1-61779-228-1

Rogers, W. A., Salomone, J. R., Tacy, D. J., Camino, E. M., Davis, K. A., Rebeiz, M., & Williams, T. M. (2013). Recurrent modification of a conserved cis-regulatory element underlies fruit fly pigmentation diversity. *PLoS Genetics*, *9*(8), e1003740. doi:10.1371/journal.pgen.1003740

Rogers, W. A., & Williams, T. M. (2011). Quantitative comparison of cis-regulatory element (CRE) activities in transgenic Drosophila melanogaster. *Journal of Visualized Experiments : JoVE*, (58). doi:10.3791/3395

Salomone, J. R., Rogers, W. A., Rebeiz, M., & Williams, T. M. (2013). The evolution of Bab paralog expression and abdominal pigmentation among Sophophora fruit fly species. *Evolution & Development*, *15*(6), 442–57. doi:10.1111/ede.12053

Sethupathy, P., & Collins, F. S. (2008). MicroRNA target site polymorphisms and human disease. *Trends in Genetics : TIG*, *24*(10), 489–97. doi:10.1016/j.tig.2008.07.004

Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R., & Flavell, R. A. (2005). Interchromosomal associations between alternatively expressed loci. *Nature*, *435*(7042), 637–45. doi:10.1038/nature03574

The ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. doi:10.1038/nature11247

Visel, A., Rubin, E. M., & Pennacchio, L. A. (2009). Genomic views of distant-acting enhancers. *Nature*, *461*(7261), 199–205. doi:10.1038/nature08451

Williams, T. M., Selegue, J. E., Werner, T., Gompel, N., Kopp, A., & Carroll, S. B. (2008). The regulation and evolution of a genetic switch controlling sexually dimorphic traits in Drosophila. *Cell*, *134*(4), 610–23. doi:10.1016/j.cell.2008.06.052

Wittkopp, P. J., Carroll, S. B., & Kopp, A. (2003). Evolution in black and white: genetic control of pigment patterns in Drosophila. *Trends in Genetics : TIG*, *19*(9), 495–504. doi:10.1016/S0168-9525(03)00194-X

Xu, N., Tsai, C.-L., & Lee, J. T. (2006). Transient homologous chromosome pairing marks the onset of X inactivation. *Science (New York, N.Y.)*, *311*(5764), 1149–52. doi:10.1126/science.1122984