Analysis Of Long Range Gene Regulation In *Drosophila;*

Insights From The *dPax2 sparking* Enhancer

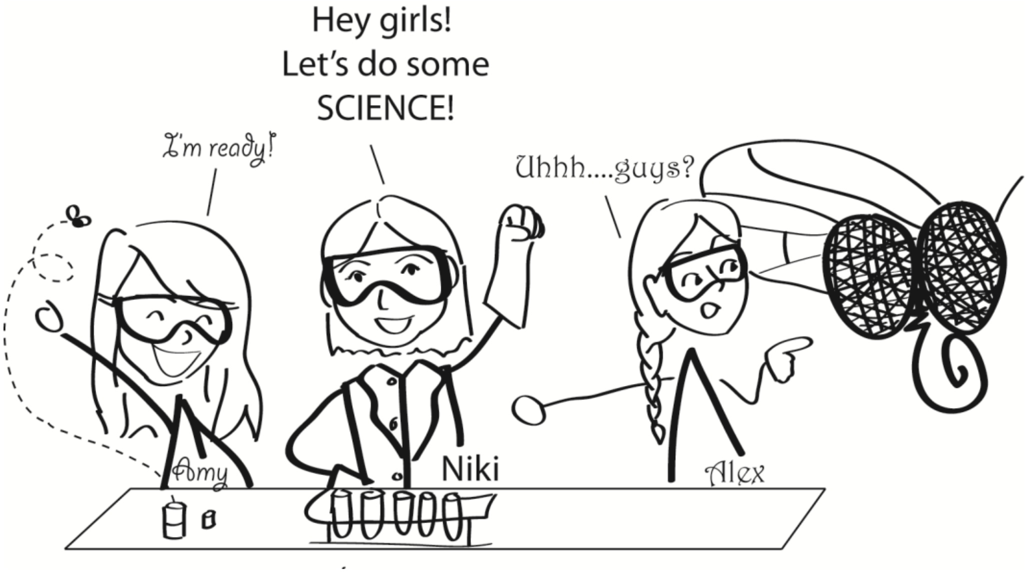
by

Nicole C. Evans

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Cell and Developmental Biology)
in the University of Michigan
2012

Doctoral Committee:

Associate Professor Scott Barolo, Chair
Professor Kenneth Cadigan
Assistant Professor Catherine Collins
Assistant Professor Kentaro Nabeshima


Hey girls!
Let's do some
SCIENCE!
I'm ready!
Uhhh....guys?
Amy
Niki
Alex

DEDICATION

To my family, who have been unfailingly supportive through my entire education and the doctoral experience. My Grandpa Vern, who assigned my sister and me our future careers at age ten, but would have loved me just the same if I had not followed his instructions. To my Grandma Judy, who can make anything positive and has always been there for me with her love and encouragement. To my sister Katelyn, whose sense of humor and sassiness makes my world bright. Especially to my parents, my Dad who taught me to approach every problem with a creative and open mind and my Mom who inspired me to learn and teach others. Without their laughter, love, and support I would not be here chasing my dreams.

ACKNOWLEDGMENTS

I deeply thank Christina Swanson for her friendship and enabling me to pursue a study of the *sparking* enhancer. I am grateful to Amy Strom and Alex Chappell whose smiling faces and eager brains remind me why I love science every day, for their contributions to this work, and for all the mini preps.

TABLE OF CONTENTS

List of Figures

List of Tables

CHAPTER 1

SPARKLING INSIGHTS INTO ENHANCER, STRUCTURE, FUNCTION, AND EVOLUTION

How does a fertilized egg develop into a complex organism, in which individual cells with the same genome take on a wide diversity of forms and functions? This is the fundamental question of developmental biology. A large part of the answer is *differential gene expression*, a process by which each cell expresses a unique subset of its genes at the correct time, and in the correct location. The discovery of a DNA fragment from the Simian Virus 40 just over thirty years ago that could induce a 200 fold increase in transcription of the rabbit β-globin gene opened the door wide for the study of gene regulation (Banerji et al., 1981). Coined an "enhancer" by the authors, this viral DNA sequence can act independent of its orientation, in conjunction with several promoters, and at significant genomic distances to drive gene transcription. The identification and characterization of the SV40 enhancer soon led to the discovery of similarly functioning cis-regulatory sequences across evolutionarily distant species (Khoury and Gruss, 1983).

1.1 The enhancer: what is it and what can it do?

In the thirty years to follow, enhancers have been recognized as critical subclass of *cis*-regulatory elements, consisting of genomic sequences that control gene transcription, both qualitatively and quantitatively, through a wide variety of mechanisms (Blackwood and Kadonaga, 1998; Bulger and Groudine, 2010; Levine, 2010). Eukaryotic enhancers are classically depicted as clusters of transcription factor binding sites (TFBS), found somewhere in the genomic neighborhood of the gene(s) they regulate, where they integrate signals from the cellular environment to direct the timing, levels, and cell-type specificity of gene expression. Enhancers can be located 5' or 3' of their target gene and can often be found within introns or UTRs of the transcription unit itself, but they do not always target the promoter nearest to them. *cis*-regulatory sequences vary greatly in size, typically ranging on average from hundreds of base pairs to a few kilobases. Extremely long or short enhancer sequences have been identified such as the 53bp testes-specific enhancer for *gonadal* expression to 5kp enhancer that drives stripes of *runt* expression in the *Drosophila* embryo (Klingler et al., 1996; Schulz et al., 1990). However, in many reported cases of large enhancers, little effort has gone toward functionally defining a truly minimal element *in vivo*. The boundaries of *cis*-regulatory elements are difficult to draw with precision: the “minimal” enhancer (i.e., the smallest fragment that is sufficient to generate a given pattern) is often weaker than larger fragments including the minimal element. Highly trimmed sequences sometimes drive a restricted or expanded pattern of gene expression, compared to larger fragments or the gene locus as a whole.

A single gene can be regulated by multiple enhancers, each responsible for a specific domain of the gene’s complete expression pattern, a characteristic referred to as enhancer modularity. A classic example of multi-modular *cis*-

regulation is the *even-skipped* gene of *Drosophila*, whose seven stripes of embryonic expression are controlled by independent enhancers, as are later aspects of gene expression in muscle precursors, the central nervous system, and elsewhere(Goto et al., 1989; Harding et al., 1989; Sackerson et al., 1999; Small et al., 1992). Another informative case study is the sea urchin *endo16* gene, which contains over 30 high-specificity binding sites spread over a 2.3-kb region (Yuh and Davidson, 1996). In other well-documented cases such as for the *Drosophila brinker*, multiple separate enhancers contribute to a single aspect of a gene's expression pattern (Frankel et al., 2010; Levine, 2010; Perry et al., 2010; Yao et al., 2008). It is unclear in these cases whether these broad enhancer regions represent multiple modular enhancers, or if the total regulatory input is simple spread out.

1.1a How do enhancers activate gene expression?

Enhancers are composed of combinations of protein binding sites, which recruit sequence-specific TFs. These TFs, in turn, recruit non-DNA-binding cofactors, which regulate transcription through a variety of mechanisms, including direct recruitment of RNA polymerase II and the basal transcription machinery, either directly or via Mediator, a large multiunit complex that promotes transcription via assembly of the basal transcription machinery (Malik and Roeder, 2005; Szutorisz et al., 2005; Wang et al., 2005). Enhancers also influence their local chromatin environment via epigenetic changes—for example, by recruiting ATP-dependent nucleosome remodeling complexes or histone acetyltransferases (HATs) or deacetylases (HDACs)—resulting in changes in chromatin structure that stimulate (or inhibit) transcription (Narlikar et al., 2002; Orphanides and Reinberg, 2002). While these biochemical activities are essential for proper transcriptional regulation in many contexts, they have not been shown to be sufficient to explain enhancer-mediated gene expression *in vivo*. In fact, growing evidence in the field supports additional mechanisms of enhancer action such as the production of non-coding RNAs that stimulate transcription, long distance communication between enhancers and promoters, and changes in the subnuclear localization of DNA (Drissen et al., 2004; Kagey et al., 2010; Kim et al., 2010; Ong and Corces, 2011; Orom et al., 2010; Spilianakis and Flavell, 2004; Tsai et al., 2010; Vakoc et al., 2005; Wang et al., 2011).

1.1b Combinatorial inputs determine transcriptional output

As all enhancers are directly bound and regulated by transcription factors and their cofactors the identity, combination, organization, and spacing of these binding sites are a basic and crucial aspect of enhancer structure. Most developmental enhancers require a specific combination of activator and repressor sites. Some of these sites are bound by the affecters of cell signaling pathways, while others are bound by locally expressed factors. The integration of these signals allows the correct level, timing, and location of gene expression – a method widely known as combinatorial control. For example, the enhancer that activates *Drosophila prospero* expression in the R7 photoreceptors and cone cells of the eye is directly regulated by the EGFR and Notch intercellular signaling cascades, via direct binding of the transcription factors Pointed/Yan and Suppressor of Hairless (Xu et al., 2000). However, EGFR and Notch signaling in the *Drosophila* eye is not limited to these specific cell types (Shilo, 2003; Voas and Rebay, 2004) which is also true of the other proteins known to interact with the *prospero* enhancer – Glass, Sine oculis, Lozenge and Seven-up (Hayashi et al., 2008). Thus, it is the combination of these signals on the enhancer that activate *prospero* expression in the correct cells.

1.1c Models of enhancer structure

The presence and regulatory significance of enhancer “structure” or “grammar” (i.e., the arrangement and spacing of TFBSs) is currently a topic of active debate in the field (Crocker and Erives, 2008; Crocker et al., 2010; Hare et al., 2008a; Kulkarni and Arnosti, 2005; Papatsenko et al., 2009; Papatsenko and Levine, 2007; Rastegar et al., 2008; Swanson et al., 2011). Two quite different views of enhancer organization, as it relates to function, are the “enhanceosome” model and the “information display” (or “billboard”) model (Figure 1.1) (Arnosti and Kulkarni, 2005). In a *cis*-regulatory sequence defined as an enhanceosome, the organization of binding sites within an enhancer is highly constrained, such that only one arrangement results in proper gene expression. Changes in spacing between TFBSs inhibit enhancer activity due to the disruption of local protein–

protein interactions results in loss of cooperative binding and synergistic activation (Figure 1.1). Well-studied examples include enhanceosomes of the *IFN-β* and *TCRa* genes which require very specific binding site spacing and organization (Giese et al., 1995; Thanos and Maniatis, 1995), however this may be a relatively rare class of enhancer.

Conversely, the "information display" model proposes that control of gene transcription is controlled more loosely, by simply acquiring the correct amount of "positive inputs" (in the absence of negative inputs). Under this model, the organization and spacing of TFBSs can be quite flexible, and a high degree of cooperativity among TFs may not be required, since the enhancer presents multiple semiredundant contact surfaces for coactivators or the basal transcription complex. Note that this view seems to imply only a single class of "activation activity," to which all activating TFs contribute: the total amount of this activity recruited to the enhancer is the critical factor determining gene activation (Figure 1.1). Evidence for this model stems from the well-studied stripe 2 enhancer of the *Drosophila even-skipped* gene (eve S2E). While *eve* S2E is fairly well conserved overall, TFBS arrangement shows significant variation among *drosophilid* and *sepsid* fly species (Hare et al., 2008b; Ludwig et al., 1998). Additionally, the effects of destroying activator sites within *eve* S2E can be rescued by adding binding sites for a heterologous activator (Arnosti et al., 1996) suggesting that the requirements for transcriptional activation can be extremely flexible. Together this information supports a model whereby simply integrating

Figure 1.1

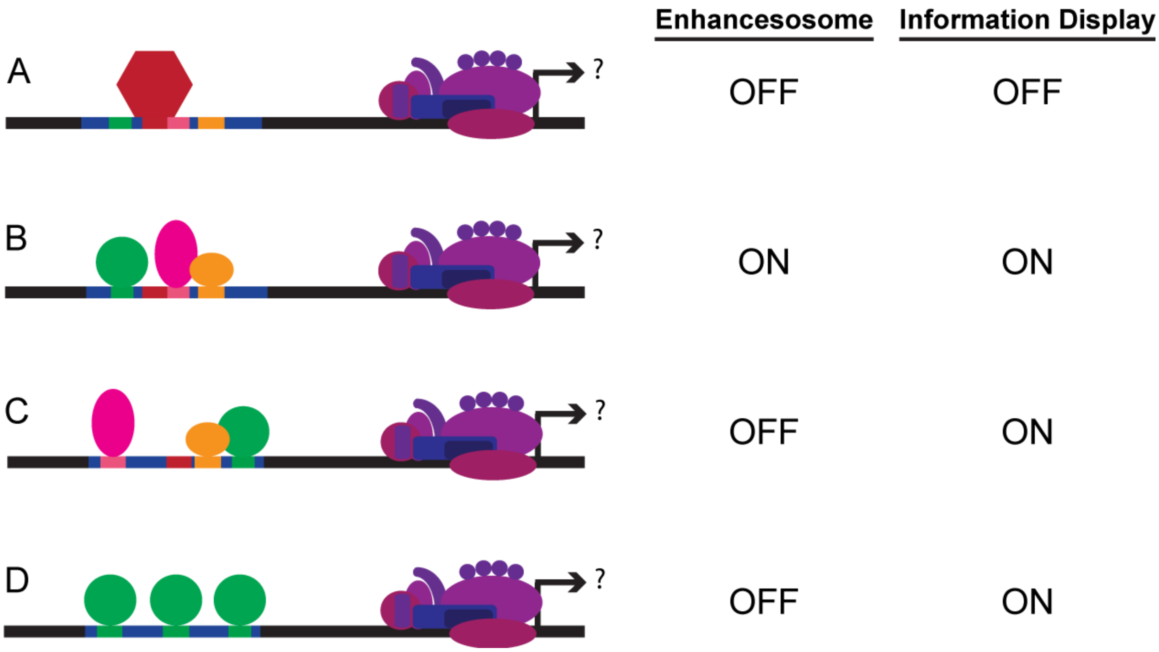


Figure 1.1 Enhancesome vs Information Display models of enhancer structure. The enhansomesome model proposes that the identity and location of transcription factor binding sites is precise. The information display model proposes that transcription factor identity and order is flexible. The presence of a repressor protein (red) inhibits transcription in both models (A). B-D show the predicted outcomes in each model when activator proteins (green, pink, yellow, and blue) are in different arrangements (B) or different identities (C)

the adequate number of inputs results in transcriptional activity in the correct time and place. While these two models are supported by a small number of endogenous enhancers, most enhancers cannot be completely or accurately characterized by either model.

1.1d Persistent questions in the field

Over the past 30 years, enhancers—that is, *cis*-regulatory genomic sequences that stimulate promoter activation—have been identified in all examined forms of life: viruses, bacteria, yeast, and multicellular animals and plants all use this strategy to control gene expression (Banerji et al., 1981; Levine, 2010; Priest et al., 2009). Nevertheless, our knowledge of the basic components and structure of the enhancer remains far from complete. This is a problem for those wishing to understand complex biological systems because, as mentioned above, enhancers are responsible for processing, integrating, and generating complex patterning information. For example, cell–cell signaling pathways (Notch, Hedgehog, BMP, Wnt, etc.) pattern developing tissues and stem cell systems, primarily by directing gene expression via signal-regulated enhancers, yet to date no signal-regulated enhancer has been fully characterized (Barolo and Posakony, 2002; Johnson et al., 2008). Even the extensively studied *eve* S2E is not completely defined with respect to its essential regulatory inputs (Andrioli et al., 2002). Until an enhancer has been characterized to the point that all regulatory sites are defined, the TFs and biochemical activities recruited by those sites are known, and the spatial relationships among those sites (if any) are understood, we cannot fully grasp the relationship between *cis*-regulatory DNA sequence and gene expression patterning. This lack of understanding hampers our ability to mine new enhancers from the genome based on their DNA sequence or binding site composition, as well as our ability to create custom enhancers as research tools and therapeutic agents.

Furthermore, our basic understanding of the steps involved in the process of transcriptional activation remains incomplete: how many different biochemical activities are required to activate transcription? In which order do they occur? Do individual TFBSs recruit distinct "activation activities," which in combination allow for transcriptional activation, or do all activating binding sites recruit similar activities, which must merely reach a quantitative threshold to trigger activation? If the latter, do different enhancers use different combinations of activation activities, or is there a universal basic recipe? In order to better understand enhancer structure and function the Barolo laboratory at the University of Michigan has undertaken an extensive characterization of cone-cell-specific eye enhancer of the *Drosophila dPax2* gene, *sparkling.*

1.2 The *sparkling* enhancer of the *dPax2* gene

The *Drosophila* compound eye consists of approximately 750 simple eyes, or ommatidia, each composed of eight photoreceptors (R1-R8) and four cone cells, surrounded by two primary pigment cells (PPCs), six secondary pigment cells, three tertiary pigment cells, and three mechanosensory bristles (Figure 1.2

Figure 1.2

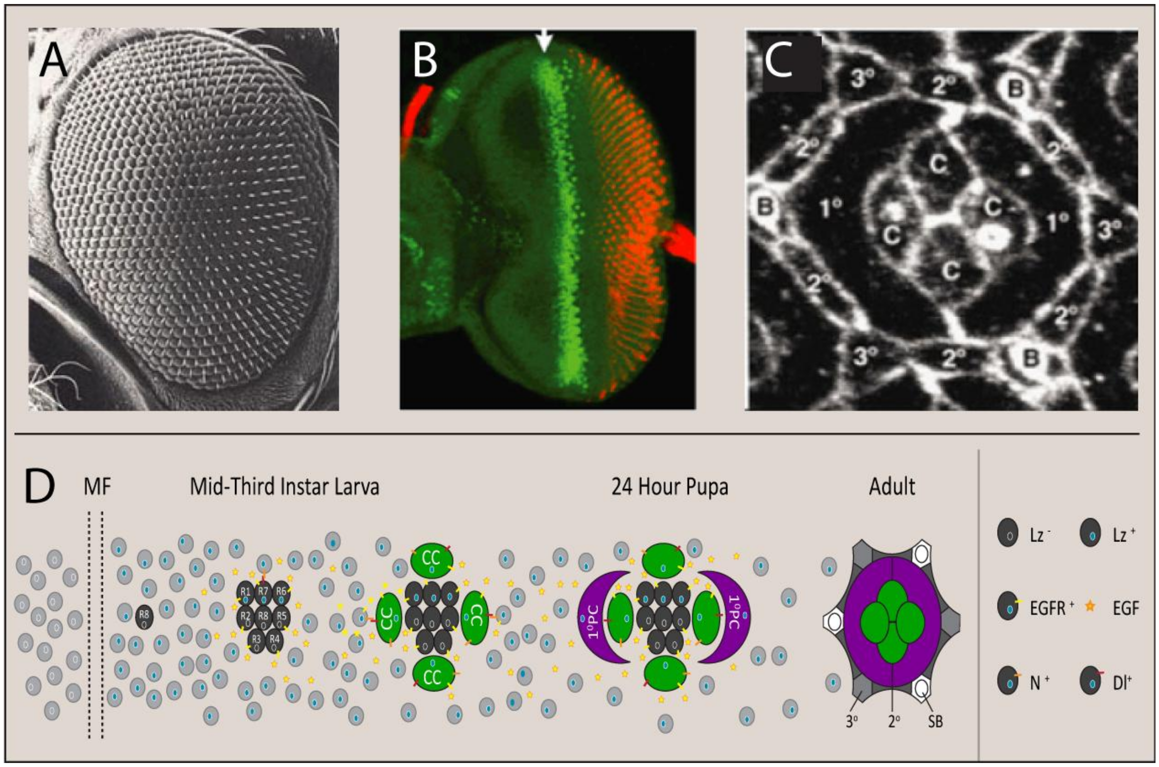


Figure 1.2 *Drosophila* eye development.(A) SEM image of the adult Drosophila eye consistiting of approximately 750 ommatidia. (A) SEM image of the adult *Drosophila* eye consistiting of approximately 750 ommatidia. (B) Cell type specification begins in the third instar larval eye disc. Differentiation occurs posterior to the morphogenic furrow (MF) which moves across the disc from the posterior to the anterior (right to left). The MF is indicated by the arrow head and green staining for *atonal* while differentiated photoreceptors (R8) are shown in red. C. By pupal stage ommatidium contains 4 cone cells (C), 2 primary pigment cells (1°), 6 secondary pigment cells (2°), 3 tertiary pigment cells (3°) and 3 bristle cells (B). The 8 photorereceptrs have also been specified, however they are located in a different plane of focus, and cannot be seen in this image. D. The eye cell types are specified sequentially via Notch and EGFR signally and Lz. First, Notch mediated lateral inhibition at the MF specifies photoreceptor 8 (R8). R2, R3, R4, R5 are then specified via EGFR signaling. R1 and R6 are then specified by EGFR signaling and the transcription factor LZ while R7 requires these factors as well as Notch signaling. Utilizing Notch, EGFR, and Lz signals, the cone cells and subsequently the primary pigment cells differentiate. Images b and c (Kumar, 2001; Voas and Rebay, 2004)

A,B) . During eye development, which occurs in the eye imaginal disc during larval and pupal stages, sequential EGFR and Notch signaling events recruit

undifferentiated retinal cells to the above cell fates (Voas and Rebay, 2004). Beginning at the morphogenetic furrow, notch mediated lateral inhibition specifies the R8 cell in each ommotidium (Figure 1.2B). Subsequently, the rest of the photoreceptors, the cone cells, and finally the pigment cells are specified. Specification of these cells is regulated by both EGFR and Notch signaling. Additionally, specification of the R1, R6, R7, cone, and PPC fates requires expression of the Runx-family TF Lozenge (Lz) (Figure 1.2). Because EGFR, Notch, and Lz are all broadly active in the retina, additional activators and repressors must act in concert with these signals to correctly determine cell-type-specific gene expression and differentiation within the eye, as will be discussed below.

dPax2 expression, which is required for proper cone cell differentiation and maintenance directly depends on EGFR and Notch signaling, as well as Lz (Flores et al., 2000; Fu and Noll, 1997; Shi and Noll, 2009). The enhancer responsible for cone-cell-specific expression of *dPax2* was identified in the fourth intron of the *dPax2* gene (Figure 1.3), thanks to spontaneous and P-element mediated mutations in that region that impair expression of *dPax2* in cone cells without disrupting its expression or function in other tissues (*dPax2* is also expressed in primary pigment cells and in sheath and shaft cells of the mechanosensory bristle, but these expression domains are under the control of different enhancers). The regulatory information for bristle cell expression lies in two enhancers that are both upstream of the *dPax2* transcription start site (Johnson et al., 2011). Meanwhile, the primary pigment cell enhancer must also lie in the fourth intron of the *dPax2* gene based on mutational analysis; however the exact location of this regulatory sequence has not been determined (Fu et al., 1998; Fu and Noll, 1997; Johnson et al., 2011). Mutations to *dPax2* that affect cone cells were originally called *sparkling* alleles, so the cone-cell enhancer was named *sparkling* (*spa*) (Fu et al., 1998).

Subsequent transgenic analyses identified a minimal 362bp sequence that is sufficient for cone-cell-specific gene expression (Flores et al., 2000), which we will refer to here as *spa* (Figure 1.3). This sequence contains five binding sites for the Notch effector Suppressor of Hairless [Su(H)] (Flores et al., 2000). In the absence of Notch signaling, Su(H) confers direct repression on its target genes, while in the presence of active Notch signaling (as occurs in cone cell precursors) it mediates direct activation, thereby acting as a signal-regulated transcriptional switch (Barolo and Posakony, 2002; Barolo et al., 2000; Bray, 2006; Morel and Schweisguth, 2000). *spa* also contains three MGGAW consensus Ets factor binding sites, which are directly bound *in vitro* by one or both of two Ets-family effectors of EGFR/MAPK signaling, Pointed P2 (PntP2), and Yan/Aop (Flores et al., 2000). In the presence of EGFR signaling, PntP2 is phosphorylated and activates gene transcription, while in the absence of

Figure 1.3

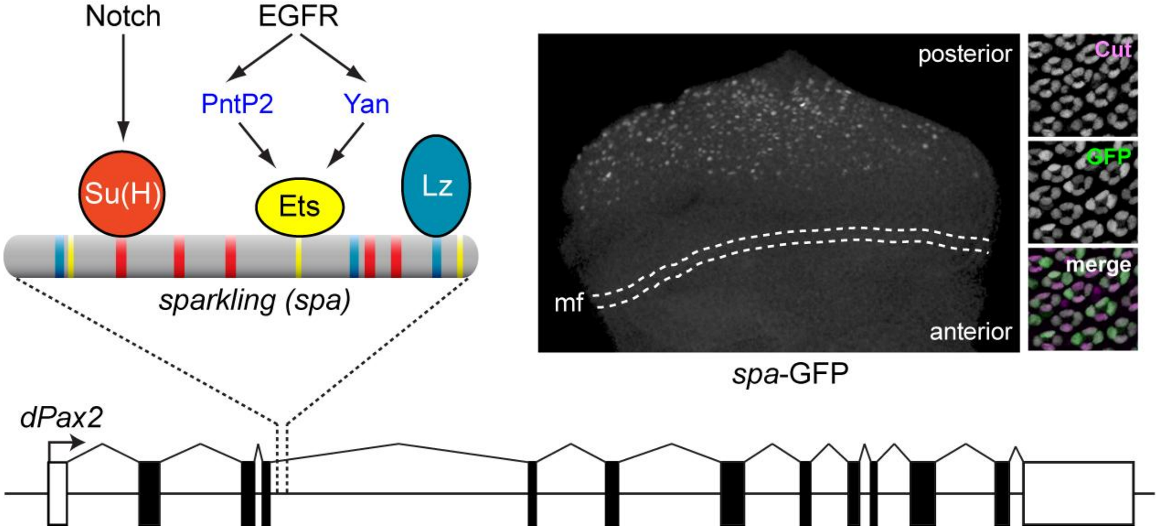
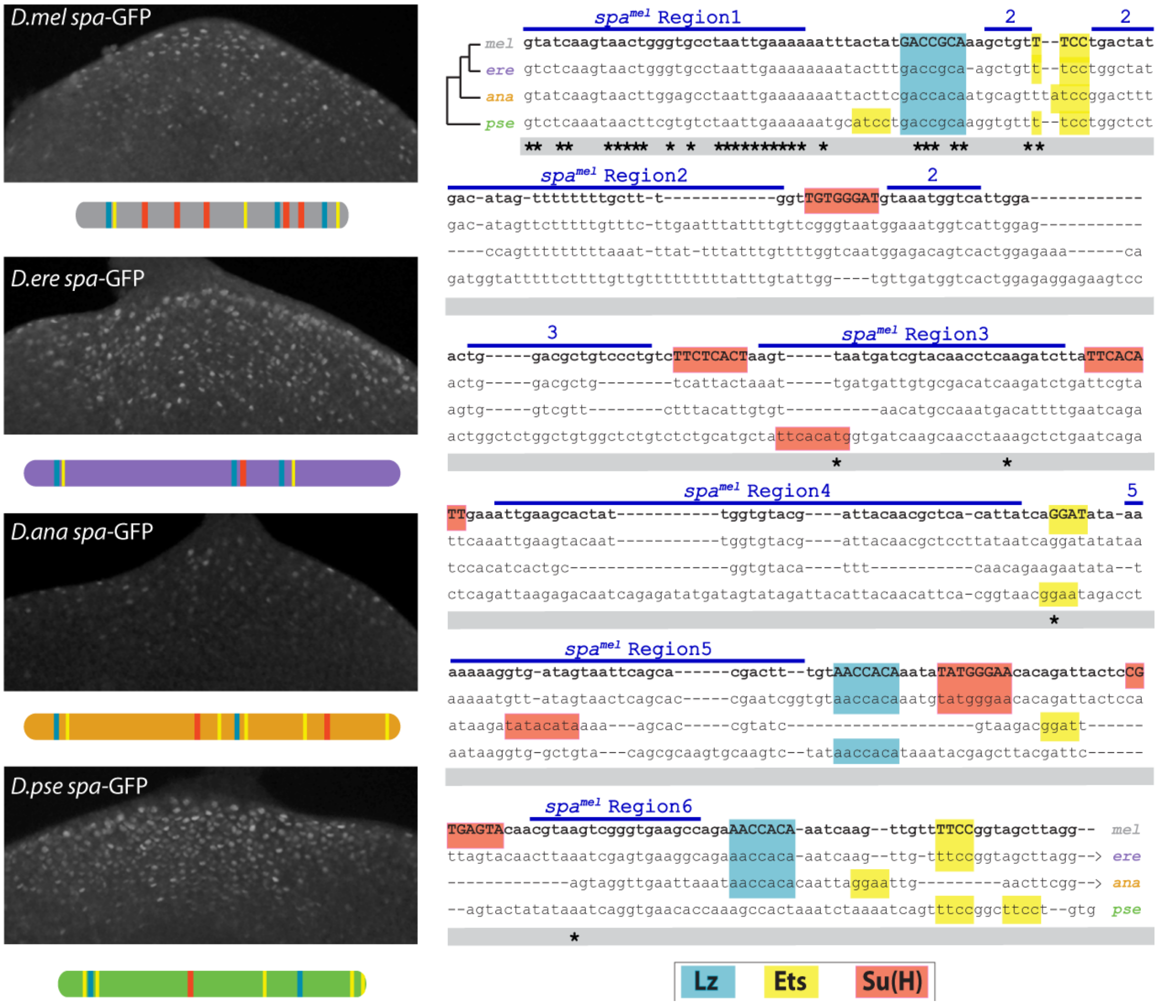


Figure 1.3 The *sparkling* (*spa*) enhancer of the *Drosophila dPax2* gene is sufficient to drive cone-cell-specific gene expression *in vivo*. Top left: diagram of the minimal 362-bp *spa* element, with Su(H), Ets, and Lz sites, as determined by Flores et al.(2000), indicated with colored stripes. Top right: fluorescence micrograph showing GFP expression, driven by the wildtype *spa* enhancer, in a transgenic third-instar eye imaginal disc. (mf, morphogenic furrow) Smaller images show coexpression of GFP with the cone cell marker Cut in pupae. Bottom: diagram of the structure of the *dPax2* gene, showing the location of *spa* in the fourth intron.

signaling, Yan binds to and represses target genes (Brunner et al., 1994; O'Neill et al., 1994; Rebay and Rubin, 1995). In addition, *spa* contains three binding sites for Lz, which is expressed in all undifferentiated progenitor cells in the *Drosophila* eye (Flores et al., 1998). All three of these regulatory inputs were confirmed to be necessary and direct. Genetic ablation of Notch signaling, EGFR signaling, or Lz abolished *dPax2* activation in cone cells, and targeted mutation of the Su(H), Ets, or Lz sites in *spa* abolished its activity in transgenic reporter and rescue assays (Flores et al., 2000).

When the DNA sequence of *spa* is compared across the genomes of other sequenced *Drosophila* species, blocks of conservation are unexpectedly few and short, compared to other enhancers (Figure 1.4) (Swanson et al., 2010; Swanson et al., 2011). However, despite very poor sequence conservation, 409bp of orthologous sequence from *D. pseudoobscura* (*D. pse spa*) was capable of driving cone-cell specific reporter gene expression in transgenic *D. melanogaster* (*D. mel*) that was indistinguishable from that of *D. mel spa* (Figure 1.4)(Swanson et al., 2010; Swanson et al., 2011). Similarly, 500bp fragments of *spa*-orthologous sequence from *D. erecta* and D. *ananase* also drove gene expression in *D. mel* cone cells, although with varying levels of GFP expression (Figure 1.4). All of these *spa* orthologs contain at least one Lz, one Su(H), and one Ets site, though most individual binding sites are very poorly conserved (Swanson et al., 2011). Therefore, while *spa* sequence is rapidly evolving, its function is conserved. The interesting expression pattern, well-defined regulatory inputs, and evolutionary properties of *spa* inspired the Barolo laboratory to further

Figure 1.4



```
mel gtatcaagtaactgggtgcctaattgaaaaaatttactatGACCGCAaagctgtT--TCCtgactat
ere gtctcaagtaactgggtgcctaattgaaaaaaaatactttgaccgca-agctgtt--tcctggctat
ana gtatcaagtaacttggagcctaattgaaaaaaattacttcgaccacaatgcagtttatccggacttt
pse gtctcaaataacttcgtgtctaattgaaaaaatgcatcctgaccgcaaggtgttt--tcctggctct
    ** **   ***** * * ********** *         *** **    **

gac-atag-tttttttgctt-t------------ggtTGTGGGATgtaaatggtcattgga-----------
gac-atagttcttttgtttc-ttgaatttattttgttcgggtaatggaaatggtcattggag----------
----ccagttttttttaaat-ttat-tttatttgttttggtcaatggagacagtcactggagaaa------ca
gatggtattttcttttgttgtttttttttatttgtattgg----tgttgatggtcactggagaggagaagtcc

actg-----gacgctgtccctgtcTTCTCACTaagt-----taatgatcgtacaacctcaagatcttaTTCACA
actg-----gacgctg--------tcattactaaat-----tgatgattgtgcgacatcaagatctgattcgta
agtg-----gtcgtt--------ctttacattgtgt----------aacatgccaaatgacattttgaatcaga
actggctctggctgtggctctgtctctgcatgctattcacatggtgatcaagcaacctaaagctctgaatcaga
                                        *                *

TTgaaattgaagcactat-----------tggtgtacg----attacaacgctca-cattatcaGGATata-aa
ttcaaattgaagtacaat-----------tggtgtacg----attacaacgctccttataatcaggatatataa
tccacatcactgc-----------------ggtgtaca----ttt-----------caacagaagaatata--t
ctcagattaagagacaatcagagatatgatagtatagattacattacaacattca-cggtaacggaatagacct
                                                              *

aaaaaggtg-atagtaattcagca------cgactt--tgtAACCACAaataTATGGGAAcacagattactcCG
aaaaatgtt-atagtaactcagcac-----cgaatcggtgtaaccacaaatgtatgggaacacagattactcca
ataagatatacataaaa---agcac-----cgtatc-------------------gtaagacggatt------
aataaggtg-gctgta---cagcgcaagtgcaagtc--tataaccacataaatacgagcttacgattc------

TGAGTAcaacgtaagtcgggtgaagccagaAACCACA-aatcaag--ttgttTTCCggtagcttagg--  mel
ttagtacaacttaaatcgagtgaaggcagaaaccaca-aatcaag--ttg-tttccggtagcttagg--> ere
------------agtaggttgaattaaataaccacacaattaggaattg---------aacttcgg--> ana
--agtactatataaatcaggtgaacaccaaagccactaaatctaaaatcagtttccggcttcct--gtg  pse
            *
```

Figure 1.4 The DNA sequence of *spa* is rapidly evolving. Left: fluorescence images of eye discs from transgenic *D. melanogaster* larvae bearing GFP reporter transgenes driven by *D. melanogaster* (*D. mel spa*)or *spa*-orthologous sequences from *D. erecta* (*D. ere spa*), *D. anannasae* (*D. ana spa*), or *D. pseudoobscura* (*D. pse spa*). Right: sequence alignment of spa orthologs from four Drosophila species. Predicted and confirmed binding sites for Lz, PntP2/ Yan, and Su(H) are indicated with colored boxes.

characterize the *cis*-regulatory logic, structural constraints, and evolutionary history of this enhancer (Johnson et al., 2008; Swanson et al., 2010; Swanson et al., 2011).

1.2a sparkling: enhanceosome or billboard?

Once Lz, EGFR/Ets, and Notch/Su(H) were identified by Flores et al. (2000) as direct regulatory inputs of the *spa* enhancer, it was plausible to imagine that these three regulatory inputs might together be sufficient to generate cone cell-specific *dPax2* expression. However, this is not the case. Wildtype *spa*-GFP reporter transgenes were active in cone cells of larval eye imaginal discs, but synthetic constructs containing only the Lz, Ets, and Su(H) binding sites from *spa* were incapable of driving gene expression, regardless of the spacing between TFBS (Figure 1.5 A) (Johnson et al., 2008; Swanson et al., 2010). Therefore, additional regulatory sequences within *spa*, besides the defined TFBSs, must also be necessary for proper *spa* activity *in vivo*. In other words, Lz+Ets+Su(H) is not the complete combinatorial code for *spa*.

To better define the full complement of regulatory sites necessary and, together, sufficient for *spa's* activity in cone cells, the DNA sequences between the Lz, Ets, and Su(H) sites were systematically mutated. The non-Lz/Ets/Su(H) sequences of *spa* were initially divided in to six regions, each of which was individually deleted. Only one of these, region 3, was found to have no significant effect on *spa* activity *in vivo* (Figure 1.5 A) (Swanson et al., 2010). Deletion of region 5 increased GFP expression in cone cells, while the individual deletion of region 1, 2, 4, or 6 resulted in severe loss of gene expression. It is important to note that these deletions resulted in loss of potential novel regulatory site, but they also altered the spacing between sites on either side of the deletion. For this reason, regions 2, 4, 5, and 6 were also subjected to sequence alterations (specifically, every second position was changed to its noncomplementary transversion; A to C, C to A, G to T, T to G) that did not affect the overall spacing or structure of the enhancer. Under these native-spacing conditions, mutating region 2 no longer strongly affected *spa* activity (Figure 1.5 A), suggesting that the length of this region is more functionally significant than its sequence. This is consistent with a structural role for region 2, rather than a role in direct TF recruitment. In contrast to the deletion of region 5, native-spacing mutation of the same region resulted in a severe loss of gene expression. This suggests that the augmented expression associated with the deletion of region 5 might be due to compressed spacing of Ets and Lz sites flanking this region: Ets and Runx factors, including PntP2 and Lz, are well known to interact with one another and to synergistically activate transcription (Akbari et al., 2007; Dittmer, 2003; Goetz et al., 2000; Jackson Behan et al., 2005; Kim et al., 1999; Liu et al., 2004). The sequence within region 5 must be essential for cone-cell-specific gene expression, which is also the case for regions 4 and 6 as the loss of these regions, either by deletion or native-spacing mutation, abolished *spa* activity (Figure 1.5 A). Similarly, smaller native-spacing mutations within regions 1, 4, 5, and 6 demonstrated that most of the sequence in these regions is required for

Figure 1.5

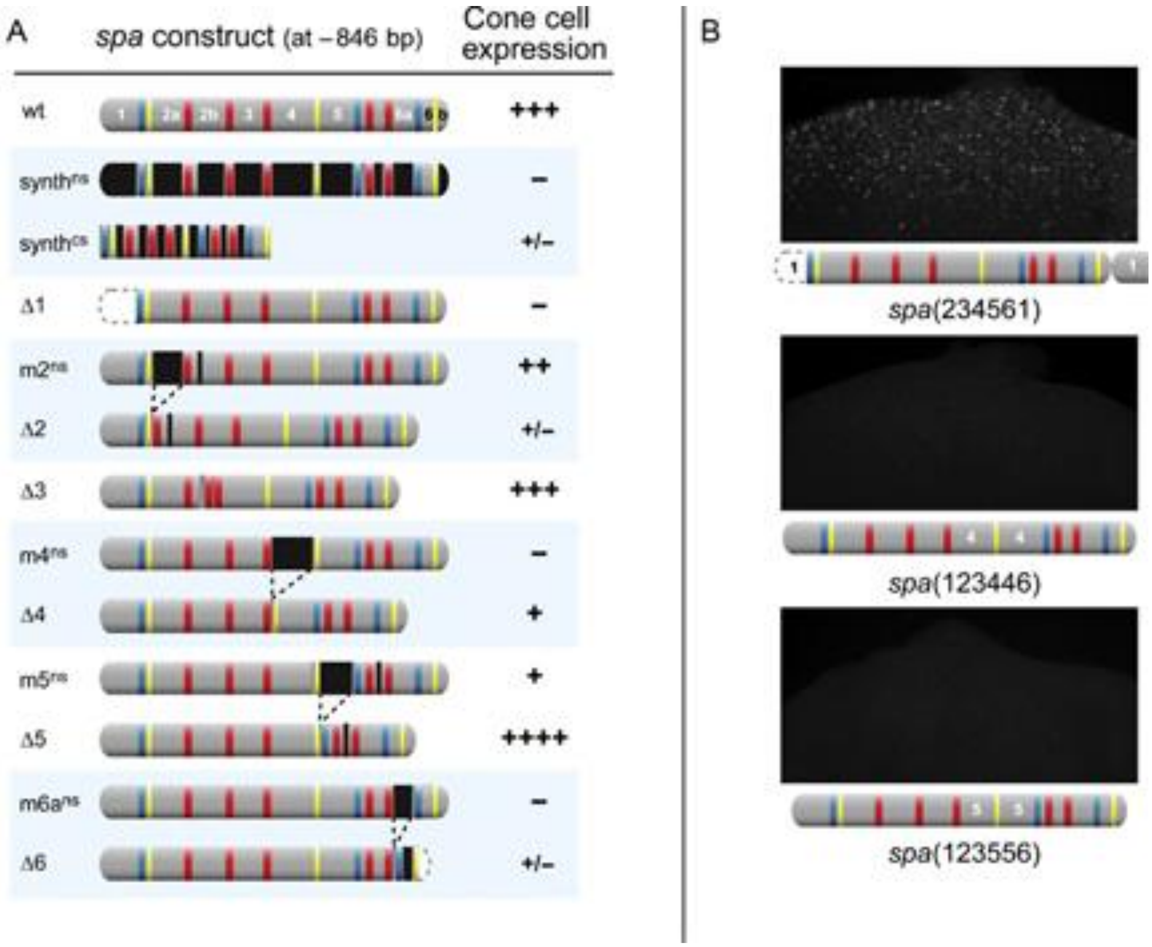


Figure 1.5 *In vivo* functional dissection of *spa* reveals high complexity, structural constraints, and multiple classes of activation activities (A) Mutational analysis of *D. mel spa*. “ns” indicates mutations that preserve native spacing within the enhancer; “Δ” indicates sequence deletions. Levels of reporter gene expression in cone cells are indicated a` la Swanson et al. (2010). (B) Relocation and substitution of various regulatory regions within *spa.*

enhancer function (Swanson et al., 2010). Thus, nearly all of the sequence of *spa* is regulatory, nonredundant, and essential for function *in vivo*.

Perhaps the strongest evidence for structural constraints on the organization of *spa* was seen when the Lz, Ets, and Su(H) sites in *spa* were ablated, and these same sites were restored, but at the 3’ end of the element. No binding sites were ultimately gained or lost, only rearranged relative to other essential sequences. The resultant enhancer construct was active in the developing eye, but in the wrong cell type: expression was lost in cone cells, while ectopic expression was observed in R1 and R6 photoreceptors, which normally do not express *dPax2* (Swanson et al., 2010). Therefore, at least for *spa*, the combinatorial code of TFs binding to the enhancer is not sufficient to specify the proper pattern of gene expression: additional essential patterning information is supplied by the arrangement of regulatory sites. In other words, a given set of inputs (binding sites) can generate multiple possible outputs (expression patterns), and enhancer structure can play an important role in determining the final pattern.

Reasoning from these and other experiments, *spa* seems to behave consistently with an enhanceosome model of action, wherein strict rules of organization and spacing of regulatory sites are enforced, at least in some regions of the enhancer. However, other evidence appears to support an information display model of *spa* function. For example, one essential region of the enhancer (region 1) can be relocated to the opposite end of the enhancer with no apparent effect on enhancer function (Figure 1.5 B) (Swanson et al., 2010). This suggests that the regulatory information within *spa* is flexible, at least with respect to the essential contribution of region 1 (more on this region later). The poor sequence conservation of *spa* could also be considered to be consistent with loose organization and therefore with the information display model.

However, evidence strongly suggests that the transcriptional activity of *spa* cannot simply depend on the presence of a sufficient number of activator binding sites, irrespective of their identity. For example, a construct that doubles the number of Lz, Ets, and Su(H) binding sites, but includes only those sites, was unable to drive cone-cell-specific gene expression, regardless of spacing between those sites (Swanson et al., 2010) (Figure 1.7). In other words, the addition of extra Lz/Ets/Su(H) inputs could not functionally compensate for the loss of the novel (i.e., non-Lz/Ets/Su(H)) regulatory inputs, which may mean that these two groups of *spa* regulatory sites provide different transcriptional activation functions. (More evidence supporting this idea will be discussed in the next section.) Similarly, the regulatory contribution of region 4 cannot be substituted with a second copy of region 5, nor can region 5 be functionally replaced by a second copy of region 4 (Figure 1.5 B). It seems that neither a straightforward information display model nor an enhanceosome model accurately represents the structure and function of *spa* (Datta and Small, 2011).

1.2b Combinatorial control of sparkling's activity

We have already seen that “Lz+Ets+Su(H)” is insufficient to describe the combinatorial code of *spa* activity; additional inputs from regions 1, 4, 5, and 6a are also necessary. While the proteins that bind to each of the essential regions of *spa* have yet to be identified, comparative analysis of *spa* sequences from multiple *Drosophila* species revealed motifs that are required for function in at least two *spa* orthologs (Swanson et al., 2010). Following the example of a highly influential evolutionary analysis of *eve* S2E (Ludwig et al., 2000), chimeras between the 5' half of *D. mel spa* and the 3' half of *D. pse spa*, or conversely, the 5' half of *D. pse spa* and the 3' half of *D. mel spa*, were assessed in transgenic *D. mel*. The *spa* chimeras provided very different results from those involving *eve* S2E: in the case of *spa* 5' *D*. mel + 3' *D. pse* was inactive, while 5' *D. pse* + 3' *D. mel* drove robust reporter gene expression (Figure 1.6 A) (Swanson et al., 2011) supporting the idea that there are differences in TFBS composition and arrangement between the two *spa* orthologs. 3' *D. pse spa* may lack regulatory inputs present in the 3'half of *D. mel spa*; inputs from these regions may lie (at least partially) in 5' *D. pse spa*. Detailed mutational analyses of these chimeras were consistent with the hypothesis that regulatory inputs into the 3' half of *D. mel spa* could be substituted with inputs into the 5' half of *D. pse spa* (Figure 1.6 A), suggesting that *spa* has been significantly reorganized over a relatively short evolutionary period. This suggests that there are differences in TFBS composition and arrangement between the two enhancers. 3' *D.pse spa* must lack regulatory inputs present in the 3' half of *D. mel spa,* however the sequence of *D. mel* regions 4, 5, and 6a, which are all present in 3' *D. mel spa* and are essential for *D. mel spa* activity, do not appear to be present in 3' *D.pse spa* based on sequence conservation. Therefore the inputs from these regions may now lie at least partially in 5' *D.pse spa*. In fact, mutations in the 5' *D.pse spa +3' D. mel spa* chimera to the *D. mel spa* regions 5 and 6a, but not region 4, do not significantly affect chimeric enhancer function, providing further support for this hypothesis.

Now armed with the knowledge that regulatory inputs from *D. mel spa* regions 5 and 6a are compensated for by regulatory inputs in 5' *D.pse spa*, motif analysis was performed using to compare these regions. MEME motif analysis identified five potential novel motifs present in *D. mel spa* region 4, 5, and 6a, and also within *D. pse spa*, but at noncorresponding positions (Figure 1.6 B). These motifs, named α through ε, are also present in *spa* orthologs from all or most of the 12-sequenced *Drosophila* species. For example what is referred to as the γ motif is found in *D. mel spa* region 4 and *D.pse spa* region E, while β is located in *D. mel spa* region 4 and *D.pse spa* regions B and E. Mutations of β and γ together in *D. mel spa* and individually in *D.pse spa* severely diminish *spa* activity. Likewise the motif α is located in *D. mel spa* region 4 and *D.pse spa* region E and mutation of these sites also decreases gene expression. *D. mel spa* region 5 contains a δ motif which can also be identified in *D.pse spa* region B and mutations in *mel* and *pse* to this motif nearly abolishes enhancer activity. (Swanson et al., 2011). Interestingly, the effects of loss of the ε motif in 3'*D. mel spa* could be rescued by transplanting the 5' *D. pse* ε input to region 2, providing

Figure 1.6

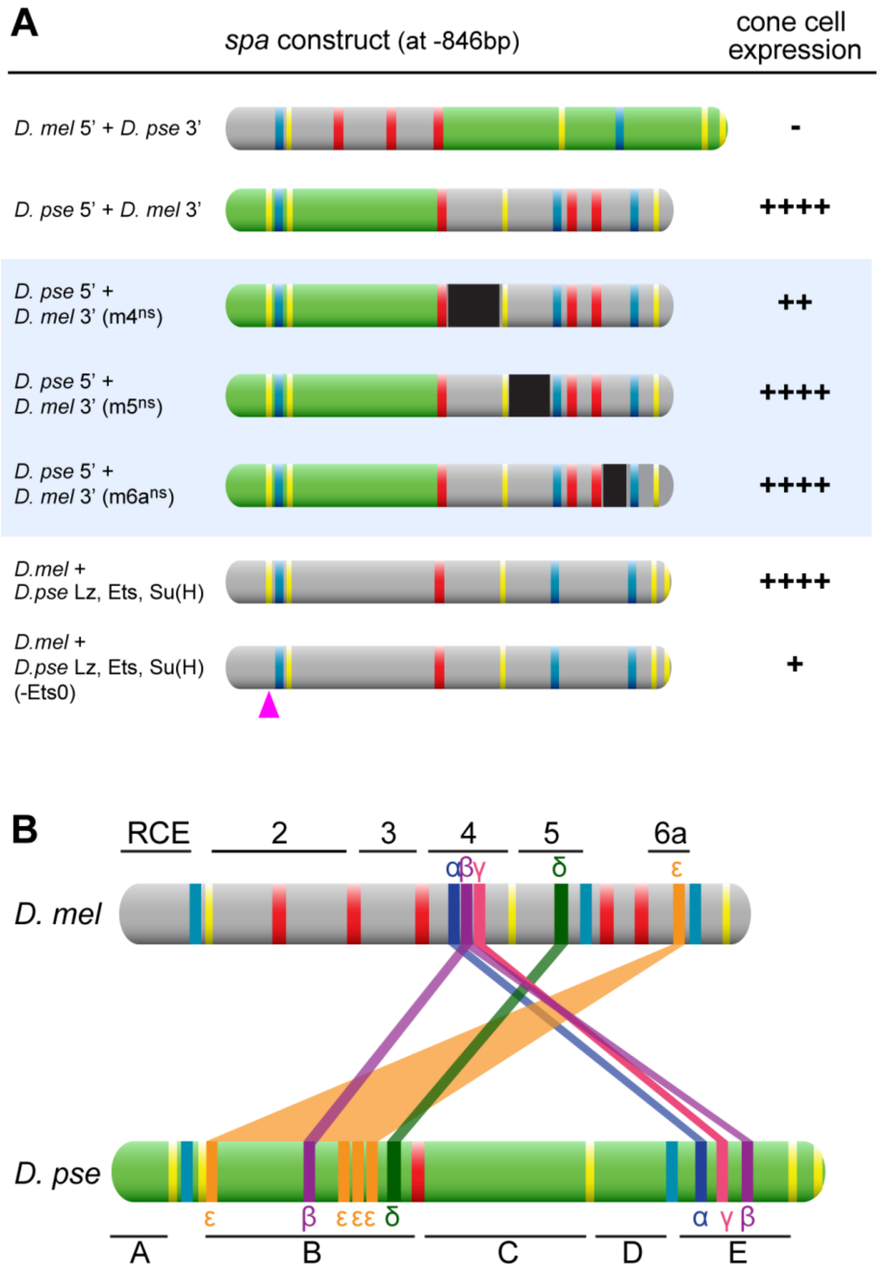

A
spa construct (at -846bp)
cone cell expression
D. mel 5' + *D. pse* 3'
-
D. pse 5' + *D. mel* 3'
++++
D. pse 5' + *D. mel* 3' (m4ns)
++
D. pse 5' + *D. mel* 3' (m5ns)
++++
D. pse 5' + *D. mel* 3' (m6ans)
++++
D.mel + *D.pse* Lz, Ets, Su(H)
++++
D.mel + *D.pse* Lz, Ets, Su(H) (-Ets0)
+
B
RCE
2
3
4
5
6a
αβγ
δ
ε
D. mel
D. pse
ε
β
εεε
δ
α
γ
β
A
B
C
D
E

Figure 1.6 Evolutionary dynamics of the cis-regulatory structure of *spa*. (A) Summary of the results of various chimeric enhancer experiments, demonstrating that regulatory information within *spa* has been reorganized since the divergence of *D. mel* and *D. pse*. (B) Diagram summarizing the proposed reorganization of *spa*, with proposed novel regulatory motifs α through ε indicated.

further evidence for the functional significance of this motif. Note that most of the motifs identified in *D. mel spa* region 4, which is not compensated for by 5' *D.pse spa* , are also present in 3' *D.pse spa*, while the motifs in *D. mel spa* regions 5 and 6 are located in 5' *D.pse spa*. The ε motif is present in increased copy number in *D. pse spa* relative to *D. mel*, which may account for the ability of *D. pse spa* to function despite its (relatively) fewer Lz and Su(H) sites. A recently derived Ets motif in 5'*D. pse spa* also seems to make an important compensatory contribution to enhancer function (Swanson et al., 2011). While the identities of the factors that bind these novel motifs and the biochemical activities they recruit are not known, the combinatorial code of *spa* can now be expanded to Lz+EGFR/Ets+Notch/Su(H)+α+β+γ+δ+ε. However, the sufficiency of these inputs *in vivo* to generate a *spa* expression pattern has not been strictly experimentally tested.

1.2c Structural and sequence constraints channel sparkling output

Evolutionary analyses of *sparkling* have also led to insights as to how the enhancer achieves cone-cell-specific gene expression. Recall that all of the identified TF inputs into *spa* activity, EGFR/Ets, Notch/Su(H), and Lz, are active in multiple cell types in the developing *Drosophila* retina, yet *spa* drives *dPax2* expression only in cone cells. How does *spa* induce gene expression in this cell type alone? The simplest explanation is that a requirement for additional DNA-binding regulatory factors (such as factors a through ε, perhaps) restricts *spa* activity to cone cells, as discussed above. However, this explanation is not sufficient, as multiple structural rearrangements of *spa* can alter the cell-type specificity of its action. For example, when the Lz, Ets, and Su(H) binding sites were mutated within *spa* and subsequently placed at the 3’ end of *spa* in compressed conformation, gene expression switched from cone cells to R1/R6 photoreceptors (Figure 1.7); Swanson et al., 2010 (Swanson et al., 2010). Yet this ectopic gene expression must be due to the new arrangement of TFBS, as restoration of the spacing among the Lz, Ets, and Su(H) sites in this construct resulted in complete loss of activity (Figure 1.5 A). Note that the only difference between these two constructs is the spacing among Lz/Ets/Su(H) sites.

The ectopic R1/R6 activity resulting from this rearrangement requires *spa* regions 1, 4, and 6 as well as Lz and Ets binding sites, but not Su(H) sites, which correlates with genetic data: the R1 and R6 cells respond to EGFR and Lz, but not to Notch signaling (Swanson et al., 2010). Ectopic expression in R1/R6 was also achieved in a different way, by mutating regions 2, 3, 5, and 6b, with all other *spa* sequences and TFBSs in their native configuration (Figure1.7A). Restoration of region 5 to this construct abolished R1/R6 expression with no gain in cone cell expression, suggesting that a regulatory site(s) within region 5 represses spa activity in photoreceptors. Genetic evidence indicates that ectopic *dPax2* expression, such as that driven by these altered enhancers, would negatively affect the fitness of the fly by disrupting cell-fate specification and differentiation in the eye (Shi and Noll, 2009).

From studies of *spa’s* cell-type specificity, we can also learn more about the structural rules governing this enhancer. For example, two copies of all of *spa's* Lz/Ets/Su(H) sites in a compressed configuration were shown to drive ectopic expression in photoreceptors, while the same sites in native spacing were incapable of driving gene expression in any cell type (Figure 1.7A). This is consistent with the idea that unrestrained Lz-Ets synergy, which is likely to be promoted by compressed TFBS spacing, results in the ectopic *spa* activity. In a similar vein, the photoreceptor-specific repressor function of region 5 must be only able to work over a short range, as all *spa* sequences (including region 5) are present in the rearranged enhancer construct *spa*(KO+synth^{cs}), which was capable of driving robust ectopic R1/R6 gene expression (Figure 1.7A). Note that the *spa*(KO+synth^{cs})and *spa*(m236b^{ns}) constructs both contain region 5, which contains a putative R1/R6 repressor binding site. However, only one confirmation, the wild-type spacing, allows for repression to occur indicating that the presence of a short-range repressor activity. When native-binding site spacing is restored to *spa*(KO+synth^{cs}), all reporter expression was lost, suggesting that the activator functions of this enhancer are also short range as both R1/R6 and cone cell expression are absent in this spread-out conformation. Together these data reveal that the activator and repressor activities of *spa* are short range in nature, allowing for precise cell-type-specific output.

1.2d Low-affinity binding sites and the Notch response

An entirely different evolutionary strategy appears to make an equally important contribution to the specificity of *spa* activity. In *spa* orthologs in all sequenced *Drosophila* species, the predicted Su(H)-binding sites are nonconsensus and generally of low-predicted affinity. The five confirmed Su(H) sites in *D. mel spa*, and the one predicted site in *D. pse spa*, all deviate by 1–4bp from the high-affinity consensus YGTGRGAAM (Crocker et al., 2010; Flores et al., 2000). Three of the *D. mel* sites and the single *D. pse* site also deviate from the looser, lower-affinity consensus RTGRGAR (Bailey and Posakony, 1995; Nellesen et al., 1999). The conserved property of low affinity for Su(H), even though the individual sites themselves are not conserved, suggested a possible regulatory mechanism. When the five low-affinity Su(H)-binding sites in *D. mel spa* were converted to high-affinity sites, representing a total change of 10bp out of 362bp, levels of cone-cell activation were increased. Possibly more significantly, ectopic gene expression was observed in photoreceptors in larval eye discs, as well as in PPCs in pupal eyes (Figure1.7B) (Swanson et al., 2011). These data again suggest that strict quantitative control of *spa's* regulatory inputs is necessary for correct *spa* activity and patterning.

1.3 *sparkling* in relation to the study of gene regulation

One lesson to be learned from the extensive characterization of the *dPax2 sparkling* enhancer by multiple laboratories is the extraordinary amount of information concerning enhancer structure, function, and evolution we can obtain by delving deeply into a single short DNA element. However, the new questions raised by these studies show how much remains to be understood about

Figure 1.7

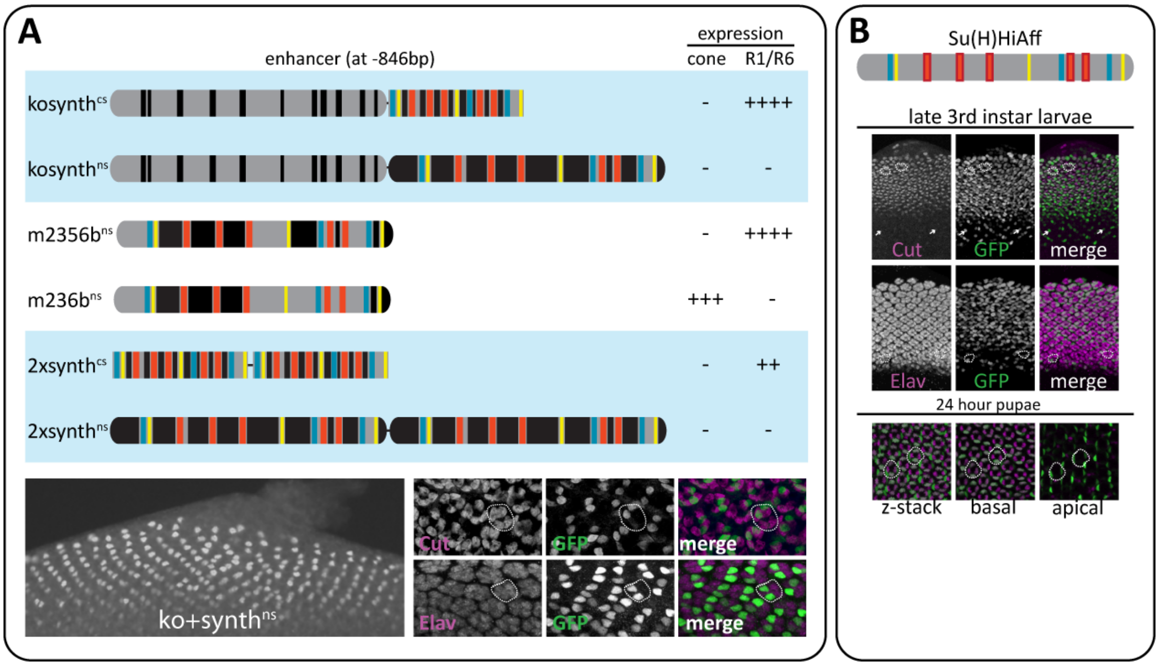


Figure 1.7 The linear arrangement, spacing, and affinity of binding sites profoundly affects the patterning of gene expression driven by *spa*. (A) Summary of results of experiments testing the effects of binding site organization on cell-type specificity of gene expression. *spa*(ko+synth^{ns}) is expressed ectopically photoreceptors, as shown by coexpression with the neuronal marker Elav in pupal eye discs (B) Increasing the affinity of Su(H) binding sites within *spa* creates an enhancer that is oversensitive to Notch signaling, resulting in ectopic gene expression in multiple Notch-responsive cell types. Top: *spa*[Su(H)HiAff]-GFP is expressed ectopically in Cut-negative photoreceptors (arrows) in the larval eye imaginal disc, as shown by coexpression with the neuronal marker Elav. Bottom *spa*[Su(H)HiAff]-GFP is also expressed ectopically in primary pigment cells in the 24-h pupal retina, identifiable by lack of Cut staining, apical position within the retina, and characteristic elongated cell shape.

transcriptional control *in vivo*. *sparkling* cannot be accurately described as either an enhanceosome or a billboard, as it contains both strict structural requirements and flexible binding site arrangements. Its sequence is quite evolutionarily labile, yet stable patterns can be detected among the rapidly shifting regulatory motifs. Mutational analysis, coupled with comparative sequence analysis, revealed that *spa* is crowded with regulatory information, suggesting that the typical enhancer may be more complex than originally anticipated. We have also seen that changes in the structure of the enhancer can result in ectopic gene expression, revealing that combinatorial control alone can be insufficient to determine the cell-type specificity of enhancer function. Furthermore, functional evidence suggests that different sites within an enhancer may mediate distinct, nonsubstitutable regulatory functions, all of which may be required for transcriptional activation *in vivo*. The most important lesson of these studies, we propose, is that the "combinatorial code" view of cis-regulatory logic, though accurate as far as it goes, does not adequately address the complexity of the enhancer. While an incredibly extensive characterization of a single transcriptional enhancer, the study of *sparkling* summarized here does not deeply examine the mechanisms by which this enhancer must to promote gene transcription. Furthermore, as *sparkling* lies in the nearly 7kb from the *dPax2* promoter it must somehow overcome this distance in order to regulate tissue and temporal specific gene expression. As with most enhancer studies, our knowledge of *spa* thus far has not addressed how the enhancer performs this critical function.

1.4 Acknowledgements

Portions of the text and figures were previously published (Evans et al., 2012). The new research findings presented here were supported by NIH grant GM076509 and ARRA supplement GM07650903S1 to S.B. Experiments were performed as described by Swanson et al. (2010, 2011). We thank the members of the Barolo lab for helpful comments on the manuscript.

1.5 References

Akbari, O.S., Schiller, B.J., Goetz, S.E., Ho, M.C., Bae, E., and Drewell, R.A. (2007). The abdominal-B promoter tethering element mediates promoter-enhancer specificity at the Drosophila bithorax complex. Fly (Austin) *1*, 337-339.

Andrioli, L.P., Vasisht, V., Theodosopoulou, E., Oberstein, A., and Small, S. (2002). Anterior repression of a Drosophila stripe enhancer requires three position-specific mechanisms. Development *129*, 4931-4940.

Arnosti, D.N., Barolo, S., Levine, M., and Small, S. (1996). The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. Development *122*, 205-214.

Arnosti, D.N., and Kulkarni, M.M. (2005). Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? J Cell Biochem *94*, 890-898.

Bailey, A.M., and Posakony, J.W. (1995). Suppressor of Hairless directly activates transcription of *Enhancer of split* Complex genes in response to Notch receptor activity. Genes Dev *9*, 2609-2622.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299-308.

Barolo, S., and Posakony, J.W. (2002). Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. Genes Dev *16*, 1167-1181.

Barolo, S., Walker, R.G., Polyanovsky, A.D., Freschi, G., Keil, T., and Posakony, J.W. (2000). A Notch-independent activity of *Suppressor of Hairless* is required for normal mechanoreceptor physiology. Cell *103*, 957-969.

Blackwood, E.M., and Kadonaga, J.T. (1998). Going the distance: a current view of enhancer action. Science *281*, 60-63.

Bray, S.J. (2006). Notch signalling: a simple pathway becomes complex. Nat Rev Mol Cell Biol *7*, 678-689.

Brunner, D., Ducker, K., Oellers, N., Hafen, E., Scholz, H., and Klambt, C. (1994). The ETS domain protein pointed-P2 is a target of MAP kinase in the sevenless signal transduction pathway. Nature *370*, 386-389.

Bulger, M., and Groudine, M. (2010). Enhancers: the abundance and function of regulatory sequences beyond promoters. Developmental Biology *339*, 250-257.

Crocker, J., and Erives, A. (2008). A closer look at the *eve* stripe 2 enhancers of *Drosophila* and *Themira*. PLoS Genet *4*, e1000276.

Crocker, J., Potter, N., and Erives, A. (2010). Dynamic evolution of precise regulatory encodings creates the clustered site signature of enhancers. Nat Commun *1*, 99.

Datta, R.R., and Small, S. (2011). Gene regulation: piecing together the puzzle of enhancer evolution. Curr Biol *21*, R542-543.

Dittmer, J. (2003). The biology of the *Ets1* proto-oncogene. Mol Cancer *2*, 29.

Drissen, R., Palstra, R.J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S., and de Laat, W. (2004). The active spatial organization of the beta-globin locus requires the transcription factor EKLF. Genes Dev *18*, 2485-2490.

Evans, N.C., Swanson, C.I., and Barolo, S. (2012). Sparkling insights into enhancer structure, function, and evolution. Curr Top Dev Biol *98*, 97-120.

Flores, G.V., Daga, A., Kalhor, H.R., and Banerjee, U. (1998). Lozenge is expressed in pluripotent precursor cells and patterns multiple cell types in the Drosophila eye through the control of cell-specific transcription factors. Development *125*, 3681-3687.

Flores, G.V., Duan, H., Yan, H., Nagaraj, R., Fu, W., Zou, Y., Noll, M., and Banerjee, U. (2000). Combinatorial signaling in the specification of unique cell fates. Cell *103*, 75-85.

Frankel, N., Davis, G.K., Vargas, D., Wang, S., Payre, F., and Stern, D.L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature *466*, 490-493.

Fu, W., Duan, H., Frei, E., and Noll, M. (1998). *shaven* and *sparkling* are mutations in separate enhancers of the *Drosophila Pax2* homolog. Development *125*, 2943-2950.

Fu, W., and Noll, M. (1997). The *Pax2* homolog *sparkling* is required for development of cone and pigment cells in the *Drosophila* eye. Genes Dev *11*, 2066-2078.

Giese, K., Kingsley, C., Kirshner, J.R., and Grosschedl, R. (1995). Assembly and function of a *TCRα* enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. Genes and Development *9*, 995-1008.

Goetz, T.L., Gu, T.L., Speck, N.A., and Graves, B.J. (2000). Auto-inhibition of Ets-1 is counteracted by DNA binding cooperativity with core-binding factor alpha2. Molecular and Cellular Biology *20*, 81-90.

Goto, T., Macdonald, P., and Maniatis, T. (1989). Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. Cell *57*, 413-422.

Harding, K., Hoey, T., Warrior, R., and Levine, M. (1989). Autoregulatory and gap gene response elements of the even-skipped promoter of Drosophila. Embo J *8*, 1205-1212.

Hare, E.E., Peterson, B.K., and Eisen, M.B. (2008a). A careful look at binding site reorganization in the *even-skipped* enhancers of *Drosophila* and sepsids. PLoS Genet *4*, e1000268.

Hare, E.E., Peterson, B.K., Iyer, V.N., Meier, R., and Eisen, M.B. (2008b). Sepsid *even-skipped* enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. PLoS Genet *4*, e1000106.

Hayashi, T., Xu, C., and Carthew, R.W. (2008). Cell-type-specific transcription of prospero is controlled by combinatorial signaling in the Drosophila eye. Development *135*, 2787-2796.

Jackson Behan, K., Fair, J., Singh, S., Bogwitz, M., Perry, T., Grubor, V., Cunningham, F., Nichols, C.D., Cheung, T.L., Batterham, P.*, et al.* (2005). Alternative splicing removes an Ets interaction domain from Lozenge during *Drosophila* eye development. Dev Genes Evol *215*, 423-435.

Johnson, L.A., Zhao, Y., Golden, K., and Barolo, S. (2008). Reverse-engineering a transcriptional enhancer: a case study in Drosophila. Tissue Eng Part A *14*, 1549-1559.

Johnson, S.A., Harmon, K.J., Smiley, S.G., Still, F.M., and Kavaler, J. (2011). Discrete regulatory regions control early and late expression of D-Pax2 during external sensory organ development. Dev Dyn *240*, 1769-1778.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Khoury, G., and Gruss, P. (1983). Enhancer elements. Cell *33*, 313-314.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S.*, et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182-187.

Kim, W.Y., Sieweke, M., Ogawa, E., Wee, H.J., Englmeier, U., Graf, T., and Ito, Y. (1999). Mutual activation of Ets-1 and AML1 DNA binding by direct interaction of their autoinhibitory domains. Embo J *18*, 1609-1620.

Klingler, M., Soong, J., Butler, B., and Gergen, J.P. (1996). Disperse versus compact elements for the regulation of runt stripes in Drosophila. Dev Biol *177*, 73-84.

Kulkarni, M.M., and Arnosti, D.N. (2005). *cis*-regulatory logic of short-range transcriptional repression in *Drosophila melanogaster*. Molecular and Cellular Biology *25*, 3411-3420.

Kumar, J.P. (2001). Signalling pathways in Drosophila and vertebrate retinal development. Nat Rev Genet *2*, 846-857.

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. Current Biology *20*, R754-763.

Liu, H., Holm, M., Xie, X.Q., Wolf-Watz, M., and Grundstrom, T. (2004). AML1/Runx1 recruits calcineurin to regulate granulocyte macrophage colony-stimulating factor by Ets1 activation. Journal of Biological Chemistry *279*, 29398-29408.
Ludwig, M.Z., Bergman, C., Patel, N.H., and Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. Nature *403*, 564-567.
Ludwig, M.Z., Patel, N.H., and Kreitman, M. (1998). Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change. Development *125*, 949-958.
Malik, S., and Roeder, R.G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. Trends Biochem Sci *30*, 256-263.
Morel, V., and Schweisguth, F. (2000). Repression by suppressor of hairless and activation by Notch are required to define a single row of single-minded expressing cells in the Drosophila embryo. Genes Dev *14*, 377-388.
Narlikar, G.J., Fan, H.Y., and Kingston, R.E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell *108*, 475-487.
Nellesen, D.T., Lai, E.C., and Posakony, J.W. (1999). Discrete enhancer elements mediate selective responsiveness of *Enhancer of split* Complex genes to common transcriptional activators. Dev Biol *213*, 33-53.
O'Neill, E.M., Rebay, I., Tjian, R., and Rubin, G.M. (1994). The activities of two Ets-related transcription factors required for Drosophila eye development are modulated by the Ras/MAPK pathway. Cell *78*, 137-147.
Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet *12*, 283-293.
Orom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q.*, et al.* (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell *143*, 46-58.
Orphanides, G., and Reinberg, D. (2002). A unified theory of gene expression. Cell *108*, 439-451.
Papatsenko, D., Goltsev, Y., and Levine, M. (2009). Organization of developmental enhancers in the Drosophila embryo. Nucleic Acids Res *37*, 5665-5677.
Papatsenko, D., and Levine, M. (2007). A rationale for the enhanceosome and other evolutionarily constrained enhancers. Current Biology *17*, R955-957.
Perry, M.W., Boettiger, A.N., Bothma, J.P., and Levine, M. (2010). Shadow enhancers foster robustness of Drosophila gastrulation. Curr Biol *20*, 1562-1567.
Priest, H.D., Filichkin, S.A., and Mockler, T.C. (2009). Cis-regulatory elements in plant cell signaling. Curr Opin Plant Biol *12*, 643-649.
Rastegar, S., Hess, I., Dickmeis, T., Nicod, J.C., Ertzer, R., Hadzhiev, Y., Thies, W.G., Scherer, G., and Strahle, U. (2008). The words of the regulatory code are arranged in a variable manner in highly conserved enhancers. Developmental Biology *318*, 366-377.
Rebay, I., and Rubin, G.M. (1995). Yan functions as a general inhibitor of differentiation and is negatively regulated by activation of the Ras1/MAPK pathway. Cell *81*, 857-866.

Sackerson, C., Fujioka, M., and Goto, T. (1999). The even-skipped locus is contained in a 16-kb chromatin domain. Dev Biol *211*, 39-52.

Schulz, R.A., Xie, X.L., and Miksch, J.L. (1990). cis-acting sequences required for the germ line expression of the Drosophila gonadal gene. Dev Biol *140*, 455-458.

Shi, Y., and Noll, M. (2009). Determination of cell fates in the R7 equivalence group of the *Drosophila* eye by the concerted regulation of D-Pax2 and TTK88. Developmental Biology *331*, 68-77.

Shilo, B.Z. (2003). Signaling by the Drosophila epidermal growth factor receptor pathway during development. Exp Cell Res *284*, 140-149.

Small, S., Blair, A., and Levine, M. (1992). Regulation of *even-skipped* stripe 2 in the *Drosophila* embryo. EMBO J *11*, 4047-4057.

Spilianakis, C.G., and Flavell, R.A. (2004). Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. Nat Immunol *5*, 1017-1027.

Swanson, C.I., Evans, N.C., and Barolo, S. (2010). Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. Dev Cell *18*, 359-370.

Swanson, C.I., Schwimmer, D.B., and Barolo, S. (2011). Rapid evolutionary rewiring of a structurally constrained eye enhancer. Curr Biol *21*, 1186-1196.

Szutorisz, H., Dillon, N., and Tora, L. (2005). The role of enhancers as centres for general transcription factor recruitment. Trends Biochem Sci *30*, 593-599.

Thanos, D., and Maniatis, T. (1995). Virus induction of human *IFN-β* gene expression requires the assembly of an enhanceosome. Cell *83*, 1091-1100.

Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science *329*, 689-693.

Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J., and Blobel, G.A. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. Mol Cell *17*, 453-462.

Voas, M.G., and Rebay, I. (2004). Signal integration during development: insights from the *Drosophila* eye. Dev Dyn *229*, 162-175.

Wang, G., Balamotis, M.A., Stevens, J.L., Yamaguchi, Y., Handa, H., and Berk, A.J. (2005). Mediator requirement for both recruitment and postrecruitment steps in transcription initiation. Mol Cell *17*, 683-694.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A.*, et al.* (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120-124.

Xu, C., Kauffmann, R.C., Zhang, J., Kladny, S., and Carthew, R.W. (2000). Overlapping activators and repressors delimit transcriptional response to receptor tyrosine kinase signals in the Drosophila eye. Cell *103*, 87-97.

Yao, L.C., Phin, S., Cho, J., Rushlow, C., Arora, K., and Warrior, R. (2008). Multiple modular promoter elements drive graded brinker expression in response to the Dpp morphogen gradient. Development *135*, 2183-2192.

Yuh, C.H., and Davidson, E.H. (1996). Modular cis-regulatory organization of Endo16, a gut-specific gene of the sea urchin embryo. Development *122*, 1069-1082.

CHAPTER 2

LONG RANGE GENE REGULATION AND THE IDENTIFICATION OF THE *SPARKLING* RCE

2.1 Abstract

Cis- regulatory sequences known as enhancers regulate transcription of their target gets both quantitatively and qualitatively by recruiting DNA binding protein referred to as transcription factors. This “combinatorial code” of transcription factor binding sites within an enhancer is thought to provide the instructions as to when and where an enhancer is active and promotes the transcription of its target gene. In turn, enhancer structure and the organization of these sites can influence its function in tandem with the identity of the binding sites it contains. In addition to this transcriptional code, enhancers can be located downstream, upstream, and even within the non-coding regions of the genes they regulate, therefore enhancers must possess and intrinsic ability to act over large genomic distances to regulate gene transcription. Little is known about mechanisms this process occurs, although it must require active facilitation. The field recognizes six models of distal enhancer action which will be discussed in depth below: 1. Looping, 2. Linking, 3. Tracking, 4. Facilitated Tracking, 5. Change in nuclear localization, and 6. Generation of long non-coding

RNAs. In order to better understand distal gene regulation at the level of enhancer DNA sequence, we performed extensive characterization of the *dPax2* cone cell specific enhancer *sparkling* (*spa) sparking.* The *spa* enhancer is known to be regulated by the Notch and EGFR signaling pathways via binding sites for their transcriptional effectors Suppressor of Hairless (Su[H]) and the Ets factors Pointed P2 (PntP2) and Yan respectively, as well as for the transcription factor Lozenge (Lz). Additional study identified four additional, approximately 40bp sequences, within the enhancers that are required for enhancer action. We examined the role of these additional sequences, referred to here as regions 1, 4, 5, and 6 in enabling *spa* to act at a distance from its target gene's promoter. Here we identified a specific, 40bp sequence within the *spa* enhancer that is solely responsible for distal gene regulation and plays no role in patterning gene expression. This proximal promoter analysis also allowed us to elucidate the specific roles of each of the critical *spa* regions (1, 4, 5, and 6a). We found that regions 5 and 6a appear to be required for proper timing of gene expression rather than maintinance of gene expression. Additionally, region 5 contains a transcription input necessary for repression of *spa* enhancer activity in photoreceptors. Meanwhile region 4 is required for robust gene activation regardless of enhancer position with respect to the promoter.

LONG RANGE GENE REGULATION AND THE IDENTIFICATION OF THE *SPARKLING* RCE

2.2 Introduction

Enhancers, a subclass of *cis*-regulatory elements, are responsible for regulating gene transcription in a tissue and temporal specific manner. Examination of enhancers over the past thirty years has shown that the genomic sequences regulate transcription through of variety of mechanism, including first recruiting a specific combination of transcription factors. This "combinatorial code" of binding sites within an enhancer is thought to provide the instructions as to when and where an enhancer is active and promotes the transcription of its target gene. These transcription factors in turn are thought to enable enhancer action through a variety of mechanisms, including recruiting cofactors that regulate genes through a variety of mechanisms which include direct recruitment of DNA polymerase II and the basal transcription machinery. (Szutorisz et al, 2005). Alternatively, enhancer binding transcription factors can recruit Mediator, the large multi-unit complex that promotes transcription via assembly of the basal transcription machinery at a genes promoter (Malik and Roeder, 2005; Wang et al, 2005). Enhancers are also capable of influencing their local chromatin environment via their specific protein interactions. For example, they can recruit ATP-dependent nucleosome remodeling complexes (ie SWI/SNF family members) and enzymes such as histone acetyltransferases (HATs) and histone deacetylases (HDACs) which covalently modify histones, resulting in changes in chromatin structure that activate or repress transcription respectively (Narlinkar et al, 2002; Orphanides and Reinberg, 2002). While the mechanisms described here have been shown to be essential for transcription, these mechanisms alone are not sufficient to regulate gene expression. Additional work has shown that aspects of an enhancer's structure and organization can influence its function in tandem with the identity of the binding sites it contains.

While it is critical to understand the basic inputs and rules that govern an enhancers structure and binding site organization, most enhancers must play another important related role, which is to regulate their target gene from a large genomic distance from the transcriptional start site (TSS) of its target genes. Enhancers can be located upstream, downstream, or even within the intronic sequences of protein coding genes. On average, enhancers lie less than 30kb from the gene they regulate. However, one extreme example of distal enhancer activity is that of the *Shh* limb enhancer in mice which lies 1Mb upstream of the *Shh* TSS (Lettice et al., 2003). This enhancer must not only activate proper gene transcription from an incredible genomic distance, it must work over a considerable gene desert and from the intron of the *Lmbr1* gene. Surprisingly, the *Shh* limb enhancer displays high specificity for the *Shh* promoter and does not induce expression of other genes in the vicinity (Amano et al., 2009). This is not true for all enhancers, some of which are promiscuous and can influence the transcription of non-target genes. This is most often seen with the insertion of transgenes into the genome, which can result in unexpected gene activation from surrounding enhancers, otherwise known as enhancer traps (Lower et al., 2009; Ruf et al., 2011; Spitz and Duboule, 2008).

The ability of enhancers to act over a distance to regulate gene transcription and communicate with their target promoters is likely to be intrinsic to the enhancer sequence. Stated differently, each enhancer DNA sequence likely contains within it the information that allows it to distally regulate gene expression. This is especially evidenced by the ability of enhancers to drive transcription from heterologous promoters and from outside their endogenous genomic locations. By definition most enhancers act some genomic distance; however, little is known about mechanisms this process occurs, although it must require active facilitation. The field recognizes six models of distal enhancer action which will be discussed in depth below: 1. Looping, 2. Linking, 3. Tracking, 4. Facilitated Tracking, 5. Change in nuclear localization, and 6. Generation of long non-coding RNAs (Figure 2.1). These models vary widely; some with much more evidence than others. However, no single model has definitively shown to be the mechanism by which enhancers act to regulation distal gene transcription. In all likelihood, a single model cannot describe the action of all enhancers. It is also likely that the mechanisms described in these models can work in conjunction to activate gene expression.

Figure 2.1

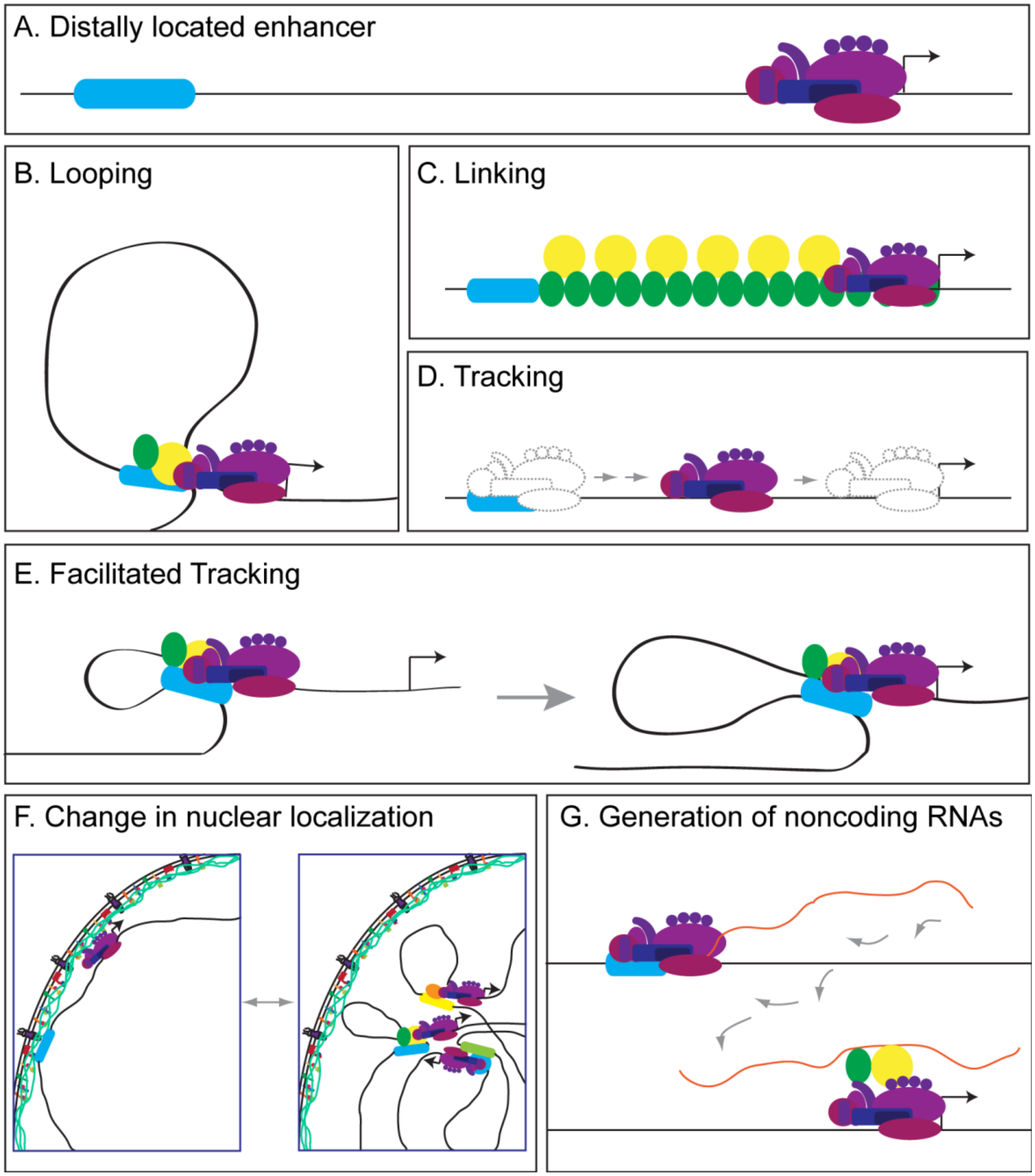

A. Distally located enhancer
B. Looping
C. Linking
D. Tracking
E. Facilitated Tracking
F. Change in nuclear localization
G. Generation of noncoding RNAs

Figure 2.1 Models of long-range enhancer action. Proposed mechanisms of distal gene regulation by enhancers. (A) Depiction of a distal enhancer (blue) with the basal transcription machinery at target gene promoter (purple). (B) Looping to bring enhancers and promoters into direct contact. (C) Linking the enhancer and promoter to each other with a large protein complex. (D) Tracking of the basal transcription machinery between the enhancer and promoter. (E) Facilitated Tracking combines looping and tracking. (F) A change in nuclear localization from the transcriptionaly inactive nuclear periphery to active locations such as transcription factories. (G) non-coding enhancer transcripts influence gene transcription.

2.2a Looping

By far the most popular model of enhancer action is that of looping. This mechanism proposes the formation of DNA loops between an enhancer and a promoter such that active gene sequences coincide within the nucleus despite being separated by significant genomic distances. As higher order nuclear architecture is already well established, this model fits well with other known actions within the nucleus, such as looping or tethering DNA by insulator sequences, or the juxtaposition of two DNA sequences by cohesion and condensin. While oft proposed as a mechanism of enhancer action, recent molecular techniques such as chromatin conformation capture (3C), have demonstrated that enhancers and promoters can indeed be located close to an active gene, suggesting a chromatin loop forms to displace the intervening sequence (Carter et al., 2002; Dekker et al., 2002; Tolhuis et al., 2002). In a classic example analyzing the globin locus, the locus control region (LCR), which is 40-60bp upstream from the genes it regulates, comes in close proximity to active genes throughout development. However, co-localization of the LCR and the globin genes is not observed in tissues where these genes are not expressed. Furthermore, the formation of these loops of DNA between the LCR and active globin genes has been shown to require several proteins including GATA-1 (Drissen et al., 2004; Vakoc et al., 2005). The LCR in T-helper type 2 cells which regulates expression of interleukins 4, 5, and 13, has been shown to form similar loops to bring the LCR into proximity with its target genes promoters (Spilianakis and Flavell, 2004).

The *Shh* limb bud enhancer located over 1Mb from its target promoter is a yet another interesting example of how linear genomic distance does not necessarily equate to spatial distance in the nucleus. 3C and FISH studies have demonstrated that the distal enhancer and the Shh promoter are brought into close proximity via formation of chromatin loops. These loops only form in limb bud cells where the enhancer is active, but not in other tissues where *Shh* is expressed in the control of different enhancers.   Interestingly, mutations can be made to this enhancer that abolish transcription of *Shh,* but do not affect the chromatin loop. This suggests the roles for looping and transcription activation and patterning information by the enhancer are controlled by separable sequences (Amano et al., 2009).

Since the identification of loops forming in chromatin between enhancers and promoters, numerous proteins have also been identified which promote the formation of chromatin loops. The insulator interacting protein CTCF has been shown to interact with both cohesion and mediator to promote the formation of DNA loops in the IFN-locus and β-globin locus (Chien et al., 2011a; Chien et al., 2011b; Hadjur et al., 2009). Interestingly, the interaction of CTCF and cohesion can induce the formation of two differentially transcribed loops in the apoliprotein locus in Hep3B cells such that an enhancer and its target genes exist in one loop together (Mishiro et al., 2009). Nipped-B-Like has been shown to colocalize with cohesion/mediator bound enhancer and promoter regions and not at CTCF/cohesion bound enhancers, suggesting that Nipped-B-like facilitates the formation of cohesion and mediator induced loops in the absence of CTCF binding sites (Kagey et al., 2010). Interestingly, the *Drosophila* homolog Nipped-B is required for long-range enhancer-promoter communication in the Cut locus, implicating a conserved role for this protein in facilitating distal gene regulation (Morcillo et al., 1996).

The majority of evidence supporting looping occurs in multigene clusters and LCRs. However there are also a few examples involving a single enhancer regulating a single gene. For example, in *Drosophila* the transcription factor Zeste has been shown to homo-oligomerize between binding sites at enhancers and promoters in the *white* and *Ubx* loci to induce the formation of DNA loops (Kostyuchenko et al., 2009; Laney and Biggin, 1997; Mohrmann et al., 2002; Qian et al., 1992). Similarly, expression of the heme oxgenase-1 (HO-1) gene in mammals has been shown to be regulated in renal epithelia cells, by a downstream intronic enhancer. Interestingly both the HO-1 promoter and the enhancer contain binding sites for specificity protein 1 (Sp1) and the interaction of this protein with each of these sequences has subsequently shown to be critical for the formation of a chromatin loop to bring the promoter and enhancer together to stimulate transcription (Deshane et al., 2010).

Importantly, looping is one of the few models of enhancer action that can account for the observation that most enhancers can act both irrespective of distance and orientation with respect to its target genes promoter. However, the model implicates the necessity of both enhancer AND promoter sequences through which the proteins can bind and promote the formation of loops. Yet, most enhancers can work to drive heterologous promoters in addition to the endogenous target promoter suggesting that enhancers can communicate remotely without the specific sequences within the promoter. That said, most enhancer studies on heterologous promoters do not study the enhancer placed at distance and in a promoter proximal position which would not require long-range sequences. In the *Engrailed* (En) locus for example, distally placed enhancers require a specific sequence within the *En* promoter to function, and cannot drive reporter expression from an hsp70 promoter unless placed in a proximal promoter location (Kwon et al., 2009). It is possible that enhancers fall into classes in which some enhancers require local promoter sequences when placed distally, while others do not, and we will not understand this distinction until more enhancers are analyzed distally.

2.2b Linking

Perhaps the oldest model of distal enhancer action is that of linking. In this model the enhancer and promoter remain separated from each other spatially and the enhancer sets up the formation of a protein complex that spreads across the chromatin between the enhancer and promoter, ultimately activating target gene transcription. Despite its early proposal, experimental evidence for this model is extraordinarily lacking.   A 1997 study demonstrated expression of the gene *Chip* is required for activation of the *Cut* gene in the wing disc margin by a distally located enhancer (Morcillo et al., 1996). The Chip protein is expressed in an apparently ubiquitous manner in all *Drosophila* cells. It is present at numerous loci along polytene chromosomes in the salivary gland and its loss is associated with diminished expression of gap and pair-rule gene expression in the *Drosophila* embryo, lending credit to a role in transcriptional regulation via chromatin interaction (Morcillo et al., 1997). Subsequent analysis showed that insertion of a gypsy insulator element intergenically between *Cut* and the enhancer only affected enhancer activity when *Chip* express was reduced (Morcillo et al., 1996). This data implicated *Chip* in a role shaping the chromatin or creating a structure on the chromatin that can influence transcription. Noteably, *Chip* has been shown to interact with LIM interaction domains found in many homeodomain binding proteins, such as Apetrous in the wing disc. Data at the time suggested homeodomain binding proteins bind to various sequences with the similar affinities and therefore homeodomains proteins were excellent candidates for binding sequences between enhancers and promoters (Carr and Biggin, 1999; Desplan et al., 1988; Walter and Biggin, 1996). In fact, unlike most transcription factors, LIM-domain proteins with homeodomains binding sites, appear to bind active gene regions ubiquitously rather than in small clustered regions (Carr and Biggin, 1999; Liang and Biggin, 1998). The affinity of Chip for LIM domains combined with its necessity in remote enhancer action at the *Cut* locus resulted in a model whereby Chip acts to recruit and crosslink LIM proteins to the DNA between enhancers and promoters, ultimately “linking” the two together through a protein complex.

Soon after the association of *Chip* with distal enhancer action in the *Drosophila Cut* locus, a linking model involving Chip was proposed for the β-globin locus. This mechanism of LCR action in the globin locus is based on homology of the factors in vertebrates. The vertebrate homologs of *Chip* are known to interact with LIM domains of homeodomain bind proteins (Breen et al., 1998). Furthermore in eyrthroid cells, where β-globin is expressed, the Chip homolog Ldb1 has been shown to form a DNA binding complex with the LIM-domain protein Lmo2, the E-box transcription factor Tal-1, and the transcription factor GATA-1 (Osada et al., 1997; Wadman et al., 1997). Together these observations lead to the hypothesis that Ldb1 facilitates homeodomain protein recruitment and crosslinking and stimulates transcription in the same manner proposed to function in the *Drosophia Cut* locus. However, there is no experimental evidence to suggest Chip homologues bind in the β-globin regions between enhancer and promoters, or that such binding is necessary for gene regulation.

2.2c Tracking

The finding that much of an organism's genome, and not just the protein coding sequences are transcribed, provides the simplest evidence for tracking as a method of enhancer action (Kapranov et al., 2007; Mercer et al., 2009; Ponting et al., 2009). In this model, enhancers act to recruit RNA polymerase II (Pol II) which then "tracks" along DNA either directionally, or bidirectionally, until it comes in contact with a promoter which it then actively transcribes. Evidence for tracking stems from data suggesting that non-coding RNAs are transcribed between some enhancers and promoters, such as at the human ε- and β-globin loci, and *Drosophila* biothorax complex (Zhu et al. 2007, Gribnau et al, 2000; Bae et al, 2002). The advent of Chromatin Immunoprecipitation (ChIP) has led to numerous studies assessing the location of phosphorylated (active) Pol II in the genome. These studies have found instances of Pol II at enhancers and promoters as well as the sequences in between (Johnson et al., 2001; Wang et al., 2005a). In the human ε-globin Pol II and TATA binding protein (TBP) localized to the HS2 enhancer and along the DNA to the promoter, where the synthesized short polyadenylated, intergenic RNAs (Zhu et al., 2007). In another study, the localization of Pol II was assessed in the *prostate specific antigen* locus in and was found to be recruited first to the enhancer, then to the intervening sequences, and finally to the genes promoter in androgen inducible manner. Furthermore, this localization and the production of non-coding RNAs was inhibited with an insulator blocking sequence was positioned between the enhancer and promoter (Wang et al., 2005b).

Tracking is a preferred model of enhancer activity as it invokes a capability all enhancers much have – the ability to recruit the basal transcription machinery to their target promoter. However, this mechanism of action seems unlikely for intronic enhancers as Pol II would have to travel in both directions along an actively transcribed gene; first upstream from the enhancer to locate the target promoter, and then downstream to transcribe the gene. This mechanism would make long term expression of a gene rather complicated. Furthermore, a tracking model does not take into account that many enhancers must activate a single promoter despite being closer to or equidistant to a different promoter. The observation that some enhancers activate expression from a single promoter, while others appear to drive “leaky” expression from other local promoters may ultimately help classify *cis*-regulatory sequences as looping or tracking enhancers.

2.2d Facilitated tracking

The distal gene regulation model of facilitated tracking merges the models of looping and tracking. Here, Pol II is recruited to an enhancer and subsequently moves along DNA while retaining contact with the enhancer such that chromatin loops are formed as enhancer/Pol II complex moves further along the DNA. Upon locating the target promoter the transcription machinery must release the enhancer and chromatin loop, and begin transcription of the target gene. While facilitated tracking is an oft cited and tempting model as it combines the two most popular models of distal enhancer action, little to no experimental evidence exists to support the model.

One study examining the human ε-globin gene locus postulates a facilitated tracking mechanism for activation of target gene expression from the HS2 enhancer over 10kb away. The authors found that short, overlapping polyadenylated RNA’s are transcribed from the intergenic sequence between the HS2 enhancer and the ε-globin gene. Accordingly, Chromatin Immunoprecipation (ChIP) showed Pol II localizes to the enhancer, the promoter, and to the intergenic sequences. Production of RNAs and intergenic and promoter localization of Pol II was blocked by insertion of an insulator sequence in the intergenic region. Looping between the HS2 enhancer and ε-globin promoter was demonstrated by 3C and was reduced in the presence of the intergenic insulator. As the insulator blocked Pol II movement and transcription, this provides tentative evidence for looping mediated by Pol II interaction (Zhu et al., 2007).

A separate study analyzed gene regulation of the prostate specific antigen (PSA) by an enhancer 4kb upstream in response to androgen stimulation. As described above, looping was demonstrated though ChIP and 3C showing that the enhancer and promoter are in close proximity after stimulation. Similarly, Pol II is found at the enhancer, promoter and intergenic sequences. What is unique about this study, is that Pol II was found to bind to the locus in a time dependent manner after androgen stimulation. Pol II was found first at the enhancer, then the intergenic sequences, and lastly at the PSA promoter. Therefore this gene locus is another example of tracking and looping occurring in concert, notably when the enhancer is active. However, when tracking is inhibited by preventing the phosophorylation of Pol II, looping still occurs between the enhancer and promoter, suggesting these two processes utilize independent mechanisms of action (Wang et al., 2005).

Importantly, neither of the studies described here identified loops between the enhancer or promoter and the intergenic sequences, which would be a hallmark of facilitated tracking if it were to act as a mechanism of long range transcriptional regulation. While both authors made legitimate attempts, it is technically challenging to separate the processes of tracking and looping and still allow for potential gene transcription, especially to inhibit looping while allowing tracking to occur. Until this can be achieved in numerous gene loci, it remains to be seen whether tracking can inducing the formation of chromatin loops, or whether tracking and looping are merely coincidal at the loci.

2.2e Change in nuclear localization

Chromatin within the eukaryotic nucleus exhibit complex levels of organization that can affect gene expression. For example gene-rich and gene-poor chromosomal territories (CTs) are typically confined to independent regions of the nucleus. Gene-poor CTs, heterochromatin, and actively repressed gene regions are restricted to the nuclear periphery (Boyle et al., 2001; Cremer and Cremer, 2010). This localization is dependent on the nuclear lamina proteins which line the inner side of the nuclear envelope (Shevelyov et al., 2009; Wilson and Berk, 2010). Reporter gene assays in which the regulatory sequences are confined to the nuclear periphery by fusion to lamina proteins can inhibit the expression of an otherwise ubiquitously expressed transcript. The necessity of the nuclear envelope for gene repression is demonstrated in the activity in *Drosophila testes-specific* genes. In somatic cells these genes are associated with B-type lamins thereby localizing them to the nuclear periphery. Loss B-lamins results in dissociation of the testes specific genes with the nuclear envelope and subsequent increase in somatic cell expression of these genes (Shevelyov et al., 2009).

Conversely, gene-rich euchromatin and actively transcribed genomic regions are typically confined to the nuclear interior (Boyle et al., 2001; Cremer and Cremer, 2010). In a further level of organization, actively transcribed genes are often extruded from their local CT such that it is spatially separate from the surrounding chromatin (Bickmore et al., 2004). For example, in developing limb cells, the *Shh* hedgehog locus moved out of its CT, but only in the cells of the posterior limb bud where *Shh* is expressed. Furthermore, movement requires the presence of the enhancer sequence 1Mb upstream (Amano et al., 2009).

In addition to moving away from its local CT, actively transcribed genes can be associated with transcription factories. It has been observed that active RNA Pol II is localized in a non-uniform manner and specifically within interchromsomal spaces (Rada-Iglesias et al., 2011). These distinct foci are known as transcription factories, as evidence suggest several independent transcriptional units can occupy a single loci. HeLa cells, for example, have about 10,000 transcription factories per nucleus, while non-cell line tissues such as erythroid cells have only 100-300 Pol II foci per nucleus. As erythroid cells can express at least 4,000 genes at the same time, many active genes must occupy the same transcription factory at the same time, perhaps as many as 13 genes per factory (Jackson et al., 1998). Different studies have shown that regulatory sequences that share similar characteristics can occupy the same Pol II foci. Sequences contain the same promoter types are frequently grouped together, while differing promoter types are confined to distinct foci. Splicing factors, such as SC35, are enriched only at some transcription factories. Accordingly, genes containing introns localize to these loci specifically (Rada-Iglesias et al., 2011). Active genes from the same chromosome frequently colocalize in the nucleus, as do tissue and temporal specific genes. In an exquisite 3D study, the localization of several eythroid specific genes was studied. During mouse eyrthroid cell differentiation, *Hbb-b1* and *Eraf*, both on the 7th chromosome, colocalized. *Hbb-b1,* is also located with a ubiquitously expressed gene also on the 7th chromosome, *Uros,* and surprisingly with another erythroid specific gene *Hba* on the 11th chromosome. Importantly, none of these genes were located near the imprinted and therefore repressed genes *Igf2 and Kcnq1ot1* on the 7th chromosome. The authors went onto show that the observed gene colocalization correlated with Pol II foci and active transcription. When the genes were not actively transcribed, they were not located in near proximity. Finally, using 3C, this study demonstrated the colocalized genes were in physical contact with one another (Osborne et al., 2004).

While there is little experimental evidence to directly implicate enhancers altering a gene's nuclear localization, the involvement of enhancers in this activity is not only plausible, it is likely. We have already seen that the *Shh* gene locus cannot move away from its local CT in the absence of its limb enhancer. Furthermore, it has been shown that reporter constructs containing regulatory sequences are found in the same factories as the endogenous gene, suggesting colocalization to a specific focus is driven by shared transcription factors.

Accordingly, the mouse *globin* genes assemble preferentially in transcription factories with hundreds of other sequences that are regulated by the transcription factor Klf1 (Schoenfelder et al., 2010). As enhancers have been shown to associate with transcription factors and chromatic remodeling proteins, it is easy to imagine how an enhancer could regulate movement of its target gene within the nucleus by moving the chromatin away from the local CT, or by forming enhancer promoter loops and associating with transcription factories simultaneously. Alternatively, an enhancer and its target promoter could move separately to the same transcription factory via interaction with the same co-factors.

2.2f Generation of noncoding RNAs

In recent years it is been observed that the genomes of mammalians and other organisms are pervasively transcribed (Kapranov et al., 2007; Mercer et al., 2009; Ponting et al., 2009). As less 5% of mammalian genomes encode proteins, much interest lies in the function of these transcripts. Many of these noncoding RNAs, such as microRNA, small interfering RNAs, and Piwi – interacting RNA, have been shown to interfere with gene regulation at the post transcriptional level by affecting mRNA stability or inhibiting translation. Another diverse class of noncoding RNAs is distinct from these and other small RNAs, not only in size (typically longer than 200 nucleotides), but also in their ability act on gene regulation at the transcriptional level. Long non-coding RNAs,or lncRNA, have been found in organisms from yeast to humans and have been shown to regulate chromatin modifications, transcription, splicing, mRNA processing, translation, the production of endogenous short interfering RNAs (Cabili et al., 2011; Wyers et al., 2005).

With respect to transcription, lncRNAs were thought to only inhibit gene transcription until recently. The classic example of transcriptional repression through lncRNAs is in activation of the female X-chromosome in mammals by the lncRNA *Xist*. *Xist* RNA is transcribed from an X-chromosome target for inactivation and subsequently coats the chromosome and inducing irreversible chromatin modifications that inhibit transcription (Heard and Disteche, 2006). Similarly, lncRNAs such as *Air*, which is a single, polyadenylated, unspliced, transcript covering for than 100kb, can bind to chromatin even several hundred kilobases away, ultimately resulting in genomic imprinting of genes such as insulin-like growth factor 2 receptor (Sleutels et al., 2002). In another example, the *HOTAIR* lncRNA is expressed from the HOXC locus in mice and subsequently inhibits transcription of the HOXD locus on another chromosome via interaction with the Polycomb Repressor Complexes (PRCs) (Wang et al., 2011b). Despite clear evidence that lncRNAs are capable of repressing transcription, more recent studies show that they can also enhance transcription. Here, the classic example also lies in the dosage compensation of the sex chromosomes. In *Drosophila*, two seemingly redundant lncRNAs *roX1* and *roX2* are responsible for activating parts of the male X chromosome (Franke and

Baker, 1999). Ironically, *Xist* expression is positively regulated by another lncRNA *Jpx* which lies upstream of *Xist (Tian et al., 2010).*

lncRNA that positively regulate transcription have been classified by several characteristics into lincRNA (long intervening non coding RNAs), eRNA (enhancer RNA), uaRNA (upstream-antisense RNA), and CUT (cryptic unstable transcript). uaRNAs are 50 to 1000bp transcripts from promoter regions whose function is generally unknown, although they may act as chromatin tethers (Flynn et al., 2011). Likewise CUTs are expressed near promoters in yeast but generate divergent, both sense and anti-sense transcripts. Anti-sense CUTs, appear to positively regulate nearby gene transcription, while sense CUTs, which are made preferentially, antagonize transcription of the same gene (Wyers et al., 2005).

Unlike uaRNAs and CUTs, lincRNAs and eRNAs are expressed from locations distal to transcriptional start sites (TSS). eRNAs are divergent RNAs transcribed from known enhancers. The existence of eRNAs has also been used as the primary source of evidence for the tracking model of long range gene regulation as well (Kim et al., 2010). Studies in prostate cancer cell lines suggest that the expression of eRNAs is inducible and correlates with cellular differentiation. Additionally, expression levels of eRNAs correlates with levels of target gene transcription (Wang et al., 2011a). Furthermore, active enhancers that express eRNA also coincide with H3K4me1 and H3L27ac marks and p300 and Med12 residence, indicators of active transcription, at the enhancer and target genet promoters (Creyghton et al., 2010; Ogryzko et al., 1996; Rada-

Iglesias et al., 2011). Together this leads to a model by which eRNAs interact with the chromatin to establish or maintain these active chromatin marks. Alternatively, eRNAs could act by creating a scaffold for protein bind which allows transcription factor and Pol II association with the enhancer and target promoters, perhaps even affecting spatial arrangements such as looping through interactions with the chromatin. eRNAs were identified and characterized in mice, however similar sequences have been identified in *Drosophila* and are referred to as TSS-distal DHSs, and due to the compact nature of the *Drosophila* genome are found more often in intronic sequences than seen in mammals but are similarly bidirectional transcribed and marked as active (Kharchenko et al., 2011). lincRNAs differ from eRNAs in that they are not expressed from known enhancers, and like protein-coding RNAs they are poly adenylated, spliced (typically 1 to 2 exons), capped, and are rarely bi-directional (Cabili et al., 2011). eRNAs are abundantly prevalent at active enhancers and correlate with changes in chromatin suggesting the RNAs themselves are important in regulating gene transcription; however, they could simply be a byproduct of active polymerase at enhancers.

lincRNAs on the other hand have been shown to be critical for the transcription of neighboring genes (Orom et al., 2010). One study assessed the ENCODE annotation of the human genome and identified a set of 3019 lincRNAs. They subsequently showed that 15% of the lincRNAs expressed in keratinocytes are differentially expressed and observed that positive lincRNA levels correlated with positive mRNA levels for neighboring genes. Using siRNA targeted for these lincRNA sequences the authors showed that transcription of these neighboring genes was dependent on lincRNA transcription. Intriguingly, when lincRNAs were used in luciferase reporter assays, the presence of wildtype lincRNA sequences resulted in an increase in luciferase reporter activity, which was lost when lincRNA production was inhibited through targeted siRNAs or the insertion of polyadenylation cassettes in the middle of the lincRNA sequence. Notebly, these luciferase assays utilized a heterologous promoter rather than the endogenous neighboring genes promoter, suggesting the lincRNA affects the chromatic structure or environment rather than targeting a specific promoter. These lincRNAs also functioned to regulate luciferase activity regardless of their position either upstream or downstream of the heterologous promoter (Orom et al., 2010). In another recent study, a lincRNA in the HOXA locus known as *HOTTIP,* was found to be transcribed from a distal genomic region, but was necessary for expression of the downstream 5' *HOXA* genes. The *HOTTIP* linc RNA acts by interacting with directly with WDR5, a component of the mixed-lineage leukemina (MLL) -containing complexes which have been shown to affect H3K4me3 marks in the *HOXA* locus. Loss of the *HOTTIP* RNA, or inhibition of the WDR5 interaction, results in diminished *HOXA* gene expression (Trievel and Shilatifard, 2009; Wang et al., 2011b). While lincRNAs are not associated with known enhancers, their activity appears to replicate that of transcriptional enhancers in that they become active upon an inductive differentiation signal, they can act on heterologous promoters, and they can act regardless of orientation with respect to the promoter. It is possible then that lincRNAs are transcribed from classic enhancers or that genomic sequences that encode lincRNAs represent a specific class of enhancers.

2.2g The dPax2 sparkling enhancer as a model distally located enhancer

Despite the numerous models of distal enhancer action and the identified proteins that influence long-range gene transcription, no element within an enhancer alone has been identified to play this role specifically. Examples of DNA sequences within transcriptional units that can facilitate long-range transcription to data involve sequence within both the enhancer and its target promoter such as binding sites for Zeste or Sp1 (Deshane et al., 2010; Kostyuchenko et al., 2009; Laney and Biggin, 1997; Qian et al., 1992). Alternatively, single sequence elements are all involved tethering sequences with a specific promoter or insulator bypass elements. For example the *Abdominal-B* and *Antenepedia* loci in *Drosophila* contain approximately 200bp DNA sequences near their respective promoters, which provide no patterning information but are required for upstream enhancers to contact the correct promoter. As such they are called promoter-tethering elements to denote their function in restricting an enhancer to a single distal promoter (Akbari et al., 2008; Akbari et al., 2007; Calhoun et al., 2002). In the *Drosophia Cut* locus, both Chip and Nipped-B are required for a distally placed enhancer to activate transcription despite the presence of an intervening insulator sequence (Morcillo et al., 1996).

To date, no enhancer sequence that does not also require a specific promoter element to act distally has been identified. Yet, enhancers must be able to function at a distance without target promoter specific binding sequences as most enhancers can be removed from their genomic context and act in reporter constructs to regulate gene transcription from a heterologous, not the endogenous, promoter. In order to better understand distal gene regulation at the level of enhancer DNA sequence, we turned to our laboratories previous extensive characterization of the *dPax2* cone cell specific enhancer *sparkling* (*spa) sparking,* which is critical for *dPax2* expression in cone cells of the developing *Drosophila* eye (Fu et al., 1998; Fu and Noll, 1997). The enhancer is known to be regulated by the Notch and EGFR signaling pathways via binding sites for their transcriptional effectors Suppressor of Hairless (Su[H]) and the Ets factors Pointed P2 (PntP2) and Yan respectively, as well as for the transcription factor Lozenge (Lz) which is critical for the specification of the ommotidial cell types (Flores et al., 1998; Flores et al., 2000; Voas and Rebay, 2004). The *sparkling* enhancer is located in the 4th intron of the *dPax2,* putting it 7kb downstream of the gene's promoter. *sparkling's* location with respect to its target genes promoter means that this enhancer must be able to function at a distance to interact with and stimulate transcription in the correct time and place. This genomic location, combined with the significant work already performed to identify the minimal enhancer and at least three of its transcriptional inputs, made the *sparkling* candidate a promising candidate for the study of long-range transcriptional regulation at the level of enhancer sequences.

Subsequent analysis of the *sparkling* enhancer demonstrated that when *spa* is placed at a moderate distance from a heterologous promoter, hsp70, driving GFP the wildtype enhancer is capable of inducing GFP expression in cone cells in the developing imaginal eye disc in pattern reminiscent of *dPax2* expression (Swanson et al., 2010). Mutational enhancer analysis of the enhancer revealed that the known transcription factor binding sites for Su(H), Ets, and Lz are not sufficient for enhance activity. The additional inputs for necessary for *spa* action were traced to four additional sub elements within the *spa* enhancer. These sequences, which we call regions 1, 4, 5 and 6a due to their position within the enhancer, were shown to be critical for *spa* function. Mutation of any of these sequences individually results in loss of enhancer activity. Regions 1, 4, 5, can subsequently be divided into 3, approximately 10bp, sub-regions that are also all individually critical for *spa* activity. The size of these sub-elements suggests that each essential *spa* region (1, 4, 5, and 6a) contains at least one additional transcription factor binding site (Swanson et al., 2010).

Having established the necessary components of the *sparkling* enhancer at such a fine scale was a critical step in determining the potential role of *spa* sequences in long range gene regulation. Few enhancers are analyzed in such depth; however, the role of the entire enhancer sequence needs to be established in order to design a useful set of experiments to analyze distal gene regulation. The method we used to examine long-range enhancer activity in the *spa* enhancer is remarkably basic. We simply made the same mutations the abolished enhancer activity when *spa* was placed at 846bp from the TSS but with the enhancer placed in the more promoter proximal position of 121bp from the TSS (Figure 2.2 A). The optimal distance for passive interactions between two sequences is less than 260-300bp. Interaction greater than this distance require active facilitation, most likely through interaction with DNA binding proteins/protein complexes (Rippe, 2001). This experimental design places our *spa* enhancers on either side of this optimal length, with -121bp for proximal promoter interaction and -846bp for distal promoter interaction. Using this method, we define enhancer elements involved in long-range transcription regulation as those that are necessary when the enhancer is in the distal position, but no longer required when the enhancer is in the promoter proximal position.

Using this method we identified a specific, 40bp sequence within the *spa* enhancer that is solely responsible for distal gene regulation and plays no role in patterning gene expression. This proximal promoter analysis also allowed us to elucidate the specific roles of each of the critical *spa* regions (1, 4, 5, and 6a). We found that regions 5 and 6a appear to be required for proper timing of gene expression rather than maintinance of gene expression. Additionally, region 5 contains a transcription input necessary for repression of *spa* enhancer activity in photoreceptors. Meanwhile region 4 is required for robust gene activation regardless of enhancer position with respect to the promoter.

2.3 Results

2.3a Evidence for a special type of regulatory site, specifically mediating action at a distance

The wild-type *spa* enhancer drives the same pattern from −121 bp as from −846 bp (Figure 2.2), although activation is noticeably more robust from the more proximal position. A mutant *spa* enhancer lacking region 1 [*spa*(Δ 1)], which is transcriptionally dead at −846 bp (Figure 2.2 C), is completely rescued by placement at position −121, driving robust gene expression in the normal pattern (Figure 2.2 F). By contrast, enhancers with mutations in regions 4, 5, or 6a remain unable to drive wild-type levels or patterns of gene expression at −121 (Figure 2.2).

To our knowledge, this is the first case of a regulatory element found within an enhancer that specifically mediates action from a remote position, with no apparent role in patterning of gene expression or other basic activation functions. We therefore refer to region 1 as a “remote control” element to functionally distinguish it from patterning elements within *spa*, which include the known transcription factor binding’s sites as well as regions 4, 5 and 6, of the enhancer. Future experiments will test the range, potential promoter preferences, and functional properties of this intriguing regulatory element.

Having identified *spa* region 1, hereafter referred to as the RCE, as essential for long-range enhancer activity, we decided to examine the flexibility of

Figure 2.2

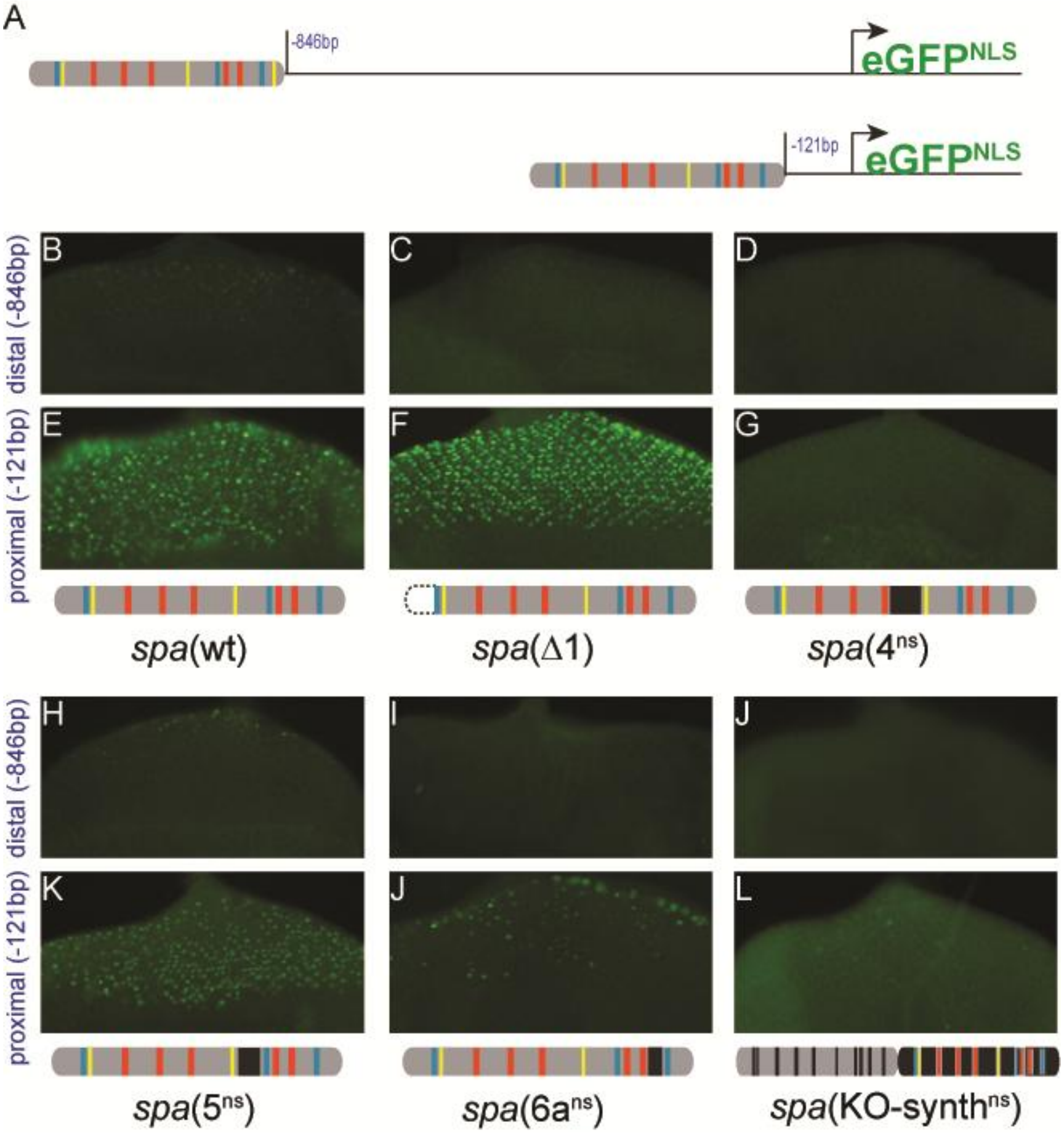


Figure 2.2 The effect of promoter position on *sparkling* activity. We examined *sparkling* reporter constructs in both the promoter distal position of 846 bp from the transcriptional start site, and the promoter proximal position of 121bp from the transcriptional start site. We found that the wildtype enhancer drives increased expression in the proximal position (B, E). *sparkling* requires regions 1, 4, 5, and 5 in when it is located distally (C, D, H, I, J). However in the promoter proximal position *sparkling* no longer requires region 1 in order to activate reporter gene expression, suggesting it is critical for long-range enhancer action only (F). Region 4 sequence is absolutely required for *spa* action regardless of position (G). Meanwhile, loss regions 5 and 6 result in diminished *spa* action at -121bp (K, J). Drastic rearrangements of *spa* sequence are not tolerated in either position (J, L)

this sequence by moving it from the 5′ end to the 3′ end of the enhancer. This rearranged enhancer performs normally at −846 bp (data not shown), which indicates that the precise position of the RCE, relative to the other regulatory sites within *spa*, is not a critical factor in its remote activation function. Future experiments will determine the distance, relative to the enhancer and to the promoter, over which the RCE can act.

2.3b sparkling regions 5 and 6 are required more for initiation than maintenance of gene expression

The amplification of gene expression we saw when the wildtype enhancer in the promoter proximal position allowed us to further detail the specific roles of each of the *spa* enhancer regions. When the enhancer lacking region 5, which is the only enhancer construct tested here to drive low levels of activity from -846bp, was moved to the proximal position, it drives only slightly diminished levels of gene expression; however the pattern of expression is grossly disrupted (Figure 2.2 H, K). Likewise, an enhancer lacking region 6 is also capable of driving GFP expression from the proximal position, albeit at significantly decreased levels. Interestingly, in 24 hour pupa, less than a day later, GFP is expressed in a complete or nearly complete wildtype pattern of gene expression from both these enhancers (Figure 2.3 .T-V and data not shown). In fact, even in the distal position both these enhancers, which drive low or no gene expression in larva, are capable of activating gene expression in the pupal eye discs at slightly lower than wildtype levels (data not shown). Together this data suggests that regions 5 and 6 of the *sparkling* enhancer are essential for initiation of *spa* activity at the correct developmental time point. However, that without either of these sequences, enhancer activity is delayed but can ultimately function. Furthermore, given the levels of gene expression in seen in pupal eye discs, it is unlikely either of these sequences are required for maintenance of enhancer activity after initiation.

2.3c sparkling region 5 contains sequences necessary for enhancer repression in photoreceptors

As mentioned above, the GFP expression driven by the *spa* enhancer lacking region 5 exhibited an unusual pattern of gene expression. This observation prompted us to examine the cell type specificity of this GFP expression. The only other specified cell type at the larval stage of eye development is the photoreceptors (Voas and Rebay, 2004). As such we performed antibody staining in larval discs with antibodies against GFP and Cut to mark cone cells, or Elav to mark photoreceptors. Using this co-staining we saw that GFP is expressed in both larval cone cells and photoreceptors (Figure 2.3 N-S). The gene expression in photoreceptors is even easier to visualize during pupal development when the photoreceptors and cone cells have separated into two distinct layers within the disc tissue. When examining co staining at this age it is clear that in addition to expression in the 4 cone cells,

Figure 2.4

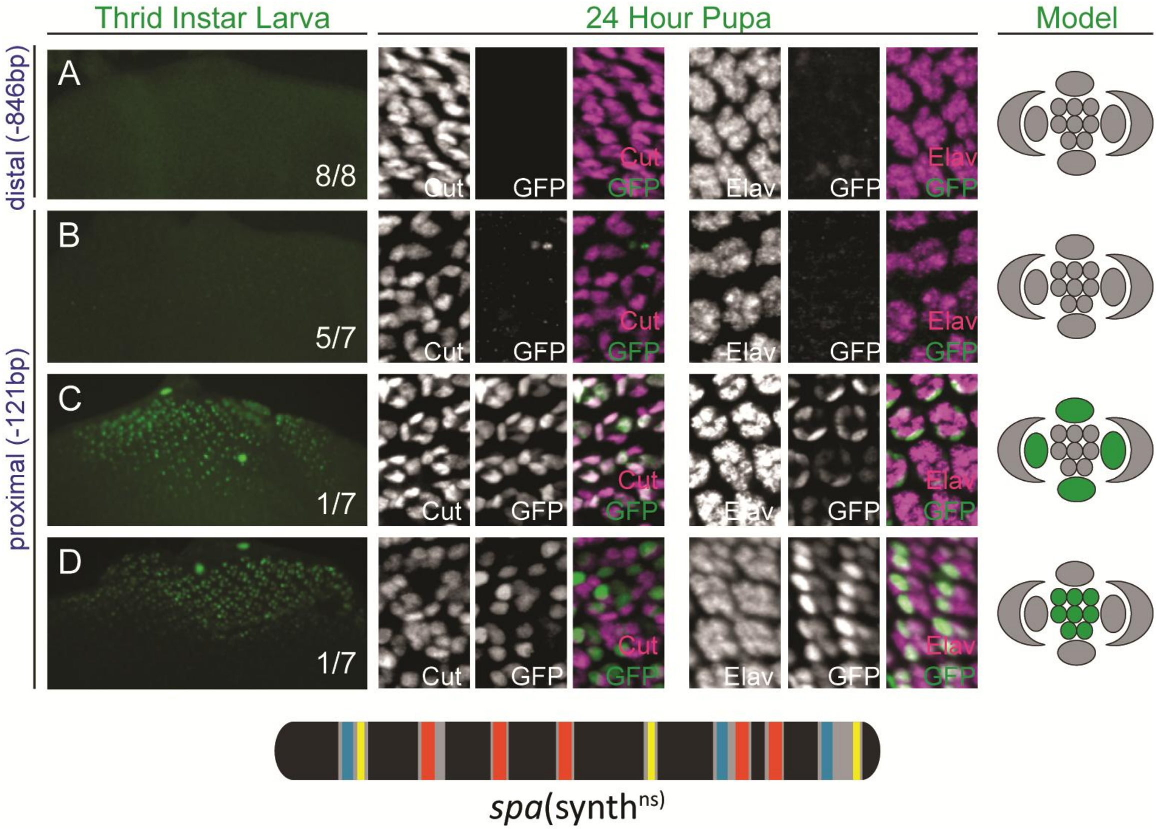


Figure 2.3 *sparkling* is poised for activation in multiple eye cell types. *spa*(synth^{ns}) contains intact binging sites for the known regulators of *sparkling*: Suppressor of Hairless (red), Ets (yellow), and Lozenge (blue). While most of the time we do not see any expression from this construct (A,B), we do see expression in cone cells (C) and photoreceptors (D) in two of our promoter proximal lines. This indicates that *sparkling* enhancer contains information for activation in multiple cell types, and the correct combination of sequences allows for proper enhancer action.

GFP is also expressed in at least one additional cell that is in the center of the cone cell rosette (Figure 2.3 T-V). This additional GFP expression colocalizes with Elav staining (Figure 2.3 W-Y). Cone cells and photoreceptors are both specified by EGFR, Notch, and Lz signaling. Therefore, *spa* activity in cone cells specifically must be regulated by additional factors, likely including additional cone cell specific activators as well repressors expressed in photoreceptors. The data presented here, along with others from our lab, suggest one such repressor binding site lies within *spa* region 5 as *spa* activity is depressed in photoreceptors upon its loss.

2.3d The known transcription factor binding sites and spa region 4 are crucial for robust enhancer activity

We have already seen that the *spa* RCE is not required when the enhancer is placed in the promoter proximal position and that loss regions 5 and 6 do not completely abolish enhancer activity from this position. Contrary to these results loss of region 4 eliminates *spa* function at both the distal and proximal positions (Figure 2.2 D, G). Furthermore, enhancer activity is not recovered during pupal development as we saw with regions 5 and 6 (data not shown). These observations suggest that region 4 is critical for initiation, patterning, and maintenance of *spa* activity.

Likewise the Su(H), Ets, and Lz binding sites are required at both distances from the promoter further demonstrating the necessity of these

Figure 2.3

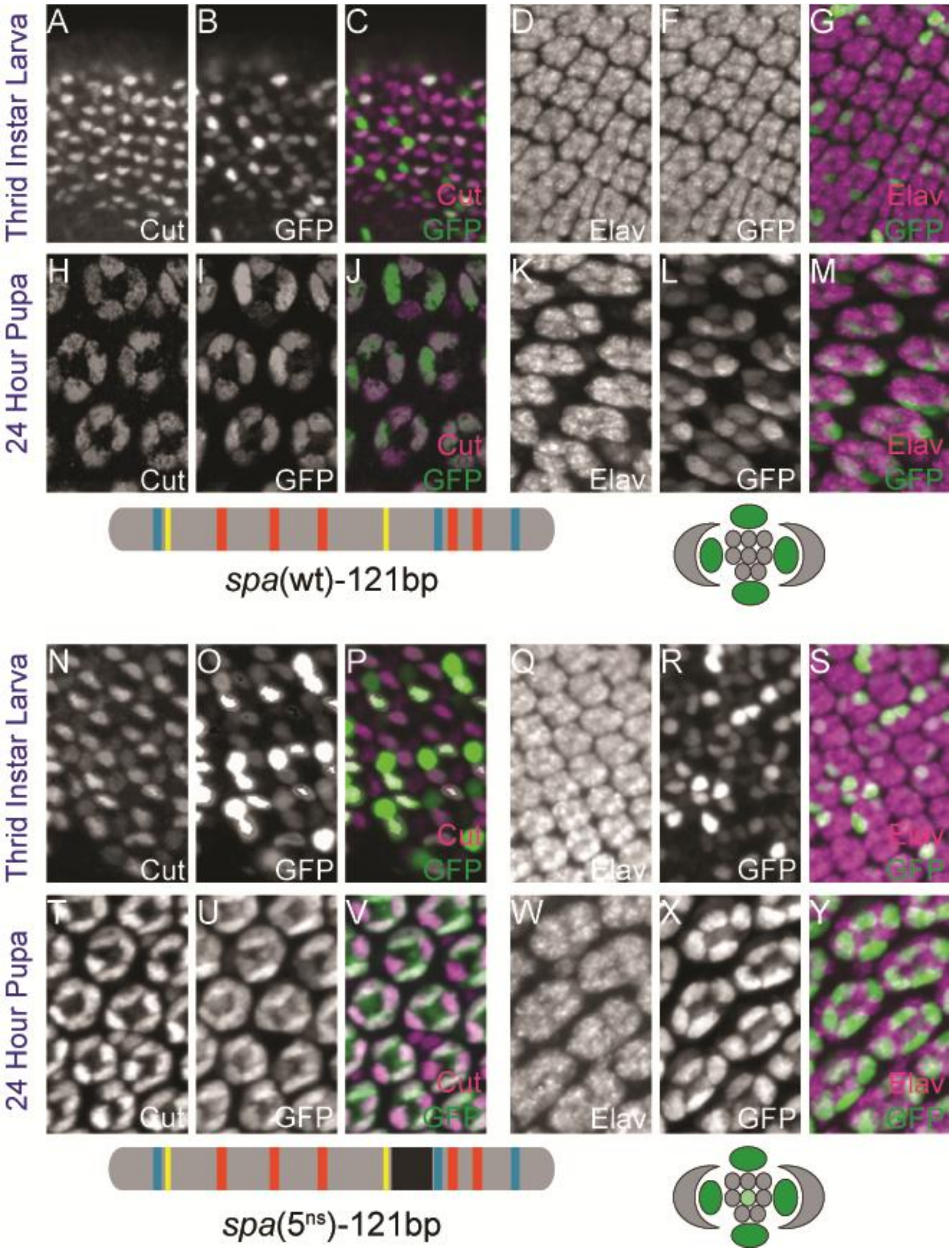


Figure 2.4 Loss of *sparking* region 5 results in ectopic GFP in photoreceptors. Wildtype *sparkling* at -121bp drives GFP expression in cone cells in both the third instar larva and in 24 hour pupa (A-M). Co-labeling with Cut, but not with Elav demonstrates that GFP is expressed in cone cells, but not in photoreceptors. Conversely, loss of region 5 results in expression in both cone cells (N-P, T-V) and photoreceptors (Q-S, W-Y), suggesting the wildtype region 5 sequence contains a repressor binding which inhibits enhancer activity in photoreceptors.

transcriptional inputs (Figure 2.4 A, B). While it is clear that the known transcription factor binding sites are necessary for *spa* activity, we did observe gene activity in two situations. Our *spa* reporter constructs are inserted into the *Drosophila* genome using random P-element mediated transposition and as such we analyze our reporter constructs in multiple lines to overcome any position affect on enhancer activity. We observed no gene expression in all of our *spa*(synth^{ns}) constructs at -846bp. Similarly, in 5 independent insertions of the same construct at -121bp we also saw no gene expression. However, in two independent insertions we observed GFP expression. In one insertion, the reporter construct drove GFP expression in cone cells, while in the other GFP was expressed in photoreceptors. This observation leads to the hypothesis that the regulatory information of Su(H)+Ets+Lz within *spa,* leave it to poised for action in both cone cells and photoreceptors, and a that small amount of additional information from either the remaining sequences within *spa* or from the genomic context of the insertion site can stimulate *spa* activity in either cell type.

2.4 Discussion

We have undertaken a promoter proximal and distal analysis of the *dPax2.* In the course of this work we found that the *sparkling* enhancer at 846bp from the TSS is actually a relatively weak enhancer. Especially compared to the same enhancer placed 121bp from the TSS, which drives significantly higher levels of reporter gene expression. This finding is not unprecedented however, as other enhancers studied at multiple distances exhibit similar transcriptional profiles, including the SV40 enhancer (Banerji et al., 1981). By analyzing an enhancer in this manner we able to identify not only the activation and pattering inputs into the enhancer, but also sequences responsible specifically for long-range gene regulation.

2.4a sparkling regulatory sequences each contribute unique roles to enhancer function

The *dpax2 sparkling* enhancer requires contributions from the transcription factor binding sites for Su(H), Lz, and the Ets factors Pointed P2 and Yan (Flores et al., 2000). In our investigation of the proximal and distal properties of this enhancer, we found that these sites are not sufficient to drive long-range or promoter proximal gene expression. However, we did find that these transcription factors, which are expressed in photoreceptors, cone cells, and primary pigment cells, leave the enhancer poised for action in photoreceptors and cone cells, but likely primary pigment cells as well. One way we demonstrated this is through reporter construct insertion site position affects. While these binding sites cannot normally drive expression by themselves, the contribution of the local genomic sequence allowed one reporter to be expressed in photoreceptors, and another to be expressed in cone cells. This suggests that the additional regulatory sequences within *sparkling* provide the correct combination of activation and patterning information to allow for cone cell specific gene expression.

Indeed, we were able to further characterize the function of each region by analyzing its function in both the promoter distal and promoter proximal position. We found that region 4 is absolutely critical for cone cell specific gene expression in larval and pupal development regardless of position with respect to the promoter. Conversely, loss of regions 5 or 6 does not completely abolish reporter gene activity in the promoter proximal position. This observation led to the finding that neither region 5 or 6 is required for *sparkling* activity in the pupal eye disc at the distal or proximal position. This suggests that these sequences are required for initiation of *sparkling* activity in the correct time and place; however the enhancer can ultimately recover from their loss. In addition to its role in initiating *spa* activity, region 5 is also required to repress *spa* activity in photoreceptors.

As each of these sequences plays a unique role within the enhancer, we expect each is able to interact distinct transcription factors to facilitate their function. We plan to assess the ability of these regions to interact with protein and potentially identify the factors required for their action. Additionally, knowing the role of each of these sequences will further allow us to examine the structural and organization rules that govern the enhancer. To examine *sparkling* as enhanceosome vs information display model enhancer, we will make enhancer rearrangements and region substitutions.

2.4b Functional evidence for a special enhancer regulatory element, mediating remote interactions but not patterning

Enhancers are often located many kilobases from the promoters they regulate. Enhancer-promoter interactions over such distances are very likely to require active facilitation(Rippe, 2001). Even so, few studies have focused specifically on transcriptional activation at a distance, and the majority of this work involves locus control regions (LCRs) and/or complex multigenic loci, which are not part of the regulatory environment of most genes and enhancers (Carter et al., 2002; Chien et al., 2011; Li et al., 2006; Osborne et al., 2004; Tolhuis et al., 2002; Vakoc et al., 2005). Like *spa*, many developmental enhancers act at a distance in their normal genomic context, yet can autonomously drive a heterologous promoter in the proper expression pattern. However, in nearly all assays of enhancer function, the element to be studied is placed immediately upstream of the promoter. In such cases, regulatory sites specifically mediating remote interactions cannot be identified. Because our initial mutational analysis of *spa* was performed on enhancers placed at a moderate distance from the promoter (−846 bp), we were able to screen for sequences required *only* at a distance, by moving crippled enhancers to a promoter-proximal position.

Only one segment of *spa*, region 1, was absolutely essential at a distance but completely dispensable near the promoter. This region, which contains the only block of extended sequence conservation within *spa*, plays no apparent role in patterning or in basic activation at close range. We therefore call this segment of *spa* a “remote control” element (RCE). The remote enhancer regulatory activity described here differs from previously reported long-range regulatory mechanisms in two important ways. First, the remote function of *spa* does not require any sequences in or near the *dPax2* promoter. This functionally distinguishes *spa* from enhancers in the *Drosophila* Hox complexes that require promoter-proximal “tethering elements” and/or function by overcoming insulators (Akbari et al., 2008; Akbari et al., 2007; Calhoun et al., 2002). This distal activation mechanism also likely differs from enhancer-promoter interactions mediated by proteins that bind at both the enhancer and the promoter, as occurs in looping mediated by Zeste, CTCF, and Sp1 (Chien et al., 2011; Deshane et al., 2010; Hadjur et al., 2009; Kostyuchenko et al., 2009). Second, studies of distant enhancers of the *cut* and *Ultrabithorax* genes have revealed a role for the cohesion-associated factor Nipped-B, especially with respect to bypassing insulators (Misulovin et al., 2008) it has not been demonstrated that Nipped-B, or any other enhancer-binding regulator such as Chip (Morcillo et al., 1997; Morcillo et al., 1996), is required *only* when the enhancer is located distally.

To our knowledge, the *spa* RCE is the first enhancer sub-element demonstrated to be essential for enhancer-promoter interactions at a distance, but unnecessary for proximal enhancer function and cell-type specificity. However, the present work contains only a limited examination of this activity, as part of a broader study of enhancer function. We are currently extending these functional studies, testing for potential promoter preferences and distance limitations, and pursuing the identities of factors binding to the RCE as well as the mechanisms by which it acts to facilitate long-range gene transcription.

2.5 Experimental methods

2.5a Enhancer constructs

The 362-bp *sparkling* enhancer was amplified from w^{1118} genomic DNA with the following primers: 5′-CACCGGATCCgtatcaagtaactgggtgcctaattg-3′; 5′-GGGTCTAGAcctaagctaccggaaaacaacttg-3′. Lowercase sequence is homologous to genomic DNA. Most mutant *spa* constructs were generated by one of three PCR techniques: (1) amplification of *spa*(wt) with tagged primers to create mutations at the 5′ or 3′ end; (2) overlap extension (sewing) PCR to generate internal mutations; or (3) assembly PCR to synthesize enhancers with multiple mutations.

2.5b Mutagenesis by overlap extension PCR (Sewing PCR)

When targeting mutations in the interior of *spa*, we separately amplified 5′and 3′fragments, using overlapping tagged primers to integrate mutated sequence, and then joined the fragments using overlap extension (Swanson et al., 2008). In our sewing PCR protocol, the 5′ and 3′ fragments (which overlap by 20 bp) were separately PCR amplified and gel purified. We combined 3 µl of each gel purified fragment with 33.5 µl water, 1.5 µl of 10 µm dNTPs, and 5 µl 10X PCR buffer (Roche Expand High Fidelity PCR System). This mix was incubated at 90°C 10 min, then cooled one degree per min to 72°C. 1 µl of polymerase mix (Roche Expand High Fidelity PCR System) was then added, followed by incubation for 10′ at 72°C. Finally, 1.5 µl of each the flanking 5′ and 3′ primers (15 pmol each) was added and the full-length construct was amplified in our standard PCR program (94°C for 2′; 10 cycles of (94°C for 15″, 55°C for 30″, 72°C for 45″); 20 cycles of (94°C for 15″, 55°C for 30″, 72°C for 45″+5″/cycle); 72°C for 7′).

2.5c Assembly PCR

In constructs with extensive mutated sequence, constructs were built by annealing overlapping 40 bp oligonucleotides to create the full-length construct by assembly PCR (Swanson et al., 2008). We combined 2.5 μl of each flanking primer (10 μM), 1 μl internal primer mix (each primer at 0.25 μM), 1 μl of 10 μM dNTPs, and 18 μl sterile water in the template mix. The enzyme mix contained 19.25 μl sterile water, 5 μl 10X PCR buffer, and 0.75 μl DNA polymerase (Roche Expand High Fidelity PCR System). The template mix and enzyme mix were combined immediately before amplification in our standard PCR program (see above). In mutating previously uncharacterized enhancer sequences, we made non-complementary transversions to every other base pair. We left 2–4 bp of non-mutated sequence to either side of every TFBS (as defined by consensus sequences), to avoid interfering with TF binding.

2.5d Enhancer cloning, vectors, and transgenesis

PCR-amplified enhancer constructs were TOPO-cloned into the pENTR/D-TOPO vector (Invitrogen). Subcloned constructs were then Gateway-cloned into the Ganesh-G1 GFP reporter vector (Swanson et al., 2008) via LR recombination (Invitrogen), with the following exception: constructs placed at −121 bp from the promoter were Gateway-cloned into Ganesh-G2, which lacks the 0.7-kb spacer sequence between the recombination cloning site and the promoter (Swanson et al., 2008). P element transformation was performed essentially as described by Rubin and Spradling (1982). w^{1118} flies were used for transgenesis (Rubin and Spradling, 1982).

2.5e Tissue preparation, staining, and microscopy

Eye tissues were dissected from transgenic third-instar larvae or 24-hour pupae and fixed in 4% formaldehyde in PBS for 30 minutes at room temperature. For larval imaginal discs, GFP fluorescence was imaged with an Olympus BX51 microscope and an Olympus DP70 digital camera. Pupal eyes were stained with antibodies to GFP (see below) and imaged with an Olympus IX71 inverted microscope and an Olympus FV500 confocal system. Primary antibodies used: rabbit anti-EGFP (a gift from B. Novitch), diluted 1:100; mouse anti-Cut 2B10 (a gift from K. Cadigan), diluted 1:100; mouse anti-Elav 9F8A9 (Developmental Studies Hybridoma Bank), diluted 1:100.

2.6 Acknowledgments

Portions of this text were previously published (Swanson et al., 2010). This research was supported in part NIH grant GM076509.

2.7 References

Akbari, O.S., Bae, E., Johnsen, H., Villaluz, A., Wong, D., and Drewell, R.A. (2008). A novel promoter-tethering element regulates enhancer-driven gene expression at the bithorax complex in the Drosophila embryo. Development *135*, 123-131.

Akbari, O.S., Schiller, B.J., Goetz, S.E., Ho, M.C., Bae, E., and Drewell, R.A. (2007). The abdominal-B promoter tethering element mediates promoter-enhancer specificity at the Drosophila bithorax complex. Fly (Austin) *1*, 337-339.

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., and Shiroishi, T. (2009). Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. Dev Cell *16*, 47-57.

Bickmore, W.A., Mahy, N.L., and Chambeyron, S. (2004). Do higher-order chromatin structure and nuclear reorganization play a role in regulating Hox gene expression during development? Cold Spring Harb Symp Quant Biol *69*, 251-257.

Boyle, S., Gilchrist, S., Bridger, J.M., Mahy, N.L., Ellis, J.A., and Bickmore, W.A. (2001). The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. Hum Mol Genet *10*, 211-219.

Breen, J.J., Agulnick, A.D., Westphal, H., and Dawid, I.B. (1998). Interactions between LIM domains and the LIM domain-binding protein Ldb1. J Biol Chem *273*, 4712-4717.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev *25*, 1915-1927.

Calhoun, V.C., Stathopoulos, A., and Levine, M. (2002). Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex. Proc Natl Acad Sci U S A *99*, 9243-9247.

Carr, A., and Biggin, M.D. (1999). A comparison of in vivo and in vitro DNA-binding specificities suggests a new model for homeoprotein DNA binding in Drosophila embryos. Embo J *18*, 1598-1608.

Cremer, T., and Cremer, M. (2010). Chromosome territories. Cold Spring Harb Perspect Biol *2*, a003889.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A.*, et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A *107*, 21931-21936.

Desplan, C., Theis, J., and O'Farrell, P.H. (1988). The sequence specificity of homeodomain-DNA interaction. Cell *54*, 1081-1090.

Flores, G.V., Daga, A., Kalhor, H.R., and Banerjee, U. (1998). Lozenge is expressed in pluripotent precursor cells and patterns multiple cell types in the Drosophila eye through the control of cell-specific transcription factors. Development *125*, 3681-3687.

Flores, G.V., Duan, H., Yan, H., Nagaraj, R., Fu, W., Zou, Y., Noll, M., and Banerjee, U. (2000). Combinatorial signaling in the specification of unique cell fates. Cell *103*, 75-85.

Flynn, R.A., Almada, A.E., Zamudio, J.R., and Sharp, P.A. (2011). Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. Proc Natl Acad Sci U S A *108*, 10460-10465.

Franke, A., and Baker, B.S. (1999). The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in Drosophila. Mol Cell *4*, 117-122.

Fu, W., Duan, H., Frei, E., and Noll, M. (1998). *shaven* and *sparkling* are mutations in separate enhancers of the *Drosophila Pax2* homolog. Development *125*, 2943-2950.

Fu, W., and Noll, M. (1997). The *Pax2* homolog *sparkling* is required for development of cone and pigment cells in the *Drosophila* eye. Genes Dev *11*, 2066-2078.

Heard, E., and Disteche, C.M. (2006). Dosage compensation in mammals: fine-tuning the expression of the X chromosome. Genes Dev *20*, 1848-1867.

Jackson, D.A., Iborra, F.J., Manders, E.M., and Cook, P.R. (1998). Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. Mol Biol Cell *9*, 1523-1536.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermuller, J., Hofacker, I.L.*, et al.* (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science *316*, 1484-1488.

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T.*, et al.* (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature *471*, 480-485.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S.*, et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182-187.

Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet *12*, 1725-1735.

Liang, Z., and Biggin, M.D. (1998). Eve and ftz regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in Drosophila. Development *125*, 4471-4482.

Lower, K.M., Hughes, J.R., De Gobbi, M., Henderson, S., Viprakasit, V., Fisher, C., Goriely, A., Ayyub, H., Sloane-Stanley, J., Vernimmen, D.*, et al.* (2009). Adventitious changes in long-range gene expression caused by polymorphic structural variation and promoter competition. Proc Natl Acad Sci U S A *106*, 21771-21776.

Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. Nat Rev Genet *10*, 155-159.

Misulovin, Z., Schwartz, Y.B., Li, X.Y., Kahn, T.G., Gause, M., MacArthur, S., Fay, J.C., Eisen, M.B., Pirrotta, V., Biggin, M.D.*, et al.* (2008). Association of cohesin and Nipped-B with transcriptionally active regions of the Drosophila melanogaster genome. Chromosoma *117*, 89-102.

Morcillo, P., Rosen, C., Baylies, M.K., and Dorsett, D. (1997). Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in Drosophila. Genes Dev *11*, 2729-2740.

Morcillo, P., Rosen, C., and Dorsett, D. (1996). Genes regulating the remote wing margin enhancer in the Drosophila cut locus. Genetics *144*, 1143-1154.

Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H., and Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell *87*, 953-959.

Orom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q.*, et al.* (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell *143*, 46-58.

Osada, H., Grutz, G.G., Axelson, H., Forster, A., and Rabbitts, T.H. (1997). LIM-only protein Lmo2 forms a protein complex with erythroid transcription factor GATA-1. Leukemia *11 Suppl 3*, 307-312.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W.*, et al.* (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet *36*, 1065-1071.

Ponting, C.P., Oliver, P.L., and Reik, W. (2009). Evolution and functions of long noncoding RNAs. Cell *136*, 629-641.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Rippe, K. (2001). Making contacts on a nucleic acid polymer. Trends Biochem Sci *26*, 733-740.

Ruf, S., Symmons, O., Uslu, V.V., Dolle, D., Hot, C., Ettwiller, L., and Spitz, F. (2011). Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. Nat Genet *43*, 379-386.

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N.F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J.A., Umlauf, D., Dimitrova, D.S.*, et al.* (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet *42*, 53-61.

Shevelyov, Y.Y., Lavrov, S.A., Mikhaylova, L.M., Nurminsky, I.D., Kulathinal, R.J., Egorova, K.S., Rozovsky, Y.M., and Nurminsky, D.I. (2009). The B-type lamin is required for somatic repression of testis-specific gene clusters. Proc Natl Acad Sci U S A *106*, 3282-3287.

Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature *415*, 810-813.

Spitz, F., and Duboule, D. (2008). Global control regions and regulatory landscapes in vertebrate development and evolution. Adv Genet *61*, 175-205.

Swanson, C.I., Evans, N.C., and Barolo, S. (2010). Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. Dev Cell *18*, 359-370.

Tian, D., Sun, S., and Lee, J.T. (2010). The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. Cell *143*, 390-403.

Trievel, R.C., and Shilatifard, A. (2009). WDR5, a complexed protein. Nat Struct Mol Biol *16*, 678-680.

Voas, M.G., and Rebay, I. (2004). Signal integration during development: insights from the *Drosophila* eye. Dev Dyn *229*, 162-175.

Wadman, I.A., Osada, H., Grutz, G.G., Agulnick, A.D., Westphal, H., Forster, A., and Rabbitts, T.H. (1997). The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. Embo J *16*, 3145-3157.

Walter, J., and Biggin, M.D. (1996). DNA binding specificity of two homeodomain proteins in vitro and in Drosophila embryos. Proc Natl Acad Sci U S A *93*, 2680-2685.

Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A.*, et al.* (2011a). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature *474*, 390-394.

Wang, G., Balamotis, M.A., Stevens, J.L., Yamaguchi, Y., Handa, H., and Berk, A.J. (2005). Mediator requirement for both recruitment and postrecruitment steps in transcription initiation. Mol Cell *17*, 683-694.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A.*, et al.* (2011b). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120-124.

Wilson, K.L., and Berk, J.M. (2010). The nuclear envelope at a glance. J Cell Sci *123*, 1973-1978.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.C., Dufour, M.E., Boulay, J., Regnault, B., Devaux, F., Namane, A., Seraphin, B.*, et al.* (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell *121*, 725-737.

Zhu, X., Ling, J., Zhang, L., Pi, W., Wu, M., and Tuan, D. (2007). A facilitated tracking and transcription mechanism of long-range enhancer function. Nucleic Acids Res *35*, 5532-5544.

CHAPTER 3

STRUCTURE FUNCTION ANALYSIS OF THE *SPARKLING* REMOTE CONTROL ELEMENT

3.1 Abstract

Cis-regulartory sequences which regulate the levels and location of gene transcription known as enhancers can be located upstream, downstream, or even within the introns of the genes that they regulate. As enhancers often act at a great genomic distance from their target gene's promoter it is likely they possess intrinsic abilities in the form of specific DNA sequences which ensure they can act distally to activate the proper gene transcription. Accordingly, we previously discovered 40bp sequence within the *dPax2 sparkling* (*spa*) enhancer that is required only when *spa* was in the distal position, but is dispensable in the promoter proximal position. As this sequence is required for long-range gene regulation, we refer to it as the *sparkling* remote control element, or RCE. Further analysis of this unique enhancer sub-element has demonstrated that *spa* requires the RCE in order to activate gene transcription from a more moderated distance upstream target promoter than previously tested. Furthermore, *spa* is not sensitive to changes in RCE copy number. Interestingly, the RCE can be

moved a significant distance upstream of *spa* and still promote enhancer activity. We have also begun to characterize *spa's* ability to interact with different promoters and its possible core promoter preferences and hypothesis that the RCE may play an important role in targeting the enhancer to specific promoter elements.

STRUCTURE FUNCTION ANALYSIS OF THE *SPARKLING* REMOTE CONTROL ELEMENT

3.2 Introduction

Enhancers are *cis*-regulatory elements that control target gene expression in both a cell type specific and temporal manner. These sequences of non-coding DNA act by recruiting transcription factors, which in turn recruit cofactors that influence gene transcription through a variety of mechanism including recruiting chromatin remodeling complexes and the basal transcription machinery (Blackwood and Kadonaga, 1998; Bulger and Groudine, 2010; Levine, 2010). It is through recruiting the “correct” combination of transcription factors that a target gene’s pattern of expression is established. However, at the genomic level, most enhancers must also perform an additional crucial function. Enhancers can be located upstream, downstream, or even within the introns of the genes that they regulate. Therefore, enhancers often act at a great genomic distance from their target gene’s promoter, frequently activating a specific promoter with preference over other nearby promoters. Nuclear proteins influence the three-dimensional structure of DNA to facilitate these enhancer-promoter interactions. For example, nuclear laminins interact with inactive DNA regions, restricting them to the nuclear periphery, leaving active regions of DNA localized together towards the center of the nucleus (Wilson and Berk, 2010). Furthermore, insulator sequences act to restrict particular DNA sequences from interacting with each other such that enhancers can only interact with their target promoter (Dorsett, 1993, 1999). Yet, it is likely that enhancers also possess intrinsic abilities to ensure they can act distally to activate the proper gene transcription.

The ability of an enhancer to act over large genomic distances is a critical part of most enhancers' normal function and likely requires active facilitation through DNA-promoter interactions; however astonishing little is known about the mechanisms underlying enhancer-promoter interaction (Rippe, 2001). Current opinions in the field favor six main, non-exclusive models for how enhancers act long-range to regulate gene transcription. (1) direct enhancer-promoter chromatin looping; (2) linking: the formation of protein complexes along chromatin between an enhancer and the promoter; (3) tracking: the recruitment of RNA poll at the enhancer, followed occasionally by intergenetic transcription to the target gene's promoter; (4) facilitated tracking combines the tracking and looping models; (5) change in subnuclear localization to either a transcriptional active location, transcription factor, or away from the nuclear periphery; (6) the formation of non-coding RNAs that stimulate transcription (Li et al., 2006; Osborne et al., 2004). There is good experimental evidence to support some of these models. Several groups have shown chromatin looping through techniques such as chromosome conformation capture. For example, long range physical interactions have been observed in the *TNF* and β-globin loci (Tolhuis et al., 2002; Tsytsykova et al., 2007). Evidence for tracking stems from data suggesting that non-coding RNAs are transcribed between some enhancers and promoters, such as at the human ε- and β-globin loci, and *Drosophila* biothorax complex (Bae et al., 2002; Gribnau et al., 2000; Zhu et al., 2007). Furthermore, chromatin-immunoprecipitation has revealed the presence of Pol II along DNA between several enhancers and promoters, such as the human β-globin and *prostate specific antigen* loci (Johnson et al., 2001; Wang et al., 2005). Comparatively, there is less experiment evidence for changes in subnuclear localization, and linking; however Osborne and colleges showed that active genes are recruited to distinct locations within the nuclei called transcription factories, and Morcillo et al. describe a protein's complex containing Chip that links the cut wing margin enhancer and its promoter in *Drosophila* (Morcillo et al., 1996; Osborne et al., 2004).

There are also surprisingly few examples of specific DNA elements known to be involved in long range transcriptional regulation. These involve tethering elements specific to a promoter, or insulator bypass elements, such as those in the *AbdB*, *white*, *biothorax*, and *Cut* loci (Akbari et al., 2008; Akbari et al., 2007; Calhoun et al., 2002; Kostyuchenko et al., 2009; Laney and Biggin, 1997; Misulovin et al., 2008). No study had identified a region within an enhancer that mediates general transcription at a distance until the discovery of the *sparkling* "remote control" element (RCE) (Swanson et al., 2010).

The *sparking* (*spa*) enhancer regulates expression of *dPax2* in the developing eye imaginal disc. In third instar larvae, the eye disc contains two differentiated cell types: 8 photoreceptors and 4 cone cells per ommatidium. By the 24 hours pupal stage a third cell type is specified in each ommatidium: two primary pigment cells (Voas and Rebay, 2004). *spa* is required to activate *dPax2* expression in cone cells, while ensuring it is not expressed in photoreceptors

(Flores et al., 2000; Fu et al., 1998; Fu and Noll, 1997; Swanson et al., 2010; Swanson et al., 2011). *dPax2* is also expressed in primary pigment cell; however, these regulatory sequences have yet to be identified (Fu et al., 1998; Fu and Noll, 1997). *sparkling* lies in the 4th intron of the *dPax2* gene, where it is located 7kb downstream from the *dPax2* transcriptional start site (Flores et al., 2000; Fu et al., 1998). Therefore, *spa* must be able to act at a distance to regulate gene transcription in cone cells. In order to determine what part of this enhancer is responsible for its long-range capabilities, we analyzed *spa's* ability to drive reporter gene (GFP) expression from a moderate distance of 846 bp from the transcription start sight (TSS) as well as in the more promoter proximal position of 121 bp upstream of the TSS. By comparing mutations made to the enhancer at both distances we are able to identify regions that were only required at a distance compared to sequences required only for activation and patterning information, which are required at both distances. Indeed, we found a 40bp sequence that was required only when *spa* was in the distal position, but was dispensable in the promoter proximal position (Swanson et al., 2010). As this sequence is required for long-range gene regulation, we refer to it as the *sparkling* remote control element, or RCE.

Having identified a unique enhancer sequence responsible for distal enhancer activity, we sought to further characterize this subunit by generating additional reporter constructs to test the functional properties and limitations of the RCE. This work has shown that *spa* requires the RCE in order to activate gene transcription from 605bp upstream of the TSS. Furthermore, *spa* is not sensitive to changes in RCE copy number. Interestingly, the RCE can be moved a significant distance upstream of *spa* and still promote enhancer activity. We have also begun to characterize *spa's* ability to interact with different promoters and its possible core promoter preferences.

3.3 Results

3.3a Analysis of sparkling in multiple insertion contexts

Our analysis of *spa* to this point was performed with the enhancer cloned into the Ganesh vector backbone (Evans et al., 2012; Johnson et al., 2008; Swanson et al., 2010; Swanson et al., 2011). Using these Gateway vectors, the reporter constructs are integrated into the *Drosophila* genome pseudorandomly utilizing P-element transgenesis (Bellen et al., 2004; Rubin and Spradling, 1982; Swanson et al., 2008). We have identified a sequence with *sparkling* that is necessary for its long range functions, which we refer to as the RCE (Swanson et al., 2010). In order to continue our analysis of the RCE we decided to utilize Φ3C1 mediated site-specific integration into the genome (Groth et al., 2004; Thorpe and Smith, 1998). Using these genomic "landing sites" we can control the location in which every reporter construct is inserted into the genome. This method of integration is designed to reduce reporter construct variation from one random insertion site to another. Therefore, integration of reporter constructs at a single known location should allow us to detect subtle changes in enhancer

Figure 3.1

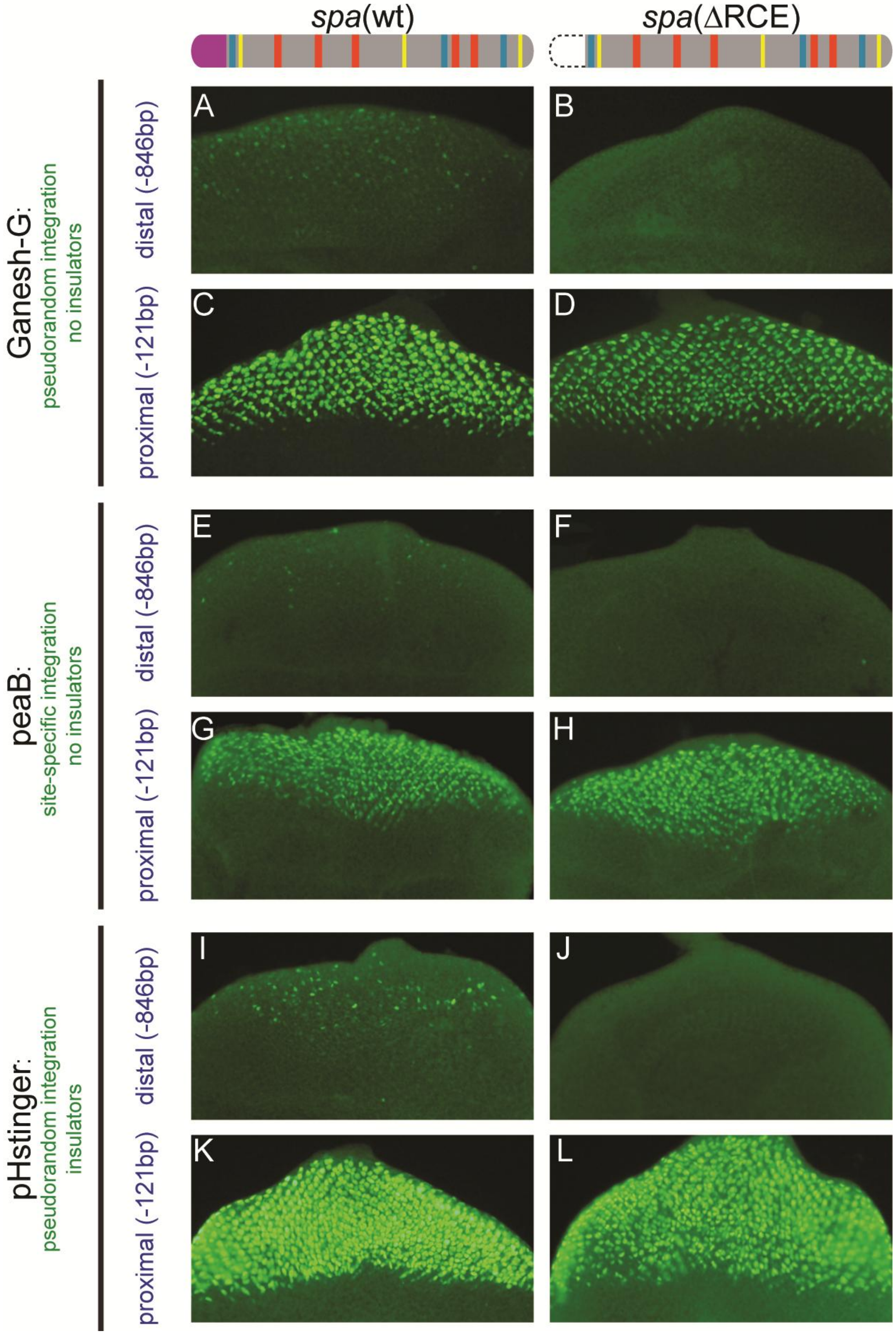


Figure 3.1 The *sparkling* remote control element is required in the distal position. The original

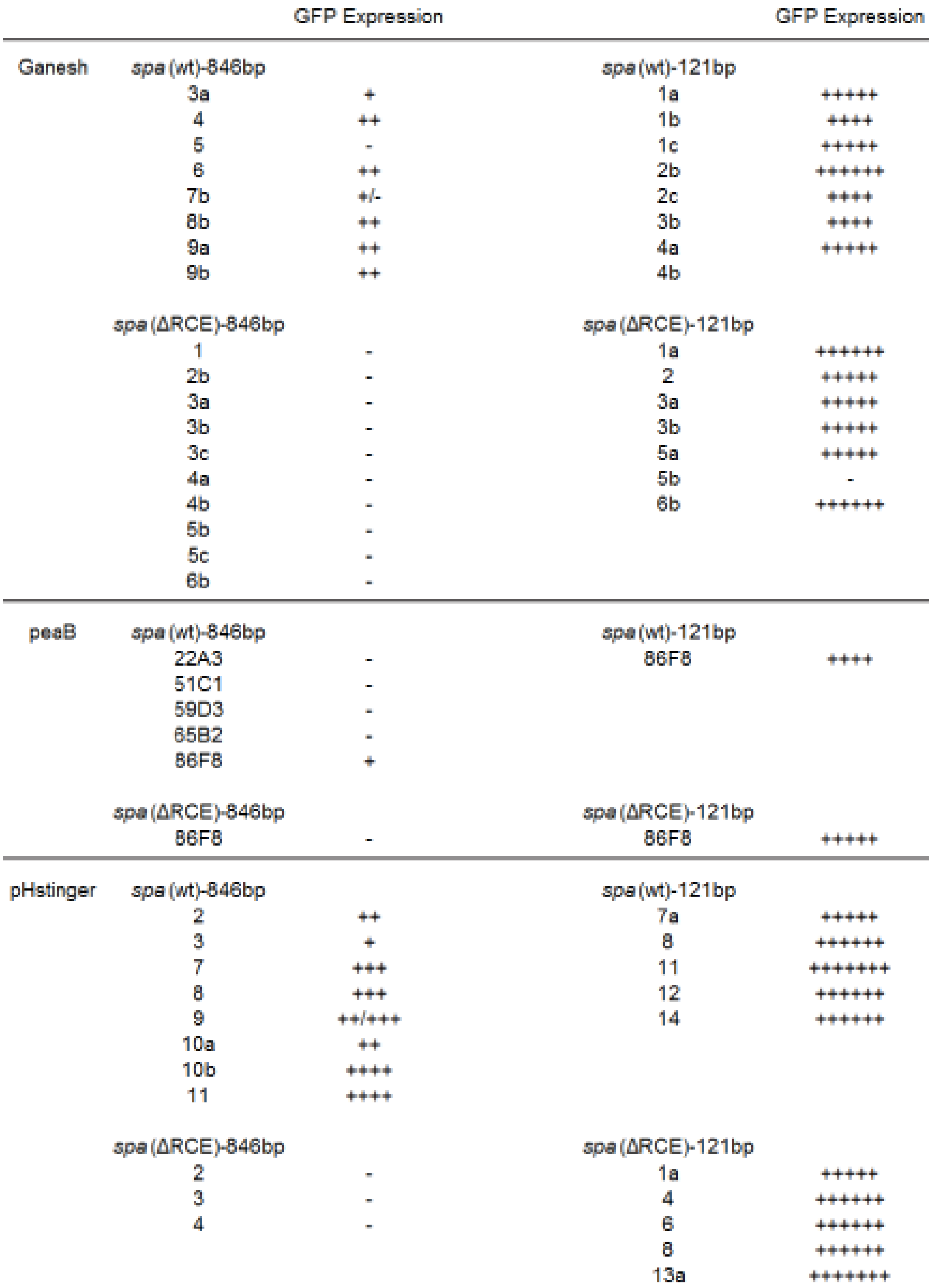
spa analysis was performed with *spa* constructs cloned into the Ganesh P-element reporter vectors (A – D). Here we see that the RCE is required at -846 bp but not at -121 bp (B, D) Upon integration into the 86F8 attP landing site, *spa*(wt)-846bp has similar levels of expression as seen previously; however, the pattern is not complete (E). *spa*(wt)-121 driven gene expression is comparable to that seen previously (G). In this landing site the RCE was again required only at a distance and dispensable in the promoter proximal position (F and H). When *spa*(wt)-846 was cloned into the insulated P-element vector, pHstinger, the average GFP expression is higher than that seen previously (H). The RCE remains essential only in the distal position (J and L).

activity. We are especially interested in this ability confidently detect small changes in gene expression as we previously noted that GFP expression driven by *spa*(ΔRCE)-121bp is on average greater than that driven by *spa*(wt)-121bp (Swanson et al., 2010). This finding has significant implications on the method of RCE function; however, it does not occur in every insertion location we identified. Therefore, the ability to analyze these two constructs in the same genomic location is ideal.

To this end, we generated a vector system for studying reporter genes in landing sites. This reporter vector, peGFPattB (peaB), contains an hsp70 promoter driving expression of a nuclear localized GFP, the transgenic selector gene *white*, and an attB site. Upon exposure to Φ3C1 integrase, the attB site recombines with an attP site which is inserted at a single known location in the *Drosophila* genome. In this study, we utilized an attP landing site on the 3rd chromosome at the cytological location 86F8. We first generated and integrated the distally placed wildtype *sparkling* enhancer, which is positioned, 846 bp from the reporter genes transcriptional start site (TSS). We were surprised to find *spa* activity differed slightly in this landing site from that previously observed. While levels of expression are comparable between the average randomly inserted levels and this site specific integration, the pattern is erratic and disrupted (Figure 3.1 A, E). Postulating this limited pattern may be due to the insertion site used, we integrated this same reporter construct into additional landing sites. However, in four independent landing sites, *spa* did not drive any GFP expression (Table 3.1). Our laboratory has previously established a clear pattern of spa(wt)-846 expression activity (Swanson et al., 2010). Using the Ganesh vector to randomly integrate the wildtype reporter construct, we found that five out of eight insertions drove nearly identical patterns and levels of gene expression. Furthermore, only in one insertion site was *spa* incapable of driving any expression (Table 3.1). These results are in stark contrast to *spa*(wt)-846 activity when integrated into attP landing sites, where only one out of five lines had any activity (Figure 3.3 E). Despite the incomplete pattern driven by *spa* in the 86F8 landing site, antibody staining for GFP and the cone cell marker Cut demonstrates that expression is at least cone cell specific (data not shown).

Given the activity of *spa*-846 bp in 86F8, we were somewhat surprised to find that when *spa* was moved to the promoter proximal position of 121 bp from the TSS, the enhancer drives GFP expression at levels comparable to those seen previously in our random insertion lines (Figure 3.1 C, G). This suggests that the limited activity of *spa* from the 86F8 location is restricted to its function at a distance. One hypothesis to explain these results is that *sparkling,* possibly though the RCE, is able to interact with multiple promoters and not simply the

Table 3.1

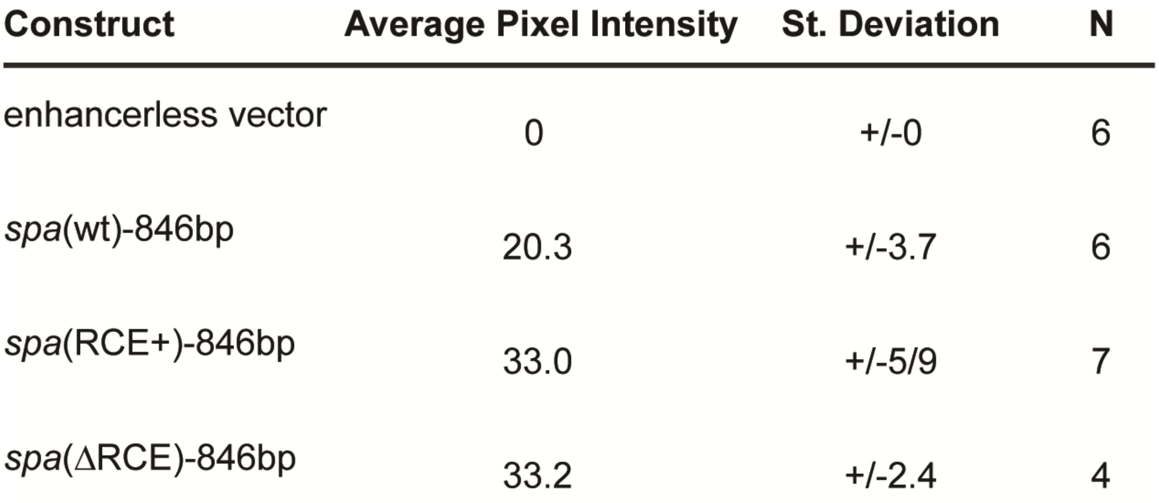


		GFP Expression		GFP Expression
Ganesh	*spa* (wt)-846bp		*spa* (wt)-121bp	
	3a	+	1a	+++++
	4	++	1b	++++
	5	-	1c	+++++
	6	++	2b	++++++
	7b	+/-	2c	++++
	8b	++	3b	++++
	9a	++	4a	+++++
	9b	++	4b	
	spa (ΔRCE)-846bp		*spa* (ΔRCE)-121bp	
	1	-	1a	++++++
	2b	-	2	+++++
	3a	-	3a	+++++
	3b	-	3b	+++++
	3c	-	5a	+++++
	4a	-	5b	-
	4b	-	6b	++++++
	5b	-		
	5c	-		
	6b	-		
peaB	*spa* (wt)-846bp		*spa* (wt)-121bp	
	22A3	-	86F8	++++
	51C1	-		
	59D3	-		
	65B2	-		
	86F8	+		
	spa (ΔRCE)-846bp		*spa* (ΔRCE)-121bp	
	86F8	-	86F8	+++++
pHstinger	*spa* (wt)-846bp		*spa* (wt)-121bp	
	2	++	7a	+++++
	3	+	8	++++++
	7	+++	11	+++++++
	8	+++	12	++++++
	9	++/+++	14	++++++
	10a	++		
	10b	++++		
	11	++++		
	spa (ΔRCE)-846bp		*spa* (ΔRCE)-121bp	
	2	-	1a	+++++
	3	-	4	++++++
	4	-	6	++++++
			8	++++++
			13a	+++++++

Table 3.1 Qualitative representation of GFP expression levels in *sparkling* reporter constructs.

most proximal one. This explanation is not unreasonable as in our reporter constructs *spa* acts from upstream to regulate the TATA containing hsp70 promoter; whereas, in its endogenous location it acts from downstream to regulate the TATA-less *dPax2* promoter. The 86F8 insertion site is comparably promoter rich. The peaB integration vector contains the transgenic selector gene, *white*, and a 3xP3-RFP is integrated adjacent to the attP site to mark the landing site. Furthermore, the 86F8 landing site lies in the second intron of the *Chlorine channel a* gene, which puts it 5, 4, and 2kb away from the genes primary and secondary promoters; all notably closer than the *dPax2* promoter is to the endogenous *spa* enhancer. It is possible then that distal *spa* activity is limited in this landing site because the enhancer also interacts with these nearby promoters. Accordingly, *white* and RFP expression are both increased upon integration of *spa*(wt)-846bp compared to integration of an enhancerless, or empty peaB. *spa*(wt)-121bp would not be as dramatically affected if, due to its location, the enhancer interacted primarily with the hsp70 promoter. Alternately, the relatively weak GFP expression driven by *spa* from -846, even in randomly inserted constructs, could be more susceptible to a non-permissive transcriptional environment than the stronger enhancer *spa*(wt)-121 bp. *spa*(wt)-846bp's susceptibility to non-permissive transcriptional environments could also explain the complete lack of activity in four additional landing sites.

Table 3.2



Construct	Average Pixel Intensity	St. Deviation	N
enhancerless vector	0	+/-0	6
spa(wt)-846bp	20.3	+/-3.7	6
spa(RCE+)-846bp	33.0	+/-5/9	7
spa(∆RCE)-846bp	33.2	+/-2.4	4

Table 3.2 GFP expression levels in third instar imaginal discs. We quantified the GFP expression in 886F8 reporter constructs. Pixel intensity was measured posterior to the morphogeneic furrow, normalized to the area of the disc, and averaged across N samples. *spa*(ΔRCE)-846bp drives significantly greater (p=0.015) GFP expression than *spa*(wt)-846bp. *spa*(RCE+)-846bp will be discussed in Chapter 6.

Next we analyzed loss of the RCE in these landing sites. As *spa*(wt)-846 was only active in a single landing site, we continued to use 86F8 for this study. We found once again that loss of the RCE results in a complete absence of GFP expression when the distal enhancer is integrated at 86F8 (Figure 3.1 D). However, when *spa* lacking the RCE is moved to the promoter proximal position, (-121 bp), GFP expression is completely recovered to wildtype levels (Figure 3.1 F). Therefore, despite the limited expression pattern driven by *spa* in this insertion site, the function of the RCE as a long-range transcriptional facilitator remains consistent. This provides further support not only for the existence of the RCE, but also that *spa* does function in this landing site; however, it is just susceptible to the genomic environment of the insertion.

Recall that one of the reasons we switched to using site-specific integration in landing sites to study the RCE was to address the level of GFP expression driven by *spa*(wt)-121bp and *spa*(ΔRCE)-121bp. With the two constructs integrated into the same location we again see an increase in expression from *spa*(wt) to *spa*(ΔRCE). We subsequently quantified GFP levels in these eye discs and found that *spa*(ΔRCE)-121bp drives 50% more GFP expression than *spa*(wt) in the same position (Table 3.2). Again this observation can be explained in two ways. Perhaps the simplest explanation is that the RCE sequence contains a short range repressor site in addition to its long range abilities, and loss of the RCE relieves this repression. A second explanation invokes the hypothesis that *sparkling* can interact with additional promoters. If this capability is facilitated by the RCE, we would anticipate that even in the presence of the proximal hsp70 promoter, *spa* spends some amount of time locating and attempting regulate other promoters. Meanwhile, without the RCE, distal *spa* is unable to interact with any promoter resulting in no GFP expression, and the proximal *spa* can only interact with the hsp70 promoter resulting in higher levels of GFP expression.

As we saw strikingly different expression profiles between the Ganesh mediated randomly inserted reporter genes and those integrated at six individual attP sites, we decided to also analyze *spa* activity in yet another insertion context. We cloned our *spa* constructs into the pHstinger reporter construct (Barolo, 2000). This vector has the same transgenic marker, *white,* and hsp70 promoter driving nuclear GFP as Ganesh and peaB; in fact, both Ganesh and peaB were built from pHstinger (Swanson et al., 2008). Like with Ganesh, enhancers ligated into pHstinger are integrated into the *Drosophila* genome pseudorandomly via P-transposase. However, the enhancer, hsp70 promoter, and GFP coding sequence are flanked by gypsy insulators. The presence of these insulators sequences should inhibit genomic sequences in the insertion site from interacting with *spa* or hsp70 regulatory sequences and vise versa (Barolo, 2000). This should minimize the line-by-line variation seen with Ganesh insertions while allowing us to analyze *spa* activity in even more locations. Remarkably, despite our anticipation of less variation, we actually found that *spa*(wt)-846bp regulated GFP expression varies from line-to-line more than when the reporter construct is uninsulated (Table 3.1).

Despite the variability we again observed only one location with extremely low expression (Table 3.1). Additionally, the average expression driven by *spa*(wt)-846 bp in pHstinger is higher than that seen previously (Figure 3.1 A, H). This supports the hypothesis that *spa* can interact with additional promoters in the insertion site other than hsp70. As the insulators would minimize these interactions, GFP expression driven by *spa* would be brighter than an uninsulated *spa* reporter construct. The increased levels seen here could also explain the variability between insertion sites as it is possible this variation was simply below our ability to detect previously, given the relatively weak nature of uninsulated *spa*(wt)-846bp activity.

RCE activity in the insulated vectors is similar to that seen in Ganesh and peaB mediated integration (Figure 3.1 C, G, K). Once again we see that loss of the RCE results in absence of gene expression when *spa* is in the promoter distal position (Figure 3.1 J). This loss is abolished when *spa* lacking the RCE is moved near the promoter (-121bp) where it recapitulates the wildtype expression pattern (Figure 3.1 L). Interestingly, for the first time, *spa*(wt)-121bp and *spa*(ΔRCE)-121bp expression levels were consistently the same across all lines analyzed (Table 3.1). These levels are also distinctly greater than the wildtype levels in Ganesh or peaB mediated integration events. These observations provide even more support for the hypothesis that *spa* interacts with multiple, and not simply the closest, promoters. The insulators inhibit *spa* from interacting with promoters other than hsp70, raising the level of GFP expression overall and rendering the wildtype and ΔRCE enhancers equivalent. These results also make the alternative explanation of a repressor binding site in the RCE unlikely. Regardless of the genomic insertion context and wildtype levels of expression, the RCE is required to mediate only the long-range function of the *spa* enhancer.

3.3b The effect of position and copy number on RCE activity

Despite the limited expression pattern driven by *sparking* in the 86F8 landing site, it clearly exhibits a *spa* like mode of function including requiring the RCE for distal gene regulation. As such, we were able to address a few of our questions about the RCE's functional properties in this insertion location. One of the most obvious questions pertaining to *spa* activity is; at what distance does *spa* require a DNA element to facilitate long-range transcriptional regulation? We know that *spa* requires the RCE when it is 846 bp from the reporter gene's TSS. However, it is no longer necessary when the enhancer is placed 121bp from the TSS (Figure 3.1) (Swanson et al., 2010). *spa* also requires the RCE from the position of 605 bp from the transcriptional start site (Figure 3.2 B). This supports the hypothesis that the RCE interacts with a specific protein or protein complex if the RCE functions by looping. Work studying the flexibility of DNA suggests that two points on DNA can contact each other freely when the distance between them is between 100 and 300 bp; however any distance greater than 300 bp requires active facilitation (ie: through interaction with proteins) (Rippe, 2001). It is unsurprising then that the RCE sequence is required when *spa* is 605bp away from the point it needs to contact.

Figure 3.2

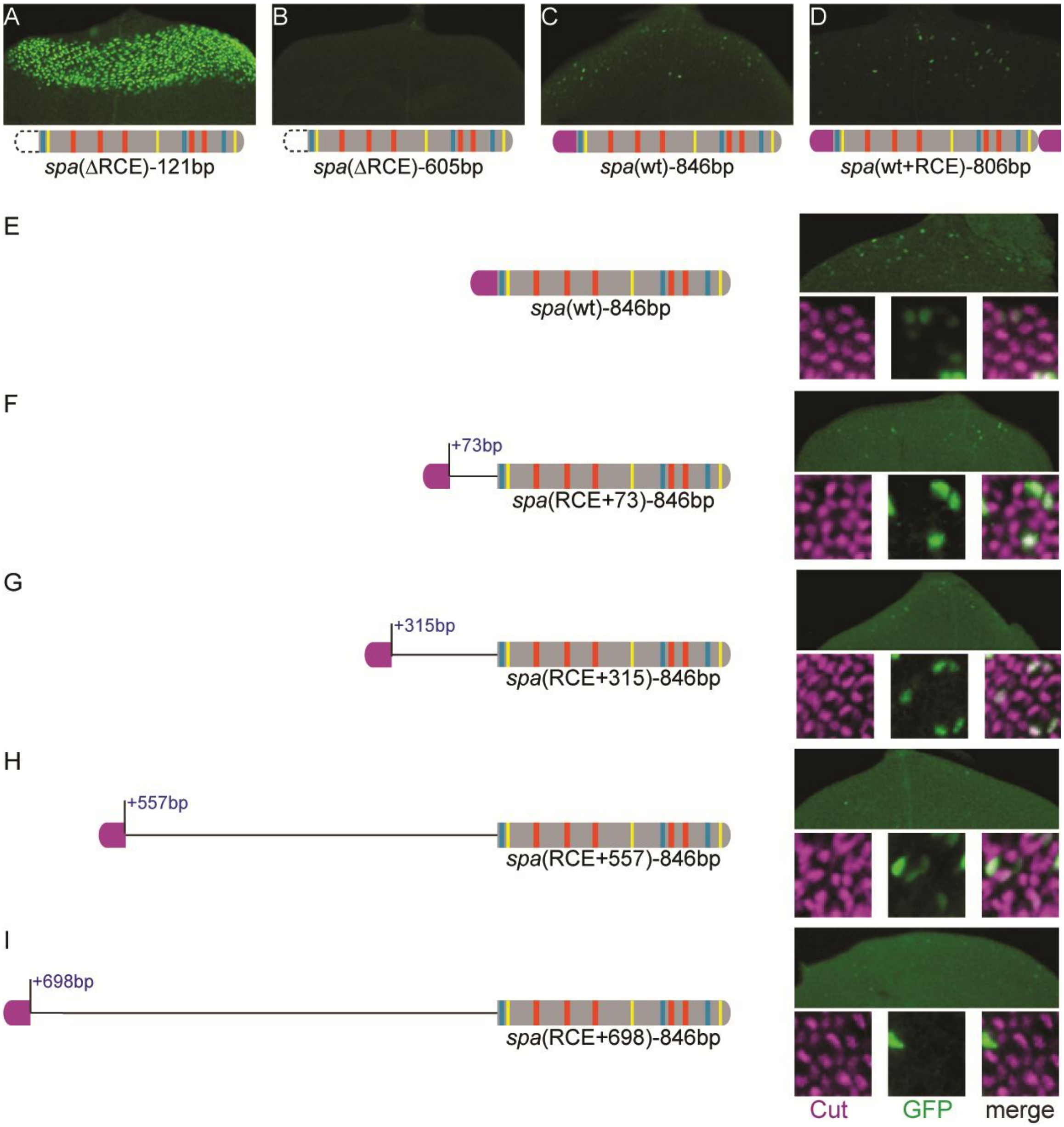


Figure 3.2 Functional analysis of the *sparkling* RCE. *spa* requires the RCE when it is 846 bp from the TSS, but not at 12 bp from the TSS (A). We now know that *spa* also requires the RCE when it is placed 605 bp from the TSS (B). When *spa* is in the distal position a second copy of the RCE placed 3' to the enhancer does not affect reporter gene transcription (C and D). If the RCE is moved 73 bp upstream of the *spa* enhancer placed at -846 bp from the TSS, we observe a slight decrease in gene GFP expression (E and F). Similarly, as the RCE is moved 315, 557, and 698 bp upstream of *spa*, GFP expression only decreased minimally (G – I). This expression remains cone cell specific as seen by colabeling at GFP with Cut; a cone cell marker (E – I).

We also examined the effect of RCE copy number on *spa* activity. Previous work demonstrated that the RCE can perform its long-range function when it is placed at the 3' end of the enhancer rather than the 5' end of the enhancer (Swanson et al., 2010). Therefore, we decided to place a copy of the RCE at both the 5' and 3' ends of the enhancer. When *spa* possesses two copies of the RCE, we do not see a discernible difference in GFP levels or pattern compared to wildtype levels of expression (Figure 3.2 C, D). Some proteins such as Zeste are capable of binding specific sequences within an enhancer (ie: of *ultrabiothorax* and *white*), and also at the gene's promoter (Kostyuchenko et al., 2009; Laney and Biggin, 1997; Mohrmann et al., 2002; Qian et al., 1992). The proteins bind at each location then interact with each other to form a complex that forces DNA to loop between these sites. The ability of these proteins to homo-oligermize or to form multiprotein complexes loop DNA is a prominent model for how enhancers act at a distance (Li et al., 2006). This mechanism of action is unlikely for *spa* as in its enogenous locus it interacts with the *dPax2* promoter, and in our reporter constructs it regulates the hsp70 promoters. It is unlikely that the promoters contain the same sequence that would allow this mechanism; however it should be noted that both the RCE and the 846 bp spacer sequence contain putative Zeste binging sites. If *spa* were inducing DNA looping through the same sequence at both the enhancer and promoter, we would expect that the presence of a second RCE downstream of the second copy could induce the formation of exclusionary DNA loops that would restrict gene activation. Not only is this not the case, but the second copy does not boost gene transcription either (Figure 3.2 D). This indicates that, at least in this location, a second copy of the RCE provides no additional aid in long-range function. Furthermore, this could mean that if *spa* is working by forming DNA loops, it is likely not working through a homo-oligomerzation mechanism.

When studying enhancers, researchers commonly strive to identify the minimal DNA sequence required to recapitulate target gene expression. However, it is important to remember that correct identification of patterning information is insufficient for most endogenous enhancers, which must find a way to regulate a specific, distally located promoter. As most “minimal enhancers” are only examined in the promoter proximal context, we do not know if these enhancers contain sequences involved in distal gene regulation or not. We decided to test the RCE’s ability to function at various distances upstream of the rest of *sparkling* with the enhancer remaining 846 bp from the GFP transcriptional start site. This will address the question of where does a DNA sequence that facilitates long-range enhancer activity need to be with respect to the patterning information of the enhancer. This may also help with *in silico* searches for enhancers, as well as the identification of “RCEs” for known minimal enhancers. When we move the RCE only 73 bp upstream of *spa*, we observed a slight decrease in GFP expression compared to wildtype (Figure 3.2 E and F). When the RCE is subsequently moved further from *spa*, 315, 557, and 698 bp upstream, GFP expression remains, with only a slight decrease in expression levels and number of cells expressing GFP (Figure 3.2 G-I). Remarkably, *spa* itself cannot function to drive gene expression without the RCE when it is 605 bp upstream from the TSS (Figure 3.2 B), yet the RCE can function to drive enhancer activity even when it is 698 bp upstream from the promoter. Due to the sporadic nature of GFP expression, we were uncertain of the cell type in which this expression occurs; cone cells are typically easy to identify as they form distinctive groups of four that resemble rosettes. However, colabeling with Cut demonstrates that this activity is indeed cone cell specific (Figure 3.2 E-I).

Given the already limited expression driven by wildtype *sparkling* in 86F8, the continued expression we see when the RCE is moved away from the enhancer suggest that in other enhancers, long-range facilitating sequences may lie a considerable distance outside the patterning information. The ability of the RCE to act upstream of *spa* is consistent with *spa* acting through a tracking mechanism, recruiting the basal transcription machinery and initiating transcription. This mechanism is unlikely however, as in its endogenous location, *spa* is downstream of its target promoter, but in our reporter constructs it is upstream of the target promoter. In order to function via tracking, the Pol II transcription initiated by the RCE would have to be bidirectional.

3.3c DNA sequences other than the RCE can convey long-range activity on sparkling

We have on occasion substituted the RCE sequence within *spa* for other DNA sequences. One such sequence, which will be discussed in greater detail

Figure 3.3

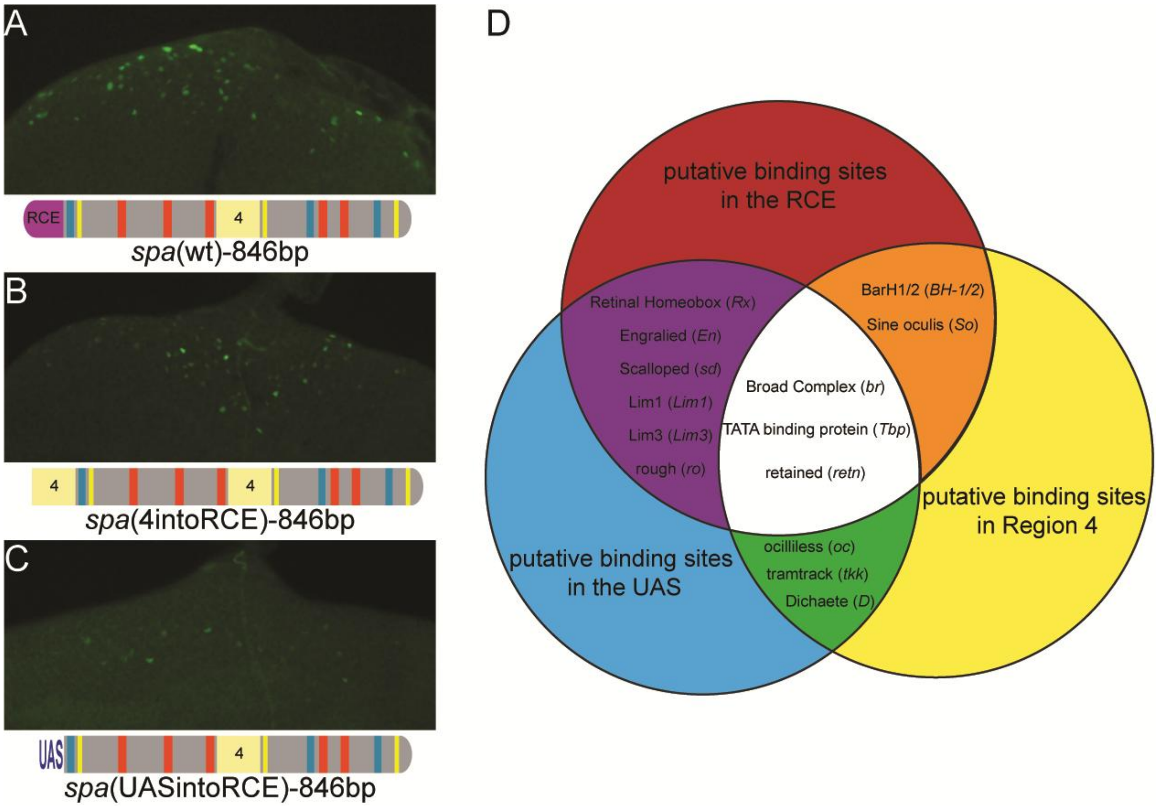


Figure 3.3 Additional DNA sequences are capable of promoting RCE activity. We have seen that the RCE DNA sequence enables *spa* to function at the distal position of -846 bp (A). At this same position, the RCE sequence can be exchanged for either *spa* Region 4 (B) or an upstream activator sequence (UAS) (C) and *spa* is still able to drive GFP expression. Using evoprinter to compare the RCE, Region 4, and RCE sequence, we generated a list of proteins that could putatively interact with two or three of these sequences (D).

in later chapters, is *sparking* Region 4. The sequence within *spa* Region 4 is absolutely required for robust cone cell specific gene expression regardless of position with respect to promoter (Swanson et al., 2010). Interestingly, when Region 4 is substituted for the RCE, such that *spa* contains two copies of Region 4, we observe wildtype levels of gene activity at a distance (Figure 3.3 A, B). This indicates that Region 4 has RCE activity, but only when it is present in two copies, as a single copy is insufficient to compensate for loss of the RCE when it is deleted from *spa* at -846 bp.

While it was surprising to us that a sequence we previously thought contributed only to patterning information is capable of providing RCE activity, at least this sequence is from a known, distally regulated *Drosophila* enhancer. Astonishingly, when a UAS site was placed 5' of *sparkling* in the place of the RCE, we observed GFP expression in the *Drosophila* eye imaginal disc (Figure 3.3C). A UAS, or upstream activating sequence, is the DNA binding site for the yeast transcriptional activator Gal4. UAS binding sites are found in promoter regions in the yeast genome (Struhl, 1987) so they would not traditionally be suspected for long-range activity. Gal4 is not expressed in any organism other than yeast, so it cannot be functioning to regulate *sparkling* in *Drosophila*. This suggests that a *Drosophila* protein is able to interact with a sequence within the UAS site and stimulate *sparkling*'s long-range activity.

We now know that the RCE, *spa* Region 4, and a UAS site can all facilitate promoter distal gene transcription in conjunction with the *sparkling* enhancer. Hypothesizing that they do so by interacting with the same protein, or related protein, we used evoprinter to compare the RCE, Region 4, and UAS DNA sequences for common motifs. From this analysis, we generated a list of proteins whose putive binding sites lie in two or three of these sequences (Figure 3.3 D). Some of the proteins are good candidates for regulating *spa* activity based on their know expression patterns in eye discs such as, tramtrack, scalloped, rough, and Sine oculis (Campbell et al., 1992; Halder et al., 1998; Voas and Rebay, 2004). Although none of these proteins posses known functions that directly implicate them in long-range activity. We also found putative binding sites for known transcriptional repressors engrailed and Bar H1/2, which could be modulating *spa* activity (Jaynes and O'Farrell, 1991; Laughon, 1991). Finally, we identified proteins, which based on know molecular function, could facilitate RCE activity. For example, Dichaete, which is expressed in the eye disc, although not in ommatidial cells, has been shown to promote DNA bending (Pil et al., 1993). The Lim proteins have been shown to form complexes on chromatin involved in both looping and linking (Morcillo et al., 1996). The presence of TATA binding protein sites in sequences associated with RCE function is consistent with a tracking mechanism of RCE activity (Blackwood and Kadonaga, 1998). Together, the discovery of new DNA sequences that promote long-range gene regulation has opened the door to new avenues of experimentation and candidate proteins to test for RCE binding potential.

3.3d   Analysis of sparkling’s promoter preferences

Figure 3.4

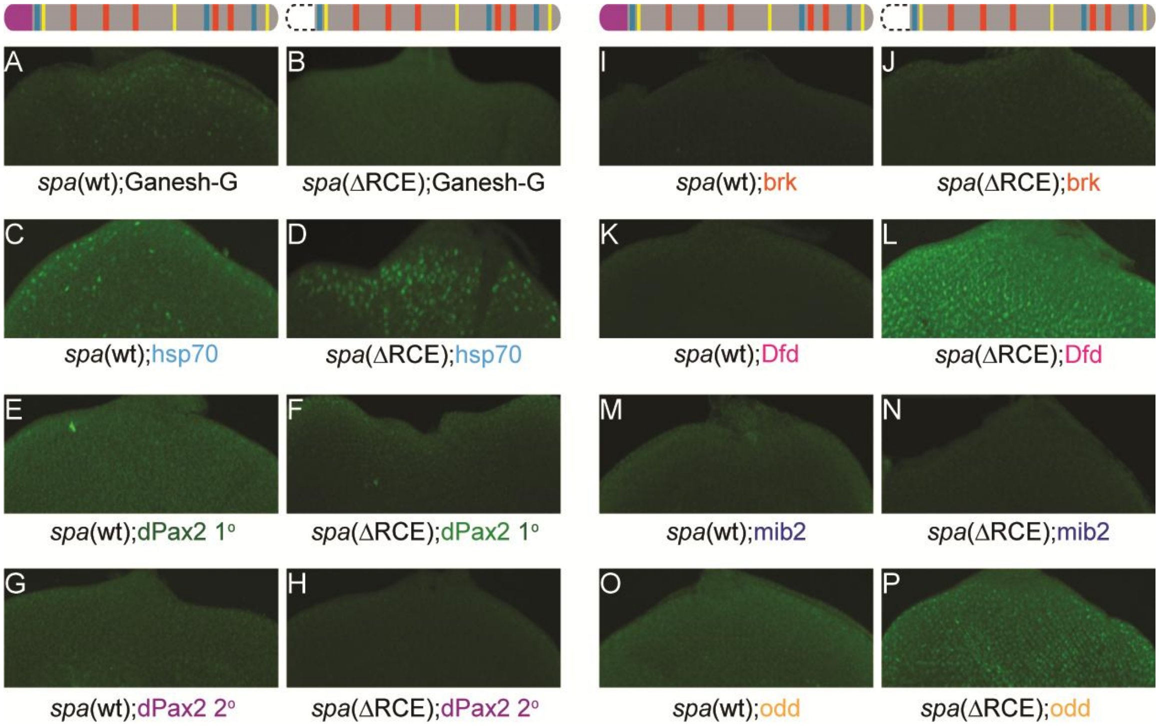


Figure 3.4 s*parkling* lacking the RCE is capable of driving gene expression from the distal position in the attP2 landing site. We know that *spa*(wt)-846bp drives reporter gene expression in cone cells (A and B). When similar constructs driving expression from the hsp70 promoter are inserted at the attP2 locus, we observed GFP expression in both constructs (C,D). The wildtype *spa* enhancer did not drive GFP expression from any other promoter (E, G, I, K, M, O). Similarly, *spa*(ΔRCE) does not drive expression from the primary or secondary *dPax2* promoters, or from the *mib2* promoter (F, H, N). However, *spa*(ΔRCE) stimulated gene transcription at high levels from the *Dfd* promoter (L) and low levels from the *brk* promoter (J) in non-cone cell types. *spa*(ΔRCE) also promotes GFP expression from the *odd* promoter in cone cells and undifferentiated cells extending to the morphogenic furrow (P).

We have long postulated that *spa* can interact with multiple different promoters. Case in point, we know it can activate transcription from both the hsp70 and native *dPax2* promoter. Circumstantial evidence also suggests wildtype *spa* can interact with local promoters in the context of the reporter construct insertion site. All distally located enhancers must interact with their target gene's promoter, sometimes with preference over other, more closely located promoters suggesting strict regulation of promoter- enhancer communication. Yet many enhancers, including *sparkling* can work with heterologous promoters like the hsp70 promoter, indicating an enhancer's promoter preference is relatively flexible. In order to investigate these contradictory observations and the potential promoter preferences of *sparkling*, we began collaboration with Marc Halfon's laboratory at State University of New York (SUNY, Buffalo). In this study, the wildtype *sparkling* and *spa* lacking the RCE were placed 846 base pairs upstream of various promoters driving GFP. First, a simple reporter construct to those used previously with the enhancer placed upstream of the hsp70 promoter was generated. Recall that the hsp70 promoter contains a TATA box. Next *spa* constructs were placed upstream of the *dPax2* primary promoter, which is 101 bp upstream of the *dPax2* TSS, and the *dPax2* secondary promoter, which is 1947 bp downstream of the *spa* enhancer sequence in the *dPax2* fourth intron. As the *dPax2* 1st–4th exons are poorly conserved, it is plausible that the essential *dPax2* transcripts in the *Drosophila* eye stem from this secondary promoter, and that *spa* actually targets this promoter primarily. Note, this would revitalize tracking as an RCE function as bidirectional transcription would no longer be necessary. We also examined several promoters the Halfon lab is interested in as representatives of different promoter classes; *brinker* (*Brk*), which contains binding sites for the Shnurri/Madmadea (SMM) complex, *Deformed* (*Dfd*) which contains a polycomb response element (PRE), *mindbomb2* (*mib2*) which contains an E-box, and *odd skipped* (*odd*) which is repressed by hairy (Jimenez et al., 1996; Park et al., 2000; Ringrose et al., 2003; Yao et al., 2008).

These reporter constructs were integrated into the attP2 landing site at 68A4, which we had not previously analyzed for *spa* activity. Due to our previous difficulties with landing sites and *spa* activity we decided to study these reporter constructs in conjunction with randomly integrated controls. As the vectors the Halfon lab uses are derivatives of the Ganesh cloning vectors, we use our *spa*-846 constructs in Ganesh (Figure 3.4 A, B). Interestingly, we saw that upon insertion into the attP2 landing site at 68A4 (Groth et al., 2004), both *spa*(wt) and *spa*(ΔRCE) drove GFP expression in the *Drosophila* eye in conjunction with the hsp70 promoter (Figure 3.4, C and D). This is the first and only time in all of our experiments that loss of the RCE has not abolished *spa* activity from a distance; in fact, the expression seen here with loss of the RCE is higher, and in a more complete pattern than either wildtype enhancer (Figure 3.4 A, C, D). Conversely, neither wildtype *spa* nor *spa* lacking the RCE was capable of stimulating transcription from either of the *dPax2* promoters (Figure 3.4 E-H). Neither construct was able to drive transcription from the mib2 promoter (Figure 3.4 M,

Figure 3.5

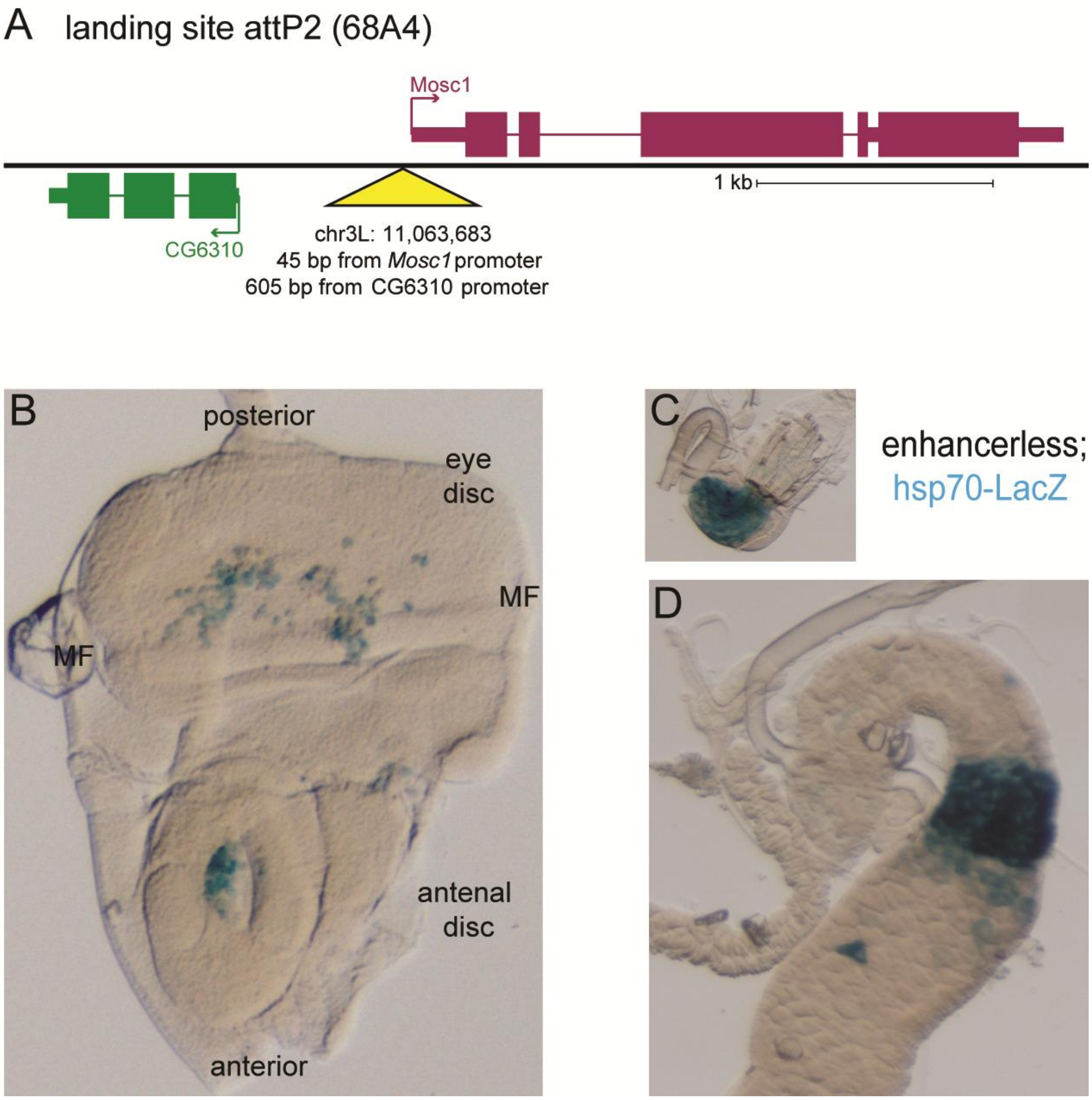


Figure 3.5 *Mocs1* regulatory information can influence reporter gene activity. The attP site in the attP2 integration locus is 684 bp downstream of CG6310 and only 45 bp upstream of Mocs1 (A). As *Mocs1* is known expression in the eye disc we analyzed the expression of an “enhancerless” hsp70-lacZ reporter construct in this landing site. We found the lacZ is indeed expressed from the hsp70 promoter alone in both eye and antennal discs (B). LacZ is also expressed from the hsp70 promoter in the posterior portion of the spherical and in the posterior larval midgut (C and D).

N). While we saw no GFP expression driven by the wildtype *spa* enhancer at any of the remaining promoters, *brk*, *Dfd*, and *odd*, we did see expression from the *spa*(ΔRCE) construct (Figure 3.4). From the *Dfd* promoter we did see bright expression in non-cone cells, possibly peripodial cells, and we see a similar but much lighter pattern driven from the *brk* promoter (Figure 3.4 J, L). With the *odd* promoter, *spa* drives GFP expression in cone cells and additional cell types which extends anteriorly to the morphogenic furrow, where *spa* is not active (Figure 3.4 P) (Fu and Noll, 1997).

Given the high levels the expression driven by the *spa*(ΔRCE) construct compared to *spa*(wt), the insertions were first sequenced to confirm the enhancer sequences were correct. Next we decided to examine the genomic environment of the attP2 insertion site. The attP2 in this landing site is 685 bp downstream of CG6310 and only 45 bp upstream of *Mocs1* (Figure 3.5 A). Mocs1 has known expression in the *Drosophila* eye (Graveley et al., 2011). As we observed dramatically different expression patterns driven by the wildtype *sparkling* and *spa* lacking the RCE depending on whether the reporter construct was integrated into the attP2 landing site or by random integration (Figure 3.4, A-D), we hypothesized the *Mocs1* regulatory information can affect reporter gene transcription from the hsp70 promoter. To this end, we analyzed the expression of an enhancerless reporter construct with the hsp70 promoter driving expression of LacZ in larval tissue. We found that LacZ expression is indeed driven by the hsp70 promoter alone in both the eye and antennal discs (Figure 3.5 B). Interestingly, LacZ it is also expressed in the posterior spherical and in posterior midgut (Figure 3.5 C, D). While the LacZ expression driven by *Mocs1* regulatory sequence alone does not recapitulate any of the expression patterns, we observed with the *spa*(ΔRCE) construct, this additional information could substantially influence enhancer and promoter activity and therefore the results of any of our reporter gene studies.

We decided to repeat several of the reporter constructs of interest, hsp70, the primary *dPax2* promoter, and *Dfd* in a second landing site - attP40 at the cytological location 25C6. The expression patterns of these reporter constructs were again compared to *spa* -846bp in Ganesh (Figure 3.6 B, C). This time we looked first at an enhancerless vector, with hsp70 driving GFP. Here we see no inappropriate gene expression resulting from the local insertion site (Figure 3.6 A). Unlike constructs inserted into attP2, the *spa*(wt) and *spa*(ΔRCE) did not drive GFP expression from any promoter, even the hsp70 promoter from which we know *spa* can stimulate transcription (Figure 3.6 D-I). As we have encountered problems with *spa* activity in other landing sites, it is not too surprising that we found yet another landing site in which *spa* does not promote gene activation.

3.4 Discussion

The *dPax2 sparkling* enhancer is responsible for driving cone cell specific gene expression in the developing *Drosophila* eye. We have identified a sequence at the 5' end of *spa* that specifically mediates the enhancer's ability to activate gene expression from a distance in the genome. This sequence,

Figure 3.6

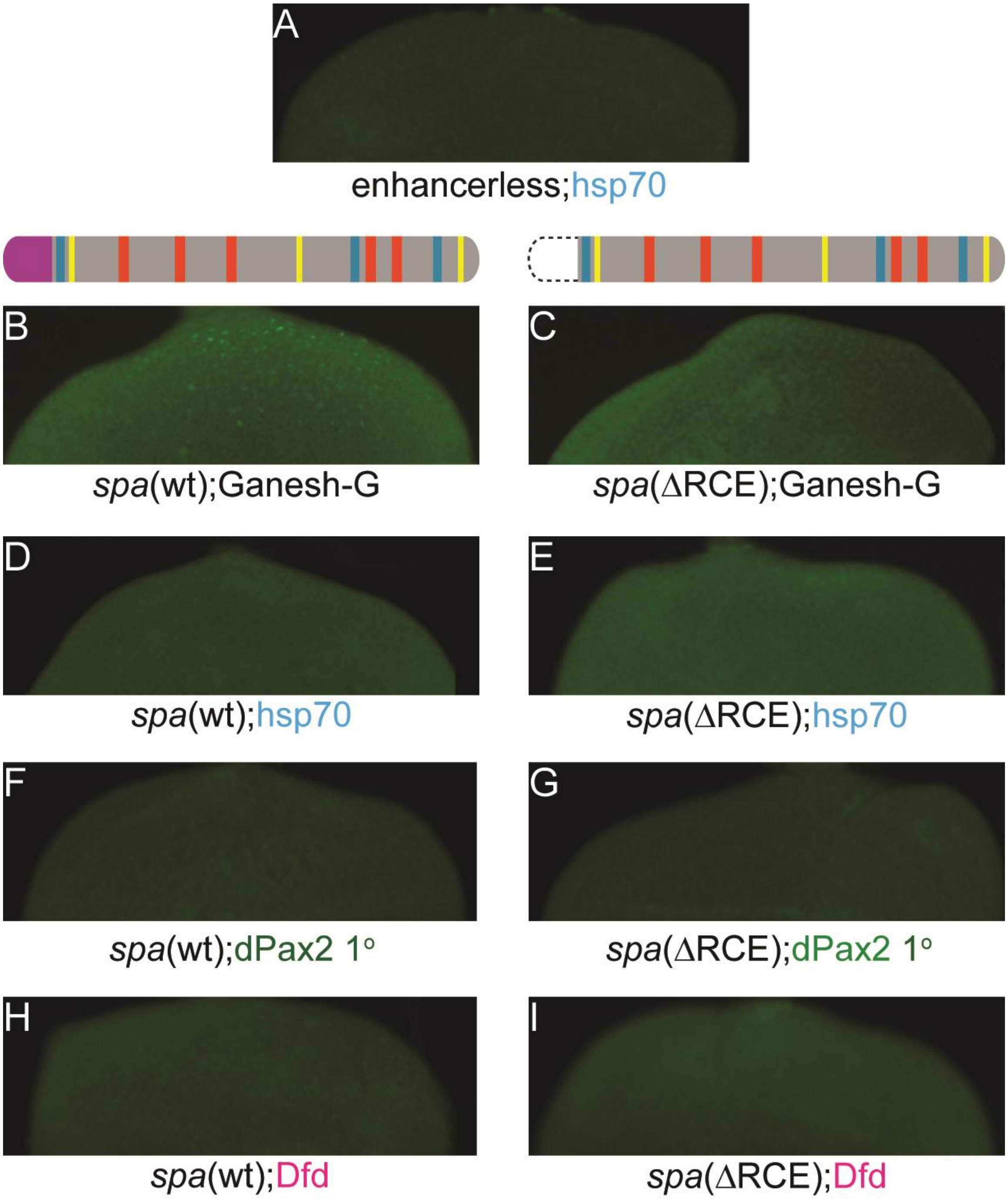


Figure 3.6 *sparkling* is inactive in the attp40 landing site. Previous studies have demonstrated that *spa* drives reporter gene expression in cone cells of the developing *Drosophila* eye (B and C). In the attp40 landing site, no GFP expression is driven by the hsp70 promoter alone (A). Neither the *spa*(wt) nor the *spa*(ΔRCE) constructs drove GFP expression from the hsp70 (D, E) or from any other promoter in this landing site (F-I).

sparkling RCE, does not convey any patterning information when the enhancer is in the promoter proximal position; although its loss can increase reporter expression. In the course of this work we have examined the mechanisms by which the RCE might act, performed motif analysis to address what proteins might facilitate RCE activity, and the ability of *spa* to activate different core promoters.

3.4a Observations on sparkling and promoter regulation

We have integrated *sparkling* enhancers into the *Drosophila* genome using 3 different integration contexts. Our previous work with *spa* utilized pseudorandom integration of an uninsulated reporter construct into the genome, using P-element mediated transposition. In this study we also integrated an insulated version of the same reporter constructs, again utilizing P-element mediated transposition. We also used site-specific integration to integrate the *spa* enhancers into the same known genomic locations via Φ3C1 mediated recombination of attB and attP site DNA elements. The pattern and levels of *sparkling* activity have been well established by us and others (Evans et al., 2012; Flores et al., 2000; Swanson et al., 2010; Swanson et al., 2011). Therefore, we were surprised to find that *spa*(wt)-846bp is unable to drive reporter expression from four independent landing sites and activates limited expression from the 86F8 landing site (Table 3.2). We have come up with two possible explanations for this observation. First, as *spa*(wt)-846 is a relatively weak enhancer, it is plausible that this enhancer is especially sensitive to non-permissive chromatin environments. It has been well documented that P-transposes has a strong preference for promoter and other regulator sequences, regions of DNA likely to be permissive for gene transcription (Bellen et al., 2004). Meanwhile, piggybac transposase, which was used to integrate attP sites for site-specific integration, does not demonstrate an insertion preferences (Bellen et al., 2004). The 86F8 landing site is in the second intron of the Chloroform channel a gene, and in fact it is the only insertion site we analyzed that lies within an intron. Conversely, the other sites we analyzed all lay outside of gene regions (between 2 and 19 kb).   It is possible then that the reason we observed expression in random integration is because *spa*(wt) is integrated into a permissive environment, while it is not in the case with the attP landing sites.

A second explanation invokes the hypothesis that *sparkling* can interact strongly with other promoters in the genomic insertion locus. The 86F8 locus is unusually promoter rich, containing the 3xP3 RFP promoter, *white* promoter, and three local CIC-a promoters. The ability to interact with these promoters may explain the limited expression seen when *spa*(wt)-846bp is integrated into the 86F8 landing site. To test this we removed the 3xP3 and *white* promoters from the insertion locus, via cre mediated recombination. We found that loss of these promoters did not substantially affect reporter gene activity (data not shown). However, we cannot rule out the interaction with the CIC-a promoters. Additional observations support the hypothesis that *spa* can interact with local promoters outside the hsp70 promoter driving GFP expression. We have observed that *spa*(wt)-121bp drives slightly decreased levels of expression compared to *spa*(ΔRCE)-121bp in both randomly inserted, uninsulated, reporter genes, and from the 86F8 landing site (Table 3.1). This result could be due to the potential promiscuity of wildtype *sparkling*. If *spa*, potentially through the RCE, interacts with multiple local promoters, it would spend less "time" activating GFP expression. Meanwhile *spa*(ΔRCE)-121bp can only act at a short range, due to loss of the RCE, and therefore is faithful to the hsp70 promoter and drives increased GFP expression. An alternative explanation of these results is that in addition to containing long range activity, the RCE also contains a binding site for a short range transcriptional repressor. Loss of the RCE would then result in derepression of promoter proximal *spa* activity. However, wildtype levels of reporter gene expression are no longer decreased compared to *spa*(ΔRCE)-121bp when gypsy insulator sequences flank the enhancer and GFP reporter gene indicating that the RCE does not contain a repressive element. The primary function of insulators in this context is to minimize interaction of the reporter constructs with the local genomic environment. Therefore, loss of interaction with local regulatory sequences could explain why *spa*(wt)-121bp and *spa*(ΔRCE)-121bp drive the same levels of GFP expression, and even why *spa*(wt)-846bp activity is higher than that driven by uninsulated reporter genes.

The ability of *spa*(wt) but not *spa*(ΔRCE) to interact with multiple promoters might also explain some of our results regarding *sparkling*'s activity with different promoter sequences. We found that the attP2 landing site is not suitable for studying as regulatory sequences from the 68A4 locus are able to drive reporter expression from hsp70 promoter alone in not only eye discs, but several other tissues as well (Figure 3.5). Interestingly, this is the primary line used to integrate RNAi constructs into the *Drosophila* genome (Ni et al., 2008). Therefore, it would be unsurprising that *spa* drives GFP expression in this locus in conjunction with the Mosc1 regulatory sequences and specific promoters. However, we actually saw that *spa*(wt)-846bp was only able to drive GFP expression from the hsp70 promoter; yet *spa*(ΔRCE)-846bp drove GFP expression from the hsp70, Dfd, brk, and odd promoters (Figure 3.6). *spa*(ΔRCE)-846bp lacks the sequences for long-range gene regulation; however it is possible it can act with Mosc1 regulatory sequences to drive GFP expression. But then, why wouldn't *spa*(wt) activate GFP expression from these same promoters? It is possible that wildtype *sparkling*, through the RCE sequence interacts strongly with the hsp70 and Mocs1 promoters at the exclusion of all other promoters. The ability of an enhancer to activate one promoter at the exclusion of another has been seen previously with the autoregulatory element (AE1) enhancer which regulates the fushi tarazu (ftz) promoter, but not the equidistant Sex combs reduced promoter (Scr). The ftz promoter contains a TATA box, while the Scr promoter does not. However, when the ftz promoter is absent, the AE1 can activate transcription from the TATA-less promoter (Ohtsuki et al., 1998).

We know that *sparkling* is capable of regulating the primary *dPax2* promoter (Flores et al., 2000). However, in these studies *spa* was driving expression of a cDNA, and not from a distance. In our hands, the only *spa* construct we found to drive GFP expression from the *dPax2* promoter, was *spa*(ΔRCE) in a promoter proximal position. (Christina Swanson, unpublished data). To us, this supports both the ideas the *spa* might regulate the *dPax2* secondary promoter and not the primary promoter, as well as the idea that wildtype *sparkling*'s can interact with local promoters affects reporter gene read out. If the regulatory information directing *spa* to stimulate the secondary promoter lies within the RCE, the construct lacking this sequence would simply activate the nearest promoter (ie: *dPax2* primary promoter or hsp70), while the wildtype does not activate any transcription.

This work has raised significant questions regarding the ability of *spa* to interact with multiple promoters, sometimes with preference for specific promoter elements. While our observations indicate this may be true, at this time we lack experimental evidence to thoroughly test these questions. One technique to test *sparkling'*s ability to interact with different promoters would be to generate reporter constructs in which *spa* is allowed to regulate two promoters, each driving different reporter genes. By varying the combinations of promoter elements, we could elucidate *spa*'s promoter preferences, which may indicate how *spa* acts from a distance. Furthermore we would like to determine which *dPax2* isoform is prevalent in the eye imaginal disc (transcribed from the primary or secondary promoter?) and whether or not *spa* is actually capable of stimulated gene expression from the *dPax2* secondary promoter.

3.4b Mechanisms of sparkling action

We designed a series of experiments to test the possible mechanisms by which the RCE acts within the *spa* enhancer to facilitate long-range gene activation. One of the prominent proposed mechanisms of distal enhancer action is through looping of DNA to bring the enhancer and its target promoter into close proximity and stimulate gene transcription. This is thought to occur by the formation of protein complexes between the enhancer and promoter that alter the 3D structure of the chromatin. We hypothesize that if the RCE worked by looping in a manner that required binding of the same protein to both the enhancer and promoter, the addition of a second copy of the RCE at the 3' end of the enhancer would promote the formation of exclusive loop and inhibit transcription. However, we found that addition of the second copy of the RCE had no affect on *spa* activity (Figure 3.2). Therefore it is unlikely that if the RCE functions to form loops through protein homo-oligomerization. This is not unexpected, as we know that *spa* can activate both the endogenous *dPax2* promoter and the hsp70 promoter in our reporter vector, which is separated from the enhancer by a completely independent DNA sequence. In order to further test the role of intervening sequence between the enhancer and promoter we can change the DNA sequence that generates the 846bp spacer. This would also address the possibility of linking as a mechanism of *spa* activity. These results also indicate that multiple copies of the RCE do not boost GFP expression levels either. However, we need to test the effect of two RCE sequences placed immediately adjacent to each other in order to determine whether proteins are capable of binding cooperatively to the RCE and increase reporter gene activity. Ideally, to test looping as a mechanism of *spa* activity we would prefer to utilize chromatin capture assays. However, we are limited by both the small amount of tissue we have to work with (cone cells of the developing *Drosophila* eye) and the distance between *spa* and the promoter; even in its endogenous location, looping between the *spa* enhancer and the *dPax2* promoter would be difficult to detect above background levels (Dekker, 2006).

We observed that the RCE can act to stimulate distal *spa* activity even when the RCE is separated from the enhancer by distances between 73 and 698bp (Figure 3.2). This finding will not only be helpful in discovering new enhancers using *in silico* techniques, but also demonstrates that we may have to look outside of known minimal enhancer sequences to find the sequences responsible for long range enhancer activity. This observation is also consistent with a tracking mechanism for *spa* activity. As *spa* can act from downstream to activate *dPax2* transcription and upstream to activate GFP transcription the RCE would have to recruit the basal transcription machinery and stimulate transcription in both directions in order for tracking to be a plausible mechanism - unless, or course, *spa* actually activates *dPax2* from the secondary, downstream, promoter. Notably, *spa* would also have to be able to activate transcription using a different mechanism, although only from the promoter proximal position, as under this model *spa*(ΔRCE) would not be able to recruit the basal transcription machinery. In order to further test tracking as a mechanism RCE activity, we could analyze intergenic transcription in both the endogenous locations and from our reporter constructs. We would also like to perform Chromatin Immunoprecipitation (ChIP) for PolII at the *spa* enhancer and intervening sequences; however we are again limited by the amount of tissue available to we have to work with.

We have not performed any experiments that specifically address the mechanisms of altering sub-nuclear localization and the production of non-coding RNAs. If intergenic RNA is detected from the *spa* enhancer or nearby, we could decrease levels of this RNA using siRNA, and assess the effect on endogenous, or reporter gene transcription (Orom et al., 2010). Again, this mechanism is unlikely as *spa* stimulates transcription from both the *dPax2* and hsp70 heterologous promoter over different intervening sequences. However, long non-coding RNAs have previously been shown to stimulate transcription from both endogenous and heterologous promoters (Orom et al., 2010). We can also assess the nuclear localization of *spa* and its target promoter in both active (cone cells) and inactive (photoreceptor or a non-eye cell type) nuclei by tagging the locus (ie through fluorescent in situ hybridization, FISH) and visualizing the location of *spa* within these nuclei (Osborne et al., 2004). The *Drosophila* third instar eye imaginal disc can be removed and grown in culture for at least 24 hours, during which time cone cell development proceeds (N. Evans unpublished observation). Therefore, it is possible that we can visualize the location of *spa* and its target promoter dynamically during cell type specification.

3.4c Identification of putative protein binding sites that may facilitate RCE activity

The RCE is a 40bp DNA sequence at the 5' end of the *spa* enhancer. We know this sequence conveys distal enhancer activity to the *sparkling* enhancer. As such we expected this RCE would contain unique protein binding sites compared to the rest of the enhancer, and that this sequence would always be required when *spa* is placed in a promoter distal position. Therefore, we were surprised to find that at least two DNA sequences can replace the RCE and promote distal *spa* activity. First, region 4, which lies in the center of the *spa* enhancer and is critical for enhancer function, is able to replace the RCE sequence at the 5' end of the enhancer and drive GFP expression (Figure 3.3). Note that either region 4 must be present in two copies, or must be in this 5' position in order to perform long-range function, as *spa*(ΔRCE)-846bp contains a copy of region 4 in its wildtype position but does not drive any reporter gene expression. Similarly, an upstream activating sequence (UAS) - the yeast Gal4 binding site - is able to contribute to distal *spa* activity from the position of the RCE (Figure 3.3). The yeast genome is very compact, and therefore little distal gene regulation is necessary (Vassarotti and Goffeau, 1992). As such, UAS sites are typically found near promoters. However, in rare cases such as for the HO gene, regulatory UAS sites have been identified 1 to 2bp away from promoters, suggesting they are capable of acting at a distance (Breeden and Nasmyth, 1987). As we suspect the RCE binds to and interacts with proteins in order to facilitate long-range promoter communication, we hypothesized that the RCE, *spa* region 4, and the UAS contain similar protein binding motifs that might allow us to identify these proteins.

To this end we performed motif analysis between these three sequences and generated a list of putative binding candidates that are expressed in the eye disc and have potential binding sites in two or three of these DNA sequences.

The known molecular functions of some of the candidates make them better or worse candidates for RCE function. For example, a few proteins can be easily eliminated. BarHI/2 and Engrailed are unlikely to possess distal enhancer activities, as both are traditionally thought to be repressors (Jaynes and O'Farrell, 1991; Laughon, 1991). Furthermore, BarH1/2 and rough are not expressed in cone cells where *spa*, and therefore the RCE, are active (Hayashi et al., 1998; Voas and Rebay, 2004). It remains possible BarH1/2 does bind to these sequences and helps repress *spa* activity in photoreceptors. However, such an action must be redundant with other sequences within the enhancer as we do not see ectopic gene expression in photoreceptors when the RCE or region 4 are lost (Swanson et al., 2010). Similarly, while Tramtrack is a zinc finger containing transcription factor that has previously been shown to bind to developmental enhancers, it has only been show to act as a transcriptional repressor (Harrison and Travers, 1990; Xiong and Montell, 1993). While ocilliless is expressed in the eye disc, it is thought to act much earlier in development to specify the eye and antennal discs than *dPax2* is expressed, making it a less likely candidate (Royet and Finkelstein, 1996). Scalloped is also expressed in the eye imaginal disc; however vestigial, scalloped’s known activation partner, is not, which would require the presence of a unique activating cofactor in the eye (Kurata et al., 2000). Of the known transcriptional activator proteins identified by this analysis, only Sine oculis remains as a good candidate for RCE interaction as it is expressed in cone cells, regulates known developmental eye enhancers, and both the RCE and region 4 contain near consensus binding sites for the protein (Pauli et al., 2005) .

Additional proteins identified by this analysis have not necessarily been implicated in transcriptional activation, but instead perform functions that could directly influence enhancer-promoter interactions. The Lim proteins have been shown to bind homeodomain binding sites, which both *spa* region 4 and the RCE contain, and aid in long-range gene transcription via an interaction with Chip (Morcillo et al., 1996); however, experimental data on this action is extraordinarily limited. Similarly, an interaction with TATA binding proteins (Tbp) would be consistent with a tracking mechanism of long-range gene regulation in which Tbp proteins bind to the enhancer via the RCE and subsequently recruit the Pol II RNA transcription complex (Blackwood and Kadonaga, 1998). Interestingly, Dichaete is HMG family member (High Mobility Group) which has been shown to bind to and bend DNA (Pil et al., 1993). This protein is not expressed in the correct cells to be involved in *spa* activity; however, it is possible a family member with similar function and binding affinity is (Mukherjee et al., 2000).

We have identified a putative list of RCE interacting proteins that can be further analyzed by electron mobility shift assays, gene knockdown, and binding site mutation in reporter constructs. On a broader scale, this study requires far more reporter constructs to further our understanding the RCE’s capabilities. Additionally, the reporter genes analyzed here, especially those regarding *sparkling*’s ability to interact with different core promoters, must be repeated using P-element mediated random integration instead of site-specific integration.

3.5 Experimental methods

3.5a Vector construction, reporter gene generation, and transgenesis

peGFPattB(peaB) was constructed by swapping the UAS-MCS-S40 cassette from pUASattB (GenBank EF362409) (Bischof et al., 2007) with annealed oligos containing HindIII, SphI, XhoI, XbaI, EcoNI restriction sites via BamHI digest and BhlII overhangs. Oligo sequences are listed below:

Top: 5' gatctaagcttgctagcatgcatctcgagattctagacctacgtaagga 3'

Bottom: 5' gatctccttacgtaggtctagaatctcgagatgcatgctagcaagctta 3'

Subsequently, the hsp70 promoter and cGFP-NLS sequence was digested from pHstinger (Gen Bank AF24246S) (Barolo, 2000) with SphI and SpeI and ligated into the above plasmid after digestion with SphI and XbaI to generate peaB.

Sparkling enhancer sequences with the 846, 605, and 121 bp spacer sequences where generated by sewing PCR and cloned into peaB or pHstinger with ECORI and BamHI digestion. Alterations to the RCE sequence were generate by standard PCR, while constructs with multiple RCE copies or the RCE moved upstream were generated by sewing PCR and ligate into peaB via EcoRI and BamH1 digestion. The 5' UAS site was added to the *spa*(ΔRCE)-846 sequence already ligated into pEAB via ligation of annealed oligos containing the UAS sequence and a 5' HindIII and 3' EcoRI overhang. Oligo sequences are as follows:

Top: 5’ agcttggtcggagtactgtcctccgagg 3’

Bottom: 5’ tggtcggagtactgtcctccgaggaatt 3’

Sparking sequences in Ganesh cloning vectors were generated as previously described (Swanson et al., 2008). To generate the promoter constructs, *sparking* enhancers were digested from pENTR-D-Topo (Invitrogen) with BamHI and XbaI and ligated into pattBnucGFP. Promoter cassettes were then exchanged using Gateway attL/attR mediated recombination

Reporter vectors containing attB sites were integrated into the *Drosophila* genome using vasaintDm(Φ3C1) and integrated at attP landing sites obtained from the Bloomington Stock Center (Bischof et al., 2007; Groth et al., 2004; Venken et al., 2006). P-element transformation was performed in w^{1118} flies as described previously (Rubin and Spradling, 1982).

3.5b Tissue preparation, antibody staining, microscopy

Eye disc tissues were dissected from third instar larvae. Disc tissues were then fixed in 4% paraformaldehyde at room temperature for 30 minutes. Discs were then washed 3x 5 minutes in 1x PBS and mounted in Prolong Gold with 4’, 6’-diamidino-2 phenylidole (DAPI) (Invitrogen). Imaging was performed on an OlympusBX51 microscope with an Olympus DP70 digital camera.

Immunohistochemistry was performed on dissected eye discs from 24 hour pupa. Discs were fixed in 4% paraformaldehyde for 30 minutes at room temperature and then washed 3x10 minutes in PBS-Tx (1x PBS + 0.1% Triton x-100). Fixed discs were then incubated in PBS-Tx + 2% for 1 – 3 hours and then incubated overnight in primary antibodies against GFP (Invitrogen) and Cut (*Drosophila* studies Hydrodoma Bank) diluted 1:100. The next day, tissues were washed 3x10 minutes with PBS-Tx and then incubated in secondary antibodies; goat anti-mouse 568 nm and goat anit-rabbit 488 nm (Invitrogen) diluted 1:1000. Finally, the discs were washed 3x 20 minutes in PBS-Tx and mounted in Prolong Gold with DAPI (Invitrogen). Stained discs were imaged on an Olympus FLUO View 500 Laser Scanning Confocal microscope mounted on and Olympus 1x71 inverted microscope.

X-gal staining of larval tissues was performed as described previously (Current Protocols in Molecular Biology) and imaged using a LiecaMZ12.5 dissecting microscope equipped with LeicaFireCam software.

3.5c Quantification of GFP expression

In order to quantify GFP expression within third instar eye imaginal discs, we first imaged the discs using Confocal microsopy using the same settings and number of z images per sample. We controlled for disc age by analyzing the entire area posterior to the morphogenic furrow. As 86F8 lines also have GFP expression in the optic nerve, we excluded this region from analysis. An average GFP pixel intensity greater than disc background and adjusted for the area of the disc was obtained using a custom MatLab program. Details available upon request.

3.5d DNA motif analysis

Potential transcription factor binding sites for candidate proteins were identified using genepallete (Rebeiz and Posakony, 2004). Novel DNA motifs were discovered using evoprinterHD to first identify clusters of conserved sequences within the d*ppD* enhancer (Odenwald et al., 2005). This was followed using *cis*-decoder to determine which sequences from these conserved clusters are likely to be transcription factor binding sites (Brody et al., 2007). TomTom from the Meme Suite was then used to find transcription factors whose binding sites resemble the motifs these programs identified (Gupta et al., 2007).

3.6 Acknowledgments

We would like to thank Christina Swanson for cloning *spa*(wt) and *spa*(ΔRCE)-846 in Ganesh, Amy Strom for cloning *spa*(423456)-846 in peaB, Marc Halfon and Qianqian Zhu for cloning the *spa*(wt) and *spa*(ΔRCE) promoter constructs, and Christian Morrison for assistance with MATLab programming.

3.7 References

Akbari, O.S., Bae, E., Johnsen, H., Villaluz, A., Wong, D., and Drewell, R.A. (2008). A novel promoter-tethering element regulates enhancer-driven gene expression at the bithorax complex in the Drosophila embryo. Development *135*, 123-131.

Akbari, O.S., Schiller, B.J., Goetz, S.E., Ho, M.C., Bae, E., and Drewell, R.A. (2007). The abdominal-B promoter tethering element mediates promoter-enhancer specificity at the Drosophila bithorax complex. Fly (Austin) *1*, 337-339.

Bae, E., Calhoun, V.C., Levine, M., Lewis, E.B., and Drewell, R.A. (2002). Characterization of the intergenic RNA profile at abdominal-A and Abdominal-B in the Drosophila bithorax complex. Proc Natl Acad Sci U S A *99*, 16847-16852.

Barolo, S., Carver, L.A., Posakony, J.W. (2000). GFP and beta-galactosidase transformation vectors for promoter/enhancer analysis in Drosophila. BioTechniques *29*, 726-732.
Bellen, H.J., Levis, R.W., Liao, G., He, Y., Carlson, J.W., Tsang, G., Evans-Holm, M., Hiesinger, P.R., Schulze, K.L., Rubin, G.M.*, et al.* (2004). The BDGP gene disruption project: single transposon insertions associated with 40% of Drosophila genes. Genetics *167*, 761-781.
Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. Proceedings of the National Academy of Sciences of the United States of America *104*, 3312-3317.
Blackwood, E.M., and Kadonaga, J.T. (1998). Going the distance: a current view of enhancer action. Science *281*, 60-63.
Breeden, L., and Nasmyth, K. (1987). Cell cycle control of the yeast HO gene: cis- and trans-acting regulators. Cell *48*, 389-397.
Brody, T., Rasband, W., Baler, K., Kuzin, A., Kundu, M., and Odenwald, W.F. (2007). cis-Decoder discovers constellations of conserved DNA sequences shared among tissue-specific enhancers. Genome Biol *8*, R75.
Bulger, M., and Groudine, M. (2010). Enhancers: the abundance and function of regulatory sequences beyond promoters. Developmental Biology *339*, 250-257.
Calhoun, V.C., Stathopoulos, A., and Levine, M. (2002). Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex. Proc Natl Acad Sci U S A *99*, 9243-9247.
Campbell, S., Inamdar, M., Rodrigues, V., Raghavan, V., Palazzolo, M., and Chovnick, A. (1992). The scalloped gene encodes a novel, evolutionarily conserved transcription factor required for sensory organ differentiation in Drosophila. Genes Dev *6*, 367-379.
Dekker, J. (2006). The three 'C' s of chromosome conformation capture: controls, controls, controls. Nat Methods *3*, 17-21.
Dorsett, D. (1993). Distance-independent inactivation of an enhancer by the suppressor of Hairy-wing DNA-binding protein of Drosophila. Genetics *134*, 1135-1144.
Dorsett, D. (1999). Distant liaisons: long-range enhancer-promoter interactions in Drosophila. Curr Opin Genet Dev *9*, 505-514.
Evans, N.C., Swanson, C.I., and Barolo, S. (2012). Sparkling insights into enhancer structure, function, and evolution. Curr Top Dev Biol *98*, 97-120.
Flores, G.V., Duan, H., Yan, H., Nagaraj, R., Fu, W., Zou, Y., Noll, M., and Banerjee, U. (2000). Combinatorial signaling in the specification of unique cell fates. Cell *103*, 75-85.
Fu, W., Duan, H., Frei, E., and Noll, M. (1998). *shaven* and *sparkling* are mutations in separate enhancers of the *Drosophila Pax2* homolog. Development *125*, 2943-2950.
Fu, W., and Noll, M. (1997). The *Pax2* homolog *sparkling* is required for development of cone and pigment cells in the *Drosophila* eye. Genes Dev *11*, 2066-2078.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W.*, et al.* (2011). The developmental transcriptome of Drosophila melanogaster. Nature *471*, 473-479.

Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R., and Fraser, P. (2000). Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. Mol Cell *5*, 377-386.

Groth, A.C., Fish, M., Nusse, R., and Calos, M.P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics *166*, 1775-1782.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol *8*, R24.

Halder, G., Callaerts, P., Flister, S., Walldorf, U., Kloter, U., and Gehring, W.J. (1998). Eyeless initiates the expression of both sine oculis and eyes absent during Drosophila compound eye development. Development *125*, 2181-2191.

Harrison, S.D., and Travers, A.A. (1990). The tramtrack gene encodes a Drosophila finger protein that interacts with the ftz transcriptional regulatory region and shows a novel embryonic expression pattern. Embo J *9*, 207-216.

Hayashi, T., Kojima, T., and Saigo, K. (1998). Specification of primary pigment cell and outer photoreceptor fates by BarH1 homeobox gene in the developing Drosophila eye. Dev Biol *200*, 131-145.

Jaynes, J.B., and O'Farrell, P.H. (1991). Active repression of transcription by the engrailed homeodomain protein. Embo J *10*, 1427-1433.

Jimenez, G., Pinchin, S.M., and Ish-Horowicz, D. (1996). In vivo interactions of the Drosophila Hairy and Runt transcriptional repressors with target promoters. Embo J *15*, 7088-7098.

Johnson, K.D., Christensen, H.M., Zhao, B., and Bresnick, E.H. (2001). Distinct mechanisms control RNA polymerase II recruitment to a tissue-specific locus control region and a downstream promoter. Mol Cell *8*, 465-471.

Johnson, L.A., Zhao, Y., Golden, K., and Barolo, S. (2008). Reverse-engineering a transcriptional enhancer: a case study in Drosophila. Tissue Eng Part A *14*, 1549-1559.

Kostyuchenko, M., Savitskaya, E., Koryagina, E., Melnikova, L., Karakozova, M., and Georgiev, P. (2009). Zeste can facilitate long-range enhancer-promoter communication and insulator bypass in Drosophila melanogaster. Chromosoma *118*, 665-674.

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

Laney, J.D., and Biggin, M.D. (1997). Zeste-mediated activation by an enhancer is independent of cooperative DNA binding in vivo. Proc Natl Acad Sci U S A *94*, 3602-3604.

Laughon, A. (1991). DNA binding specificity of homeodomains. Biochemistry *30*, 11357-11367.

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. Current Biology *20*, R754-763.

Li, Q., Barkess, G., and Qian, H. (2006). Chromatin looping and the probability of transcription. Trends Genet *22*, 197-202.

Misulovin, Z., Schwartz, Y.B., Li, X.Y., Kahn, T.G., Gause, M., MacArthur, S., Fay, J.C., Eisen, M.B., Pirrotta, V., Biggin, M.D.*, et al.* (2008). Association of cohesin and Nipped-B with transcriptionally active regions of the Drosophila melanogaster genome. Chromosoma *117*, 89-102.

Mohrmann, L., Kal, A.J., and Verrijzer, C.P. (2002). Characterization of the extended Myb-like DNA-binding domain of trithorax group protein Zeste. J Biol Chem *277*, 47385-47392.

Morcillo, P., Rosen, C., and Dorsett, D. (1996). Genes regulating the remote wing margin enhancer in the Drosophila cut locus. Genetics *144*, 1143-1154.

Mukherjee, A., Shan, X., Mutsuddi, M., Ma, Y., and Nambu, J.R. (2000). The Drosophila sox gene, fish-hook, is required for postembryonic development. Dev Biol *217*, 91-106.

Ni, J.Q., Markstein, M., Binari, R., Pfeiffer, B., Liu, L.P., Villalta, C., Booker, M., Perkins, L., and Perrimon, N. (2008). Vector and parameters for targeted transgenic RNA interference in Drosophila melanogaster. Nat Methods *5*, 49-51.

Odenwald, W.F., Rasband, W., Kuzin, A., and Brody, T. (2005). EVOPRINTER, a multigenomic comparative tool for rapid identification of functionally important DNA. Proc Natl Acad Sci U S A *102*, 14700-14705.

Ohtsuki, S., Levine, M., and Cai, H.N. (1998). Different core promoters possess distinct regulatory activities in the Drosophila embryo. Genes Dev *12*, 547-556.

Orom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q.*, et al.* (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell *143*, 46-58.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W.*, et al.* (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet *36*, 1065-1071.

Park, H.C., Kim, C.H., Bae, Y.K., Yeo, S.Y., Kim, S.H., Hong, S.K., Shin, J., Yoo, K.W., Hibi, M., Hirano, T.*, et al.* (2000). Analysis of upstream elements in the HuC promoter leads to the establishment of transgenic zebrafish with fluorescent neurons. Dev Biol *227*, 279-293.

Pauli, T., Seimiya, M., Blanco, J., and Gehring, W.J. (2005). Identification of functional sine oculis motifs in the autoregulatory element of its own gene, in the eyeless enhancer and in the signalling gene hedgehog. Development *132*, 2771-2782.

Pil, P.M., Chow, C.S., and Lippard, S.J. (1993). High-mobility-group 1 protein mediates DNA bending as determined by ring closures. Proc Natl Acad Sci U S A *90*, 9465-9469.

Qian, S., Varjavand, B., and Pirrotta, V. (1992). Molecular analysis of the zeste-white interaction reveals a promoter-proximal element essential for distant enhancer-promoter communication. Genetics *131*, 79-90.

Rebeiz, M., and Posakony, J.W. (2004). GenePalette: a universal software tool for genome sequence visualization and analysis. Developmental Biology *271*, 431-438.

Ringrose, L., Rehmsmeier, M., Dura, J.M., and Paro, R. (2003). Genome-wide prediction of Polycomb/Trithorax response elements in Drosophila melanogaster. Dev Cell *5*, 759-771.

Rippe, K. (2001). Making contacts on a nucleic acid polymer. Trends Biochem Sci *26*, 733-740.

Royet, J., and Finkelstein, R. (1996). hedgehog, wingless and orthodenticle specify adult head development in Drosophila. Development *122*, 1849-1858.

Rubin, G.M., and Spradling, A.C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. Science *218*, 348-353.

Struhl, K. (1987). Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast. Cell *49*, 295-297.

Swanson, C.I., Evans, N.C., and Barolo, S. (2010). Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. Dev Cell *18*, 359-370.

Swanson, C.I., Hinrichs, T., Johnson, L.A., Zhao, Y., and Barolo, S. (2008). A directional recombination cloning system for restriction- and ligation-free construction of GFP, DsRed, and lacZ transgenic *Drosophila* reporters. Gene *408*, 180-186.

Swanson, C.I., Schwimmer, D.B., and Barolo, S. (2011). Rapid evolutionary rewiring of a structurally constrained eye enhancer. Curr Biol *21*, 1186-1196.

Thorpe, H.M., and Smith, M.C. (1998). In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. Proc Natl Acad Sci U S A *95*, 5505-5510.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell *10*, 1453-1465.

Tsytsykova, A.V., Rajsbaum, R., Falvo, J.V., Ligeiro, F., Neely, S.R., and Goldfeld, A.E. (2007). Activation-dependent intrachromosomal interactions formed by the TNF gene promoter and two distal enhancers. Proc Natl Acad Sci U S A *104*, 16850-16855.

Vassarotti, A., and Goffeau, A. (1992). Sequencing the yeast genome: the European effort. Trends Biotechnol *10*, 15-18.

Venken, K.J., He, Y., Hoskins, R.A., and Bellen, H.J. (2006). P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. Science *314*, 1747-1751.

Voas, M.G., and Rebay, I. (2004). Signal integration during development: insights from the *Drosophila* eye. Dev Dyn *229*, 162-175.

Wang, G., Balamotis, M.A., Stevens, J.L., Yamaguchi, Y., Handa, H., and Berk, A.J. (2005). Mediator requirement for both recruitment and postrecruitment steps in transcription initiation. Mol Cell *17*, 683-694.

Wilson, K.L., and Berk, J.M. (2010). The nuclear envelope at a glance. J Cell Sci *123*, 1973-1978.

Xiong, W.C., and Montell, C. (1993). tramtrack is a transcriptional repressor required for cell fate determination in the Drosophila eye. Genes Dev *7*, 1085-1096.

Yao, L.C., Phin, S., Cho, J., Rushlow, C., Arora, K., and Warrior, R. (2008). Multiple modular promoter elements drive graded brinker expression in response to the Dpp morphogen gradient. Development *135*, 2183-2192.

Zhu, X., Ling, J., Zhang, L., Pi, W., Wu, M., and Tuan, D. (2007). A facilitated tracking and transcription mechanism of long-range enhancer function. Nucleic Acids Res *35*, 5532-5544.

CHAPTER 4

A POTENTIAL ROLE FOR THE *SIX* FAMILY MEMBER, *SINE OCULIS,* IN *SPARKLING* ACTIVITY

4.1 Abstract

The *sparkling* enhancer is responsible for regulating *dpax2* expression in the developing cone cells of the *Drosophila* eye imaginal disc. We have previously demonstrated that independent sub-elements, the remote control element (RCE) and region 4, within this enhancer are capable of regulating distal gene transcription from this the *spa* enhancer, albeit in slightly different manners. Given the size of these sequences, 40bp, and their overlapping functions, we postulated that these DNA sequences may interact with the same protein to facilitate long-range gene expression. Motif analysis demonstrated that both the RCE and region 4 contain putative binding sites for the transcription factor Sine oculis (So), a protein that is required for *Drosophia* eye development. Given the sequence and functional homology between So and the vertebrate Six proteins a potential role for Sine oculis in long-range enhancer activity would have far reaching implications. Here, we show that So can indeed bind to each these DNA sequences *in vitro*. Futhermore *In vivo* mutations to the enhancer sub

regions that contain So binding sites abolish enhancer activity. However, targeted mutations to the So sites specific does not seem to affect *spa's* ability to drive long range gene expression in cone cells leading us to examine a potential additional interaction of So with *spa* region 5. Further analysis of the potentially partially redundant function of these *spa* enhancer sub elements led to the observation that a surprising number of *spa* sequences are able to substitute for RCE activity and drive distal gene expression. Furthermore, some sequences within the enhancer can fully substitute for each other, while other substitutions are not tolerated. Finally, location of some functions, such as those encoded for by region 4 require specific placement within the enhancer, while others such as the RCE can be relocated.

A POTENTIAL ROLE FOR THE *SIX* FAMILY MEMBER, *SINE OCULIS,* IN *SPARKLING* ACTIVITY

4.2 Introduction

The *dpax2 sparkling* enhancer is capable of driving *dpax2* cDNA and reporter gene expression in the developing cone cells of the *Drosophila* eye imaginal disc (Flores et al., 2000; Johnson et al., 2008; Swanson et al., 2010; Swanson et al., 2011). This 362bp minimal enhancer has been shown to require inputs from the Notch and EGFR signaling pathways, via binding sites for Surpressor of Hairless, Su(H) and the Ets factors Pointed P2, PntP2, and Yan, as well through binding sites for the transcription factor Lozenge, Lz (Flores et al., 2000). In addition to these essential DNA binding sites, further analysis of the enhancer has shown that *spa* contains at least four additional regulatory sequences (Swanson et al., 2010). We refer to these regions of *sparking* as regions 1, 4, 5, and 6a based on their locations within the enhancer. Regions 1, 4, and 5, which are about 40bp in size, can be divided into 3 critical subelements (a, b, c). We have also further characterized the function of each of these sequences. s*parkling* (*spa*) region 4 is critical for regulation gene expression in cone cells. Regions 5 and 6 are required for proper initiation of gene expression, but not necessary maintenance of reporter activity. Furthermore region 5 contains sequences responsible for repressing *spa* activity in photoreceptors. Finally, region 1 of the enhancer is responsible for distal gene transcription activated by the *spa* enhancer. This sequence is required only when the enhancer is placed at a moderate distance (846bp) from the reporter genes transcriptional start site (TSS) and is dispensable for *spa* activity in the promoter proximal position (121bp). Therefore, we refer to it as the remote control element or RCE.

We have previously demonstrated the ability of the RCE to drive distal gene expression is not constrained to the RCE sequence alone. If *spa* region 4 is placed in the position of the RCE (at the 5' end of *spa*) the enhancer drives wildtype levels of gene expression. Similarly, when an upstream activating sequence (UAS) is substituted for the RCE sequence, the enhancer retains the ability to drive GFP expression in the eye imaginal disc. We performed motif analysis to compare these sequences and identify potential protein binding sites that might be responsible for enabling long-range gene regulation (Chapter 3). Some of these candidate binding sites are more likely to regulate *spa* activity than others. For example BarHI/2, Engrailed, and tramtrack all act primarily as transcriptional repressors (Harrison and Travers, 1990; Jaynes and O'Farrell, 1991; Laughon, 1991; Xiong and Montell, 1993). While these proteins may indeed bind the enhancer and modulate *spa* activity in cone cells and repress its expression in photoreceptors, they are unlike to facilitate long-range gene regulation. We also identified putative binding sites for proteins that, based on previous molecular function enhancer promoter interactions, such as the Lim proteins, have been shown to bind homeodomain binding sites and aid in long-range gene transcription via an interaction with Chip (Morcillo et al., 1996). An interaction with TATA binding proteins (Tbp) would be consistent with a tracking mechanism of long-range gene regulation in which Tbp proteins bind to the enhancer via the RCE and subsequently recruit the Pol II RNA transcription complex (Blackwood and Kadonaga, 1998). We also identified a binding site for an HMG family member (High Mobility Group) which have been shown to bind to and bend DNA (Pil et al., 1993). While these proteins are all potential candidates for RCE function, it is important to remember that *spa* region 4 can substitute for RCE activity in the *spa* enhancer. Region 4 is absolutely essential for activating the correct levels and pattern of reporter expression, suggesting it likely interacts with one or more transcription factors. Given its ability facilitate RCE activity it is possible that binding of a transcription factor to this sequence, rather than a chromatin binding protein or the basal transcription machinery, is responsible for distal gene regulation. The only likely transcription factor binding site identified by our motif analysis is for the Six family transcription factor Sine oculis.

The *sparkling* RCE contains a sequence motif that differs from the Sine oculis (So) consensus binding site GTAANYNGANAYS by only two base pairs (Pauli et al., 2005). Similarly, *spa* region 4 contains a sequence motif that differs from the So binding site by only one base pair. Along with *eyeless*, *optimoter blind,* and *eyes absent*, *So* is critical for early eye development. In fact, So is required for the specification of the eye primordium, a group of cells designated during embryogenesis that later migrates to the larval head. As So is required for specification of the eye primordium, it is unsurprising that only a small number of ommatidia ever form when *So* expression is lost. During eye imaginal disc development, So is expressed in the undifferentiated cells and helps establish the morphogenic furrow (Cheyette et al., 1994; Daniel et al., 1999). So is later required to specify at least the photoreceptor cells (Blanco et al., 2010; Bovolenta et al., 1998; Serikaku and O'Tousa, 1994). As photoreceptors are required to specify the other cell types of the *Drosophila* eye, the subsequent role of So in their development is unknown (Voas and Rebay, 2004). Throughout the entire processes of eye development, So and Eyes absent (Eya) have been shown to form a complex that is necessary for proper eye development (Blanco et al., 2010; Halder et al., 1998; Pignoni et al., 1997). Interestingly, the two proteins can act together to induce the formation of ectopic eyes in other *Drosophila* tissues (Halder et al., 1995).

The interaction between So and Eya is facilitated by the conserved Six, protein-protein interaction domain (SD). While the vertebrate Six proteins are homologs of the *Drosophila* So protein, So is most closely related to the murine Six 1 and Six 2. These proteins share 84% similarity in the SD and an astounding 95% similarity in the homeodomain (HD) (Seo et al., 1999). Interstingly, Six 1 and 2 are not expressed in the mouse eye at all. However, Six 3, 4, 5, and 6 are all expressed in the murine eye and play roles in its development (Kumar, 2009). The binding specificity of the entire Six family is highly conserved however, as the *Drosophila* and vertebrate family members share 64% homology. As the vertebrate Six proteins are critical for eye, brain, kidney, muscle, and gonadal development, and have been implicated in tumorigenesis, a potential role for Sine oculis in long-range enhancer activity has far reaching implications (Kumar, 2009).

Notably, *So* expression, and its transcription factor binding sites have been found to be crucial for the activity of several eye specific enhancers. For example, it is required for expression of Hedgehog in the eye disc through two different enhancers (Pauli et al., 2005). *So* expression is required for expression of Lozenge, the local transcription factor essential for specification of the ommatidial cell types. This action has been attributed to the presence of So binding sites in an eye specific enhancer of Lz (Yan et al., 2003). So and Lz in turn regulate the *prospero* enhancer that regulates gene expression in the R7 photoreceptors and cone cells (Hayashi et al., 2008). In an interesting feedback loop, eyeless initiates *So* expression in the eye primordium, meanwhile So binding sites are required for continued eyeless expression (Czerny et al., 1999; Hauck et al., 1999). Furthermore, Sine oculis binds its own autoregulatory element (SoAE) to maintain gene expression (Pauli et al., 2005). Interestingly, while So does contain a homeodomain binding region, the typical So binding sites do not contain a classic homeodomain binding site, TAAT, although they occasionally do (Hazbun et al., 1997; Pauli et al., 2005).

In order to determine whether Sine oculis interacts with and regulates the *sparkling* enhancer through the RCE and or region 4, we examined So's ability to interact with these sequences *in vitro*, and found that So can indeed bind to these DNA sequences. *In vivo* mutations to the enhancer sub regions that contain So binding sites, RCE sub region a and region 4 sub region b, abolish enhancer activity. However, targeted mutations to the So sites specific does not seem to affect *spa's* ability to drive long range gene expression in cone cells. This lead us to examine whether So can interact with any other regions of the *spa* enhancer. We observed that So can potentially interact with two additional sequences within *spa* region 5. Finally, we found that the So homeodomain binding domain is insufficient to bind DNA alone, suggesting that the protein requires additional sequence in order to interact with DNA stably.

Our investigation into the potential of So interaction with additional *spa* sequences allowed us to continue to examine the structural constraints of the *sparking* enhancer. Two models of enhancer activity are frequently proposed and debated in the field (Arnosti and Kulkarni, 2005). The first, the enhancesome model, predicts that enhancer structures and inputs are rigid. In this scenario, the identity of transcription factor binding sites within and enhancer is specific and they are strictly organized (Giese et al., 1995; Thanos and Maniatis, 1995). In the second model, information display or billboard, enhancer structure and inputs are more fluid. Here, transcription factor binding sites can be rearranged within the enhancer, and multiple transcription factors can perform the same function (Hare et al., 2008; Ludwig et al., 1998). In the loosest definition of this model only the levels of activation and repressor input matters, not the identity of each input (Arnosti et al., 1996). To test which model best describes *spa* activity, we created reporter constructs in which the critical regions of *spa* (RCE, 4, 5, and 6) are rearrange and substituted for one another. These experiments are designed to test the importance of the specific identity of the transcription factor inputs into the enhancers function, as well as if the arrangement of these sites is important. We found that a surprising number of *spa* sequences are able to substitute for RCE activity and drive distal gene expression. Furthermore, some sequences within the enhancer can fully substitute for each other, while other substitutions are not tolerated. Finally, location of some functions, such as those encoded for by region 4 require specific placement within the enhancer, while others such as the RCE can be relocated.

4.3 Results

4.3a Sine oculis interacts with the sparkling RCE in vitro

The Six family transcription factor Sine oculis (So) is a viable candidate for interacting with the RCE based on its known expression pattern on the presence of a conserved putative binding site in the RCE (Figure 4.1 A). In order to test this potential interaction we first generated full length So protein by *in vitro* transcription and translation. The *spa* RCE can be subdivided into three sub regions (RCE a, b, c) which are individually essential for *spa* activity (Swanson et al., 2010). The putative So binding site lies completely in RCE sub region a. Therefore, we tested the ability of So to interact with this portion of the RCE using Electron Mobility Shift Assays (EMSA), or gel shift assays. We first demonstrated that *in vitro* generated So protein interacts with a known So

Figure 4.1

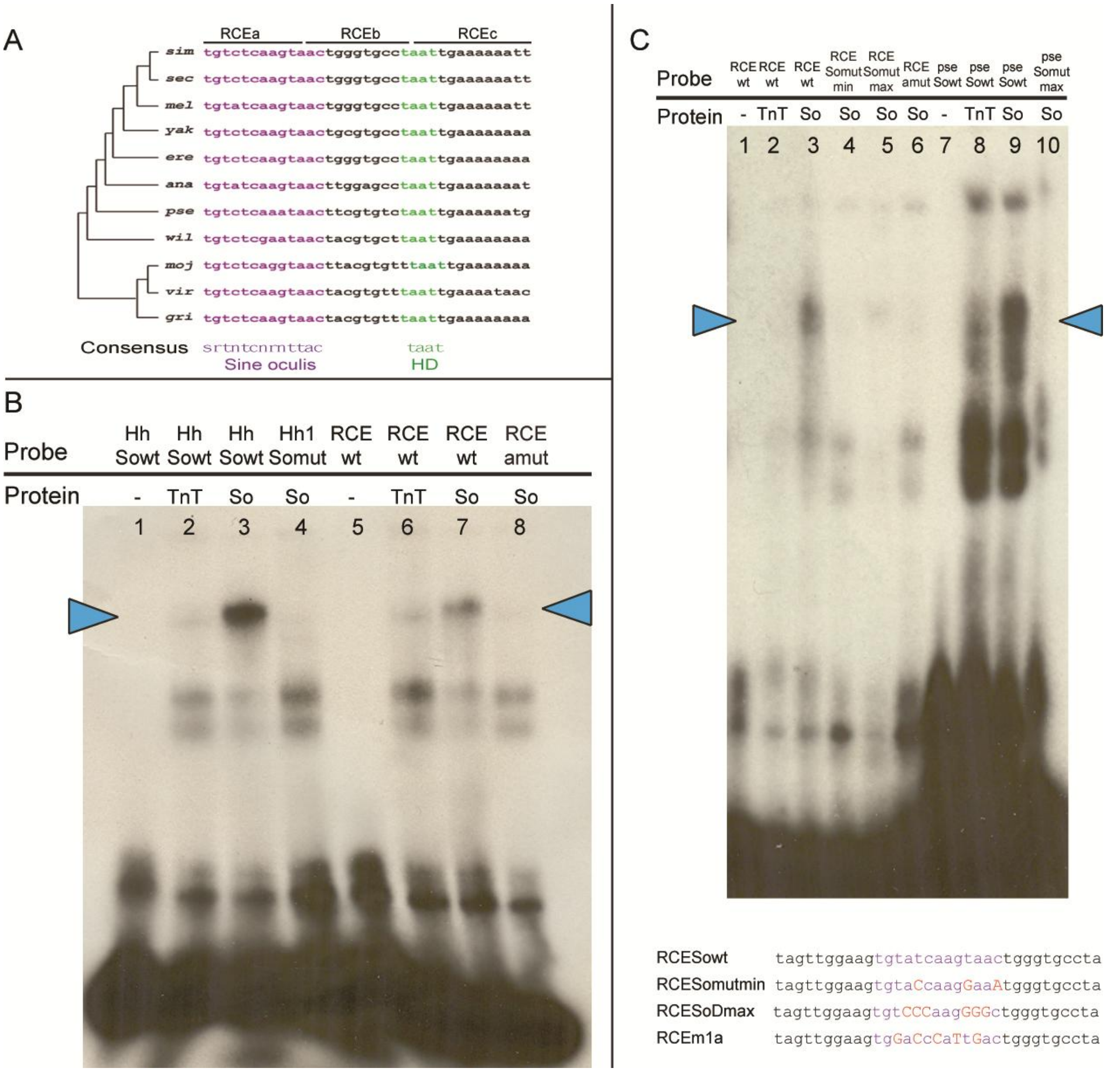

A
RCEa
RCEb
RCEc
sim tgtctcaagtaactgggtgcctaattgaaaaaatt
sec tgtctcaagtaactgggtgcctaattgaaaaaatt
mel tgtatcaagtaactgggtgcctaattgaaaaaatt
yak tgtctcaagtaactgcgtgcctaattgaaaaaaaa
ere tgtctcaagtaactgggtgcctaattgaaaaaaaa
ana tgtatcaagtaacttggagcctaattgaaaaaaat
pse tgtctcaaataacttcgtgtctaattgaaaaaatg
wil tgtctcgaataactacgtgcttaattgaaaaaaaa
moj tgtctcaggtaacttacgtgtttaattgaaaaaaa
vir tgtctcaagtaactacgtgtttaattgaaaataac
gri tgtctcaagtaactacgtgtttaattgaaaaaaaa
Consensus srtntcnrnttac taat
Sine oculis
HD
B
Probe
Hh Sowt
Hh Sowt
Hh Sowt
Hh1 Somut
RCE wt
RCE wt
RCE wt
RCE amut
Protein
-
TnT
So
So
-
TnT
So
So
1
2
3
4
5
6
7
8
C
Probe
RCE wt
RCE wt
RCE wt
RCE Somut min
RCE Somut max
RCE amut
pse Sowt
pse Sowt
pse Sowt
pse Somut max
Protein
-
TnT
So
So
So
So
-
TnT
So
So
1
2
3
4
5
6
7
8
9
10
RCESowt tagttggaagtgtatcaagtaactgggtgccta
RCESomutmin tagttggaagtgtaCcaagGaaAtgggtgccta
RCESoDmax tagttggaagtgtCCCaagGGGctgggtgccta
RCEm1a tagttggaagtgGaCcCaTtGactgggtgccta

Figure 4.1 Sine oculis interacts with the *sparkling* RCE *in vitro*. The *spa* RCE sequence contains conserved Sine oculis (So) and homeodomain (HD) binding sites (A). In order to determine whether Sine oculis binds to the RCE, we performed *in vitro* gel shift assays (EMSA). *In vitro,* transcribed and translated So binds to a control probe from a Hedgehog eye enhancer (HhSo; B lane 3) but not to other proteins in the TnT reaction (B, lane 2). This interaction is abolished when the control So site is mutated (HhmutSo; B lane 4). Similarly, the RCE wildtype probe interacts with So protein (B, lane 7) and not with TnT proteins (B, lane 6). This shift is lost when the RCE sub region a, which contains the putative So binding site, is abolished (B, lane 8). In order to further assess potential So binding to the RCE we generated targeted mutations to the So binding site. Again, the RCE wildtype probe binds to and shifts So (C, lane 3), but the mutation to RCE region a cannot (C, lane 6). A 6 bp mutation to the So site nearly abolishes the RCE/So interaction (RCESomutmin; C, lane 4). The ability of So to interact with the *spa* RCE *in vitro* is conserved in that So can also interact with corresponding sequence from *D.psuedobsura* (*D.pse*). The *D.pse* wildtype sequence shifts So protein and not protein from the TnT (pseSowt; C, lanes 7-9). However, targeted mutation of the So site in this sequence disrupts the gel shift

(pseSomutmox; C, lane 10).

binding site from an enhancer that drives Hedgehog (Hh) expression in the eye imaginal disc (Pauli et al., 2005). Here, we see that So protein, and not other proteins from the TnT reaction mixture, shift the labeled HhSo probe, and that this shift is lost upon mutation of the So binding site within this sequence (Figure 4.1 B). These results indicate that we have successfully generated So protein that can interact with DNA in our *in vitro* system. Next we observed that the labeled RCE probe can interact with So protein resulting in a gel shift (Figure 4.1 B). This shift is not observed when the labeled RCE probe contains the mutation in sub-region a, which abolishes gene expression *in vivo* (Figure 4.1 B). These data suggest that the RCE interacts with So through the a sub region. In the RCE region a mutation alters the 10bp sequence by non complementary transversion of every other base pair. To further test the ability of So to interact with the RCE, we made mutations to the RCE EMSA probes designed to specifically abolish So binding. Two different mutations were made. The first, “So mut max,” alters 6 base pairs that are known to abolish So binding in the So binding sites of the Sine oculis auto regulatory element (SoAE). The second, “So mut min,” alters only three base pairs determined important for So DNA binding based on sequence comparison of the So binding sites in two Hh enhancers, a Lozenge enhancer, an eyeless enhancer, and in the SoAE as well as the affect of single base pair mutations in SoAE gel shifts (Pauli et al., 2005). Again, the wildtype RCE probe interacts with So resulting in a gel shift (Figure 4.1 C). However, both targeted So binding site mutations abolish this interaction with the So mut max probe retaining a small amount of So binding (Figure 4.1 C).

Together, these observations indicate that, at least *in vitro*, So is capable of interacting with the RCE. This interaction is likely conserved as So also binds the corresponding RCE sequence within the *D.psuedoobscura sparkling* enhancer. (Figure 4.1 C).

4.3b sparkling region 4 can convey RCE activity which is potentially mediated by Sine oculis

We have previously demonstrated that *sparkling* region 4 can substitute for the RCE in the promoter distal position (-846 bp) (Chapter 3). The ability of region 4 to compensate for the loss of the RCE is reliant on two copies of region 4 DNA sequence. Recall that loss of both the RCE and *spa* region 4 individually abolish *spa* activity at -846 bp (Swanson et al., 2010). However, when a copy of region 4 sequence is placed in the position of the RCE, the enhancer drives wildtype levels of gene activity (Figure 4.2 A-C). Interestingly, *spa* region 4 and the RCE contain similar motifs. As mentioned previously, the RCE contains a putative Sine oculis binding site as well as a homeodomain binding site. *spa* region 4 not only contains these same protein binding sites, but they are separated by similar basepair spacing, 7 and 8 bp respectively (Figure 4.2 A). Like the RCE, region 4 can be subdivided into 3 essential sub elements (4a, b, c) (Swanson et al., 2010). To better understand the role of region 4 in distal gene activation, we made mutations to the a, b, and c sub regions in the 5' copy of region 4, while leaving the wildtype region 4 intact. Loss of the 5' 4a sequence does not significantly alter reporter gene expression (Figure 4.2 D). However, loss of the 5' region 4b eliminates enhancer activity, while loss of 4c

Figure 4.2

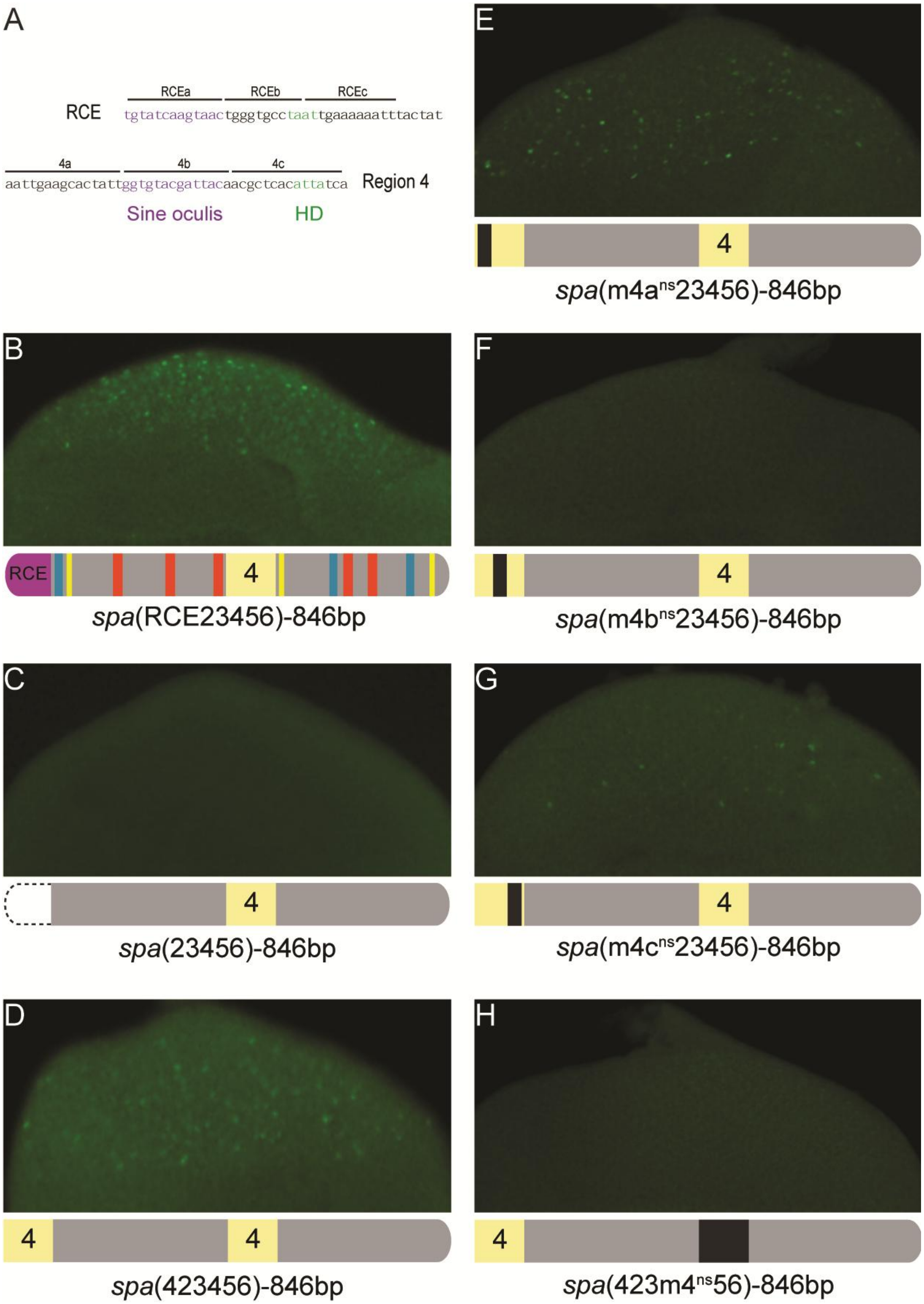

A
RCEa
RCEb
RCEc
RCE
tgtatcaagtaactgggtgcctaattgaaaaaatttactat
4a
4b
4c
aattgaagcactattggtgtacgattacaacgctcacattatca
Region 4
Sine oculis
HD
B
RCE
4
spa(RCE23456)-846bp
C
4
spa(23456)-846bp
D
4
4
spa(423456)-846bp
E
4
spa(m4a^{ns}23456)-846bp
F
4
spa(m4b^{ns}23456)-846bp
G
4
spa(m4c^{ns}23456)-846bp
H
4
spa(423m4^{ns}56)-846bp

Figure 4.2 *sparkling* region 4 can stimulate long-range transcription, potentially through Sine oculis and homeodomain binding sites. We know that *spa* requires the RCE sequence to drive GFP expression when the enhancer is placed 846 bp from the transcriptional start site (B and C). The RCE and region 4 share two DNA motifs: a Sine oculis (So) binding site and a homeodomain (HD) binding site (A). *spa* region 4 can be subdivided into three essential sub-regions; 4a, b, and c. Mutation of 4a in the 5' copy of region 4 does not alter enhancer activity (E). However, loss of 4b which contains the So site, and 4c which contains the HD site, result in a decrease of enhancer activity (F and G). The ability of region 4 to compensate for loss of the RCE is dependent on two copies of region 4 as *spa* (m4a234m4^{rs}56) is unable to drive GFP expression (H).

significantly reduces gene expression (Figure 4.2 E, F). Notably, *spa* region 4b contains the putative So binding site while 4c contains the homeodomain binding site (Figure 4.2 A), further implicating Sine oculis in binding both RCE and region 4 activity, potentially facilitating long-range gene regulation. We have postulated here that region 4 must be present in two copies in order to perform its roles in both long-range gene activation and cone cell specific gene activity. However, it remains possible that it could perform both roles from the 5' enhancer position. This is unlikely though as a construct containing a mutation to the 5' 4a sequence, which does not affect gene expression, and a mutation of the entire endogenous region 4 sequence, has no GFP expression (Figure 4.2 H). We will discuss this possibility in more depth later.

In our original experimental design to identify the RCE, we defined an element involved in long-range gene regulation as any DNA sequence necessary at a distance, by dispensable in the promoter proximal position. As such, we moved our region 4 constructs to the proximal position (-121 bp) and assessed reporter gene activity. We found that the presence of a 5' copy of region 4 did not alter gene expression compared to wildtype *spa* activity (Figure 4.3 A-C). Similarly, loss of the 5' 4c individually does not alter reporter gene expression

(Figure 4.3 E). These results are unsurprising, as deletion of the RCE in this position does not affect *spa* activity (Figure 4.3 B). Unless region 4 contained a repressor input that only acts at short-range, we would expect these *spa* sequence alterations to be inert or to enhance *spa* function. It is interesting then that loss of the 5’ 4a sequences results in increased GFP expression, suggesting region 4 sequences may indeed contain repressive input in this position (Figure 4.3 D).

We also examined the *spa* 5’ region 4 and its sub regions in the context of loss of the region 4 sequence from its wildtype position. In no scenario, (5’ wt4, m4a^{rs}, m4b^{rs}, or m4c^{rs}) is the enhancer capable of driving GFP expression (Figure 4.3 F-J). As the *spa*(423m4^{rs}56) construct contains all of the same DNA sequence input as *spa*(23456) we know that region 4 cannot act from the 5’ position to activate gene expression. This informs us that *spa* region 4 cannot perform either the long-range or cone cell specification capabilities from the 5’ position alone.

To test the ability of So to interact with *sparkling* region 4 we again turned to *in vitro* gel shift assays (EMSA). Again, we saw that the So binding site from the Hh enhancer binds to and shifts So protein (Figure 4.4 lanes 1 – 3). This shift can be competed for with the addition of 100x unlabeled wildtype HhSo probe, but not with the 100x mutant HhSo probe, demonstrating the specificity of this interaction (Figure 4.4, lanes 4 and 5). Similarly, labeled region 4 probe interacts with, and shifts So protein (Figure 4.4, lanes 6 – 8). This shift is lost with the addition of 100x unlabeled region 4 probe, but not when the entire

Figure 4.3

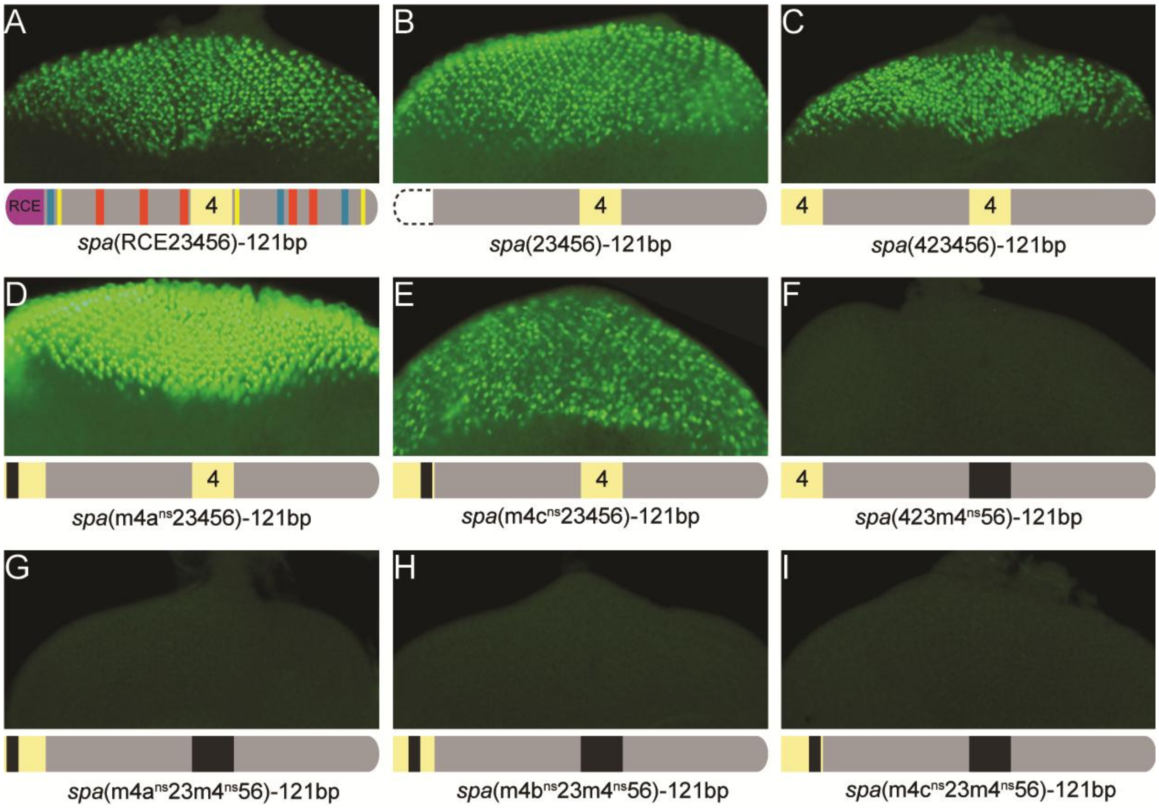

A
RCE
4
spa(RCE23456)-121bp
B
4
spa(23456)-121bp
C
4
4
spa(423456)-121bp
D
4
spa(m4a^{ns}23456)-121bp
E
4
spa(m4c^{ns}23456)-121bp
F
4
spa(423m4^{ns}56)-121bp
G
spa(m4a^{ns}23m4^{ns}56)-121bp
H
spa(m4b^{ns}23m4^{ns}56)-121bp
I
spa(m4c^{ns}23m4^{ns}56)-121bp

Figure 4.3 *sparkling* requires region 4 sequence input in its native location to drive cone cell specific gene expression. As *spa* does not require the RCE in the promoter proximal position of 121 bp from the transcriptional start site (A and B) it is not surprising that the presence of a 5' copy of region 4 (C) or mutation of the a and b sub regions (D and E) does not affect gene expression. The ability of region 4 to drive gene expression in cone cells relies on a copy of region 4 in its native position as constructs with a 5' copy of region 4 but lacking the wildtype region 4 were unable to activate any transcription (F).

sequence is mutated (Figure 4. 4, lanes 9 and 10). However, competitor reactions containing mutations to the 4a or 4c sequences do not have gel shift, suggesting So can bind to these probes, and doesn't require the a or c sequence information (Figure 4.3, lanes 11 and 13). Only loss of the 4b sequence results in failure to compete for the region 4 wildtype interaction (Figure 4.4, lane 12). Together, these data suggest So is capable of interacting with *spa* region 4 and does so through the 4b DNA sequence. The RCE sequence can also compete for region 4 So interaction (Figure 4.4, lane 14). This interaction is lost when the probe contains a three bp targeted mutation to the So site (Figure 4.4, lane 15). In summation, we can conclude that So is capable of interacting *in vitro* with sub region a of the RCE and sub region b of region 4.

4.3c sparkling activity requires specific DNA sequence input in the position of region 4

We observed that *spa* region 4 is capable of compensating for loss of the RCE when it is present in both its wildtype position and in the position of the RCE (Figure 4.2 C). As such, we wondered at the additional similarities between the RCE and region 4. We know that *spa* region 4 is required in its wildtype location regardless of enhancer position with regards to the promoter, while we also know the RCE is only necessary in the distal position (Swanson et al., 2010). We first asked if the RCE can compensate for loss of region 4. Indeed, *spa*(RCE23RCE56)-846 bp drives cone cell specific gene expression (Figure 4.5 A, B). These results indicate that DNA sequences within the RCE are capable of

Figure 4.4

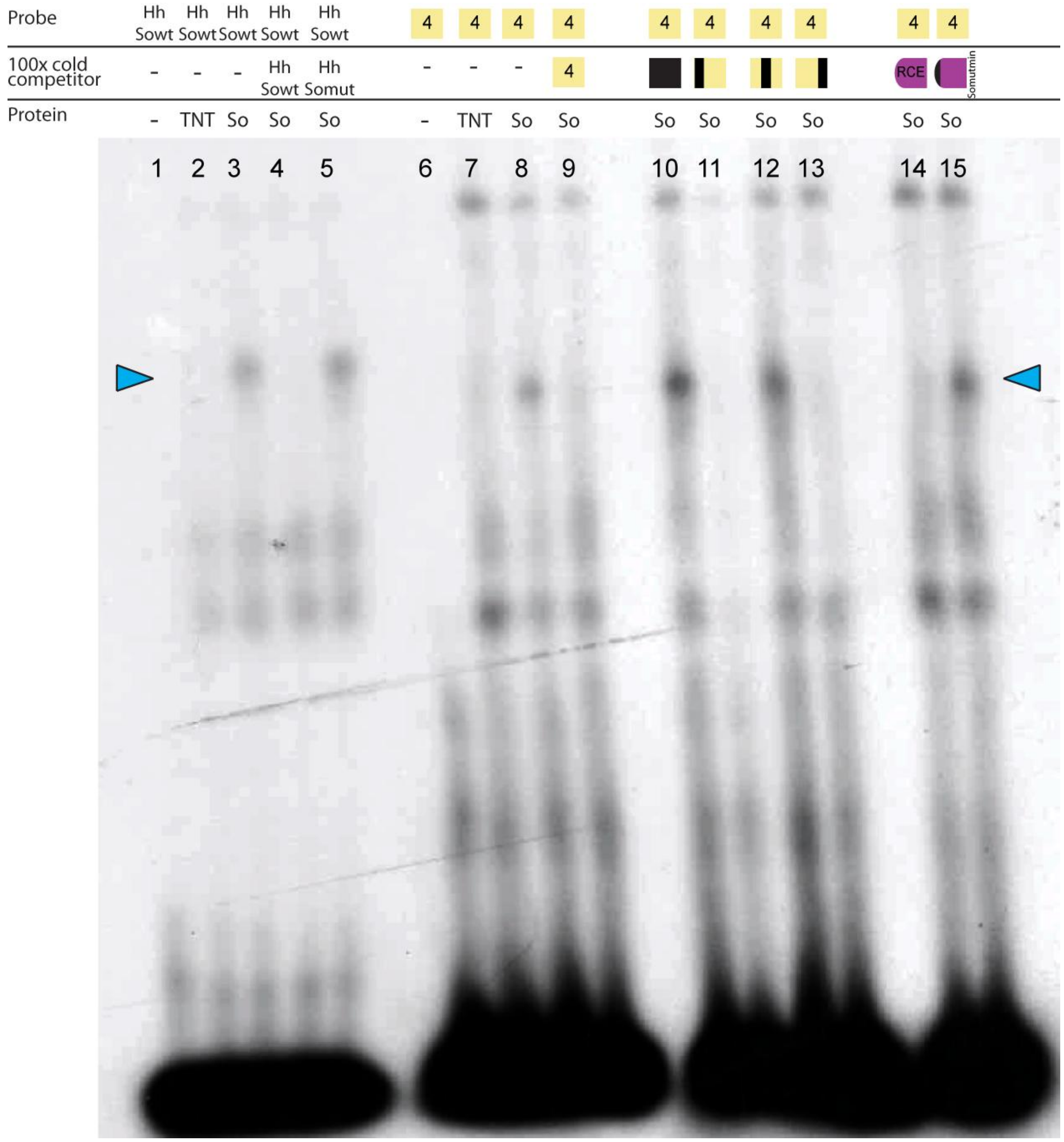


Figure 4.4 Sine oculis interacts with *sparkling* region 4 *in vitro*. We have seen that *spa* region 4 can compensate for loss of the RCE. In gel shift assays the control HhSo probe shifts So protein (lane 3). This shift is competed for with the addition of 100xunlabeled wildtype probe (lane 4) but not with the mutant probe (HhmutSo, lane 5). *spa* region 4 binds to the So protein, but not other TnT proteins, resulting in a gel shift (lanes 7 and 8). This shift is fully competed for by the wildtype region 4 sequence (lane 9) but not by a mutant region 4 sequence (lane 10). Probes containing mutations to regions 4 a and c can still compete for the So/region 4 interaction (lane 11 and 13). However loss of the b sub region eliminates this competition (lane 12) suggesting So binds region 4 through the b sub region, which contains the putative So binding site. The RCE sequence, but not the So mutant, can also compete for the ability of region 4 to interact with So

(lanes 14 and 15).

influencing the cone cell pattern of *spa* (region 4's role), and that it can act from the position of region 4 within the *spa* enhancer. The ability of the RCE to substitute for region 4 is unique, as region 5 is unable to substitute for region 4 (Evans et al., 2012). We know that region 5 also encodes a cone cell activation input (Swanson et al., 2010); therefore the specific identity of the protein binding input must be essential for the *spa* region 4/RCE mediated gene expression rather than activation ability alone (enhancesome vs information display Chapter 1, Figure 4.1). Knowing the RCE can act from region 4's position and region 4 can act from the RCE's position, we assessed the ability of *spa* to act when these sequences are exchanged such that the *spa* sequence order is 423RCE56. This construct actually drives increased levels of expression compared to wildtype (Figure 4.5 C), demonstrating that these sequences can indeed be exchanged. Interestingly, at both the distal (-846) and proximal position *spa*(423RCE56) drives GFP expression not only in cone cells but also in photoreceptors, possibly explaining the increased levels in reporter gene expression (Figure 4.6 B). As seen by co-staining with the photoreceptor marker Elav, in the third instar lava, this expression is sporadic with about one photoreceptor per ommotidium demonstrating ectopic expression; however the identity of the photoreceptor is not consistent across all ommatidia (Figure 4.6 B). Photoreceptor expression is unique to the 423RCE56 sequence order as RCE23456(wt) and 423456 result in *spa* activity in only cone cells (Figure 4.6, C). This ectopic expression in photoreceptors raises two distinct possibilities. One, the RCE contains a binding site for a protein(s) that activate gene expression in photoreceptors. This

Figure 4.5

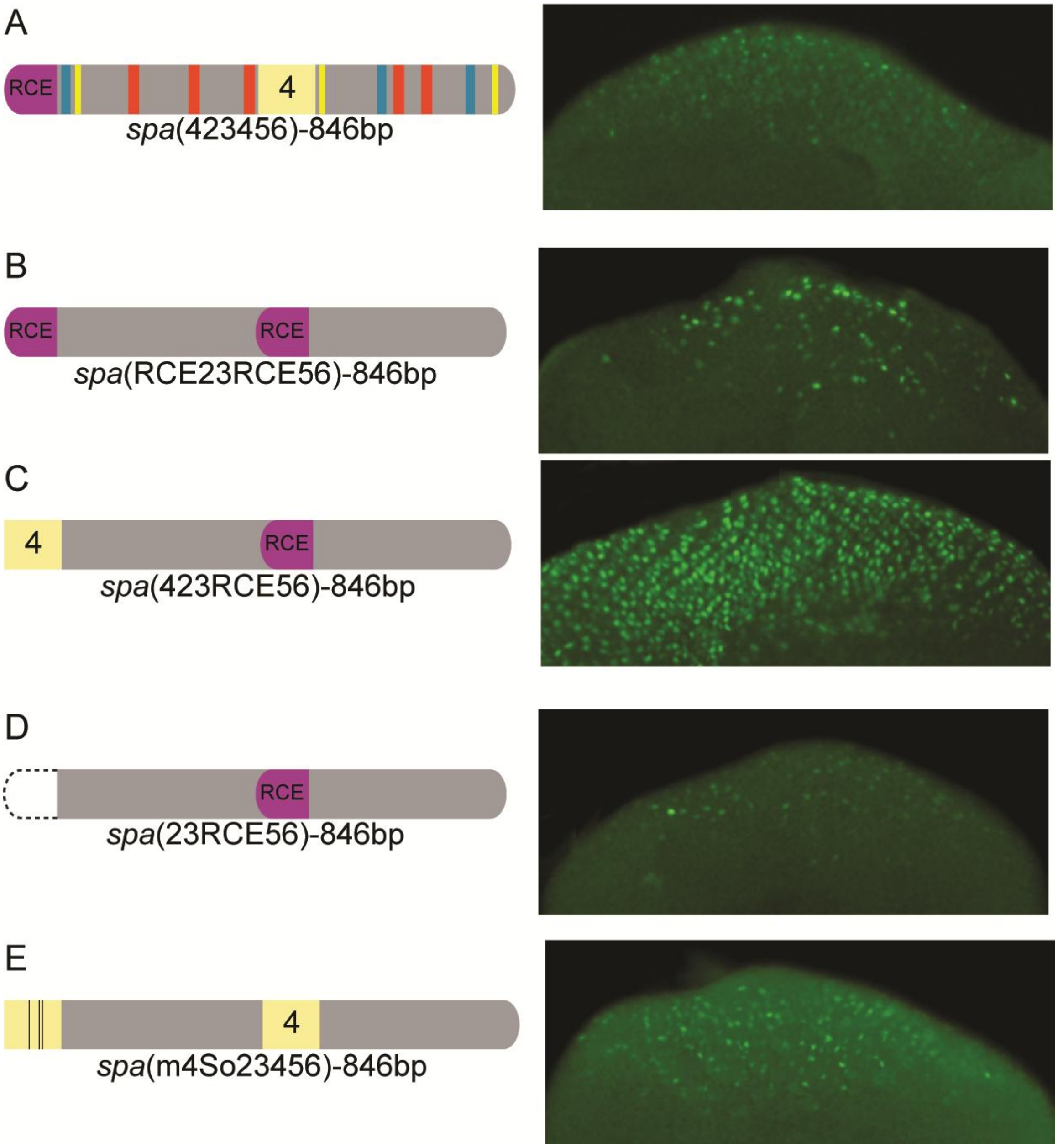


Figure 4.5 The *sparkling* RCE and region 4 can perform similar functions. We have observed that the *spa* region 4 can substitute for the RCE when present in two copies. Similarly, the RCE can substitute for region 4 when it is placed in the native region 4 location (B). Furthermore, the RCE and region 4 sequences can be exchanged and the enhancer can activate gene expression (C). In this context, the region 4 sequence can be removed and still lead to GFP expression (D) suggesting the RCE can function as a long-range element and to pattern cone cell expression from this location. The ability of region 4 to substitute for the RCE is not dependent on the 5' Sine oculis binding site (E).

activation must only be in the genomic context of the DNA sequence inputs surrounding the wildtype region 4 location. Alternatively, region 4 could possess a binding site for a protein that represses gene expression in photoreceptors. If this repression must act locally, moving region 4 to the 5' end of the enhancer would abolish this repression. Interestingly, the ectopic photoreceptor expression observed in the larval eye discs is absent in pupal eye discs (data not shown), suggesting either the proteins that stimulate photoreceptor expression are not expressed in pupal tissues, or that *sparkling* possesses other mechanisms to repress gene expression in pupal tissues. As photoreceptors and cone cells both express the known regulators of *spa* Su(H), PntP2, Yan and Lz, as well as Sine oculis, *spa* must possesses mechanisms to repress expression in photoreceptors. As *dPax2* expression in photoreceptors is detrimental to the fly, it is not unlikely that *spa* utilizes several methods to ensure this does not occur (Shi and Noll, 2009).

We have shown that the RCE and region 4 can substitute for each other (RCE23RCE56 and 423456) and that their position can be exchanged (423RCE56) in the promoter distal position. We also know that region 4 cannot act from the 5' position alone to drive cone cell specific gene expression (423m4^{rs}56). Together, these data combined with what we already know about *spa* function, suggests that at a distance *sparkling* requires an input from either of these sequences at both positions. To further test this model of *spa* activity

Figure 4.6

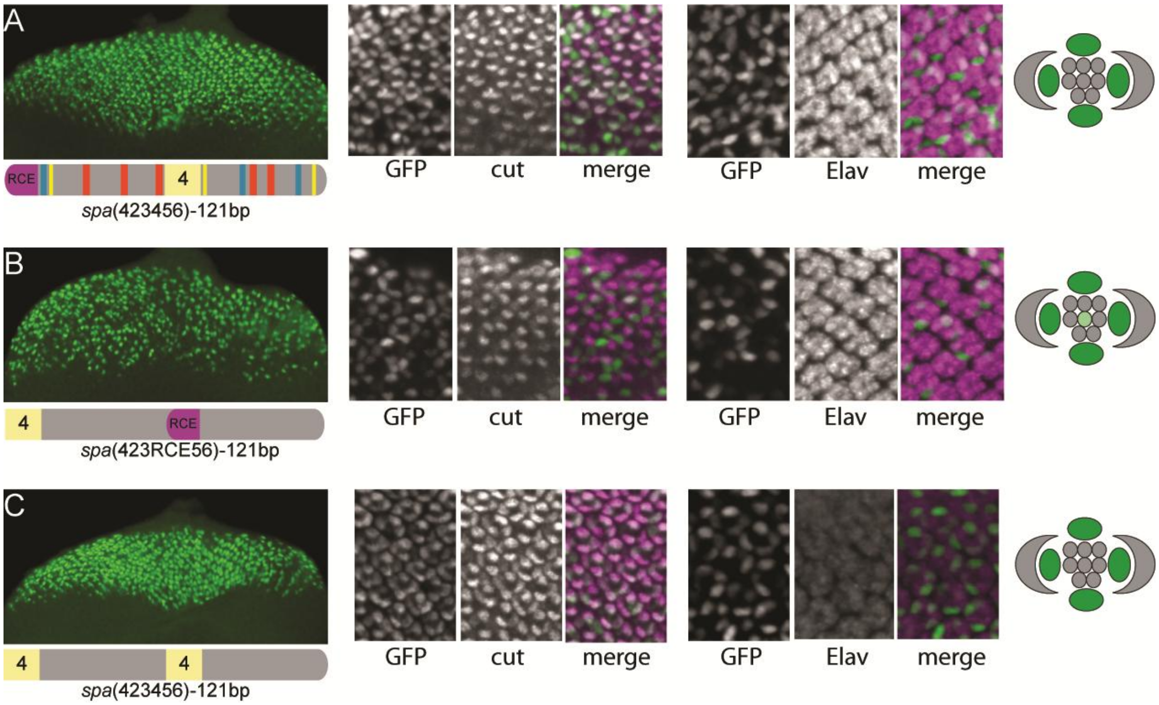


Figure 4.6 When the *sparkling* RCE and region 4 sequences are exchanged, the enhancer drives ectopic expression in photoreceptors. Regardless of position (promoter proximal depicted here) the *spa* (423RCE56) constructs drives expression in cone cells and in sporadic photoreceptors (B). Co-staining of larval eye discs with GFP and Cut demonstrates activity in cone cells while co-staining of GFP with Elav demonstrates enhancer activity in photoreceptors. Neither the wildtype *spa* (RCE23456) nor *spa* (423456) drive inappropriate GFP expression (A and C).

we examined a reporter construct lacking 5’ RCE or 5’ region 4 with the RCE in the place of the wildtype region 4 (23156). We were surprised to find that *spa*(23RCE56) -846 is capable of driving GFP expression (Figure 4.5D). These results indicate that the RCE is capable of performing its long-range function from the middle of the *sparkling* enhancer. Additionally, when the RCE is in this position it can also provide proper patterning information. Recall that *spa*(RCE23m4^{rs}56) -846bp does not function to activate reporter gene expression (Swanson et al., 2010). Therefore, in order to activate gene expression, *sparkling*, requires input shared between region 4 and the RCE in the position of region 4. This suggests proper enhancer activity depends on interaction between the input into this sequence and those in the surrounding DNA sequences. Furthermore, for its long-range activity *spa* requires EITHER the RCE in this position OR two copies of region 4.

We have postulated that the RCE and region 4 act at least in part through interaction the transcription factor Sine oculus. In order to assess the role of this interaction *in vivo* we mutated the So site in the 5’ copy of region 4 in *spa*(423456) construct. We found that mutation of the So site does not affect enhancer activity (Figure 5.5 E). This result raises several possibilities. First, the enhancer does not require So to activate distal gene expression. Second, the three base pair targeted mutation does not abolish So binding *in vivo*, despite inhibiting DNA interactions *in vitro*. Third, So, or another protein, can bind to a different region of the enhancer in the absence of this 5’ site and compensate for its loss.

4.3d Sine oculis interacts with spa region 5 possibly through a homeodomain binding site

In order to determine whether Sine oculis can interact with the *sparkling* enhancer outside of the region 4 and RCE interactions we designed seven additional EMSA probes that together with the RCE and region 4 probe span the entire *spa* enhancer (Figure 4.7). We next performed gel shifts in which the labeled probe is always the So binding site from the Hh eye disc enhancer. Then we asked what DNA sequences could compete for interaction with this probe. As seen previously, the HhSo probe binds to and shifts So protein (Figure 4.7, lanes 1 – 3). This interaction is specific as it is competed for by the wildtype HhSo probe, but not the mutant probe (Figure 4.7, lanes 4 and 7). The RCE and region 4 probes do compete for So binding (Figure 5.7, lanes 6 and 11). However *spa* probes 1 – 3 and 7 and 8, which cover *spa* regions 2, 3, 6a, and 6b, are incapable of competing for So binding (Figure 5.7, lanes 8 – 10 and 14 – 15). This suggests So does not interact with the non-essential *spa* regions 2, 3, and 6b, or the essential sequences of region 6a. *spa* probes 5 and 6 did not compete for So binding, but rather the position of the gel shift moved higher on the gel upon addition of these two probes (Figure 5.7, lanes 12 and 13). As the labeled probe and protein source are the same in all of these reactions, the super shift we observed must be due to the identity of the cold competitors. The HhSo labeled probe is still binding So resulting in a gel shift; however the cold competitors must also interact with the labeled probe resulting in further retardation on the gel. Such an interaction between the labeled probe and the

Figure 4.7

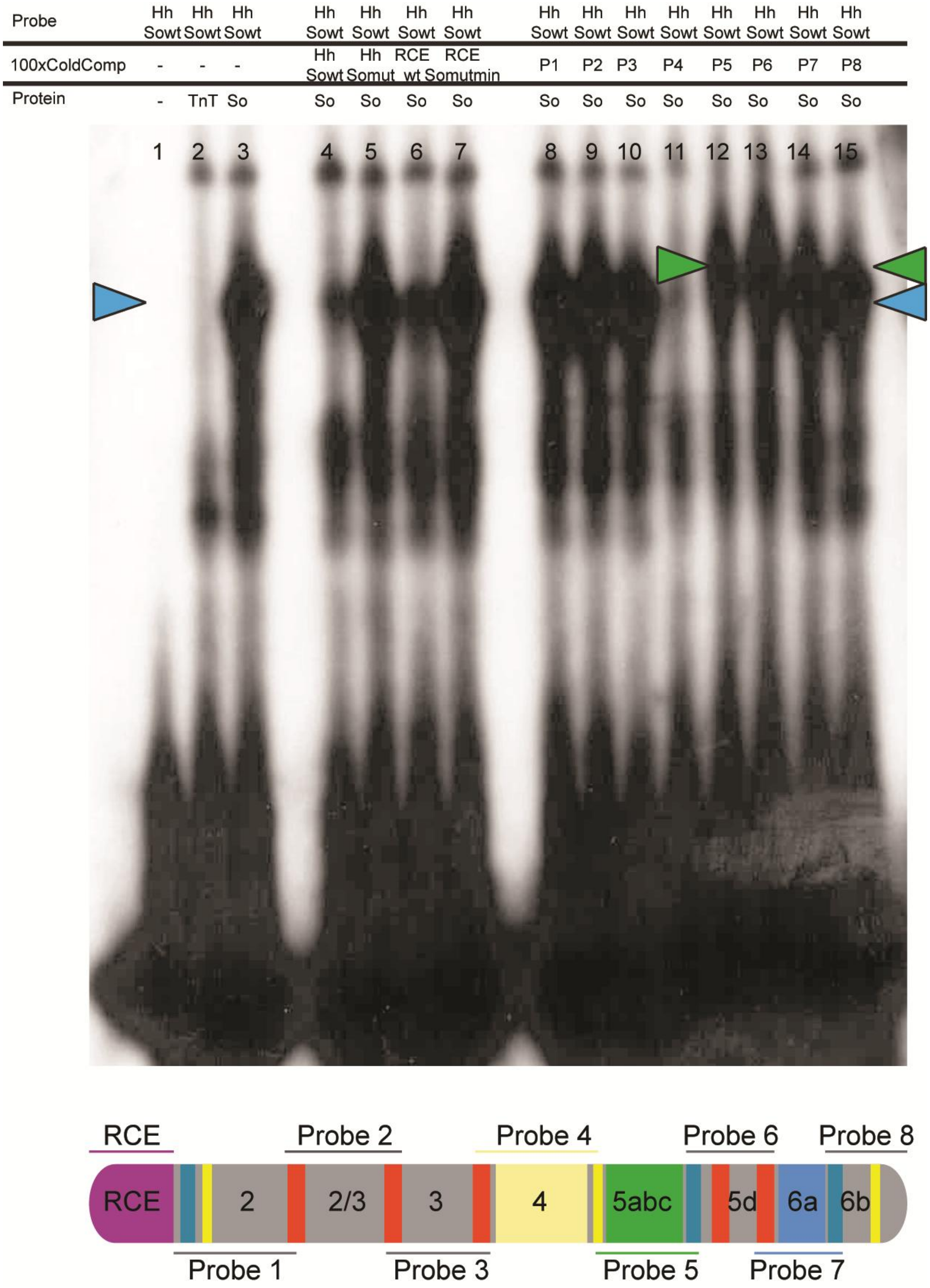

Probe
Hh Sowt Hh Sowt Hh Sowt
Hh Sowt Hh Sowt Hh Sowt Hh Sowt
Hh Sowt Hh Sowt Hh Sowt Hh Sowt Hh Sowt Hh Sowt Hh Sowt Hh Sowt
100xColdComp
- - -
Hh Sowt Hh Somut RCE wt RCE Somutmin
P1 P2 P3 P4 P5 P6 P7 P8
Protein
- TnT So
So So So So
So So So So So So So So
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
RCE
Probe 2
Probe 4
Probe 6
Probe 8
RCE
2
2/3
3
4
5abc
5d
6a
6b
Probe 1
Probe 3
Probe 5
Probe 7

Figure 4.7 *sparkling* region 5 can interact with Sine oculis protein in conjunction with a So binding site. We designed gel shift probes to span the entire *spa* enhancer (B), and assessed their ability to compete for So interaction with the HhSo wildtype probe. The HhSo probe binds to and shifts So protein (A, lane 3). This shift can be competed for by the HhSo wildtype, RCE, and region 4 probes, but not the mutant versions of each probe (A, lanes 4-7 and 11). *spa* probes 1 – 3 and 7 and 8 are unable to compete for ability of the HhSo probe to interact with So protein (lanes 8, 9, 10, 14, 15). As these probes span *spa* regions 2, 3, 6a, and 6b, it is unlikely these regions contain So input (B). *spa* probes 5 and 6 do not compete for So protein binding either, but do cause the labeled HhSo probe to supershift on the gel (A, lanes 14 and 15). s*pa* probes 5 and 6 span regions 5 and 5d which both contain homeodomain binding sites (HD; C).

cold competitors is likely mediated through the binding of So to both sequences either through two DNA binding domains in one protein, or the formation of dimers between two So molecules. Notably, the *spa* probes 5 and 6, which span the essential region 5 and nonessential region 5d, both contain homeodomain binding sites. Recall that *spa* region 4 and the RCE contain conserved homeodomain binding sites 7-8 bp away from the putative So binding sites (Figure 4.2 A). As So possesses a homeodomain DNA binding domain, it is plausible So could bind to DNA through the traditional homeodomain binding site (TAAT). Then, if So were bound to the HhSo site and the region 5 and 5d homeodomain sequences at the same time, we would observe a supershift in EMSA's. This explanation of our gel shift results is predicated on three assumptions. First, that in the presence of a "good" So binding site the DNA/So interaction is completely competed away. The RCE and region 4 probes also contain homeodomain binding sites in addition to putative So binding sites. When these probes are used to compete for HhSo interactions we see a complete competition rather that a super shift (Figure 5.7). Second, So binding to homeodomain sites must depend on more information than the classic TAAT sequence, as other probes we have used, such as region 4 mutant b or RCE mutant So contain intact HD binding sites but cannot compete for So binding and do not result in a supershift of the labeled probe (Figure 5.4). Third, in the DNA interaction that results in the observed supershift, So must interact first with a "good" So binding site and second with the homeodomain binding site as region 5 probes are unable to shift So alone (data not shown).

The observation that So may interact with two DNA elements at the same time combined with its large consensus DNA binding site (13 bp) led us to question whether the known DNA binding domain of So can bind DNA alone and if any other regions of the protein have DNA binding capabilities. To this end, we generated an *in vitro* transcribed and translated So homeodomain only (Hazbun et al., 1997). Side by side with this reaction, we also performed a TnT reaction with a fluorescent labeled lysine, which is incorporated into the nascent protein during translation. Running this reaction on an SDS-page gel followed by fluorescent imaging showed a band at the same approximate location as the So homeodomain based on predicted molecular weight, indicating we successfully generated So DNA binding domain protein (Figure 5.8 A, lanes 5-7). Unfortunately, the full length So protein runs at the same size as a known protein in the TnT reaction that autofluoresces at the same wave length as the fluorescent lysine (ref, Figure 5.8 A, lanes 1-4). Intriguingly, the wildtype HhSo probe does not interact with the So homeodomain, indicating the DNA binding domain alone is insufficient to bind DNA (Figure 5.8 B).

Figure 4.8

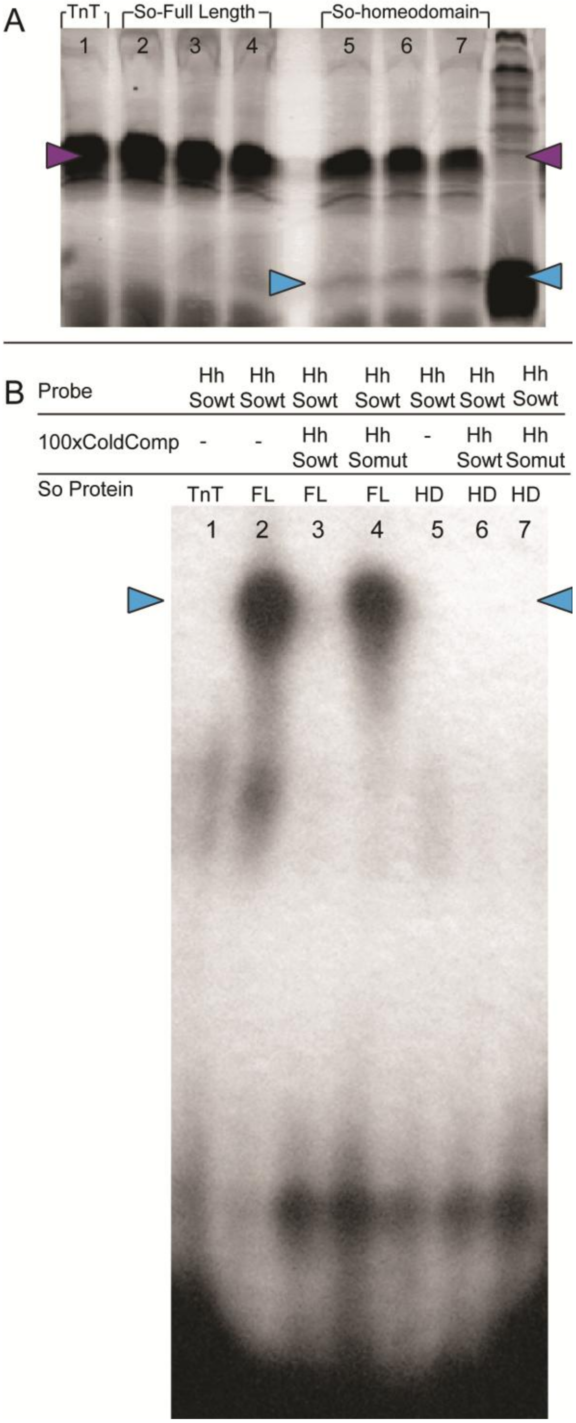

A
TnT
So-Full Length
So-homeodomain
1 2 3 4 5 6 7
B
Probe
Hh Sowt Hh Sowt Hh Sowt Hh Sowt Hh Sowt Hh Sowt Hh Sowt
100xColdComp
- - Hh Sowt Hh Somut - Hh Sowt Hh Somut
So Protein
TnT FL FL FL HD HD HD
1 2 3 4 5 6 7

Figure 4.8 The Sine oculis homeodomian is insufficient to bind DNA alone. We performed side-by-side *in vitro* transcription and translation reactions where one reaction was supplemented with fluorescently labeled lysines, which were integrated into newly translated proteins. After running on an SDS-PAGE gel and imaging, we confirmed So homeodomain protein is made by the TnT reaction (A, lanes 5 -7). The full-length So protein is obscured by an auto-flourescent band know to be present in the TnT protein mixture (A, lanes 1-4). We then utilized the So homeodomain protein (SoHD) in *in vitro* gel shift assays. The HhSo wildtype probe interacts with full-length Sine oculis protein (B, lane 2). This shift is abolished with the addition of 100x unlabeled wildtype probe (lane 3), but not with the mutant competitor (lane 4). However, the HhSo probe does not interact with the So homeodomian alone (B, lane 6).

4.3e sparkling regions 5 and 6 can also activate gene transcription at a distance

The observation that So might interact with *spa* region 5 prompted us to examine region 5's ability to compensate for loss of the RCE in the same manner as region 4. We found that *spa*(523456) is indeed able to drive GFP expression from the promoter distal position (-846), albeit at slight decreased levels compared to *spa*(wt) (RCE23456) and *spa*(423456) (Figure 5.9 G). Interestingly, this same arrangement drives GFP expression in the promoter proximal position at slightly diminished levels as well (Figure 5.9 J). This data suggests *spa* region 5 can compensate for loss of the RCE at a distance. However, as *spa*(523456) drives less expression than *spa*(23456) which contains no upstream sequence, region 5 must also contain sequences that are detrimental to *spa* activity.

We also examined the ability of *spa* region 6, which has no indicated So input, to drive reporter gene activity in the place of the RCE (623456). We again

Figure 4.9

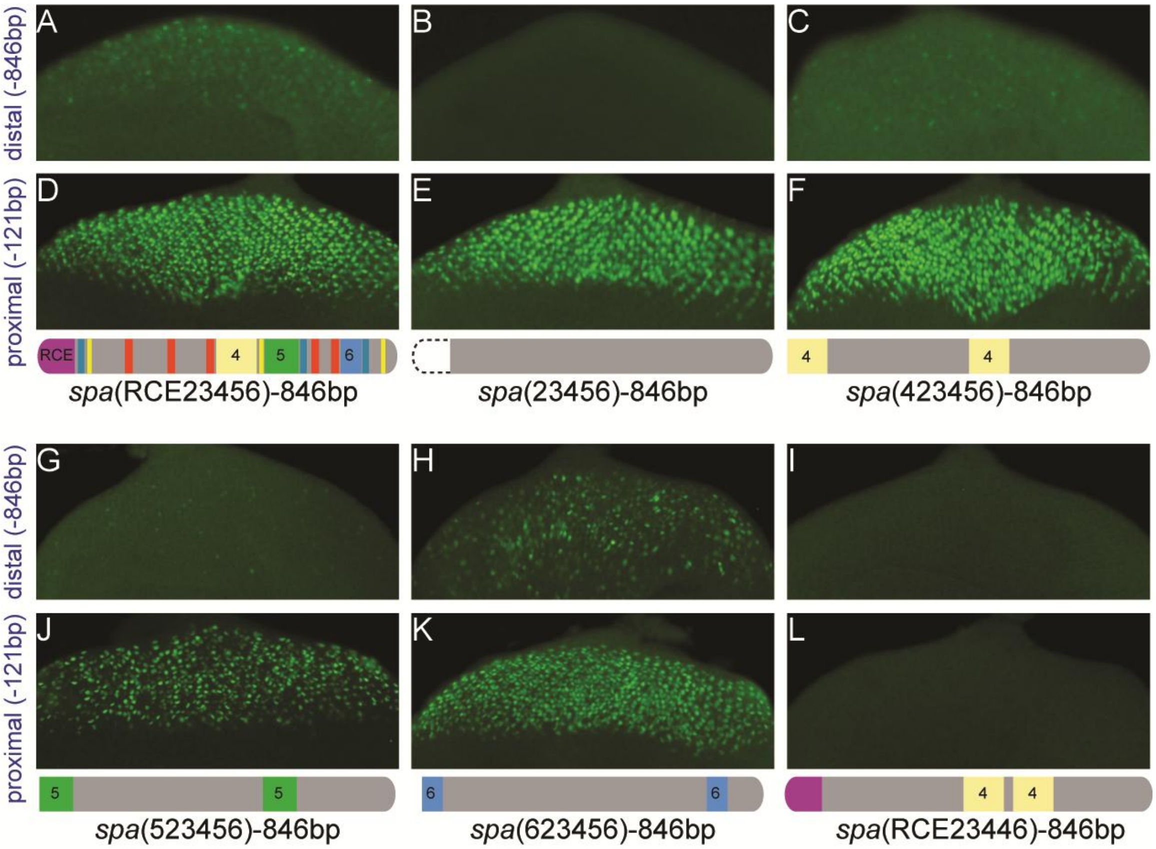


Figure 4.9 *sparkling* regions 5 and 6 can enable distal enhancer activity. We know that *sparkling* region 4 can substitute for the RCE's distal enhancer activities when present in two copies (Figure 9 A – F). Using the same experimental approach, we found that *spa* region 5 can also substitute for the RCE at a distance, although at lightly decreased levels (G). Even though *spa* does not require and 5' sequence (RCE) at the promoter proximal position (-121 bp) *spa*(523456) actually drives diminished expression in this position (J). Similarly, *spa* region 6 can also drive even higher levels of GFP expression than *spa*(wt) -846 or *spa*(423456) -846 (H). At -121 bp *spa*(623456) drives only slightly higher levels of gene expression than the wildtype enhancer (K). The ability for *spa* regions to compensate for each other is not universal as region 4 cannot substitute for region 5, regardless of position with respect to promoter (I and L).

found that *spa*(623456) is capable of activating gene expression from the distal position, although here it is at higher levels that *spa*(RCE23456) or *spa*(423456) (Figure 4.9 H). In the promoter proximal position these expression levels do not increase dramatically, but are only slightly less than wildtype levels at expression (Figure 4.9 K). Based on expression pattern alone, *spa*(623456) is expressed in both cone cells and a subset of photoreceptors. This is most easily observed by the presence of GFP positive cells anterior to the distinctive cone cell rosettes (Figure 4.9H). Together, our observations indicate that *sparkling* regions 4, 5, and 6 can all perform RCE-like activities. It should be noted however, that not all *spa* sequences can compensate for each other as region 4 cannot substitute for region 5 and region 5 cannot substitute for region 4 (Figure 4.9 K,L and data not shown).

4.4 Discussion

Motif analysis comparing *spa* region 4 and the RCE identified a putative Sine oculis binding site in each of these essential *spa* sequences. This observation led us to assess the ability of So to interact with the *sparkling* enhancer *in vitro* and the role of these sites in the *in vivo.* During the course of this work we also characterized the structural flexibility of the *spa* enhancer in both the promoter proximal and promoter distal locations.

4.4a Both strict and loose rules govern sparkling enhancer structure

There are two major models of the rules that govern enhancer structure, the enhancesome model which predicts structure and input identity is rigid, and information display which predicts binding site identify and arrangement are flexible (Arnosti and Kulkarni, 2005). In order to test these models with respect to the *sparkling* enhancer we generated reporter constructs in which the essential *spa* regions, RCE, region 4, region 5, and region 6, were substituted for one another and/or swapped in location. We already know that the RCE can be moved from the 5' to the 3' end of the enhancer and well as a significant distance upstream of the enhancer and still facilitate enhancer activity, suggesting that the location of the long-range inputs into the enhancer is very flexible (Swanson et al., 2010). As *spa* region 4 and the RCE are responsible for drastically different roles within the enhancer, we were astounded to find that these sequences are surprisingly interchangeable. Region 4 can substitute for the RCE (423456). The RCE can substitute for region 4 (RCE23RCE56). The position of the two sequences can be exchanged within the enhancer (423RCE56). These observations suggest that the arrangement of region 4 and the RCE sequences is flexible implicating an information display model of *spa* function. However, *spa*(423RCE56) drives expression in photoreceptors, suggesting that this rearrangement is not tolerated for repression of *spa* in non-cone cell types. It also appeared that *spa* requires both the two copies of either the RCE or region 4 as a distal enhancer lacking either sequence is unable to drive reporter gene activity. Interestingly we found that this is not the whole story. When the RCE sequence is in the wildtype region 4 position, the enhancer does not need a second input from either the RCE or region 4 (23RCE56). Therefore, the RCE sequence alone is able to drive both long-range transcription and proper cone cell patterning, but only through short range interactions with DNA inputs near the region 4 enhancer position. Conversely, in order for region 4 to drive long-range transcription and proper cone cell pattering it must be present in two copies.

The previous observations suggest that the structure of the *spa* enhancer is flexible with a small number of inputs that must be ordered correctly to ensure repression in photoreceptors. *sparkling* regions 5 and 6 play additional unique roles in enhancer activity. Both sequences are required for robust initiation of gene expression. Meanwhile, region 5 is required for the repression of *spa* activity in photoreceptors. Given the roles role of each of these regions we were surprised to find that both of the sequences can stimulate transcription from the position of the RCE. Unlike region 4 however, these reporters did not drive wildtype levels of expression as *spa*(523456) drives decreased GFP expression and *spa*(623456) drives increased levels of GFP expression. The observation that all of the sequences within *spa* known to posses the ability to activate gene expression, as well as a UAS site, can substitute at least in part for the RCE function at a distance leaves us to wonder whether the RCE performs long-range functions as we previously thought, or whether the enhancer is an additional activation input at this distance that is not required in the promoter proximal position. We can test this by substituting the binding sites for transcription activators expressed in the eye, such as grainyhead, for the RCE sequence (Uv et al., 1997). This conclusion is unlikely however as the region 5 and 6 do not drive proper GFP expression from the position of the RCE. Furthermore, if *sparkling's* ability to function at a distance compared to in the promoter proximal position were simply dictated by the number of activator inputs, we would expect that any construct at -121bp that is missing only one of these essential sequences (i.e. region 4, 5, or 6) would drive the same, wildtype, levels of gene expression as *spa*(ΔRCE)-121bp, which is simply not the what we see in our reporter constructs (Swanson et al., 2010).

While we have observed flexibility within the *spa* structure, certain enhancer rearrangements are not tolerated. For example, region 4 cannot take on the role of region 5 (RCE23446). Similarly, while both region 4 and 5 can act for the RCE when present in a second copy, region 5 cannot take on the role of region 4. As regions 5 and 6 can act in the position of the RCE, it will be interesting to see whether the RCE can also take the place of region 5 or 6 within the enhancer (RCE234RCE6 and RCE2345RCE). Our observations suggest that *spa* structural organization is flexible for some inputs and constrained for others, indication that *spa* takes on both enhanceosome and information display modes of function. Furthermore, while several of our rearrangements drive GFP expression, it is not always at wildtype levels. These changes in expression may not be tolerated in the endogenous context of *sparkling* activity.

4.4b Sine oculis may regulate the sparkling enhancer

The ability of *spa* region 4 to substitute for the RCE, led us to examine the shared motifs between these two sequences. We found that both sequences contain putative binding sites for the transcription factor Sine oculis. Interestingly, the *spa* sequence containing the So binding sites within these regions had each been mutated previously and shown to be essential for *spa* activity (RCEa and 4b) (Swanson et al., 2010) further implicating an So interaction with these regions. Indeed, So protein does bind to both region 4 and the RCE *in vitro*. We were therefore surprised to find that the So binding site in *spa*(423456)-846bp, is not required for enhancer activity. As the region 4 So mutation and the 4b mutation only differ by 3 basepairs, we were surprised that loss of this binding site had no affect on reporter gene expression. It is possible that So is still able to bind to and facilitate this enhancers activity, but it is equally, if not more likely, that the enhancer does not require So input to function. We will further investigate the role of So binding in RCE function by mutating all of the So binding sites within the wildtype enhancer as well as making the larger (mut max) targeted mutation to ensure loss of So interaction (Pauli et al., 2005).

We observed a clear association between Sine oculis binding sites and the classic homeodomain binding site (TAAT) in the *sparkling* enhancer. We were able to identify So binding sites, that each deviate from the consensus binding site by 2 basepairs, in the 5' end of the DNA sequence we suspect corresponds to *sparkling* enhancer sequence in 11 *Drosophila* species. In every species other than *D. mojavensis,* ab HD binding site lies 7 basepairs downstream, while in *D.*

mojavensis the HD site is 8 basepairs downstream. The HD site in *spa* region 4c is also 8 basepairs downstream from the So site in region 4b. Given the relationship between these sites, it is possible that So interacts cooperatively with one of the many retinal homeodomain proteins such as eyeless or twin of eyeless, to bind the *sparkling* enhancer. We can test the ability of these proteins to interact with *sparkling* together in *in vitro* gel shift assays, and the importance of the spacing between the sites using *in vivo* reporter constructs. At least *in vitro* So is able to interact with the *spa* enhancer alone; however this may not be true *in vivo* or So may be required to recruit a homeodomain binding protein to the *spa* enhancer.

It is also possible that Sine oculis itself is able to interact with both the So consensus site and the HD site. This could occur through the formation of So dimers on the DNA. Alternatively, So, like other homeodomain binding proteins, may contain two DNA binding domains (Czerny et al., 1999), which could interact with the So and HD sites together. We have seen further evidence that So might interact with HD sites in the *spa* enhancer in the ability of region 5 and 5d probes to supershift the So protein interaction with HhSo probe in gel shift assays. We require further evidence that this is indeed occurring in gel shifts using targeted mutations to each these probes. However, if So is able to bind the *spa* HD sites in addition to the putative So binding sites, it must require an interaction with the So binding site first, as several probes containing HD sites do not shift So protein alone in gel shift assays. As we were curious about the mechanisms by with the So protein binds DNA, we performed gel shift assays using full length So and homeodomain on So. We found that while full length So can interact with the HhSo probe, the homeodomain alone cannot, suggesting So requires additional protein sequences in order to interact with DNA. We will perform the same gel shift assays with the So N-terminal domain, C-terminal domain, and Six interaction domain, as well as combinations of these domains, to determine the DNA binding requirements of the protein.

While we hypothesize that So interacts with the RCE and region 4 to facilitate *spa* enhancer activity, there is no experimental evidence to implicate So in long-range gene regulation, or known protein interactions to suggest a mechanism by which it would act to facilitate RCE activity. If we conclude So protein is vital for *sparkling* activity and RCE function we will investigate the mechanisms by which it acts by first performing affinity purification with So protein to determine possible protein interacting partings. So is known to interact with the non-DNA binding protein Eya to regulate transcription and cell differentiation in the eye disc. As such we can also asses the role of Eya in *spa* activity, *dPax2* expression, and cone cell development.

4.5 Experimental methods

4.5a Preparation of in vitro transcribed and translated proteins

Full length *Sine oculis* (So) cDNA was obtained from the Drosophila Genomics Resource Center. In order to express So full length and homeodomiain proteins, SP6 promoter and Kozak sequences were added to the 5' end of cDNA and two stop codons and a polyA tail were added to the 3' end by PCR. The PCR primers are listed below:

Fwd primers: 5'-gtaatatatttaggtgacactatagaacagaccacc-20bpcDNA specific sequence – 3'

Rev primers:5'-ttttttttttttttttttttttttttttctatca-20bpcDNA specific sequence – 3'

Additionally a point mutation in the cDNA sequence that resulted in a truncated form of the protein was corrected by sewing PCR.

Next proteins were expressed from the PCR products using the TnT®SP6 High-Yield Wheat Germ Protein Expression System according to the kit's protocol (Promega). Protein was aliquoted and stored at -80^{o}C.

The *in vitro* transcription and translation reactions were also performed in the presence of the a BODIPY®-FL conjugated lysine using the FluoroTect^{TM} Green$_{Lys}$ in vitro Translation Labeling System according to the kit's protocol (Promega) combining 0.4ul Green$_{Lys}$ reagent to 1ul, 2ul, or 3.6ul of PCR product in a 10μl TnT reaction. After translation, samples were combined with 40ul 1xSDS-PAGE loading buffer (50 mM Tris-HCl pH 6.8, 100 mM DTT, 2% SDS, 0.1% Bromophenol Blue, 10% glycerol), denatured at 70°C for two minutes and run on a 18% SDS-PAGE resolving gel, with a 5% stacking gel. Fluorescent protein was imaged utilizing a Typhoon®8800 (GE Healthcare) with a 532nm excitation laser.

4.5b Electrophoretic mobility shift assays

Labeled probes were annealed and labeled by incubating 37µl dH_2O, 5µl10xPNK buffer, 1µl each top and bottom strand oligos (2µM), 5µl γ^{32}P-ATP, and 1µl T4PNK (New England Biolabs) at 37^{o}C, boiling at 80^{o}C for 5 minutes, and allowing samples to cool to room temperature (about 1 hour) to allow oligos to anneal. Labeled double stranded probes were then purified twice using GE ProbeQuant G-50 spin columns. 100 fold excess cold competitors were prepared by combining 39µl dH_2O, 5µl10xPNK buffer, and 1µl each top and bottom strand oligos (200µM). Competitors were then prepared as with the labeled probe except they were not purified after annealing. Probe sequences used are shown below, mutated sequences are depicted in uppercase.

HhSo: 5’ tataaacaatgatatcgaattaccagagtttcg 3’

HhSomut: 5’ tataaacaatgataCcgaaGtaAcagagtttcg 3’

RCESowt: 5’ tagttggaagtgtatcaagtaactgggtgccta 3’

RCEm1a: 5’ tagttggaagtgGaGcCaTtCactgggtgccta 3’

RCESomutmax: 5’ tagttggaagtgtCCCaagGGGctgggtgccta 3’

RCESomutmin: 5’ tagttggaagtgtaCcaagGaaAtgggtgccta 3’

pseSo: 5’ ttgtatgaaatgtctcaaataacttcgtgtcta 3’

pseSomutmax: 5’ ttgtatgaaatgtCCCaaaGGGcttcgtgtcta 3’

Region 4wt: 5’ ttgaaattgaagcactattggtgtacgattacaacgctcacattatcagg 3’

Region4mut:5’ttgaaCtGgCaTcCcGaGtTgGgGaAgCtGaAacCaAgAtAaAaGtCtcagg 3’

Region 4bmut: 5’ ttgaaattgaagcactattTgGgGaAgCtGaAacgctcacattatcagg

Region 4cmut: 5’ ttgaaattgaagcactattggtgtacgattacCaAgAtAaAaGtCtcagg

DNA/protein interactions (gelshifts) were performed with 1µl labeled probe, 3µl of TnT generated protein, 1µ 10xgelshift buffer (0.1 M Tris HCl ph 7.5, 0.5 MNaCl, 10 mM DTT, 10 mM EDTA, 275 µg/ml salmon sperm DNA), 1µl polyd(I-C) (1mg/ml), 1µl DTT (100µM) and dH_2O to a final volume of 10µl. Reactions with cold competitors used 1µl annealed cold competitor in place of 1µl of dH_2O. Reactions were then incubated on ice for 15 minutes and then loaded on 4% polyacrylamide gels and run for 4 – 5 hours in 0.5x TBE at 120 volts. Completed gels were then vacuum dried at 80^oC for 1 – 2 hours and finally exposed to film.

4.5c Reporter gene generation construction and transgenesis

Sparkling enhancer sequences with 846 or 121bp spacers were amplified by PCR such that they contain a 5’ Cacc for directional Topo cloning and an EcoRI restriction enzyme site at their 5’ end and a BamHI restriction site at the 3’ end. Alterations the 5’ end of the enhancer sequence were generated using standard PCR primers containing the desired sequences. Internal alternations were made by sewing PCR. Enhancers were subsequently ligated into the pHStinger GFP reporter vector via EcoRI and BamHI digestion. P-element transformation was performed in w^{1118} flies as described previously (Rubin and Spradling, 1982)

4.5d Tissue preparation, antibody staining, microscopy

Eye disc tissues were dissected from third instar larvae. Disc tissues were then fixed in 4% paraformaldehyde at room temperature for 30 minutes. Discs were then washed 3x5 minutes in 1x PBS and mounted in Prolong Gold with 4’, 6’- diamidino-2 phenylidole (DAPI) (Invitrogen). Imaging was performed on an OlympusBX51 mircroscope with an Olympus DP70 digital camera.

Immunohistochemistry was performed on dissected eye discs from 24 hour pupa. Discs were fixed in 4% paraformaldehyde for 30 minutes at room temperature and then washed 3x 10 minutes in PBS-Tx (1xPBS + 0.1% Triton x-100). Fixed discs were then incubated in PBS-Tx + 2% for 1-3 hours and then incubated overnight in primary antibodies against GFP (Invitrogen) and Cut or Elav (*Drosophila* studies Hydrodoma Bank) diluted 1:100. The next day, tissues were washed 3x10 minutes with PBS-Tx and then incubated in secondary antibodies; goat anti-mouse 568 nm and goat anit-rabbit 488 nm IInvitrogen) diluted 1:1000. Finally, the discs were washed 3x20 minutes in PBS-Tx and mounted in Prolong Gold with DAPI (Invitrogen). Stained discs were imaged on an Olympus FLUO View 500 Laser Scanning Confocal microscope mounted on and Olympus 1x71 inverted microscope.

4.5e Sequence alignment and DNA motif analysis

The *sparkling* multi-species alignment is based on BLASTZ alignments and was taken from the UCSC genome browser (http://genome.ucsc.edu) .

Subsequently, potential transcription factor binding sites for candidate proteins were identified using genepallete (Rebeiz and Posakony, 2004). Finally, Sine oculis and homeodomain binding sites were mapped in *D. simulans, D, sechelichellia, D. melanogaster, D. yakuba, D. erecta, D. ananassae, D. psuedoobscura, D. persmililis, D. willistoni, D. mojavensis, D. virilis,* and *D. grimshawi* using a custom MATLAB program to identify specified deviant binding sites (details available upon request).

4.6 Acknowledgements

We would like to thank Amy Strom for her contributions to the *in vivo* aspects of this work and Christian Morrison for assistance with MATLAB programming.

4.7 References

Arnosti, D.N., Barolo, S., Levine, M., and Small, S. (1996). The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. Development *122*, 205-214.

Arnosti, D.N., and Kulkarni, M.M. (2005). Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? J Cell Biochem *94*, 890-898.

Blackwood, E.M., and Kadonaga, J.T. (1998). Going the distance: a current view of enhancer action. Science *281*, 60-63.

Blanco, J., Pauli, T., Seimiya, M., Udolph, G., and Gehring, W.J. (2010). Genetic interactions of eyes absent, twin of eyeless and orthodenticle regulate sine oculis expression during ocellar development in Drosophila. Dev Biol *344*, 1088-1099.

Bovolenta, P., Mallamaci, A., Puelles, L., and Boncinelli, E. (1998). Expression pattern of cSix3, a member of the Six/sine oculis family of transcription factors. Mech Dev *70*, 201-203.

Cheyette, B.N.R., Green, P.J., Martin, K., Garren, H., Hartenstein, V., and Zipursky, S.L. (1994). The drosophila sine oculis locus encodes a homeodomain-containing protein required for the development of the entire visual system. Neuron *12*, 977-996.

Czerny, T., Halder, G., Kloter, U., Souabni, A., Gehring, W.J., and Busslinger, M. (1999). twin of eyeless, a second Pax-6 gene of Drosophila, acts upstream of eyeless in the control of eye development. Mol Cell *3*, 297-307.

Daniel, A., Dumstrei, K., Lengyel, J.A., and Hartenstein, V. (1999). The control of cell fate in the embryonic visual system by atonal, tailless and EGFR signaling. Development *126*, 2945-2954.

Evans, N.C., Swanson, C.I., and Barolo, S. (2012). Sparkling insights into enhancer structure, function, and evolution. Curr Top Dev Biol *98*, 97-120.

Flores, G.V., Duan, H., Yan, H., Nagaraj, R., Fu, W., Zou, Y., Noll, M., and Banerjee, U. (2000). Combinatorial signaling in the specification of unique cell fates. Cell *103*, 75-85.

Giese, K., Kingsley, C., Kirshner, J.R., and Grosschedl, R. (1995). Assembly and function of a *TCRα* enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. Genes and Development *9*, 995-1008.

Halder, G., Callaerts, P., Flister, S., Walldorf, U., Kloter, U., and Gehring, W.J. (1998). Eyeless initiates the expression of both sine oculis and eyes absent during Drosophila compound eye development. Development *125*, 2181-2191.

Halder, G., Callaerts, P., and Gehring, W.J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. Science *267*, 1788-1792.

Hare, E.E., Peterson, B.K., Iyer, V.N., Meier, R., and Eisen, M.B. (2008). Sepsid *even-skipped* enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. PLoS Genet *4*, e1000106.

Harrison, S.D., and Travers, A.A. (1990). The tramtrack gene encodes a Drosophila finger protein that interacts with the ftz transcriptional regulatory region and shows a novel embryonic expression pattern. Embo J *9*, 207-216.

Hauck, B., Gehring, W.J., and Walldorf, U. (1999). Functional analysis of an eye specific enhancer of the eyeless gene in Drosophila. Proc Natl Acad Sci U S A *96*, 564-569.

Hayashi, T., Xu, C., and Carthew, R.W. (2008). Cell-type-specific transcription of prospero is controlled by combinatorial signaling in the Drosophila eye. Development *135*, 2787-2796.

Hazbun, T.R., Stahura, F.L., and Mossing, M.C. (1997). Site-specific recognition by an isolated DNA-binding domain of the sine oculis protein. Biochemistry *36*, 3680-3686.

Jaynes, J.B., and O'Farrell, P.H. (1991). Active repression of transcription by the engrailed homeodomain protein. Embo J *10*, 1427-1433.

Johnson, L.A., Zhao, Y., Golden, K., and Barolo, S. (2008). Reverse-engineering a transcriptional enhancer: a case study in Drosophila. Tissue Eng Part A *14*, 1549-1559.

Kumar, J. (2009). The sine oculis homeobox (SIX) family of transcription factors as regulators of development and disease. Cellular and Molecular Life Sciences *66*, 565-583.

Laughon, A. (1991). DNA binding specificity of homeodomains. Biochemistry *30*, 11357-11367.

Ludwig, M.Z., Patel, N.H., and Kreitman, M. (1998). Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change. Development *125*, 949-958.

Morcillo, P., Rosen, C., and Dorsett, D. (1996). Genes regulating the remote wing margin enhancer in the Drosophila cut locus. Genetics *144*, 1143-1154.

Pauli, T., Seimiya, M., Blanco, J., and Gehring, W.J. (2005). Identification of functional sine oculis motifs in the autoregulatory element of its own gene, in the eyeless enhancer and in the signalling gene hedgehog. Development *132*, 2771-2782.

Pignoni, F., Hu, B., Zavitz, K.H., Xiao, J., Garrity, P.A., and Zipursky, S.L. (1997). The eye-specification proteins So and Eya form a complex and regulate multiple steps in Drosophila eye development. Cell *91*, 881-891.

Pil, P.M., Chow, C.S., and Lippard, S.J. (1993). High-mobility-group 1 protein mediates DNA bending as determined by ring closures. Proc Natl Acad Sci U S A *90*, 9465-9469.

Rebeiz, M., and Posakony, J.W. (2004). GenePalette: a universal software tool for genome sequence visualization and analysis. Developmental Biology *271*, 431-438.

Rubin, G.M., and Spradling, A.C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. Science *218*, 348-353.

Seo, H.C., Curtiss, J., Mlodzik, M., and Fjose, A. (1999). Six class homeobox genes in drosophila belong to three distinct families and are involved in head development. Mech Dev *83*, 127-139.

Serikaku, M.A., and O'Tousa, J.E. (1994). sine oculis is a homeobox gene required for Drosophila visual system development. Genetics *138*, 1137-1150.

Shi, Y., and Noll, M. (2009). Determination of cell fates in the R7 equivalence group of the *Drosophila* eye by the concerted regulation of D-Pax2 and TTK88. Developmental Biology *331*, 68-77.

Swanson, C.I., Evans, N.C., and Barolo, S. (2010). Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. Dev Cell *18*, 359-370.

Swanson, C.I., Schwimmer, D.B., and Barolo, S. (2011). Rapid evolutionary rewiring of a structurally constrained eye enhancer. Curr Biol *21*, 1186-1196.

Thanos, D., and Maniatis, T. (1995). Virus induction of human *IFN-β* gene expression requires the assembly of an enhanceosome. Cell *83*, 1091-1100.

Uv, A.E., Harrison, E.J., and Bray, S.J. (1997). Tissue-specific splicing and functions of the Drosophila transcription factor Grainyhead. Mol Cell Biol *17*, 6727-6735.

Voas, M.G., and Rebay, I. (2004). Signal integration during development: insights from the *Drosophila* eye. Dev Dyn *229*, 162-175.

Xiong, W.C., and Montell, C. (1993). tramtrack is a transcriptional repressor required for cell fate determination in the Drosophila eye. Genes Dev *7*, 1085-1096.

Yan, H., Canon, J., and Banerjee, U. (2003). A transcriptional chain linking eye specification to terminal determination of cone cells in the Drosophila eye. Dev Biol *263*, 323-329.

CHAPTER 5

BIOCHEMICAL ANALYSIS OF THE *DPAX2 SPARKLING* ENHANCER

5.1 Abstract

cis-regulatory elements known as enhancers regulate gene expression at least in part through the recruitment of DNA binding proteins known as transcription factors (TFs). For example, the *dpax2* cone cell enhancer *sparkling* (*spa*) is regulated by the TFs Suppressor of Hairless (Su[H]) the Ets factors Pointed P2 (Pntp) and Yan regulated by the Notch and EGFP signaling as well as the eye specific TF lozenge (Lz). Further study demonstrated that this enhancer contains at least four additional critical regulatory sequences, refered to here as the RCE, region 4, region5, and region 6a. Given their size, each of the regions could contain one or more TF binding sites. Therefore, *spa* is likely to contain additional transcription factor binding sites, and, as each of the regions perform unique functions, *spa* likely contains binding sites for multiple different proteins. We investigated the protein binding capabilities of each of these essential enhancer sub-elements using two biochemical approaches. 1) We performed Electron Mobility Shift Assays, EMSA, (gel shifts) with region specific probes and nuclear proteins isolated from *Drosophila melongaster* embryos or specific candidate proteins generated by *in vitro* transcription and translation. 2)

The nuclear extract was also used in affinity purification with regions specific probes and subsequent analysis by mass spectrometry. Using these techniques we have been able to further identify and characterize the precise sequences within in these regions that are likely to facilitate their essential DNA-protein interactions. Furthermore, we identified proteins that are capable of interacting the *sparkling* RCE, potentially regulating its long-range enhancer function. The most promising candidate thus far to arise from this study is Taf6, a Tata-binding faction, which implicates a tracking as method of *spa* enhancer action.

BIOCHEMICAL ANALYSIS OF THE *dPax2 sparkling* ENHANCER

5.2 Introduction

cis-regulatory elements known as enhancers regulate tissue and temporal specific gene transcription. In order to achieve these very specific patterns of gene expression enhancers are thought to function, at least in part, through the recruitment of DNA binding proteins known as transcription factors (TFs). It is then the specific combination of TF binding that enables enhancers to regulate target gene transcription in tissue and temporal specific manner. For example, in the developing *Drosophila* eye the TFs Suppressor of Hairless (Su[H]) the Ets factors Pointed P2 (Pntp) and Yan regulated by the Notch and EGFP signaling pathways respectively, are utilized repetitively to specify the twelve unique cell types present in each *Drosophila* ommatidium (Voas and Rebay, 2004). Therefore, in order to specifiy each individual cell type, these factors must act in concert with additional inputs to induce specific cell fates. Enhancers must also interact either directly or indirectly, through the transcription factors, with numerous additional proteins in order to stimulate gene transcription. These include, but are not limited to: histones, to affect nucleosome positioning, histone modifying proteins such as methylases and deacetylases in order to alter local chromatin structure, and Mediator and/or the basal transcription machinery in order to stimulate transcription at the target promoter (Malik and Roeder, 2005; Narlikar et al., 2002; Orphanides and Reinberg, 2002; Szutorisz et al., 2005; Wang et al., 2005).

dPax2 is required in the *Drosophila* eye for specification of the cone cells, primary pigment cells, and mechanoscenory bristle cells (Fu et al., 1998; Fu and Noll, 1997). The enhancer responsible for activating *dPax2* expression in the cone cells of the eye imaginal disc lies in the 4th intron of the gene and is called *sparkling* (*spa*) (Flores et al., 2000; Fu et al., 1998). The 362bp minimal *spa* enhancer is capable of driving both *dPax2* cDNA and GFP reporter expression in the developing cone cells of the *Drosophila* third instar imaginal disc (Evans et al., 2012; Flores et al., 2000; Johnson et al., 2008; Swanson et al., 2010; Swanson et al., 2011). Unsurprisingly, *spa* is known to be regulated by both Su(H) and Ets factors as well as the eye specific TF lozenge (Lz),leading to the proposition of a combinatorial code in which *spa* regulated gene expression is the result of Lz + Ets + Su(H) inputs (Flores et al., 2000). However, further characterization of the *spa* enhancer in our laboratory revealed sequences outside the identified binding sites for these factors are also required for proper *spa* function (Swanson et al., 2010). Therefore, Lz + Ets + Su(H) does not sufficiently define the regulatory code for *spa* activity.

Mutagenesis of the remaining sequences within *spa* revealed at least four subregions of the enhancer are necessary for promoter gene regulation, which we denote as regions 1, 4, 5 and 6a. Regions 1, 4, 5 can be subdivided into three individual sequences (ie: 4a, 4b, 4c) that each contribute to *spa* activity

(Swanson et al., 2010). Furthermore, each region contributes to a unique function of the enhancer. Region 1 is essential for long-range transcriptional activation and therefore we refer to it as the “remote control” element or RCE. Region 4 can also promote long-range enhancer action and is absolutely critical for cone cell specific gene transcription. Regions 5 and 6 are essential for robust initiation of *spa* activity, but not necessary for maintenance of reporter gene transcription. Furthermore, region 5 is required to repress gene expression in the R1 and R6 photoreceptors (Swanson et al., 2010). Given their size, each of the regions could contain one or more TF binding sites. Therefore, *spa* is likely to contain additional transcription factor binding sites, and, as each of the regions perform unique functions, *spa* likely contains binding sites for multiple different proteins.

We have already identified Sine oculis (So) as a potential binding partner for the RCE and region 4. However, there is no known function of So that implicates it in long-range enhancer activity. As such, we cannot rule out interaction with a second protein to mediate RCE functions. Furthermore, targeted mutation of So binding sites does not appear to affect *spa’s* ability to regulate distal gene transcription (Chapter 4). As such, we decided to begin a mostly unbiased approach to identifying protein binding sites within, not only the RCE, but the entire *spa* enhancer. In order to assess the protein binding ability of the essential *spa* regions, as well as the potential identities of these proteins that interact with these regions, we have undertaken two biochemical approaches. 1) We performed Electron Mobility Shift Assays, EMSA, (gel shifts) with region specific probes and nuclear proteins isolated from *Drosophila melongaster* embryos or specific candidate proteins generated by *in vitro* transcription and translation. 2) The nuclear extract was also used in affinity purification with regions specific probes and subsequent analysis by mass spectrometry. It's important to point out the significant limitation of these experiments – which is the embryonic protein source. However, utilizing this source yields sufficient amounts of protein for biochemical study, while obtaining sufficient levels of protein from other tissues, such as the imaginal discs, is difficult. Furthermore, many proteins expressed during larval development are also expressed in the embryo, including Su(H), Lz, and PntP2/Yan the known regulators of *sparkling* (Lebestky et al., 2000; Price and Lai, 1999; Scholz et al., 1993; Schweisguth and Posakony, 1992)

5.3 Results

5.3a Identification and characterization of RCE interaction proteins

The *sparkling* RCE specifically facilitates long-range gene transcription. However, we also know that in the correct location, the RCE also contains sequences that can regulate proper gene levels and patterning (Chapter 4). Therefore, we expect the RCE can interact with both transcriptional activators and proteins capable of influencing distal gene regulation such as those that alter the three dimensional chromatin structure or recruit the basal transcription machinery. In order to assess the ability of the *sparkling* RCE to interact with

Figure 5.1

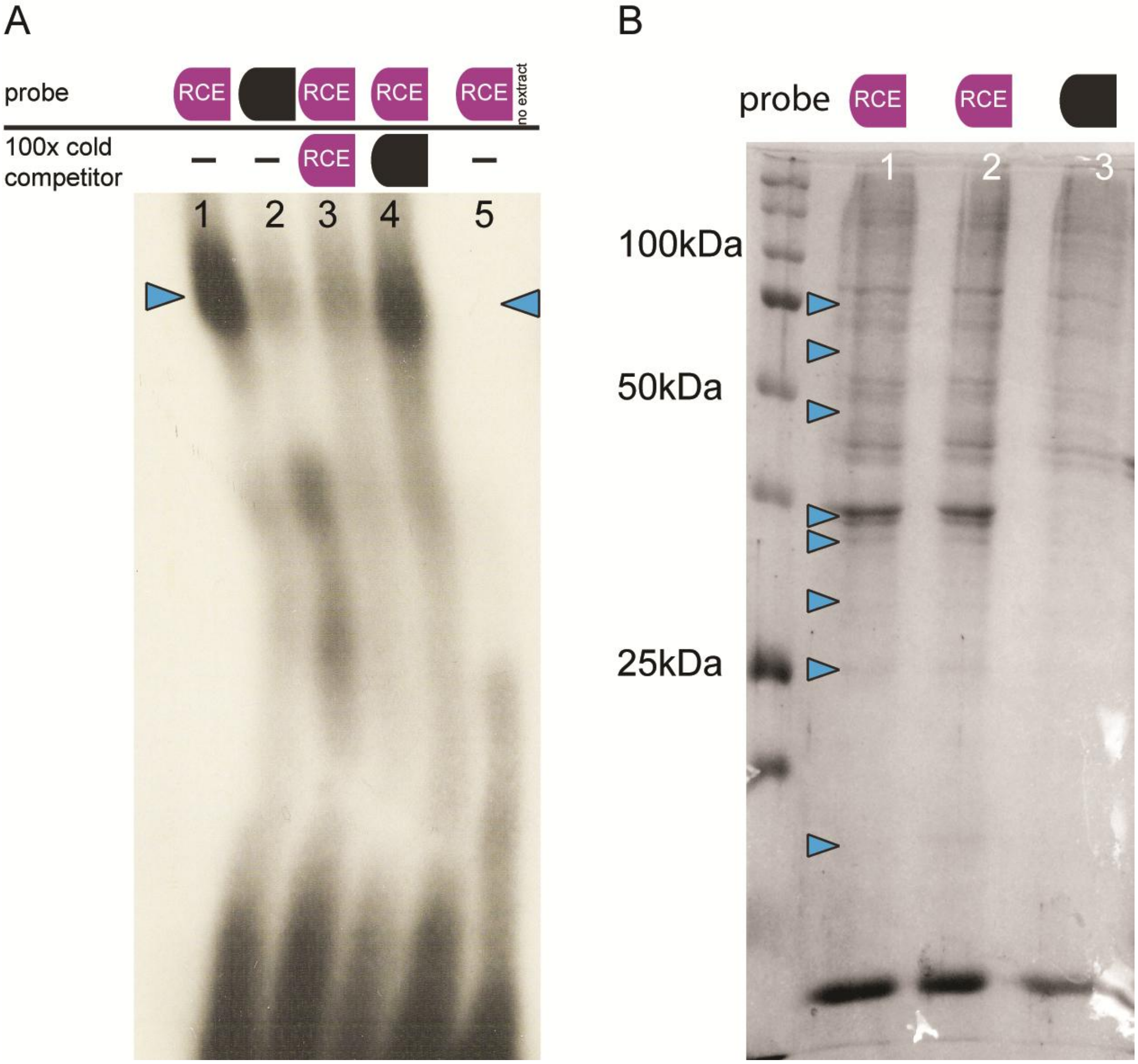


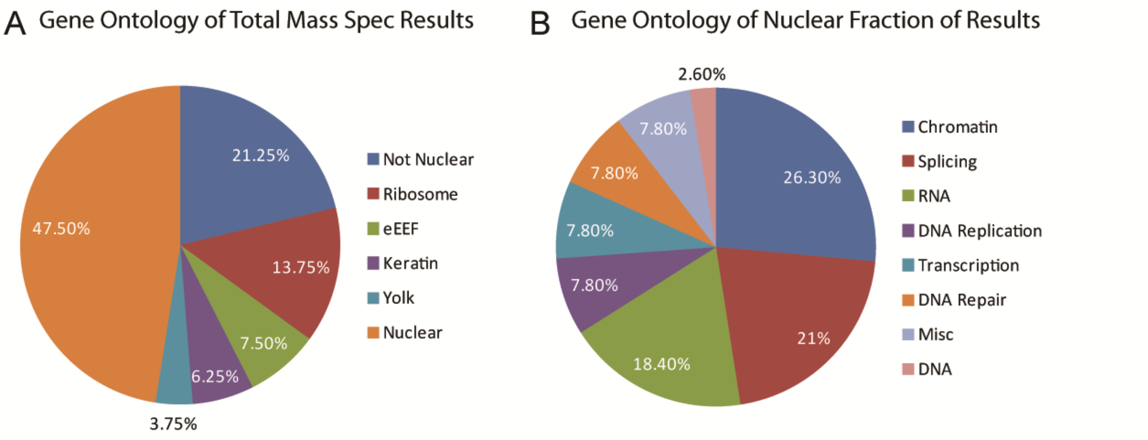
Figure 5.1 The RCE interacts with sequence specific binding components of the embryonic nuclear extract. (A) The wildtype RCE sequence is shifted by high molecular weight component in EMSA (lane 1 blue arrows). This interaction is lost when RCE sequence is mutated (lane 2). This interaction is sequence specific as the wildtype sequence can compete for RCE protein interaction, but the mutant cannot (lanes 3 and 4). (B) RCE sequences were used in affinity purification with embryonic nuclear extract. Interacting proteins were eluted, run on an SDS-PAGE gel, and visualized with Coomassie blue staining. Several protein bands were pulled down by the wildtype RCE sequence (lanes 2 & 3) that were not pulled down by the mutant sequence (lane 3, blue arrows). Eight bands of interest were excised from the gel and submitted for identification by mass spectrometry.

proteins, we first performed EMSA with radiolabeled wildtype and mutant probes and protein extracted from the nucleus of 0 – 12 hour *Drosophila* embryos. Upon addition of nuclear proteins, the RCE wildtype probe “shifts” to a higher (decreased mobility) position compared to the free probe (Figure 5.1A lanes 1 and 2). This shift is lost when the RCE sequence is abolished by mutation at every other base pair by non-complimentary transversion (Figure 5.1 Lane 3). Similarly, when 100 fold excess levels of unlabeled RCE wildtype is added to the wildtype reaction, the shift is lost. However, addition of 100 fold excess unlabeled RCE mutant probe to the wildtype reaction does not affect the ability of the RCE to interact with proteins (Figure 5.1A lanes 4 and 5). Together, this data indicates that the RCE is capable of interacting with protein, and that this interaction is specific as the wildtype probe can compete for this interaction while the mutant cannot. Furthermore, the high position of the shift on the gel suggests that a multiple protein complex is bound by the RCE.

In order to determine the identity of the protein(s) interacting with the RCE in our gel shift assay, we performed indirect affinity purification using the same wildtype and mutant probes and embryonic nuclear extract. Briefly, the extract was precleared on strepavidin coated Dynabeads (Invitrogen). Remaining protein was submitted to a second preclear with the addition of biotin tagged mutant RCE probe which was subsequently removed with the addition of the Dynabeads. This step removes the proteins that generally “stick” to DNA in a non-specific manner. Finally, the remaining proteins were mixed with either wildtype or mutant biotin tagged probes, and subsequently the DNA, and bound proteins were purified on the Dynabeads. Interacting proteins were then separated by size using SDS-PAGE electrophoresis. Several unique protein bands were present after Coomasie staining in the wildtype pull down that were absent in the mutant pull down (Figure 5.1B). We excised each unique band in the wildtype lane and the corresponding region of the mutant lane; 8 bands in total, and submitted them for protein identification by mass spectrometry at the University of Michigan Protein Structure Facility.

After identification of the proteins in each band, we first grouped the proteins by classes that represent the basic gene ontology of the data set. The largest group consisted of nuclear proteins; however we also saw a significant number of non-nuclear proteins including ribosomal proteins, eukaryotic elongation factors (eEEFs), as well as yolk proteins and keratin (a little bit of me) (Figure 5.2 A), revealing our nuclear extract protocol and does not completely eliminate non-nuclear proteins. The nuclear proteins can be further classified as chromatin interacting proteins, and proteins involved in mRNA splicing, regulation of transcription, DNA repair and replication, as well as, RNA binding proteins (Figure 5.2 B). After discarding proteins that were also present in the excised regions of the mutant pull down lane, we generated a list of 13 proteins that are likely candidates for RCE interaction based on prevalence in the protein sample and known molecular functions (Figure 5.2 C). These putative RCE interacting proteins are all compatible with a role in facilitating long range enhancer activity. For example, we identified proteins know to interact with chromatin and affect its structure and spatial organization, such as Histone H1, Bj1 and ballchen (bal)

Figure 5.2



C

		Known Larval
Candidate Protein	Gene Ontology	Expression?
Eukaryotic translation initiation factor 2 subunit (*Su[var]3-9)*	Chromatin	
Regulator of chromosome condensation (*Bj1*)	Chromatin	yes
Eukaryotic translation initiation factor 4E (*eIF-4E)*	Chromatin	yes
Transcription initiation factor TFIID subunit 6 (*taf6)*	Transcription	yes
Histone H1	Chromatin	
Nucleosomal histone kinase 1 (*ball*)	Chromatin	yes
Protein penguin OS (*pen*)	RNA	yes
Serine/threonine-protein phosphatase alpha-2 isoform (Pp1-87B)	Chromatin	yes
Polyadenylate-binding protein (*pABp*)	Splicing	yes
Chromatin-remodeling complex ATPase chain Iswi (*Iswi*)	Chromatin	yes
Transcription initiation factor TFIID subunit 9 (*e[y]1*)	Transcription	
Chromodomain-helicase-DNA-binding protein Mi-2 homolog (*mi2*)	Chromatin	yes
Protein no-on-transient A (*nonA*)	Splicing	yes

Figure 5.2 Identification of putative RCE interacting proteins. Mass spectrometry was used to identify proteins that interact with the RCE. The results were categorized by known function/cellular location. The largest fraction of proteins were nuclear (A). The nuclear fraction of proteins was further subdivided by predicted gene ontology. Large categories of note are chromatin binding proteins, splicing factors, and proteins that promote transcription (B). Finally, based on known protein function and prevalence in the sample submitted for mass spectrometry, and a list of 13 candidates were selected as putative RCE binding proteins (C).

(Frasch, 1991; Ivanovska et al., 2005). Similarly, Imitation SWI (iswi) and Mi2 have been demonstrated to interact with insulator sequences (Li et al., 2010; Mutskov et al., 2002). The protein complexes at insulators are capable of inducing the formation of DNA loops (Kadauke and Blobel, 2009). This looping action in turn can change the spatial organization of DNA such that two genomically distal sequences are brought into close proximity in the 3D space of the nucleus. As such, proteins involved in looping are strong candidates for RCE function. As RNA splicing also requires the formation of nucleotide loops, splicing proteins such as no on or off transient A (nonA) and polyA-binding protein (pABp) could potentially promote a long-range DNA interaction as well (Derry et al., 2006; Kozlova et al., 2006). Recently, non-coding RNAs have been shown to be critical in the regulation of tissue and temporal specific gene expression often acting like transcriptional enhancers (Orom et al., 2010; Tsai et al., 2010). Although we have not yet assessed the potential production of RNA's from the *sparkling* genomic region, it is possible that an RNA binding protein, like penguin (pen), could enable RNA mediated distal gene regulation (Maleszka et al., 1996). Finally, the identification of two proteins critical for the initiation transcription, TBP-associated factors 6 and 9 (Taf6 and 9), is consistent with an enhancer element, like the RCE, acting through a tracking method; recruiting the basal transcription machinery and directing it toward a target gene (Thomas and Chiang, 2006; Zhu et al., 2007).

Unfortunately, due to the nearly ubiquitous and essential nature of all of these proteins, mutant fly analysis is not likely to help clarify the potential role of these proteins in RCE activity. We also opted not to pursue a cell culture approach to study *spa* enhancer activity as we were unable to find a *Drosophila* cell line where *dPax2* is expressed, or Su(H), PntP2/Yan, and Lz. As we do not yet know all of the inputs necessary for *spa* activity in cone cells, cell culture would require complicated, uninformed, transfection experiments. Therefore, we decided to first assess the ability of these proteins to specifically bind the RCE *in vitro*. Utilizing EMSA again, we first generalized protein by *in vitro* transcription and translation from available cDNA's of those proteins with predicted nucleotide binding capabilities. We used PpI-87B as a negative control as it has no predicted direct nucleotide binding domain. Thus far, we have only been able to demonstrate that Taf6 interacts with the RCE wildtype probe, but not the mutant probe (Figure 5. 3 lanes 9 and 10). Bj1, e(y)1, eIf-4E (Figure 5.3) pen, pABp, su(var) 3-9, and iswi (Figure 5.4 A) did not shift the RCE wildtype probe uniquely from the mutant probe. We cannot however rule out any of these proteins as potential regulators for the RCE, as generating protein by this method does not allow us to determine whether or not the protein is made correctly, or if at all. Additionally, if any of these proteins bind the RCE as part of complex, individual proteins may not be able to interact and result in a gel shift alone. At this point we can only conclude that Taf6 can interact with the RCE *in vitro*.

In addition to the proteins identified by mass spectrometry we also looked at three candidate proteins not found in our screen, but based on expression and/or known function, we postulated could facilitate in RCE activity. Twin of eyeless (toy) is a critical protein in eye development and is bound to many

Figure 5.3

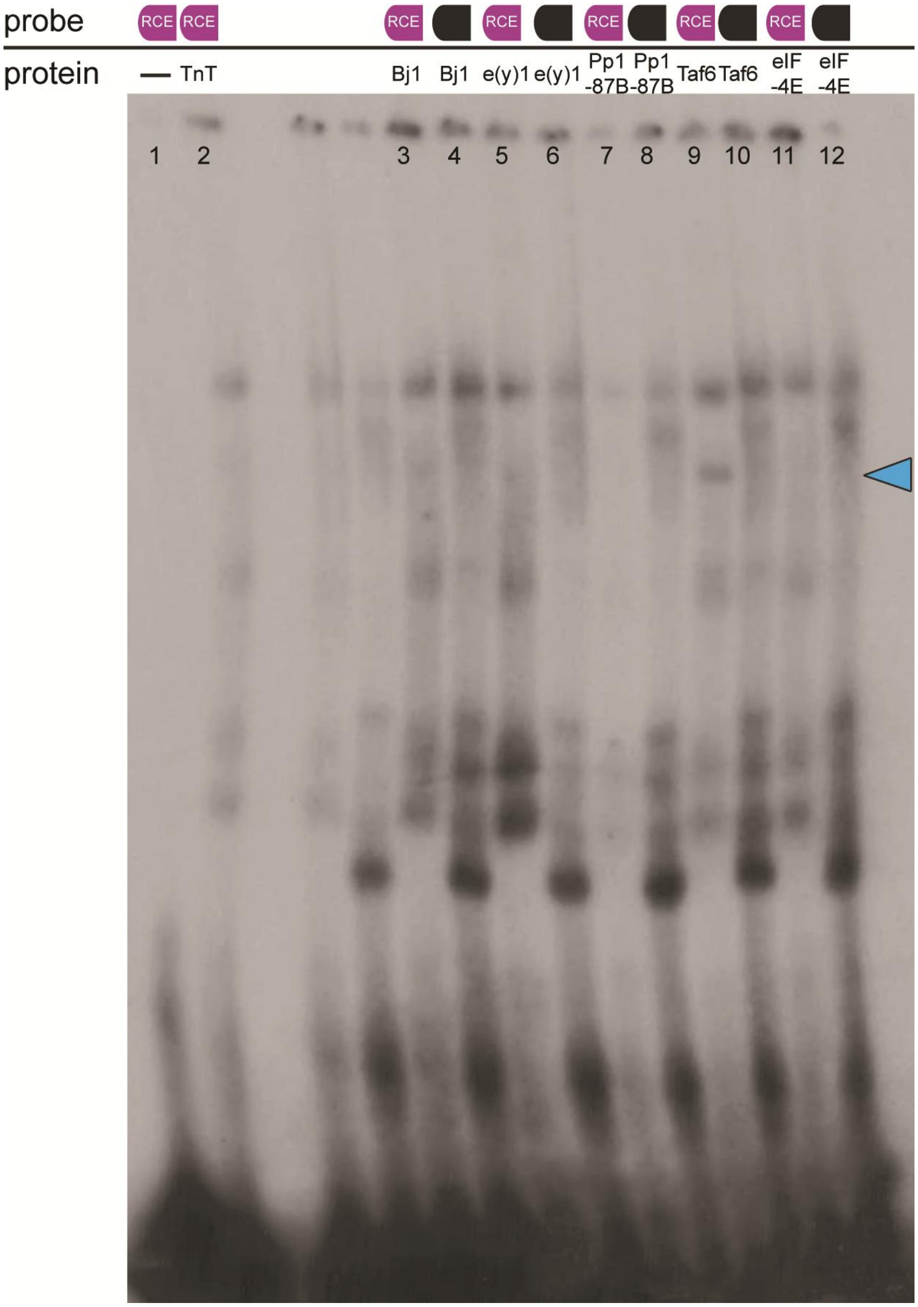


Figure 5.3 Taf6 interacts with the RCE *in vitro*. Candidate proteins were expressed from full length cDNAs using *in vitro* transcription/translation (TnT). The RCE sequence interacts with several proteins in the TnT lysate (lane2). Bj1, e(y)1, Pp1-87B, and eIF-4E lysates do not interact uniquely with the RCE wiltype or mutant sequences (lanes 3, 4, 5, 6, 7, 8, 11, 12). Only protein from the taf6 lysate shifts a unique band with the RCE wildtype sequence, but not the mutant sequence (lanes 9 and 10, blue arrow). This suggests taf6 can interact with the RCE *in vitro*.

Figure 5.4

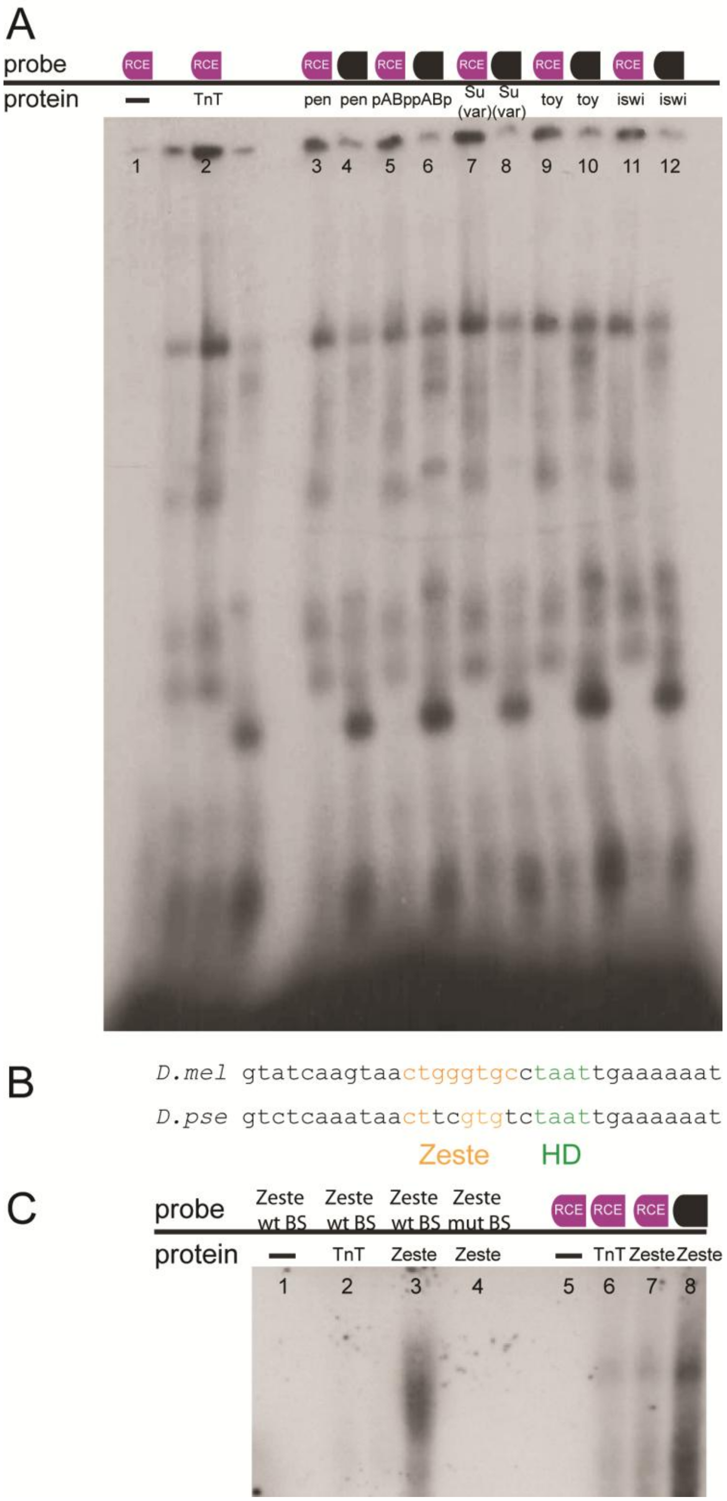


Figure 5.4 Remaining candidate proteins do not interact with the RCE *in vitro*. Candidate proteins

were expressed from full length cDNAs using *in vitro* transcription/translation (TnT). The RCE probe interacts several TnT lysate proteins (A, lane 2). pen, pABp, su(var) 3 – 9, and iswi do not interact with RCE sequence (A, lanes 3, 4 ,5, 6, 7, 8, 11, 12). As the RCE has a homeodomain binding site (B) we also looked at toy, which does not interact with the RCE (A, lanes 9 and 10). The RCE also has a conserved Zeste binding site (B). TnT generated Zeste interacts with a control binding site from an enhancer at the *white* gene, but not a mutated version of this site (C, lanes 3 and 4). However, the RCE does not appear to interact with the Zeste *in vitro* (C, lanes 7 and 8).

distally located eye specific enhancers (Czerny et al., 1999; Punzo et al., 2002; Weasner et al., 2009). Furthermore, it is a homeodomain (HD) binding protein and the RCE contains a conserved (HD) sites (Figure 5.4 B) (Czerny et al., 1999). Interestingly, Toy was also identified as a potential *sparkling binding* protein in a yeast one hybrid assay (Lisa Johnson, unpublished data). We were unable to demonstrate toy binding to the RCE in the EMSA assays (Figure 5.4 A, lanes 9 & 10). The transcription factor, Zeste, is also expressed in the *Drosophila* eye, and has been shown to homo-oligomerize between binding sites at enhancers and promoters to induce the formation of DNA loops (Kostyuchenko et al., 2009; Laney and Biggin, 1997; Mohrmann et al., 2002; Qian et al., 1992). There is also a putative Zeste binding site in the RCE (Figure 5.4 B). We saw a clear shift in our EMSAs combining *in vitro* transcribed and translated Zeste with a control labeled probe containing a Zeste binding site from a *white* enhancer. This shift is subsequently lost when this site is mutated. However, we did not see a similar shift when Zeste protein was combined with the RCE probes (Figure 5.4 C. This indicates Zeste does not bind the RCE *in vitro*. We were also unable to observe specific DNA binding by CTCF, an insulator protein that interacts with cohesion to stabilize DNA loops to either control or RCE probes (data not shown) (Sofueva and Hadjur, 2012).

5.3b Characterization of region 4 protein interactions

Region 4 has some long-range capabilities; however, we have characterized it more extensively as a crucial input for cone cell specific gene activation. In fact, it is the only region that is completely necessary in both the distal and proximal promoter positions (Chapter 2). We began our investigation of region 4 again with EMSAs. Region 4 can be subdivided into three sub regions that are each essential for *spa* activity *in vivo* (Swanson et al., 2010). We tested the ability of region 4 to bind protein(s) in nuclear embryonic extract, as well as the contribution of each of these subregions to protein binding (4a, 4b, 4c). We found that the wildtype region 4 probe consistently bound a high molecular weight component of the nuclear extract (Figure 5.5, lane 2). As with the RCE EMSA, the size of this shift suggests it is bound by a multi-protein complex, rather than a single protein. This shift is not lost upon mutation of regions 4a or 4c, but is decreased when region 4b is abolished (Figure 5.5, lanes 3 – 5). As region 4b contains a conserved homeodomain (HD) binding site, we tested the contribution of this site specifically to protein binding in two ways: 1) With a probe that mutates the HD site specifically and 2) with a probe that contains the HD site, but with all other sequences mutated. Region 4c also contains HD site, so even though region 4c did not contribute to the region 4 wildtype shift, we also generated the same HD probes for this region as for 4b. Of these probes, only

Figure 5.5

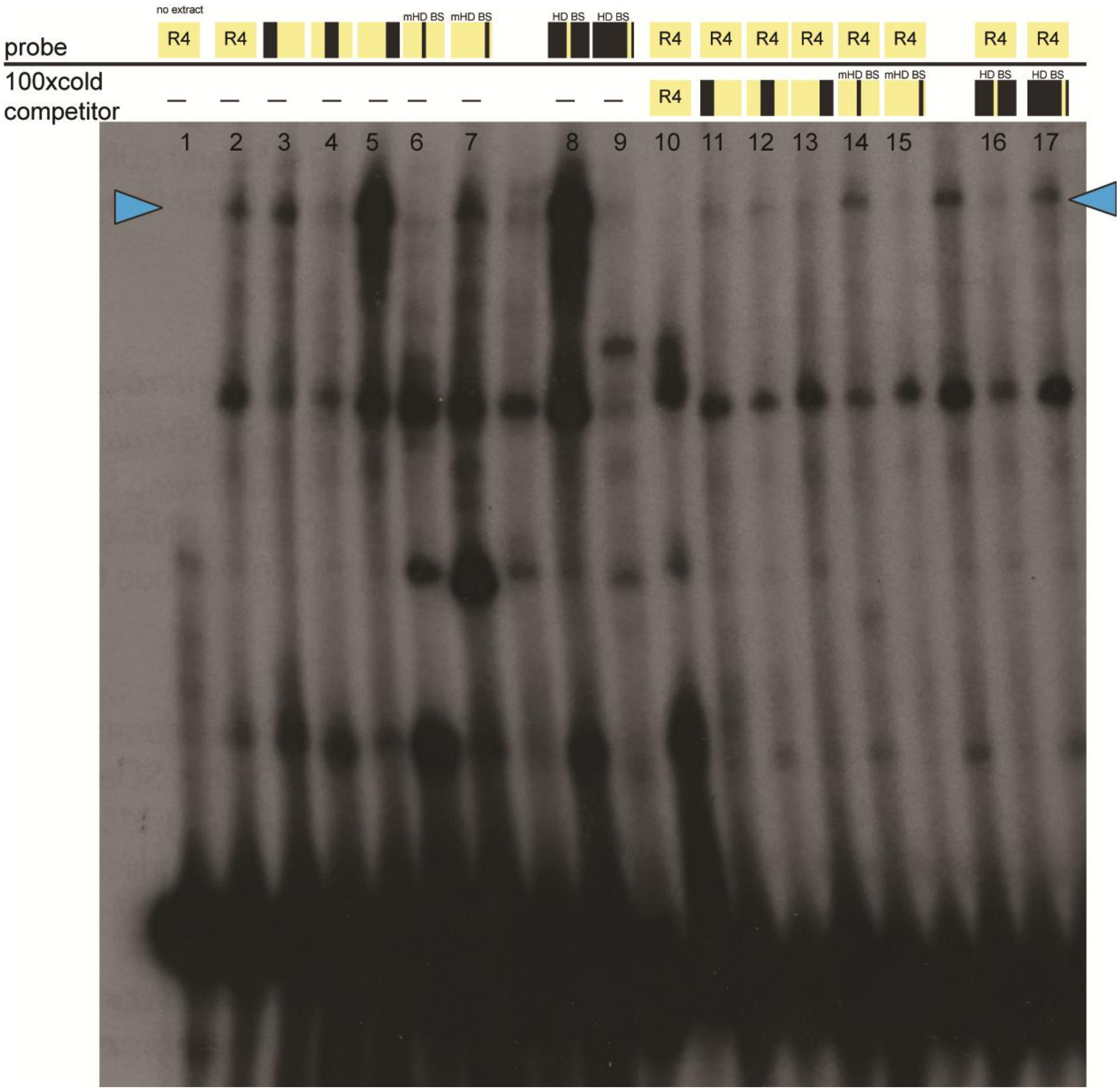


Figure 5.5 A homeodomain binding site in region 4 interacts with components of embryonic nuclear extract. Region 4 interacts with a high molecular weight entity in nuclear extract (lane 2, blue arrows). This interaction does not require Regions 4a, 4c, or the HD site in Region 4c (lanes 3, 5, 7). However, mutation of 4b or the homeodomain binding site in 4b results in loss of the shift (lanes 4, 6, 9). In fact, the 4b HD site is fully responsible for this shift (lane 8), suggesting this site is capable of interactions with protein from nuclear extract. The wildtype sequence can compete off the shift (lane 10). Loss of 4a and 4c do not affect this competition (lanes 11, 13, 15, 16). However, mutation of 4b or the 4b HD site reduces the competition (lanes 13, 15, 17).

the probe containing the intact region 4b HD site (and the remainder mutated) retained the ability to bind the nuclear protein(s), while mutation of this same site abolishes this interaction (Figure 5.5, lanes 6 – 9). Interestingly, when region 4c is lost either through mutation of the subregion, mutation of the 4c HD site, or when only the 4b HD site is left intact, the gel shift is actually significantly stronger suggesting sequences within 4c, including the HD site, decrease the ability of protein to bind to region 4b (Figure 5.5, lanes 5, 7, and 8). Accordingly, 100 fold excess of unlabeled region 4 wildtype, mutated 4a, mutated 4c, mutated 4c HD site, and intact 4b HD site probes are able to compete for the wildtype probes ability to interact with nuclear protein(s). However, the three probes lacking the region 4b HD site – mutated 4b, mutated 4b HD site, and only 4c HD site intact, do not fully compete for the wildtype DNA/protein interaction (Figure 5, lanes 10 – 17). Together these EMSA results indicate that *spa* region 4 binds embryonic nuclear protein utilizing the 4b, but not 4c, HD binding site.

We next performed indirect affinity purification using biotin tagged wildtype region 4 probe, and a mutated probe in which every other base pair of this region has been altered, which should inhibit specific protein interactions. *spa* region 4 can compensate for the RCE function when present in two copies. Therefore, we hypothesized then that region 4 and the RCE may bind the same DNA/protein complexes. As such, we simultaneously performed the affinity purification with the RCE wildtype and mutant probes. While we purified less protein overall this time, we do soo several of the expected wildtype RCE specific bands, including the characteristic doublet indicated with purple arrows (Figure

Figure 5.6

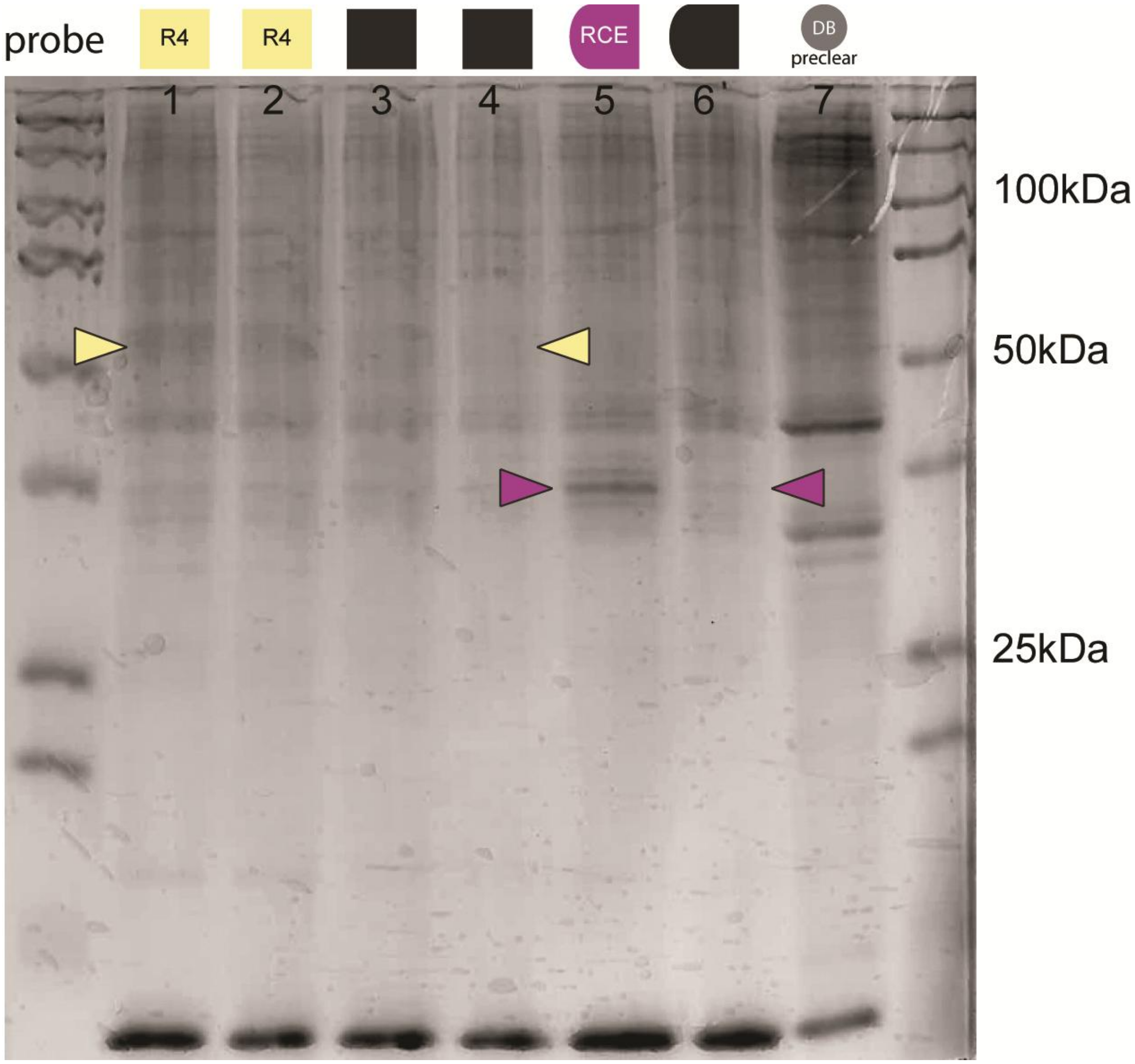


Figure 5.6 Region 4 affinity purification with embryonic nuclear extract. We performed pull downs with Region 4 and RCE sequences. As region 4 has some RCE-like activity, we compared protein bands in the RCE and region 4 purifications. We saw the consistent protein bands pulled down by the RCE (lane 5, purple arrows). While we did not see any similar bands between Region 4 and RCE pulldowns, we did identify at least 1 protein band that is pulled down by the Region 4 wildtype sequence, but not the mutant sequence (lanes 1 – 4, yellow arrows). Lane 7 shows how many proteins bind to the Dynabeads alone.

5.6, lanes 5 and 6). We do not see any protein bands that are present in the RCE and region 4 wildtype lanes, but absent in the respective mutant lanes (Figure 5.6). We do see a faint doublet of protein bands that purifies with the wildtype region 4, but the mutant probe (Figure 5.6, lanes 1 – 4, yellow arrowheads). Due to the low protein levels, we opted not to identify these proteins by mass spectrometry.

5.3c Characterization of region 5 and 6 protein interactions

sparkling region 5 is essential for initiation of cone cell specific gene expression. Interestingly, this same region is also necessary to repress *spa* activity in the R1 and R6 photoreceptors (Swanson et al., 2010). Based on size, this region likely contains one or more transcription factor binding sites, and the information for activation and repression could be found in independent or overlapping binding sites. Region 5 can be subdivided into three essential subelements (5a, 5b, and 5c). We performed EMSA with embryonic nuclear extract and wildtype region 5 probes, or probes containing mutations in each of the sub regions (a, b, and c). Wildtype region 5 interacts with a high molecular weight component of the extract (Figure 5.7 A, lane 2). Again, like the RCE and region 4, the size of this shift suggests region 5 is bound by a multiprotein complex. This shift is only lost when region 5b is mutated, but not when either 5a or 5c is lost (Figure 5.7 A, lanes 3-5). This suggests nuclear protein interacts with region 5 through the b subregion. Like region 4b, 5b contains a homeodomain binding site, so we next assessed the contribution of this HD site

Figure 5.7

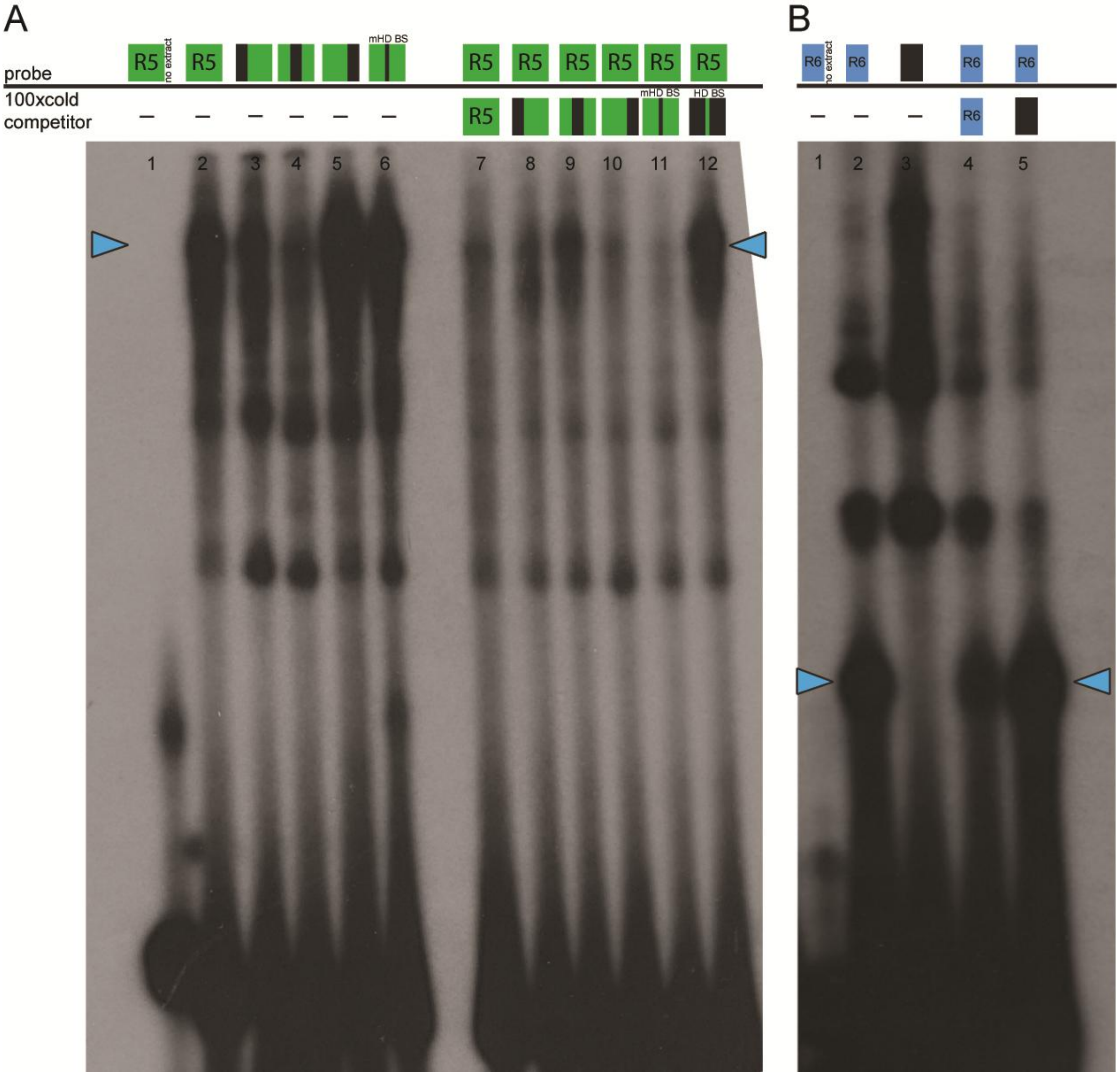


Figure 5.7 *sparkling* regions 5 and 6 interact with embryonic nuclear extract components. Region 5 wildtype sequence shifts a high molecular weight component of the nuclear extract (A, lane 2, blue arrows). This shift is not affected by the loss of Regions 5a or 5c (A, lanes 3 and 5). Mutation of 5b, however, results in loss of the shift suggesting protein binds to Region 5 through the b sequence (A, lane 4). 5b also contains a homeodomain binding site, but this sequence is not essential for the gel shift (A, lane 6). Region 5 wildtype sequence can compete for this shift (A, lane 7). The competition does not require 5a, 5c, or the 5b HD site (A, lanes 8, 10, 11), but of the remaining 5b sequence is necessary (lanes 9 and 10). Wildtype Region 6 also interacts with a low molecular weight component of the nuclear extract (B, lane 2, blue arrows). This shift is lost when the entire sequence is mutated (B, lane 3). The wildtype, but not the mutant sequence, can compete for this interaction (B, lanes 4 and 5).

to region 5 DNA protein interaction. Somewhat surprisingly, as mutation of the HD site affects 4 of the 11 bases in region 5b, loss of the 5b HD does not result in loss of the region 5 shift (Figure 5.7A, lane 6). This indicates the protein binding capabilities of region 5b does not require an intact HD binding site. Accordingly, 100 fold excess of unlabeled wildtype region 5, mutated 5a, mutated 5c, and mutated 5b HD site probes are able to compete for the wildtype region 5 protein interaction, while the mutated 5b probe and a probe with only 5b HD site intact are not capable of competing for this interaction (Figure 5.7 A, lanes 7 – 12). Together, this data suggests that region 5 is capable of interacting with protein from embryonic nuclear extract, and that this interaction requires the sequences within region 5b, but not the homeodomain binding site. We cannot however, rule out the 5b HD site as an important contributor to *spa* function in the developing *Drosophila* eye. BarH1/2 are good candidates for proteins capable of repressing *spa* in the R1 and R6 photoreceptors. These functionally redundant homeodomain proteins are expressed in a cell typea that we know *spa* is actively repressed in: R1 and R6 photoreceptors (Hayashi et al., 1998). As BarH1/H2 are thought to be transcriptional repressors, loss of the 5b homeodomain binding site might explain the why mutation of region 5 can result in depression of *spa* in R1 and R6. One of many homeodomain proteins expressed in cone cells could then bind to the same site in cone cells allowing for *spa* activity. However, to date we have been unable to demonstrate an *in vitro* interaction between BarH1 or H2 and region 5 by EMSA (data not shown). Notably, while BarH1/H2 contain homeodomains, the ideal binding site of TAATWR is not present in region 5

(Noyes et al., 2008). In fact, the only Bar consensus site in *sparkling* is in the RCE and loss of this sequence does not result in ectopic photoreceptor expression in either the promoter proximal or distal positions (Swanson et al., 2010).

sparkling region 6a is critical for robust initiation of gene expression in cone cells (Swanson et al., 2010). We performed EMSA's with region 6 wildtype and mutant probes and embryonic nuclear extract. The region 6 probe does bind protein in the extract (Figure 5.7 B, lane 2). Unlike the RCE, region 4, and region 5, this shift is relatively small, suggesting only one protein, or a low molecular weight complex interacts with region 6. Interestingly, three individual small mutations to this region did not affect the shift (data not shown). However, the region 6a protein interaction is specific, as mutation of the entire region abolishes the shift (Figure 5.7 B, lane 3). 100 fold excess unlabeled wildtype probe is able to compete for the ability of the region 6 wildtype probe to interact with nuclear protein, while the mutant probe cannot (Figure 5.7 B, lanes 4 & 5).

While both regions 5 and 6 bind proteins from embryonic nuclear extract, we have been unable to use affinity purification to identify proteins that bind to the wildtype, but not mutant probes. However, our attempts so far have utilized a different protocol than those used for region 4 and the RCE. In these attempts we covalently linked our probes to the beads prior to DNA protein interaction rather than after. Reducing the steric hindrance of the dynabeads from the DNA/protein reaction may result in the identification of interacting proteins in the future.

5.4 Discussion

The *dPax2 sparkling* enhancer is regulated by the transcription factors Su(H), Lz, and the Ets factors PntP2 and Yan. Extensive characterization of the DNA sequences within the enhancer have demonstrated that these inputs are not sufficient for enhancer activity. At least four additional sequences within the enhancer are required for *spa* activity; the RCE and regions 4, 5, and 6 (Swanson et al., 2010). Each of these regions performs overlapping and unique functions within the enhancer to contribute to its overall function. As such, we hypothesized that each of these sequences are capable of interacting with specific proteins in order to facilitate their role in *spa* activity. Indeed, using gel shift assays, we have shown that each of these sequences is able to interact with protein(s) from *Drosophila* embryonic extract.

5.4a sparkling homeodomain binding sites

The Six family transcription factor, Sine oculis, is capable of interacting with the RCE, region 4, and possibly region 5 sequences *in vitro* (Chapter 4). Sine oculis contains a homedomain DNA binding domain. It is unsurprising then that consensus So binding site GTAANYNGANAYS can contain a homeobox motif (TAAT), but most often *in vivo* So binding sites do not (Hazbun et al., 1997; Pauli et al., 2005). Interestingly, we found that RCE region a and region 4b interact with Sine oculis protein; however, the RCE region a does not interact with nuclear embryonic extract, while region 4b does. Accordingly, Sine oculis was not among the proteins identified by mass spectrometry as an RCE interacting protein. Our affinity purification results do not eliminate it as a candidate; however as in the *Drosophila* embryo, So protein is expressed at very low levels, and only in late embryo stages, which are the least represented in our extract (Graveley et al., 2011).

sparkling regions 4 and 5 and the RCE all contain homeodomain binding sites. Furthermore, the sub-elements within each of these regions that contain the HD site are all necessary for the ability of the region to interact with protein(s). We further showed that the homeodomain binding site in region 4b, but not in 4c or region 5b, are responsible for the observed DNA/protein interaction. This suggests that the activity of *spa* region 4 at least is likely to be through the binding of a homeodomain factor. Region 4b contains the putative Sine oculis binding site; however, So is unlikely to be interacting with region 4 in our gel shifts assay. This observation raises the possibility that a different protein facilitates region 4 action.

A potential candidate for this interaction is twin of eyeless, another essential protein in the *Drosophila* eye (Czerny et al., 1999). *In vitro,* this protein does not interact with the RCE, but we have not yet tested its ability to interact with region 4. dPax2 is required for expression of the homeodomain protein Cut in cone cells (Canon and Banerjee, 2003). Cut could then interact with the HD sites in *sparkling* in a feedback loop in the maintenance of enhancer activity during later stages of development. Another homeodomain binding protein that has the potential to regulate *spa* activity is the LIM family of proteins. Motif analysis has also implicated Lim protein interaction with the *sparkling* RCE (Chapter 3). Furthermore, Lim proteins have been shown to interact with Chip. This interaction is required for proper eye development in *Drosophila* and has been postulated to facilitate long-range interactions via the formation of large protein complexes "linking an enhancer and promoter" (Morcillo et al., 1996; Roignant et al., 2010). *sparkling* region 5, and possibly the RCE, contain DNA sequences capable of repressing *spa* activity in photoreceptors. This interaction is could be a result of interaction with BarH1 and BarH2, two functionally redundant homeodomain binding proteins that are expressed in the R1 and R6 photoreceptors (Hayashi et al., 1998). Each of these potential *spa* DNA protein interactions can be experimentally assessed using further gel shift assays and gene knockdown studies *in vivo.*

5.4b Taf6 interactions with the RCE

We performed affinity purification with *sparkling* RCE and nuclear embryonic extract. Using this method we identified proteins that are capable of interacting with the RCE by mass spectrometry. Interestingly, this approach did not pull out a significant number of transcription factors. This somewhat surprising result may be due to the nature of RCE activity. The RCE sequence facilitates long-range enhancer activity and is dispensable for pattering reporter gene expression. Given the proteins likely to facilitate long-range enhancer activity, for example chromatin remodeling factors and the basal transcription machinery, it is possible the RCE does not interact with any transcription factors. However, we have also seen that the RCE sequence can substitute for region 4 activity (Chapter 4). As region 4 is critical for supplying patterning information to the enhancer, this input likely requires transcription factor binding. Therefore, it is likely the RCE is at least capable of interacting with transcription factors. An alternative explanation for this observation is that transcription factors are underrepresented in our nuclear extract due to the large number of chromatin interacting proteins and general housekeeping proteins.

We did identify a few good candidates for RCE facilitator proteins using affinity purification. These candidates include proteins likely to promote looping such as iswi and Mi2 (Li et al., 2010; Mutskov et al., 2002). We also identifed putative binding parterns that can influence chromatin structure and three dimentional arrangement of DNA in the nucleus - Histone H1, Bj1, and ball (Frasch, 1991; Ivanovska et al., 2005). Interestingly, ball has been shown to interaction with nuclear laminins, which in turn can affect the localization of DNA within the nucleus (Nichols et al., 2006) (Wilson and Berk, 2010). Currently, we have been unable to confirm any of these proteins bind to and facilitate RCE function *in vitro* or *in vivo.*

We also identified two proteins critical for the initiation of transcription, the TBP-associated factors 6 and 9 (Taf6 and 9) (Thomas and Chiang, 2006; Zhu et al., 2007). Taf6 is able to interact with the RCE in *in vitro* gel shift assays in a sequence specific manner. As a member of the transcription factor IID (TFIID) the ability of this protein to interact with the RCE *in vitro* is consistent with a tracking mechanism of long-range enhancer action, whereby the enhancer recruits the Pol II complex which initiates transcription between the enhancer and target promoter. The existence of transcripts from the *spa* genomic sequence would provide additional evidence for this mechanism. We would also like to perform Chromatin Immunoprecipitation (ChIP) for PolII at the *spa* enhancer and intervening sequences; however we are again limited by the amount of tissue available to work with. *spa* mediated production of RNAs could also suggest a mechanism of enhancer action in which the non-coding RNA actually facilitates long-range gene regulation. If intergenic RNA is detected from the *spa* enhancer or nearby, we could decrease levels of this RNA using siRNA, and assess the effect on endogenous, or reporter gene transcription (Orom et al., 2010).

5.4c Future directions

sparkling regions 5 and 6 are able to interact specifically with protein in *Drosophila* nuclear embryonic extract. However, we have been unable to use affinity purification to identify uniquely interacting proteins with either of these regions. We can repeat the purification procedure using the modifications applied to the RCE and region 4 purification protocols along with further troubleshooting as necessary. We can also attempt to identify *sparkling* binding proteins through yeast one hybrid assays rather than affinity purification. Using this approach we can use a library of proteins from eye disc tissue specifically.

5.5 Experimental methods

5.5a Preparation of in vitro transcribed and translated proteins

Full length cDNA's for our candidate proteins were obtained from the Drosophila Genomics Resource Center. In order to express full length proteins, SP6 promoter and Kozak sequences were added to the 5' end of cDNA and two stop codons and a polyA tail was added to the 3' end by PCR. The following PCR primers are listed below:

Fwd: 5'-gtaatatatttaggtgacactatagaacagaccacc-20bp cDNA specific sequence – 3'

Rev: 5'-tttttttttttttttttttttttttttttctatca-20bp cDNA specific sequence – 3'

Point mutations in the eIF-4 cDNA were fixed by sewing PCR.

Next proteins were expressed from the PCR products using the TnT®SP6 High-Yield Wheat Germ Protein Expression System according to the kits protocol (Promega). Protein was aliquoted and stored at -80^{o}C.

5.5b Electrophoretic Mobility Shift Assays

We obtained embryonic nuclear extracts from 0 – 12 hour embryos collected for three days. Nuclear extract was performed according to the previously described protocol (ref). Labeled probes were annealed and labeled by incubating 37µl dH_2O, 5µl 10xPNK buffer, 1µl each top and bottom strand oligos (2µM), 5µl γ^{32}P-ATP, and 1µl T4PNK (New England Biolabs) at 32^{o}C, boiling at 80^{o}C for 5 minutes, and allowing samples to cool to room temperature (about 1 hour) to allow oligos to anneal. Labeled double stranded probes were then purified twice using GE ProbeQuant G-50 spin columns. 100 fold excess cold competitors were prepared by combining 39µl dH_2O, 5µl10x PNK buffer, and 1µl each top and bottom strand oligos (200µM). Competitors were then prepared as with the labeled probe except they were not purified after annealing. Probe sequences used are shown below. Mutated sequences are depicted in uppercase letters.

RCEWT: 5’ gtatcaagtaactgggtgcctaattgaaaaaatttactatgac 3’

RCE mut: 5’ gGaGcCaTtCaAtTgGgAcGaCtGgCaCaCaGtGaAtCtgac

Region 4wt: 5’ ttgaaattgaagcactattggtgtacgattacaacgctcacattatcagg 3’

mRegion 4a: 5’ttgaaCtGgCaTcCcGaGtggtgtacgattacaacgctcacattatcagg 3’

mRegion4b: 5’ ttgaaattgaagcactattTgGgGaAgCtGaAaacgctcacattatcagg 3’

mRegion4c: 5’ ttgaaattgaagcactattggtgtacgattacCaAgAtAaAaGtCtcagg 3’

mRegion4bHD:5’ttgaaattgaagcactattggtgtacgCGCGcaacgctcacattatcagg3’

mRegion4cHD:5’ttgaaattgaagcactattggtgtacgattacaacgctcacCGCGtcagg3’

mRegion4bHDwt:5’tGgCaCtGgCaTcCcGaGtTgGgtacgattacaaAgAtAaAaGtCtAaTg3’

mRegion4cHDwt:5’TggCaCtGgCaTcCcGaGtTgG*gGa*AgCtGaAaCcTctcacattatcaTg 3’

Region 5wt: 5’ atataaaaaaaaggtgatagtaattcagcacgactttgtaa 3’

mRegion 5a: 5’ atataCaCaCaCgTtgatagtaattcagcacgactttgtaa 3’ mRegion5b: 5’ atataaaaaaaaggtTaGaTtCaGtAagcacgactttgtaa 3’

mRegion5c: 5’ atataaaaaaaaggtgatagtaattcaTcCcTaAtGtgtaa 3’

mRegion5bHD: 5’ atataaaaaaaaggtgatagCGCGtcagcacgactttgtaa 3’

mRegion5bHDwt: 5’ aGaGaCaCaCaCgTtTatagtaattcagcCcTaAtGtTtCa 3’

Region 6 wt: 5’ agtacaacgtaagtcgggtgaagccagaaacc 3’

Region 6 mut: 5’ agGaAaCcTtCaTtAgTgGgCaTcAaTaCaAc 3’

DNA/protein interactions (gel shifts) were performed with 1µl labeled probe, 1µl nuclear extract (about 8µg of protein), or 3 – 7µl of TnT generated protein, 1µ 10xgelshift buffer (0.1 M Tris HCl ph 7.5, 0.5 M NaCl, 10 mM DTT, 10 mM EDTA, 275 µg/ml salmon sperm DNA), 1µl polyd(I-C) (1mg/ml), 1µl DTT (100µM) and dH_2O to a final volume of 10µl. Reactions with cold competitors used 1µl annealed cold competitor in place of 1µl of dH_2O. Reactions were then incubated on ice for 15 minutes and then loaded on 4% polyacrylamide gels and run for 4 – 5 hours in 0.5x TBE at 120 volts. Completed gels were then vacuum dried at 80^{o}C for 1 – 2 hours and finally exposed to film.

5.5c Affinity purification

For pulldown experiments double stranded DNA probes were generated by combining 2.5 µl biotinylated top strand oligo (200 µm) with 2.5 µl bottom strand oligo (200 µm) and dH_2O to 100 µl. This mixture was boiled for 5 minutes at 80^{o}C and then cooled slowly to room temperature (1 to 2 hours). The following probes were used. Mutated sequenced are shown in uppercase letters:

RCEWT: 5’ gtatcaagtaactgggtgcctaattgaaaaaatttactatgac 3’

RCE mut: 5’ gGaGcCaTtCaAtTgGgAcGaCtGgCaCaCaGtGaAtCtgac 3’

Region 4wt: 5’ ttgaaattgaagcactattggtgtacgattacaacgctcacattatcagg 3’

Region4mut:5’ttggaaCtGgCaTcCcGatGgTtTtCcTaGtCaCaAgAtAaAaGtCt cagg 3’

Streptavidin-coated Dynabeads (10 mg/ml, Invitrogen) were prepared for use with a magnet by washing 100 µl of beads twice with 100 µl 2xBW buffer (10 mM Tris – HCl ph 7.5 1 mM EDTA, 2M NaCl) and finally resuspended in 50 µl 1xgel shift buffer. At the same time protein was prepared for the binding reactions. 35 µl 7xcomplete protease inhibitor (Roche), 25 µl glycerol, and 150 µl nuclear extract (about 1.2 mg protein) was incubated at room temperature for 10 minutes with gentle mixing. The protein mixture was them combined with 50 µl of Dynabeads in 1xgelshift buffer and incubated at room temperature for 20 minutes with 360^{o} rotation. Beads were removed with a magnet, and protein was added to 100 µl of hybridized mutant probe an incubated at room temperature for 15 minutes with rotation, and then removed with the magnet. The remaining protein mixture was mixed with 100 µl of either hybridized wildtype or mutant probes, mixed for 20 minutes at room temperature with rotation and then combined with 50 µl of prepared Dynabeads. The beads were then removed on a magnet, washed once in 1xgelshift buffer containing 0.1 mg/ml poly d(I-C), and then washed 3 more times with 1xgelshift buffer. Beads were resuspended in 50 µl 1x SDS-PAGE gel loading buffer (50 mM Tris-HCl pH 6.8, 100 mM DTT, 2% SDS, 0.1% Bromophenol Blue, 10% glycerol) and boiled for 5 minutes. Using the magnet to remove the beads, the eluted protein (supernatant) was loaded onto an SDS-PAGE gel (12% resolving gel, 5% stacking gel) which was run in 1x Tris – Glycine buffer at 120 volts, then stained overnight with Coomassie blue and washed as described (Wong et al., 2000). Protein bands of interest were then excised from the gel and analyzed by LC-MS/MS at the University of Michigan Protein Structure Facility.

5.6 References

Canon, J., and Banerjee, U. (2003). In vivo analysis of a developmental circuit for direct transcriptional activation and repression in the same cell by a Runx protein. Genes Dev *17*, 838-843.

Czerny, T., Halder, G., Kloter, U., Souabni, A., Gehring, W.J., and Busslinger, M. (1999). twin of eyeless, a second Pax-6 gene of Drosophila, acts upstream of eyeless in the control of eye development. Mol Cell *3*, 297-307.

Derry, M.C., Yanagiya, A., Martineau, Y., and Sonenberg, N. (2006). Regulation of poly(A)-binding protein through PABP-interacting proteins. Cold Spring Harb Symp Quant Biol *71*, 537-543.

Evans, N.C., Swanson, C.I., and Barolo, S. (2012). Sparkling insights into enhancer structure, function, and evolution. Curr Top Dev Biol *98*, 97-120.

Flores, G.V., Duan, H., Yan, H., Nagaraj, R., Fu, W., Zou, Y., Noll, M., and Banerjee, U. (2000). Combinatorial signaling in the specification of unique cell fates. Cell *103*, 75-85.

Frasch, M. (1991). The maternally expressed Drosophila gene encoding the chromatin-binding protein BJ1 is a homolog of the vertebrate gene Regulator of Chromatin Condensation, RCC1. Embo J *10*, 1225-1236.

Fu, W., Duan, H., Frei, E., and Noll, M. (1998). *shaven* and *sparkling* are mutations in separate enhancers of the *Drosophila Pax2* homolog. Development *125*, 2943-2950.

Fu, W., and Noll, M. (1997). The *Pax2* homolog *sparkling* is required for development of cone and pigment cells in the *Drosophila* eye. Genes Dev *11*, 2066-2078.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W.*, et al.* (2011). The developmental transcriptome of Drosophila melanogaster. Nature *471*, 473-479.

Hayashi, T., Kojima, T., and Saigo, K. (1998). Specification of primary pigment cell and outer photoreceptor fates by BarH1 homeobox gene in the developing Drosophila eye. Dev Biol *200*, 131-145.

Hazbun, T.R., Stahura, F.L., and Mossing, M.C. (1997). Site-specific recognition by an isolated DNA-binding domain of the sine oculis protein. Biochemistry *36*, 3680-3686.

Ivanovska, I., Khandan, T., Ito, T., and Orr-Weaver, T.L. (2005). A histone code in meiosis: the histone kinase, NHK-1, is required for proper chromosomal architecture in Drosophila oocytes. Genes Dev *19*, 2571-2582.

Johnson, L.A., Zhao, Y., Golden, K., and Barolo, S. (2008). Reverse-engineering a transcriptional enhancer: a case study in Drosophila. Tissue Eng Part A *14*, 1549-1559.

Kadauke, S., and Blobel, G.A. (2009). Chromatin loops in gene regulation. Biochim Biophys Acta *1789*, 17-25.

Kostyuchenko, M., Savitskaya, E., Koryagina, E., Melnikova, L., Karakozova, M., and Georgiev, P. (2009). Zeste can facilitate long-range enhancer-promoter communication and insulator bypass in Drosophila melanogaster. Chromosoma *118*, 665-674.

Kozlova, N., Braga, J., Lundgren, J., Rino, J., Young, P., Carmo-Fonseca, M., and Visa, N. (2006). Studies on the role of NonA in mRNA biogenesis. Exp Cell Res *312*, 2619-2630.

Laney, J.D., and Biggin, M.D. (1997). Zeste-mediated activation by an enhancer is independent of cooperative DNA binding in vivo. Proc Natl Acad Sci U S A *94*, 3602-3604.

Lebestky, T., Chang, T., Hartenstein, V., and Banerjee, U. (2000). Specification of Drosophila hematopoietic lineage by conserved transcription factors. Science *288*, 146-149.

Li, M., Belozerov, V.E., and Cai, H.N. (2010). Modulation of chromatin boundary activities by nucleosome-remodeling activities in Drosophila melanogaster. Mol Cell Biol *30*, 1067-1076.

Maleszka, R., Hanes, S.D., Hackett, R.L., de Couet, H.G., and Miklos, G.L. (1996). The Drosophila melanogaster dodo (dod) gene, conserved in humans, is functionally interchangeable with the ESS1 cell division gene of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *93*, 447-451.

Malik, S., and Roeder, R.G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. Trends Biochem Sci *30*, 256-263.

Mohrmann, L., Kal, A.J., and Verrijzer, C.P. (2002). Characterization of the extended Myb-like DNA-binding domain of trithorax group protein Zeste. J Biol Chem *277*, 47385-47392.

Morcillo, P., Rosen, C., and Dorsett, D. (1996). Genes regulating the remote wing margin enhancer in the Drosophila cut locus. Genetics *144*, 1143-1154.

Mutskov, V.J., Farrell, C.M., Wade, P.A., Wolffe, A.P., and Felsenfeld, G. (2002). The barrier function of an insulator couples high histone acetylation levels with specific protection of promoter DNA from methylation. Genes Dev *16*, 1540-1554.

Narlikar, G.J., Fan, H.Y., and Kingston, R.E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell *108*, 475-487.

Nichols, R.J., Wiebe, M.S., and Traktman, P. (2006). The vaccinia-related kinases phosphorylate the N' terminus of BAF, regulating its interaction with DNA and its retention in the nucleus. Mol Biol Cell *17*, 2451-2464.
Noyes, M.B., Christensen, R.G., Wakabayashi, A., Stormo, G.D., Brodsky, M.H., and Wolfe, S.A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell *133*, 1277-1289.
Orom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q.*, et al.* (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell *143*, 46-58.
Orphanides, G., and Reinberg, D. (2002). A unified theory of gene expression. Cell *108*, 439-451.
Pauli, T., Seimiya, M., Blanco, J., and Gehring, W.J. (2005). Identification of functional sine oculis motifs in the autoregulatory element of its own gene, in the eyeless enhancer and in the signalling gene hedgehog. Development *132*, 2771-2782.
Price, M.D., and Lai, Z. (1999). The yan gene is highly conserved in Drosophila and its expression suggests a complex role throughout development. Dev Genes Evol *209*, 207-217.
Punzo, C., Seimiya, M., Flister, S., Gehring, W.J., and Plaza, S. (2002). Differential interactions of eyeless and twin of eyeless with the sine oculis enhancer. Development *129*, 625-634.
Qian, S., Varjavand, B., and Pirrotta, V. (1992). Molecular analysis of the zeste-white interaction reveals a promoter-proximal element essential for distant enhancer-promoter communication. Genetics *131*, 79-90.
Roignant, J.Y., Legent, K., Janody, F., and Treisman, J.E. (2010). The transcriptional co-factor Chip acts with LIM-homeodomain proteins to set the boundary of the eye field in Drosophila. Development *137*, 273-281.
Scholz, H., Deatrick, J., Klaes, A., and Klambt, C. (1993). Genetic dissection of pointed, a Drosophila gene encoding two ETS-related proteins. Genetics *135*, 455-468.
Schweisguth, F., and Posakony, J.W. (1992). Suppressor of Hairless, the Drosophila homolog of the mouse recombination signal-binding protein gene, controls sensory organ cell fates. Cell *69*, 1199-1212.
Sofueva, S., and Hadjur, S. (2012). Cohesin-mediated chromatin interactions--into the third dimension of gene regulation. Brief Funct Genomics *11*, 205-216.
Swanson, C.I., Evans, N.C., and Barolo, S. (2010). Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. Dev Cell *18*, 359-370.
Swanson, C.I., Schwimmer, D.B., and Barolo, S. (2011). Rapid evolutionary rewiring of a structurally constrained eye enhancer. Curr Biol *21*, 1186-1196.
Szutorisz, H., Dillon, N., and Tora, L. (2005). The role of enhancers as centres for general transcription factor recruitment. Trends Biochem Sci *30*, 593-599.
Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol *41*, 105-178.

Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science *329*, 689-693.

Voas, M.G., and Rebay, I. (2004). Signal integration during development: insights from the *Drosophila* eye. Dev Dyn *229*, 162-175.

Wang, G., Balamotis, M.A., Stevens, J.L., Yamaguchi, Y., Handa, H., and Berk, A.J. (2005). Mediator requirement for both recruitment and postrecruitment steps in transcription initiation. Mol Cell *17*, 683-694.

Weasner, B.M., Weasner, B., Deyoung, S.M., Michaels, S.D., and Kumar, J.P. (2009). Transcriptional activities of the Pax6 gene eyeless regulate tissue specificity of ectopic eye formation in Drosophila. Dev Biol *334*, 492-502.

Wilson, K.L., and Berk, J.M. (2010). The nuclear envelope at a glance. J Cell Sci *123*, 1973-1978.

Wong, C., Sridhara, S., Bardwell, J.C., and Jakob, U. (2000). Heating greatly speeds Coomassie blue staining and destaining. BioTechniques *28*, 426-428, 430, 432.

Zhu, X., Ling, J., Zhang, L., Pi, W., Wu, M., and Tuan, D. (2007). A facilitated tracking and transcription mechanism of long-range enhancer function. Nucleic Acids Res *35*, 5532-5544.

CHAPTER 6

CHARACTERIZATION OF REGULATORY SEQUENCES IN THE *dPax2* 4TH INTRON

6.1 Abstract

dPax2 is expressed in the *Drosophila* eye imaginal disc was subsequently localized to the four cone cells, two primary pigment cells, and four mechanosensory bristles in each ommatidium of the eye. Two recessive mutations, named “sparkling” and “shaven” result in loss of these cell types due to failure of *dPax2* expression. The enhancers responsible for cone cells and bristle cell specific gene expression have been identified and characterized subsequent to mapping the location of these mutations. The sparkling phenotype, which is the result of loss *dPax2* expression in cone cells and primary pigment cells, is in part the consequence of loss of the *sparkling* (*spa*) cone cell specific enhancer which lies at the 5’ end of the genes 4th intron. While the location of the sparkling mutation tells us that it is likely in the 5’ half of the *dPax2* 4th intron, the enhancer responsible for primary pigment cell expression has yet to be identified. We noticed that the addition of 60 basepairs of conserved DNA

sequence 5’ of *spa* enhancer conveyed pigment cell activity upon the enhancer. This observation led us to examine the additional regulatory potential of the entire *dPax2* 4th intron. We found that the DNA sequence spanning from the 5’ of the 4th intron to the 3’ end of *spa* represents the regulatory sequences sufficient for gene expression in primary pigment cells. Sequence and reporter gene activity implicates higher levels Notch input in this action. Furthermore, we identified a potential “shadow” or redundant regulatory sequence downstream of *sparkling* which is also capable of driving gene expression cone cells.

CHARACTERIZATION OF REGULATORY SEQUENCES IN THE DPAX2 4^{TH} INTRON

6.2 Introduction

In *Drosophila melanogaster* the *dPax2* gene is expressed in the peripheral and central nervous systems as well as in the posterior portion of the third instar eye imaginal disc (Czerny et al., 1997). It is also expressed strongly in select cells of the antennal, leg, and wing discs (Fu and Noll, 1997). As the *Pax2* gene is crucial for vertebrate eye development, the role of its homolog in the *Drosophila* eye is of special interest (Macdonald and Wilson, 1996). The expression of *dPax2* in the eye imaginal disc was subsequently localized to the four cone cells, two primary pigment cells, and four mechanosensory bristles in each ommatidium of the eye (Fu and Noll, 1997). The genomic location of this gene, as well as its regulatory regions, was identified with the help of six lethal mutations and two recessive visible mutants on the 4th chromosome (Hochman, 1971; Lindsley, 1992). The recessive mutations, named “sparkling” and “shaven” particularly facilitated the discovery of *dPax2* enhancers.

These mutant phenotypes were so named because the cone cells and primary pigment cells did form correctly and the adult eyes appeared to “sparkle”, and in another mutant fly they eyes had no bristles, making it appear “shaven”. It was later determined that these two mutant phenotypes result from loss of expression of the same gene; *dPax2* (Fu et al., 1998). Interestingly, when the endogenous mutations were mapped, it was determined that little or no coding sequence was affected by these mutations. The shaven mutation, which results in loss of *dPax2* in sensory bristle cells, affects sequences upstream of *dPax2* coding sequence. This phenotype was subsequently attributed to the loss of two bristle specific enhancers (Johnson et al., 2011). The sparkling mutation, which deletes the small 3rd and 4th exons as well as the 3rd and half of the large 4th intron, to make an in frame protein mutation, has been attributed primarily to loss of the regulatory sequences required for cone cell and primary pigment cell expression (Flores et al., 2000; Fu et al., 1998).

Extensive work has identified and characterized the enhancer responsible for regulating *dPax2* expression in cone cells, named *sparkling* (*spa*) for the mutant phenotype. The enhancer lies at the 5' end of the genes 4th intron and is regulated by EGFR and Notch signaling as well as Lozenge and other yet unidentified transcription factors (Flores et al., 2000; Swanson et al., 2010; Swanson et al., 2011). While the enhancer is named for the sparkling mutant, the phenotype of these adult eyes results from loss of both cone cells and primary pigment cells. While the location of this mutation tells us that it is likely in the 5' half of the *dPax2* 4th intron, the enhancer responsible for primary pigment cell expression has yet to be identified.

The *sparkling* enhancer has been extensively characterized by our lab and others (Evans et al., 2012; Fu et al., 1998; Fu and Noll, 1997; Swanson et al., 2010; Swanson et al., 2011), and has been refined to a 326 bp *minimal* enhancer to use in reporter constructs. However, our experiments moving the RCE upstream of *spa* and our inability to find an RCE like sequence in the *dppD* wing disc enhancer (Appendix 1) has taught us that we need to look outside minimal enhancers for important regulatory sequences. Upon examination of the DNA sequence surrounding the minimal *spa* enhancer and identified a 60 bp sequence immediately upstream of *spa* that is highly conserved. As such, we decided to generate *spa* reporter constructs that contain this sequence in order to determine whether it can contribute to *spa* activity. We were surprised to find that the addition of this sequence inhibits *spa* activity in the distal position (-846bp) but does not affect *spa* activity in the proximal position (-121bp). However, the addition of this sequence gives *spa* the ability to activate gene expression in primary pigment cells. This observation led us to examine the regulatory potential of the entire *dPax2* 4th intron and the DNA sequences flanking *spa* enhancers from other *Drosophila* species.

6.3 Results

6.3a A highly conserved sequence upstream of sparkling conveys primary pigment cell expressionbut inhibits sparklings distal activity

The *sparkling* enhancer drives cone cell specific gene expression of the GFP reporter gene in the developing *Drosophila* eye (Figure 6.1 B, E). Immediately upstream of the *spa* “remote control” element, RCE, which regulates *spa’s* ability to function at a distance (Figure 6.1C, D), is a stretch of highly conserved DNA sequence. This 60bp sequence shows greater conservation

Figure 6.1

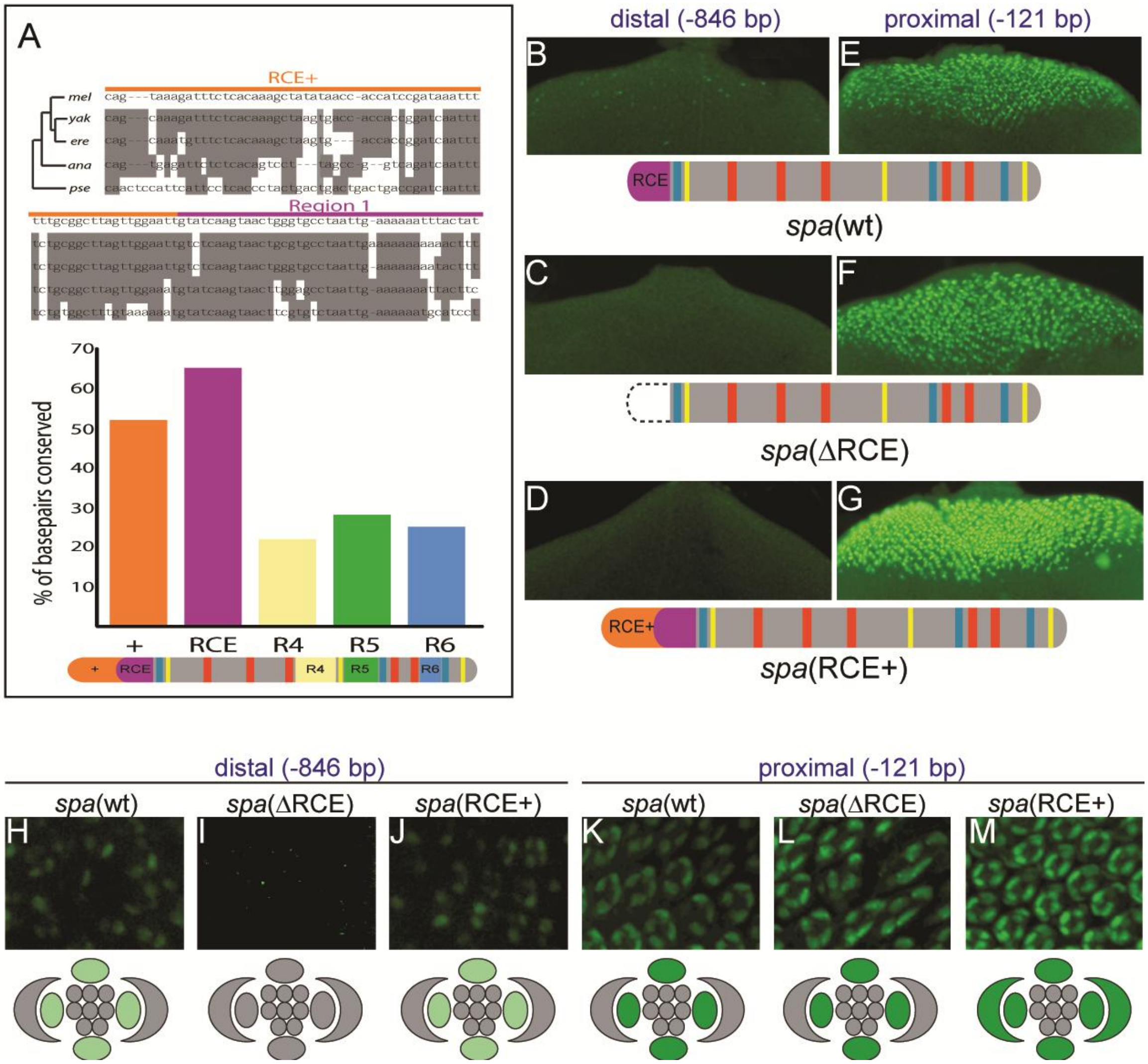

A
RCE+
mel
yak
ere
ana
pse
cag---taaagatttctcacaaagctatataacc-accatccgataaattt
cag---caaagatttctcacaaagctaagtgacc-accaccggatcaattt
cag---caaatgtttctcacaaagctaagtg----accaccggatcaattt
cag---tgagattctctcacagtcct---tagcc-g--gtcagatcaattt
caactccattcattcctcaccctactgactgactgactgaccgatcaattt
Region 1
tttgcggcttagttggaattgtatcaagtaactgggtgcctaattg-aaaaaatttactat
tctgcggcttagttggaattgtctcaagtaactgcgtgcctaattgaaaaaaaaaaacttt
tctgcggcttagttggaattgtctcaagtaactgggtgcctaattg-aaaaaaaatacttt
tctgcggcttagttggaaatgtatcaagtaacttggagcctaattg-aaaaaaattacttc
tctgtggctttgtaaaaaatgtatcaagtaacttcgtgtctaattg-aaaaaatgcatcct
% of basepairs conserved
70
60
50
40
30
20
10
+
RCE
R4
R5
R6
+
RCE
R4
R5
R6
distal (-846 bp)
proximal (-121 bp)
B
E
RCE
spa(wt)
C
F
spa(ΔRCE)
D
G
RCE+
spa(RCE+)
distal (-846 bp)
proximal (-121 bp)
spa(wt)
spa(ΔRCE)
spa(RCE+)
spa(wt)
spa(ΔRCE)
spa(RCE+)
H
I
J
K
L
M

Figure 6.1 *sparkling* upstream conserved sequence enables the enhancer to drive GFP expression in primary pigment cells in addition to cone cells. Upon analysis of the DNA sequence surrounding the *sparkling* enhancer, we noticed there is a region of sequence immediately upstream of the RCE that is highly conserved (RCE+; A top). In fact, this sequence shows higher levels of conservation than the other essential sequences within the *spa* enhancer - regions 4, 5, and 6 (A bottom). When this 60 bp sequence is added to the *sparkling* enhancer, distal gene expression is inhibited such that it acts more like *spa*(ΔRCE) than *spa*(wt) (B – D). However, at the promoter proximal position (-121 bp) all three enhancers drive GFP expression in cone cells (E – G). By pupal stages the *spa*(RCE+) construct drives expression levels similar to *spa*(wt) at -846, while loss of the RCE still results in failure of transcription (H – J). At 121 bp from the transcriptional start site the *spa*(wt) and *spa*(ΔRCE) drive GFP expression in cone cells while *spa*(RCE+) drives expression in both cone and primary pigment cells (K – M).

than the other essential sequences within *sparkling* region 4, 5, and 6 which each have less that 30% of the base pairs conserved across the 12 sequenced *Drosophila* species. Comparatively, the RCE and these 60 bp, here after referred to as RCE+, have 68% and 54% of their base pairs conserved (Figure 6.1A). We generated reporter constructs containing this additional sequence at both the promoter distal and proximal positions (-846 bp and -121 bp upstream of the TSS) with the enhancers driving GFP expression from the heterologous hsp70 promoter.

Given that *spa*(wt)-846 drives GFP expression in cone cells, we were surprised to find that the addition of these extra base pairs resulted in a complete loss of GFP expression in larval tissues at this distance (Figure 6.1D). We know loss of the RCE results in an absence of gene transcription at -846 bp (Figure 6.1 C); however, *spa*(REC+)-846 contains the RCE sequence, and we would expect that this construct would have no problems with long-range function. Intriguingly, this same construct, *spa*(RCE), drives wildtype levels at gene expression from the promoter proximal position (Figure 6.1 G). In fact, like *spa*(ΔRCE)-121bp, *spa*(RCE+)-121bp drives approximately 50% more GFP expression than *spa*(wt)-121bp (Chapter 3, Table 2). We can explain these results in two different ways: 1) The RCE+ sequence contains a binding site, or binding sites, for a protein that inhibits *spa* activity. This inhibition must only be at a distance, possibly due to the strong nature of *spa* activity in the promoter proximal position (Figure 6.2 E, F). 2) The RCE+ sequence enables *sparkling* to interact strongly with promoters in local genomic environment.

We have hypothesized that wildtype *sparkling* can interact with nearby promoters in addition to the hsp70 promoter driving GFP. It is possible that the addition of these 60 bp increases the enhancer's affinity for other promoters. The insertion site locus 86F8 is actually quite promoter rich (Figure 6.2A). The 3xP3 promoter driving RFP expression the *white* promoter, and local Chloroform channel a (CIC-a) promoter could all be potential targets for *spa*(RCE+). To test this hypothesis we removed the RFP and *white* genes from the locus via *Mos.1-cre* mediated excision (Figure 6.2A). Loss of these promoters does not affect *spa*(wt) or *spa*(ΔRCE) activity (Figure 6.2, B and C). Similarly, we still observed no GFP expression driven by *spa*(RCE+)-846 (Figure 6.2 D). While we cannot rule out the interference of a CIC-a promoter 2 kb upstream, this data suggests the 60 additional base pairs does not alter *sparkling's* promoter preferences.

As *spa*(RCE+) was able to drive reporter gene expression from the promoter proximal position, but not from the distal position, we decided to look at enhancer activity in 24 hour pupae to see if *spa*(RCE+)-846 expression is fully inhibited, or if it is delayed as we have seen with loss of other enhancer regions (Chapter 2). At this stage, *spa*(wt) drives GFP expression in cone cells while *spa*(ΔRCE) remains unable to promote transcription (Figure 6.1H,I). In pupal discs *spa*(RCE+)-846bp does drive GFP expressions in cone cells, although the pattern is not complete (Figure 6.1 J) indicating that while *spa* activity is inhibited by the RCE+ sequence early, this inhibition is overcome such that gene expression is delayed.

Figure 6.2

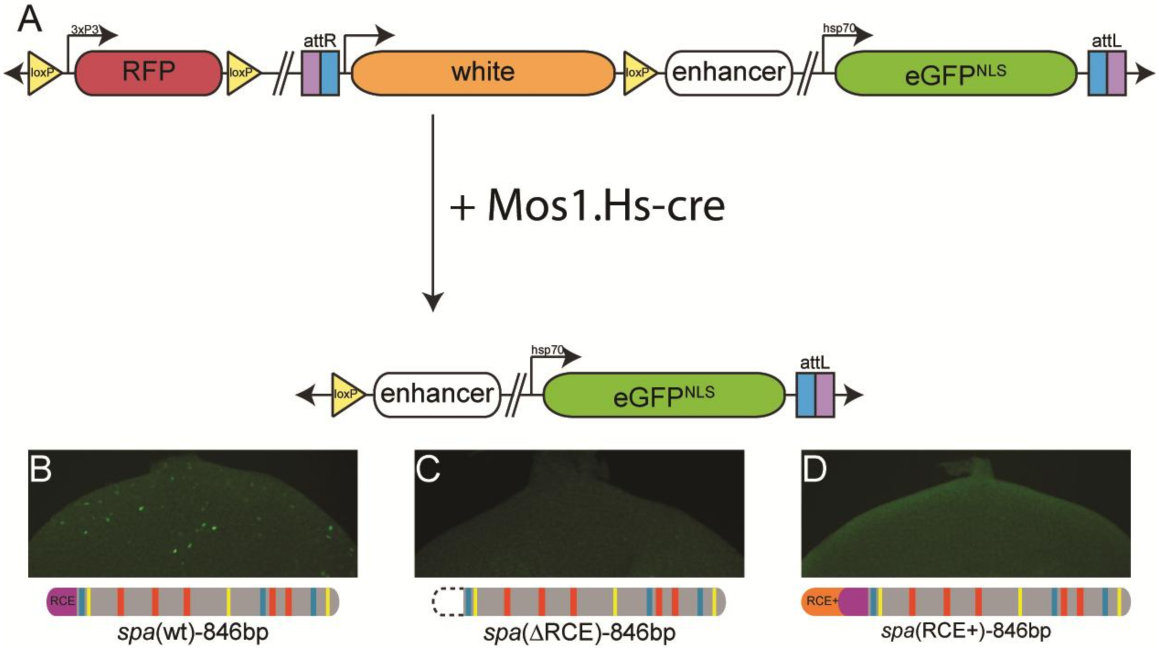


Figure 6.2 Removal of local promoters does not alter enhancer activity. The insertion locus (86F8) surrounding our reporter gene is rich in promoters (A). In addition to the hsp70 promoter driving GFP expression, there is also the 3xP3 promoter driving RFP expression and the *white* promoter. In order to rule out interference from these promoters on enhancer activity, we removed them using Mos1.HS-cre (A). After cre mediated removal of the RFP and *white* genes, *spa*(wt), *spa*(ΔRCE), and *spa*(RCE+) activity in the distal position remains unaltered (B – D).

Upon analysis of the *spa* constructs at 121 bp from the TSS. We observed GFP expression driven by *spa*(RCE+) in cone cells and primary pigment cells (PPCs) (Figure 6.1 M). Meanwhile, *spa*(wt) *and spa*(ΔRCE) drive expression in cone cells alone (Figure 6.1 K and L). As the EGFR and Notch pathways and the transcription factor Lozenge (Lz) are used to specify both the cone and primary pigment cells, it is not too surprising that the *spa* enhancer can be co-opted to act as a PPC enhancer with the addition of a small regulatory sequence (Voas and Rebay, 2004). Alternatively, as the enhancer responsible for PPC expression must also be in the 5' half of the *dPax2* 4th intron, these results indicate that the ppc enhancer may overlap, at least in part, with the *sparkling* cone cell enhancer (Fu et al., 1998; Hochman, 1971).

6.3b The dPax2 locus contains additional regulatory sequences

As we hypothesize that the *dPax2* PPC enhancer overlaps with the *sparkling* enhancer we designed a set of reporter constructs with DNA sequence starting at the beginning of the *dPax2* 4th intron and continuing through truncated portions of the *spa* enhancer (Figure 6.3 A). Using this approach, we can determine which sequences within *sparkling* are involved in PPC expression, thereby separating the PPC and cone cell regulatory sequences. Knowing that the PPC enhancer must lie in the *dPax2* 4th intron, we also generated 1 kb overlapping fragments spanning the first half of the intron (Figure 6.3 A). These reporter constructs were placed in the promoter proximal position (-121 bp) and integrated into the genome pseudorandomly using P-element transposons.

We have found that Frag 1, which starts at the beginning of the 4th intron and continues through the RCE to the 5' Lozenge and Ets binding sites, is not able to drive gene expression in cone cells or primary pigment cells (Figure 6.3 C). Frag 3 drives GFP expression in a few cells at the very posterior margin of the eye disc, which is cone cell specific as seen in 24 hour pupa (Figure 6.3 D). While this expression is poor compared to *spa*(RCE+) (Figure 6.3 B and C), it is astonishing we saw any cone cell activity at all as this sequence lacks the essential patterning sequences, *spa* regions 4, 5, and 6a. This suggests that the upstream 4th intron sequences contain additional regulatory information capable of inducing gene expression in cone cells. Similarly, Frag 5 which contains 140 more basepairs of *spa* sequence than Frag 3, also drives GFP expression in cone cells at the posterior margin of the eye disc (Figure 6.3 D). Interestingly, in pupal eye discs the expression driven by this fragment expands to fill every cone cell in the eye disc. However, we do not see any primary pigment cell expression from this construct (Figure 6.3 D). The Frag 6 sequence, which contains complete *spa* in addition to the upstream 4th intron sequence, drives GFP expression at slightly higher levels than *spa*(RCE+) supporting the observation that upstream 4th intron sequences contain binding sites for activator proteins in cone cell (Figure 6.3, E). This fragment also drives GFP expression in primary pigment cells (Figure 6.3 E). This is not unexpected as Frag 6 contains the entire *spa*(RCE+) sequence.

Our studies thus far have yet to separate the cone and PPC regulatory sequences, suggesting they may completely overlap. We have also analyzed

Figure 6.3

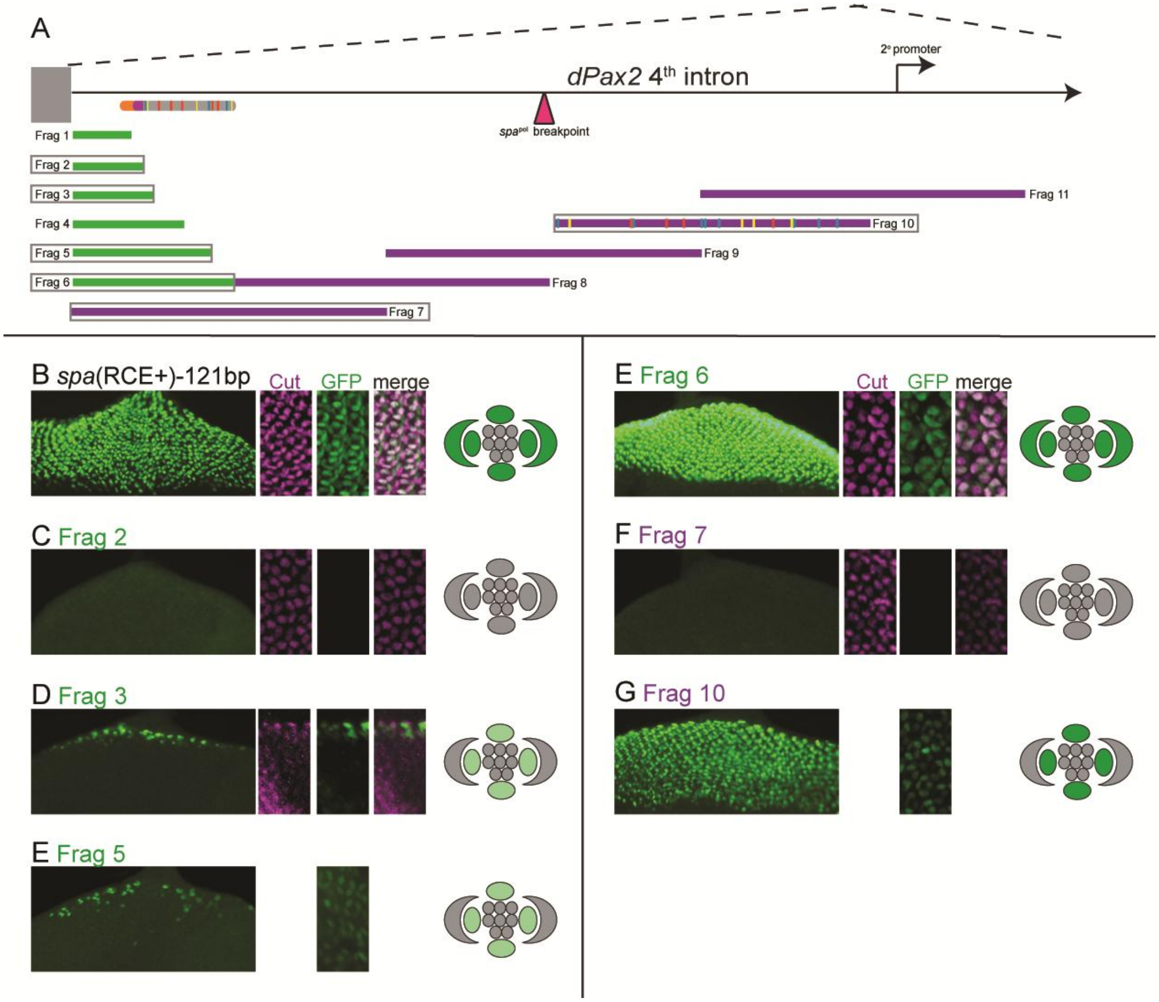


Figure 6.3 The *dPax2* 4th intron contains primary pigment cell and cone cell patterning information. In order to identify the minimal sequence required for gene expression in primary pigment cells we generated reporter constructs containing fragments of the *spa* enhancer (Frag 1 – 6, green). We also generated 1kb overlapping fragments that span the sequence deleted in the sparkling mutant (Frag 7 – 11, purple). *In vivo* results shown here are boxed in gray (A). Like *spa*(RCE+), Frag 5 is capable of driving GFP expression in cone cells and primary pigment cells (E). Frags 3 and 4 drive enhancer activity at the posterior margin of the disc (D and E). In pupa Frag 4 drives GFP expression in all cone cells (E). Frag 2 is incapable of driving any reporter gene activity (C). Similarly Frag 6 drives no transcription in larva or pupa (F). Finally, Frag 10, which does not include any *spa* sequence, activates reporter gene expression in cone cells (G). Putative protein binding sites are depicted, Suppressor of Hairless (red), Lozenge (blue), and Ets (yellow).

two of the 1 kb intronic fragments (Figure 6.3A). Interestingly, Frag 7 is unable to drive GFP expression in cone or primary pigment cells (Figure 6.3 F). At first glance, this is surprising as Frag7 contains the entire *spa* sequence. However, the additional sequences downstream of *spa* in this 1 kb fragment put *spa*, 0.7 kb upstream of the TSS. Recall that *spa*(RCE+)-846 does not drive GFP expression either. Unlike *spa*(RCE+)-846, GFP expression doesn't recover during pupal eye development (Figure 6.3, F). This suggests the upstream sequence contains additional inhibitory sequences that affect distal gene regulation.

Finally, Frag 10 which is downstream of *spa* surprisingly drives cone cell specific gene expression (Figure 6.3 G). This expression is weaker than that driven by *spa*, but it extends further toward the morphogenic furrow. From this we can conclude that Frag10 contains regulatory information capable of driving cone cell specific gene expression. In fact, this sequence contains 4 Su(H) sites (RTGRGAR) and 5 Ets (GGAW) sites(Bailey and Posakony, 1995; Flores et al., 2000; Nellesen et al., 1999). There are also two pairs of clustered Lz (RACCRCA) and Ets sites, which is especially interesting as these two proteins have been shown to interact and bind DNA cooperatively to stimulate strong transcription (Figure 6.3 A)(Dittmer, 2003; Goetz et al., 2000; Kim et al., 1999). Furthermore, the *spa* enhancer requires the linked association of these sites for its function (Swanson et al., 2010; Swanson et al., 2011). However, this sequence alone cannot regulate *dPax2* expression as the "sparkling" mutant mutation does not affect this sequence, with the mutation breakpoint about 100 bp upstream. Notably, Frag 10 is immediately upstream of, but does not include, the secondary *dPax2* promoter. Therefore, it is possible *sparkling* and/or Frag 10 regulatory sequences activate *dPax2* expression from this secondary promoter.

6.3c A potential role for Su(H) binding sites in the primary pigment cell enhancer

Our investigation into the DNA sequences flanking *sparkling* inspired us to look at the flanking sequence of putative *spa* orthologs. We know that conserved sequence from *D.erecta* and *D.annanase* are all capable of driving GFP expression in cone cells in the developing eye disc from the promoter distal position of -846 bp from the TSS (Evans et al., 2012). While each of these enhancers is active in cone cells they do not drive GFP expression in primary pigment cells (data not shown). To test the role of the flanking sequences, we generated new *spa* enhancer sequences from these species as well as from *D.virilis,* which contain 75 bp of 5' and 3' flanking sequence. The *D. ere* and *D. ana en*hancers each drive GFP expression in cone cells to various extents (Figure 6.4 C, E, and G). *D.virilis spa* was unable to activate gene transcription at -846; however, at -121 bp we see expression in some cone cells (Figure 6.4 G, H). As this is the most distantly related species we have looked for *spa* orthologs in, it is possible that we have not tried a large enough fragment yet, or the *dPax2* cone cell regulatory sequences lay elsewhere in *D.virilis*.

Like *spa*(RCE+)-121, the *D.ana* and *D.eve* enhancers containing 150 extra base pairs are active in cone and primary pigment cells (Figure 6.4 A-F).

Figure 6.4

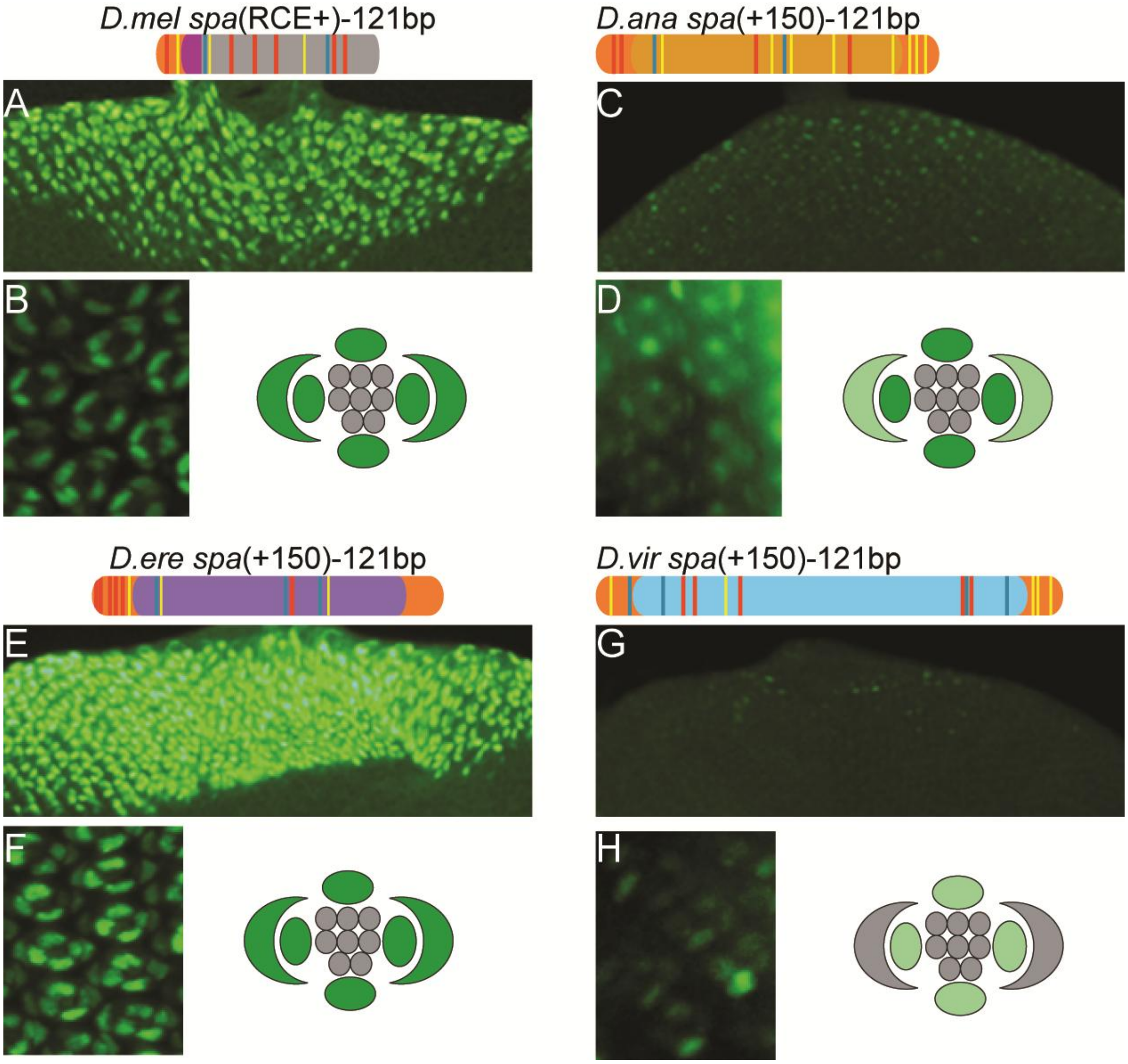


Figure 6.4 Additional sequences also enable ortholgous *sparkling* sequences to activate gene expression in primary pigment cells. We examined the role of additional flanking DNA sequences on the *spa* enhancers from *D.ananasse* (*D.ana*), *D.erecta* (D.ere), and *D.virilis* (*D.vir*). All three enhancers drive expression in cone cells to various extents from the promoter proximal position (C – H). In pupa, the *D.ana* and *D.ere*, but not *D.vir* sequences also drive GFP expression in primary pigment cells (D, F, and H). Flanking enhancer sequences are show in orange. Putative protein binding sites are depicted; suppressor of hairless (red), lozenge (blue), and Ets (yellow).

This suggests these orthologous enhancers also contain ppc regulatory information. One input all three constructs share is increased Su(H) input. The RCE+ sequence contains a high affinity Su(H) site, while the 5' flanking sequences in *D.ana* contains two Su(H) sites in addition to the two within the minimal enhancer sequence. *D.ere* 5' flanking sequence contains 5 Su(H) sites in addition to the single site within the minimal enhancer. Meanwhile, the *D.virilis spa*, which is not expressed in PPCs, contains only 1 Su(H) binding site (Figure 6.4 H). Primary pigment cell specification is known to require higher levels of Notch signaling than cone cells (Voas and Rebay, 2004). In fact, if the Su(H) sites within wildtype *spa* are converted to high affinity binding sites, *spa* drives gene expression in cone and primary pigment cells (Swanson et al., 2011). Together, this data suggests that increased Su(H) input may be important to the PPC regulatory sequence.

6.4 Discussion

The *sparkling* enhancer regulates the cone cell specific gene expression of the *dPax2* gene in the developing *Drosophila* eye disc. We found a conserved sequence immediately upstream of the enhancer that affects both proximal and distal *spa* activity.

6.4a The RCE+ sequence inhibits distal sparkling activity

When the 60bp of conserved sequence is added to the *spa* enhancer acting at -846bp, the reporter gene is inhibited (Figure 6.1). As this enhancer contains all the sequences necessary for *spa* activity, we proposed two explanations for this surprising result. The first explanation invokes the hypothesis that *sparkling* can interact with promoters from the local genomic environment. We have observed that *spa*(wt)-121bp drives slightly decreased levels of expression compared to *spa*(ΔRCE)-121bp in both randomly inserted and site specific integration of uninsulated, reporter genes, which could be explained by *spa,* potentially through the RCE, interacting with multiple local promoters and therefore, spending less "time" activating GFP expression. *spa*(RCE+)-121bp drives expression levels similar to those activated by *spa*(ΔRCE)-121bp suggesting that sequences within this region allow for increased, short range, activation or enable the enhancer to interact more strongly with the hsp70 promoter. We also have evidence that wildtype *sparkling* can interact preferentially with specific promoters at the expense of other local promoters. Given the lack of reporter expression from *spa*(RCE+)-846bp, it is possible that the extra 60bp enable the enhancer to interact so strongly with other local promoters that no GFP expression is seen. *spa*(RCE+) enhancers are integrated at the 86F8 landing site via Φ3C1 mediated recombination. We removed the 3xP3-RFP and *white* gene from the locus to assess this possibility. However, we did not see any change in reporter activity (Figure 6.2). We cannot rule out potential promoter interactions with the CIC-a promoters that remain in the locus, however this observation makes an alternative explanation more likely.

It is possible that the RCE+ sequence contains one or more binding sites for a protein that is capable of inhibiting enhancer activity. This repression only occurs when the enhancer is distally located however, as this enhancer is fully functional at the proximal position (Figure 6.1). This inhibitor input must also be inhibited in the context of the endogenous *spa* location as well, as all of this sequence is present in the native locus. In support of an inhibitory sequence, we found that GFP expression is recovered in our reporter constructs by the 24 hour pupal stage. This observation suggests that the early inhibition of *sparkling* activity can be overcome resulting in delayed reporter gene expression. Alternatively, the protein responsible for repressing *spa*(RCE+)-846bp in larva is no longer expressed at this stage. In order to further asses the inhibitory affects of *spa*(RCE+)-846bp we can make mutations to this sequence to localize the DNA sequence responsible for this activity. We can also test the ability of *spa*(RCE+)-846bp to regulate transcription from different core promoters utilizing promoter competition assays in which two promoters drive different reporter genes.

6.4b The RCE+ allows for sparkling activity in primary pigment cells

Unlike *spa*(RCE+)-846bp, the same construct at -121bp from the TSS, drives reporter gene expression in cone cells and primary pigment cells. As the wildtype *spa* enhancer does not drive GFP expression in ppcs, this additional DNA sequence must contain information that allows for this expression. *dPax2* is expressed in primary pigment cells, and loss of this expression contributes to the sparkling mutant phenotype (Fu et al., 1998; Fu and Noll, 1997). As the primary pigment cell enhancer has yet to be identified, we wondered if it could overlap with the *sparkling* cone cell enhancer. As such we generated fragments containing this upstream sequence and portions of the *spa* enhancer. While we have not completed analysis of these fragments, can make educated predictions of their activity based in the results we do have. For example, as Frag 2 does not drive any GFP expression in larva or pupa, it is likely that Frag 1 will also not be capable of activating reporter gene expression. Likewise, as Frag 3 and Frag 5 drive similar levels and patterns of GFP expression in larval eye discs, it is likely that the Frag 4 will resemble this expression. However, Frag 5, which contains more of the essential *sparkling* sequences, drives GFP expression in all of the cone cells of the pupal eye disc; a dramatic difference from Frag 3 which only drives GFP expression in the very posterior cells (Figure 6.3). Therefore, it will be interesting to analyze Frag 4 expression in pupal discs.

Only Frag 6, which contains all of the *spa*(RCE+) sequence is capable of activating reporter gene expression in both cone cells and primary pigment cells (Figure 6.3). These results indicate that they ppc and cone cell enhancers for *dPax2* expression may completely overlap. Such enhancer action would not be unsurprising as the photoreceptors, cone cells, and primary pigment cells are specified utilizing the same signaling pathways, Notch and EGFR, as well as the transcription factor Lozenge. We know that *sparkling* contains sequence information critical for its repression in photoreceptors (Swanson et al., 2010). It is known that specification of the primary pigment cells requires higher levels of

Notch signaling than other eye cell types (Voas and Rebay, 2004). Interestingly, there is a high affinity Suppressor of Hairless, Su(H), binding site (YGTGR-GAAM) in the RCE+ sequence, suggesting increased Notch input may be responsible for *sparkling* activity in primary pigment cells (Crocker et al., 2010; Flores et al., 2000). The 5 Su(H) binding sites within the *spa* enhancer also deviate from the looser, lower-affinity consensus RTGRGAR (Bailey and Posakony, 1995; Nellesen et al., 1999). When these sites are converted to high affinity Su(H) binding sites the *spa* enhancer drives reporter gene expression in both cone cells and primary pigment cells (Swanson et al., 2011). Together, with our observations that only *spa* enhancer orthologs containing increased Su(H) input act as PPC regulatory sequences, these data suggest that increased Notch signaling is critical for primary pigment cell enhancer activity. As such, we will make targeted mutations that abolish or alter affinity to these additional Su(H) binding sited and assess the effect on primary pigment cell enhancer activity.

6.4c A second cone cell enhancer exists in the dPax2 4th intron

Our investigation of the regulatory sequences within the *dPax2* 4th intron also identified an additional sequence capable of activating reporter gene expression in cone cells. This expression is decreased compared to that driven by *sparkling,* but it also extends further toward the morphogenic furrow, suggesting the enhancer is active earlier than *sparkling* (Figure 6.3). It is important to note that while Frag 10 can activate reporter expression in cone cells, it is not sufficient for *dPax2* expression *in vivo* as this sequence is present in the sparkling mutant flies which lack *dPax2* expression in the cone and primary pigment cells (Fu et al., 1998). It is possible that the enhancer sequence within Frag 10 acts early to activate *dPax2* expression, and the *spa* enhancer is required for further expression and maintenance of the gene's expression. The cooperative action of these regulatory sequences can be assessed by analysis of mutations to the enhancers using BAC transgenesis. The proximity of Frag 10 to the secondary *dPax2* promoter raises the possibility that it may regulate gene transcription from this promoter. The first three exons of *dPax2* are poorly conserved; therefore, it is possible the primary or essential transcript in the *Drosophia* eye is actually from the secondary promoter. We will perform *in situ* analysis for each transcript to determine which transcript is more prevalent in the eye. We will also generate *spa* and Frag 10 reporter constructs using this promoter. Overall, this preliminary work provides exciting new directions for the study of *dPax2* regulatory sequences in our laboratory.

6.5 Experimental methods

6.5a Reporter gene construction, transgenesis, and genetics

Sparkling enhancer sequences where generated by sewing PCR and cloned into peGFPattB (peaB) via with EcoRI and BamHI digestion. *dPax2* 4th intron fragments and *spa* orthologous enhancer sequences were amplified by PCR from genomic DNA of each species and inserted into Ganesh-G2 cloning vectors as previously described (Swanson et al., 2010). *dPax2* intronic fragment sequences are as follows:

Frag 2
GTAAGATATTTCTATAGATATACATATATGTATACTTACTAGTAACGTG
TGACCTATCTCCTACCTAATCATTCACACTGATTTTCGCATCAGTAAA
GATTTCTCACAAAGCTATATAACCACCATCCGATAAATTTTTTGCGGC
TTAGTTGGAATTGTATCAAGTAACTGGGTGCCTAATTGAAAAAATTTA
CTAT

Frag 3
GTAAGATATTTCTATAGATATACATATATGTATACTTACTAGTAACGTG
TGACCTATCTCCTACCTAATCATTCACACTGATTTTCGCATCAGTAAA
GATTTCTCACAAAGCTATATAACCACCATCCGATAAATTTTTTGCGGC
TTAGTTGGAATTGTATCAAGTAACTGGGTGCCTAATTGAAAAAATTTA
CTATGAC:CGCAAAGCTGTTTCCTGACTATGACATAGTTTTTTTTGCTT
TGGTTGTGGGATGTAA

Frag 5
GTAAGATATTTCTATAGATATACATATATGTATACTTACTAGTAACGTG
TGACCTATCTCCTACCTAATCATTCACACTGATTTTCGCATCAGTAAA
GATTTCTCACAAAGCTATATAACCACCATCCGATAAATTTTTTGCGGC
TTAGTTGGAATTGTATCAAGTAACTGGGTGCCTAATTGAAAAAATTTA
CTATGAC:CGCAAAGCTGTTTCCTGACTATGACATAGTTTTTTTTGCTT
TGGTTGTGGGATGTAAATGGTCATTGGAACTGGACGCTGTCCCTGTC
TTCTCACTAAGTTAATGATCGTACAACCTCAAGATCTTATTCACATTGA
AATTGAAGCACTATTGGTGTACGATTACAACGCTCACATTATCAGGAT
ATAAAAAAAAGGTGATAGTAATTCAGCACGACTTTGTAACCACAAATA
TATGGGAACACAGATTACTCCGTGAGTACAACG

Frag 6
GTAAGATATTTCTATAGATATACATATATGTATACTTACTAGTAACGTG
TGACCTATCTCCTACCTAATCATTCACACTGATTTTCGCATCAGTAAA
GATTTCTCACAAAGCTATATAACCACCATCCGATAAATTTTTTGCGGC
TTAGTTGGAATTGTATCAAGTAACTGGGTGCCTAATTGAAAAAATTTA
CTATGACCGCAAAGCTGTTTCCTGACTATGACATAGTTTTTTTTGCTTT
GGTTGTGGGATGTAAATGGTCATTGGAACTGGACGCTGTCCCTGTCT
TCTCACTAAGTTAATGATCGTACAACCTCAAGATCTTATTCACATTGAA
ATTGAAGCACTATTGGTGTACGATTACAACGCTCACATTATCAGGATA
TAAAAAAAAGGTGATAGTAATTCAGCACGACTTTGTAACCACAAATAT
ATGGGAACACAGATTACTCCGTGAGTACAACGTAAGTCGGGTGAAGC
CAGAAACCACAAATCAAGTTGTTTTCCGGTAGCTTAGG

Frag 7

GTAAGATATTTCTATAGATATACATATATGTATACTTACTAGTAACGTG
TGACCTATCTCCTACCTAATCATTCACACTGATTTTCGCATCAGTAAA
GATTTCTCACAAAGCTATATAACCACCATCCGATAAATTTTTTGCGGC
TTAGTTGGAATTGTATCAAGTAACTGGGTGCCTAATTGAAAAAATTTA
CTATGACCGCAAAGCTGTTTCCTGACTATGACATAGTTTTTTTTGCTTT
GGTTGTGGGATGTAAATGGTCATTGGAACTGGACGCTGTCCCTGTCT
TCTCACTAAGTTAATGATCGTACAACCTCAAGATCTTATTCACATTGAA
ATTGAAGCACTATTGGTGTACGATTACAACGCTCACATTATCAGGATA
TAAAAAAAAGGTGATAGTAATTCAGCACGACTTTGTAACCACAAATAT
ATGGGAACACAGATTACTCCGTGAGTACAACGTAAGTCGGGTGAAGC
CAGAAACCACAAATCAAGTTGTTTTCCGGTAGCTTAGGTATCTACTTC
CGGTGCTAAAGGACTTTTTAATTTAAGTAACAATTTTCAATTTATGTGC
AAAAAAGTCTTAATCATATATAATGTGTACAAAACTAATGATCAATGCT
AGGCAATAACTTTCAAAAACATGAATATTCTAAAAAATTCTTAAATAGG
GCGGGTTGGATTTATAGAAACCAAAAGTTAAATTTCAAAAGTCCGTTG
AGTAAACAATTTTTTCTGACGAAAATAAAGTATTCTCATTAGCCTATAC
AAAATATTTAAATTTTTAACCGGCCACTTTCTATTTAAGGGTCCCGAAA
GAGTATATGTATACATTGTAGTTATAGTAAAGTTAGCAGAAATCAAAAT
TTTTTCCACTGTATGTTTAAATATACATTGATACTATGACCTTAAGTAG
CTCGCGTAAGCCTTTTATATATTGATAACTACTCGAGTACTCAACTAG
TGGGAAAGCGAACAATAAATGTTAATATAGTA

Frag 10

TTTTTGCAACGTAGGTGGTTTAGGTTTGCTAAACAAACGCGAAGTAGT
TTCCTTGATATCATTGCTATAAACATAATTTACAAATAAAATTACAAAT
CCCGATATTTAAATAGAAGTGAAGGGGGATTTTATTGGACGGGAAAG
CATTGTCTATTTACATACGAATTGATGCTTTTTTAGGCTTGACTTTGAA
ATTATTATATTTAAAAGAAAAATACCTGGAAATAAGTTACATTTATTAC
GTTATTTCCCTGAAGTAAATCAGTTCACTATTAAACTGATCACATTTCT
AAATACAGTCTTGAAACATATTAGCCCTTGTTGAAAATTTTGCAGAGA
TGAAAATATTTTTTAAATTCTAATGGGTTGGTACGGGATTGATGCAAA
GAAGCGGCTGCGTGGTATTCAAGAGTGAGAAAGAGATGTCGATAAG
CTCAGTGGCTACACGTTAATGAAGTAGGCATGAAGTTGTGCTGTGAT
TGGTGCCGTTGCGCACGATGACAGACCTTCTTTGTGGCTATTGGTTG
ATTTCGAGGCCGCTCGCTTGCTCATTCCAATCCGAAATCTGTACATTT
CAGTCCAAGACGAGGAAACTAAAAATTCTATAAATGTTCTTGGTATTT
CAAAATGGGAGTTTAGTGTTTTTTTGGCTTTTCAACATAATACAATTAA
TTTATAGAGAGTATTATAAATTAATTGTTATATGTTGAAAGCCCAAATT
ATGCCAGCACAAAACCTAAAATTGCCTAAAAATAAACCTTATTCCGTC
CACACAAATGCTGGCATGCTTTGAGCGATTCGATTGGCCACTACAAG
AGCAGACACCCAGTATTGTCGTTTGGCGATACTACGCCTTGACCGTT
CCCCGGCATTAGACCGTCTTCATTGGCTGACTGCACTAAGATGAAAA
CGCCAACCGATACGAATGGGCGGAGCGGTGTCGCTGTAAGTGTAAA
GAGAACGATGTATTGCCAAGGACATGCATACGCCCGCGCATCTC

Reporter vectors containing attB sites were integrated into the Drosophila genome vasaintDm(Φ3C1) and integrated in the attP site at 86F8 obtained from the Bloomington Stock Center (24749) (Bischof et al., 2007). P-element transformation was performed in *w*^{1118} flies as described previously (Rubin and Spradling, 1982).

Mos1.HS-cre mediated recombination was performed as described previously (Siegal and Hartl, 1996). Excision of the RFP and *white gene* promoters was verified by loss of *white* expression in the adult eye and absence of RFP in the optic nerve of dissected eye discs.

6.5b Tissue preparation, antibody staining, microscopy

Eye disc tissues were dissected from third instar larvae. Disc tissues were then fixed in 4% paraformaldehyde at room temperature for 30 minutes. Discs were then washed 3x5 minutes in 1xPBS and mounted in Prolong Gold with 4', 6' – diamidino –2 phenylidole (DAPI) (Invitrogen). Imaging was performed on an OlympusBX51 microscope with an Olympus DP70 digital camera.

Immunohistochemistry was performed on dissected eye discs from 24 hour pupa. Discs were fixed in 4% paraformaldehyde for 30 minutes at room temperature and then washed 3x10 minutes in PBS-Tx (1xPBS + 0.1% Triton x-100). Fixed discs were then incubated in PBS-Tx + 2% BSA for 1 – 3 hours and then incubated overnight in primary antibodies against GFP (Invitrogen) and Cut (*Drosophila* studies Hydrodoma Bank) diluted 1:100. The next day, tissues were washed 3x10 minutes with PBS-Tx and then incubated in secondary antibodies;

goat anti-mouse 568 nm and goat anti-rabbit 488 nm (Invitrogen) diluted 1:1000. Finally, the discs were washed 3x20 minutes in PBS-Tx and mounted in Prolong Gold with DAPI (Invitrogen). Stained discs were imaged on an Olympus FLUO View 500 Laser Scanning Confocal microscope mounted on and Olympus 1x71 inverted microscope.

6.6 Acknowledgments

We would like to thank Alex Chapell for her contribution to the cloning, dissection, and staining of the *dPax2* intronic constructs.

6.7 References

Bailey, A.M., and Posakony, J.W. (1995). Suppressor of Hairless directly activates transcription of *Enhancer of split* Complex genes in response to Notch receptor activity. Genes Dev *9*, 2609-2622.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. Proceedings of the National Academy of Sciences of the United States of America *104*, 3312-3317.

Crocker, J., Potter, N., and Erives, A. (2010). Dynamic evolution of precise regulatory encodings creates the clustered site signature of enhancers. Nat Commun *1*, 99.

Czerny, T., Bouchard, M., Kozmik, Z., and Busslinger, M. (1997). The characterization of novel Pax genes of the sea urchin and Drosophila reveal an ancient evolutionary origin of the Pax2/5/8 subfamily. Mech Dev *67*, 179-192.

Dittmer, J. (2003). The biology of the *Ets1* proto-oncogene. Mol Cancer *2*, 29.

Evans, N.C., Swanson, C.I., and Barolo, S. (2012). Sparkling insights into enhancer structure, function, and evolution. Curr Top Dev Biol *98*, 97-120.

Flores, G.V., Duan, H., Yan, H., Nagaraj, R., Fu, W., Zou, Y., Noll, M., and Banerjee, U. (2000). Combinatorial signaling in the specification of unique cell fates. Cell *103*, 75-85.

Fu, W., Duan, H., Frei, E., and Noll, M. (1998). *shaven* and *sparkling* are mutations in separate enhancers of the *Drosophila Pax2* homolog. Development *125*, 2943-2950.

Fu, W., and Noll, M. (1997). The *Pax2* homolog *sparkling* is required for development of cone and pigment cells in the *Drosophila* eye. Genes Dev *11*, 2066-2078.
Goetz, T.L., Gu, T.L., Speck, N.A., and Graves, B.J. (2000). Auto-inhibition of Ets-1 is counteracted by DNA binding cooperativity with core-binding factor alpha2. Molecular and Cellular Biology *20*, 81-90.
Hochman, B. (1971). Analysis of chromosome 4 in Drosophila melanogaster. II. Ethyl methanesulfonate induced lethals. Genetics *67*, 235-252.
Johnson, S.A., Harmon, K.J., Smiley, S.G., Still, F.M., and Kavaler, J. (2011). Discrete regulatory regions control early and late expression of D-Pax2 during external sensory organ development. Dev Dyn *240*, 1769-1778.
Kim, W.Y., Sieweke, M., Ogawa, E., Wee, H.J., Englmeier, U., Graf, T., and Ito, Y. (1999). Mutual activation of Ets-1 and AML1 DNA binding by direct interaction of their autoinhibitory domains. Embo J *18*, 1609-1620.
Lindsley, D.L.a.Z., G. G. San Diego, CA: Academic Press. (1992). The Genome of Drosophila melanogaster. San Diego, CA: Academic Press.
Macdonald, R., and Wilson, S.W. (1996). Pax proteins and eye development. Curr Opin Neurobiol *6*, 49-56.
Nellesen, D.T., Lai, E.C., and Posakony, J.W. (1999). Discrete enhancer elements mediate selective responsiveness of *Enhancer of split* Complex genes to common transcriptional activators. Dev Biol *213*, 33-53.
Rubin, G.M., and Spradling, A.C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. Science *218*, 348-353.
Siegal, M.L., and Hartl, D.L. (1996). Transgene Coplacement and high efficiency site-specific recombination with the Cre/loxP system in Drosophila. Genetics *144*, 715-726.
Swanson, C.I., Evans, N.C., and Barolo, S. (2010). Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. Dev Cell *18*, 359-370.
Swanson, C.I., Schwimmer, D.B., and Barolo, S. (2011). Rapid evolutionary rewiring of a structurally constrained eye enhancer. Curr Biol *21*, 1186-1196.
Voas, M.G., and Rebay, I. (2004). Signal integration during development: insights from the *Drosophila* eye. Dev Dyn *229*, 162-175.

CHAPTER 7

SITE SPECIFIC INTEGRATION OF DEVELOPMENTAL ENHANCERS; EVERY POSITION HAS AN EFFECT

7.1 Abstract

The ability to generate and integrate foreign DNA *in vivo* is critical to the study of molecular and developmental biology, including for the investigation of regulatory sequences within the genome that influence gene expression. In recent years, a new system for genomic integration has been developed for use in *Drosophila* which allows for the integration of transgenes into a single known location utilizing the bacteriophage ΦC31. Site-specific methods of integration provide numerous experimental benefits compared to random integration processes and therefore poised to become the prominent and expected method of genomic integration in *Drosophila.* As such we began a survey of enhancer activity for several developmental enhancers in multiple. Here, we present some of our experiences and observation regarding the underappreciated limitations associated with site-specific integration. This work reveals that landing sites are not free from the position affects and enhancer trapping that complicates transgene analysis in randomly integrated DNA constructs. Furthermore, transgenes integrated at one particular locus, 51D9, shows both enhancer trapping and latent enhancer trapping dependent on the tissue type and the identity of the insertion. Finally, for at least one extensively characterized

enhancer we were unable to identify a single landing site in which reporter constructs recapitulated the documented wildtype enhancer activity. Together, this work demonstrates the importance of understanding the influence of the local genomic environment regardless of integration method, and demonstrates that random methods of integration should not be eliminated as useful experimental tools.

SITE SPECIFIC INTEGRATION OF DEVELOPMENTAL ENHANCERS; EVERY POSITION HAS AN EFFECT

7.2 Introduction

Genetic manipulation, from gene deletion to the expression of transgenes, is a cornerstone of developmental and disease research. The tool kit to create these alterations is large and spans model organisms. For example, in mice, gene knockout is often the result of homologous recombination targeting the gene of interest; however, the insertion of transgenes can be achieved either by random integration into the genome or targeted insertion into a “neutral locus” such as Rosa26 or β-actin (Jagle et al., 2007; Smithies et al., 1985; Soriano, 1999; Thomas and Capecchi, 1987). Similarly, both random and location specific integration of transgenes can be performed in zebrafish and *c. elegans* (Liu et al., 2007; Mello et al., 1991; Robert and Bessereau, 2007; Zhu and Sun, 2000); meanwhile, only random integration is currently utilized to generate transgenic chicken embryos (Sato et al., 2007). The ability to generate and integrate foreign DNA *in vivo* is critical to the study of gene function in organ development, life span, and disease progression. Furthermore, the integration of reporter genes *in vivo* allows for the investigation of regulatory sequences within the genome that influence gene expression.

For more than 30 years the integration of transgenes in the model organism *Drosophila melanogaster* has been achieved using the powerful tool of retro transposition, including piwi, piggyback, and most commonly P-element transposases (Rubin and Spradling, 1982; Spradling and Rubin, 1982). Using the later method, P-transpose recognition sites flanking the coding sequence or reporter construct of interest, allows for random integration of the transgene into the *Drosophila* genome. In practice P-element mediated integration events insert into the genome with a preference for promoter elements and other regulatory regions (Bellen et al., 2004). Therefore, this *pseudorandom* integration has proved useful for generating genomic mutations and deletions as well as for the analysis of transgenes and *cis*-regulatory sequences. However, there are also many limitations to transgene analysis after P-element insertion including: *i.* moderate to low transformation efficiency, especially when integrating large (>40kb) DNA fragments, *ii.* integration into regulatory regions can subtly affect endogenous gene expression, thereby affecting any anylasis and occasionally resulting in homozygous lethal trangenics, *iii.* in-depth analysis requires considerable effort to map the location of the insertion, and *iv.* as in any model organism, the local chromatin environment can cause *position effects* that strongly influence activity of the transgene - and these effects are different at each integration location necessitating the analysis of several independent insertions for each study (Bellen et al., 2004; Levis et al., 1985; Venken et al., 2006).

In recent years, a new system for genomic integration has been developed for use in *Drosophila* which allows for the integration or transgenes into a single, known location. This system utilizes of the bacteriophage ΦC31, a serine integrase. ΦC31 mediates recombination between a bacterial attachment site (attB) DNA element and a phage attachment site (attP) DNA element (Thorpe and Smith, 1998). Initially, an attP site was integrated randomly into the genome using P-element transposition at two independent locations, referred to as attP2 and attP40 (Groth et al., 2004). Subsequently, these stably integrated attP sites allowed for targeted insertion of a transgene by recombination with a plasmid containing an attB site into either of these independent landing sites (Cande et al., 2009; Kalay and Wittkopp, 2010; Markstein et al., 2008; Ni et al., 2008; Pfeiffer et al., 2010; Potter and Luo, 2010). Since the advent of this tool, over 100 attP sites have been integrated into the *Drosophila* genome using P-element, mariner, and piggyback transposases (Bateman et al., 2006; Bischof et al., 2007; Venken et al., 2006). This system allows researchers to overcome many of the limitations of P-element transposition as the location of every singly integration event is known and controlled. Therefore, position effects are thought to be minimized. Furthermore, piggyback transposase does not seem to possess the same preferences for promoters and regulatory sequences as P-transposase (Bellen et al., 2004). As such, most attP sites should be integrated in a more evenly distributed manner. Accordingly, the genomic location of every attP site inserted using mariner transposase was subsequently mapped. Only those that integrated at intergenic locations and were homozygous viable were selected for further experimental use. Transgene insertion into the genome also occurs with a higher transformation efficiency using ΦC31 mediated recombination compared to P-element transposition even for large DNA fragments and those so large as to be considered outside the previous range for integration (Bischof et al., 2007; Groth et al., 2004; Venken et al., 2006). Due to the simplicity of both creating and analyzing attB/attP inserted transgenes, the ΦC31 transformation system is fast becoming the standard and expected method for generating transgenic *Drosophila.*

Our laboratory has utilized P-element transposition over the past decade to extensively study several developmentally regulated enhancers in Drosophila (Evans et al., 2012; Johnson et al., 2008; Parker et al., 2011; Swanson et al., 2010; Swanson et al., 2011). As such, before switching our analysis of *cis*-regulatory sequences to the promising integration method of Φ3C1 mediated recombination, we began a survey of the wildtype enhancer activity for the regulatory sequences we study in several independent landing sites. Here, we present some of our experiences and observation regarding the underappreciated limitations associated with site-specific integration. Focusing primarily on the *dPax2* cone cell enhancer *sparkling* (*spa*) in combination with other minimal enhancers, our survey reveals that landing sites are not free from the position affects and enhancer trapping that complicates transgene analysis. Furthermore, transgenes integrated at the 51D9 locus show both enhancer trapping and latent enhancer trapping dependent on the tissue type and the identity of the insertation. Finally, despite the fact that the expression pattern of *sparkling* has been well documented by us and others using P-element transposition, we were unable to identify a single attP landing site in which reporter constructs recapitulated the wildtype pattern of *sparkling* activity. Together, this work demonstrates the importance of understanding the influence of the local genomic environment regardless of integration method, and the occasional necessity of utilizing multiple insertion loci in order to best understand a transgenes activity which may possibly be easiest achieved via random methods of integration.

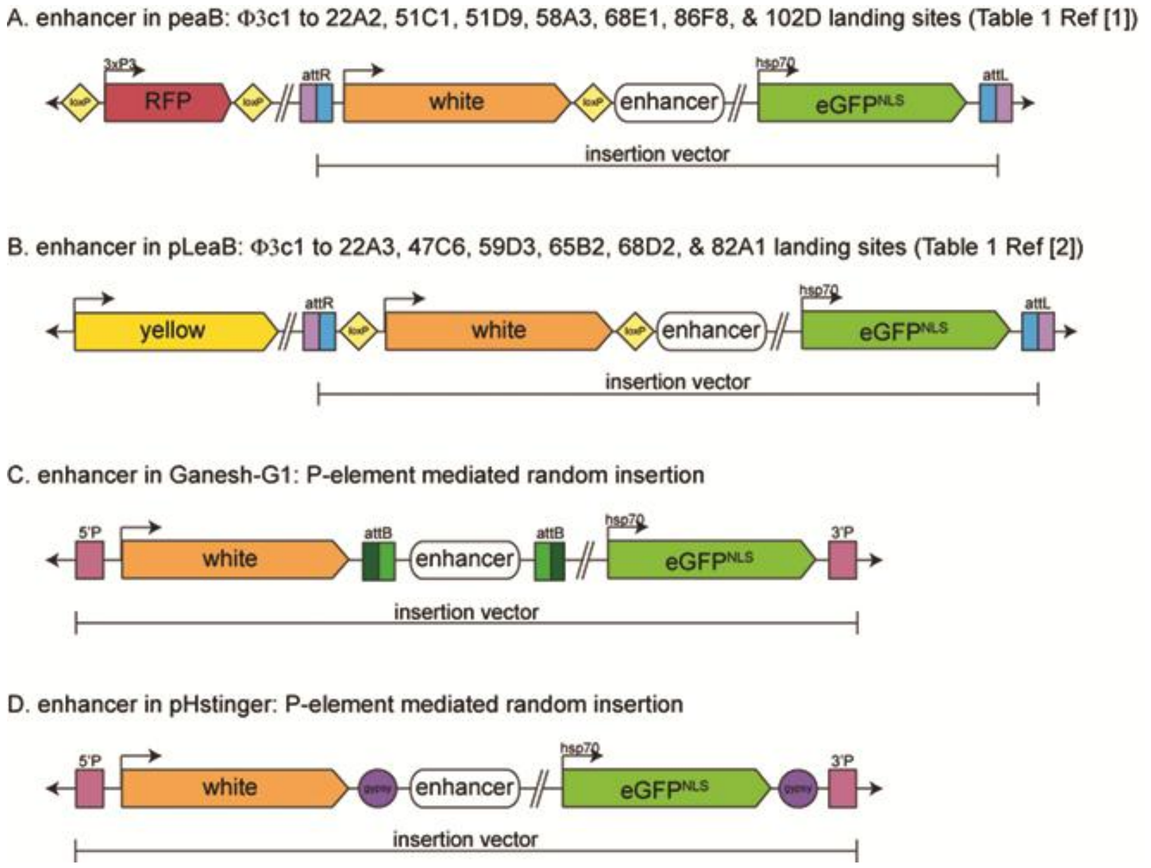
7.3 Results

7.3a We are unable to identify a landing site which allows for wildtype sparking activity

The *Drosophila* compound eye is composed of approximately 750 ommatidia, or simple eyes. Each ommatidium consists of eight photoreceptors (R1-R8), four cone cells, two primary pigment cells, six secondary pigment cells, three tertiary pigment cells, and three mechanosensory bristles (Voas and Rebay, 2004). During eye development, *dPax2* expression is required for proper cone cell differentiation and maintenance. The expression of *dPax2* specifically in cone cells is controlled in the eye imaginal discs by the *sparkling* enhancer, which is located in the 4th intron of the *dPax2* gene (Flores et al., 2000; Fu et al., 1998; Fu and Noll, 1997; Shi and Noll, 2009). In order to study the gene regulation mediated by *sparkling,* our laboratory uses reporter constructs in which *spa* is placed upstream of a minimal hsp70 promoter driving a nuclear localized eGFP (Figure 7.1 and Figure 7.2 A). Our initial studies investigating *spa* were performed with the enhancer cloned into the Ganesh-G1 vector backbone (Evans et al., 2012; Evans, 2012; Johnson et al., 2008; Swanson et al., 2010; Swanson et al., 2008). When using this Gateway vector, the reporter construct is integrated into the *Drosophila* genome utilizing P-element transgenesis (Figure 7.1C) (Rubin and Spradling, 1982). We have previously demonstrated that the wildtype *spa* enhancer placed 846bp upstream of the GFP transcriptional start site (TSS), *spa*(-846bp) drives reporter gene expression in cone cells during larval and pupal eye development (Figure 7.2 C,D) (Swanson et al., 2010). It is worth noting, however, that this expression is substantially decreased compared to that seen when *spa* is placed in the more proximal position of 121 base pairs upstream of the TSS, *spa*(-121bp) (Evans et al., 2012; Swanson et al., 2010). This suggests that the 362bp minimal *spa*(-846bp) is a relatively weak enhancer.

In order to continue our investigation of *spa* activity we decided to utilize the “new” method of site specific integration into the *Drosophila* genome utilizing Φ3C1 mediated recombination of attB and attP attachment sites (Bischof et al., 2007; Groth et al., 2004; Venken et al., 2006). As such, we first generated a reporter construct vector suitable for studying enhancers in landing sites. Starting with the pUASattB vector (Bischof et al., 2007), we swapped the UAS-MCS-SV40 cassette for the multiple cloning site, hsp70 promoter, and eGFP-NLS sequences from the pHstinger reporter vector (Barolo, 2000). This

Figure 7.1



A. enhancer in peaB: Φ3c1 to 22A2, 51C1, 51D9, 58A3, 68E1, 86F8, & 102D landing sites (Table 1 Ref [1])

B. enhancer in pLeaB: Φ3c1 to 22A3, 47C6, 59D3, 65B2, 68D2, & 82A1 landing sites (Table 1 Ref [2])

C. enhancer in Ganesh-G1: P-element mediated random insertion

D. enhancer in pHstinger: P-element mediated random insertion

Figure 7.1 Model of cloning vectors integrated into the relevant genomic context. AttP landing sites at 22A2, 51C3, 51D9, 58A3, 68E1, 86F8, and 102D were generated by integrating the attachment site with a 3xP3 RFP marker flanked by LoxP sites. We generated the cloning vector peGFPattB (peaB) which contains the transgenic marker *white* with a 3' LoxP site, a multiple cloning site (MCS) for enhancer ligation, and hsp70-eGFP reporter gene. Recombination of an attB and attP sites create an attR and attL site (A). attP landing sites at 22A3, 37C6, 59D3, 65B2, 68D2, and 82A1 were generated by integrating the attachment site with a yellow gene marker We generated the cloning vector pLoxPeGFPattB (pLeaB) which contains an additional LoxP site 5' of the *white* gene (B). We also used pseudorandom P-element mediated transposition utilizing two established vector systems to clone these reporter constructs. Both vectors, Ganesh-G1 and pHstinger, contain 5' and 3' P-element arms (P), the *white* gene to identify transgenics, and the same hsp70-eGFP cassette present in peaB and pLeaB (C and D). In Ganesh-G1 enhancer sequences are inserted into the vector via Gateway recombination resulting in two flanking attB sequences (C). In pHstinger the enhancer and GFP reporter are flanked by gypsy insulator sequences (D).

combination created the cloning vector, peGFPattB (peaB) that contains the transgenic selector gene, *white*, a single LoxP site, a MCS for enhancer ligation, eGFP-NLS, and attB site for genomic integration (Figure 7.1A). We also generated a second vector, pLoxPeGFPattB (pLeaB), which contains an additional LoxP site upstream of the *white* gene (Figure 7.1B).

Using these reporter vectors we began to integrate *spa*(-846bp) into twelve different published landing sites (Bischof et al., 2007; Venken et al., 2006). First, we found that five lines had transformation rates substantially too low for use in large scale enhancer mutation analysis (Table 1, no transformants). We were very surprised to find that *spa*(-846bp) did not drive any GFP expression in four of the remaining seven lines tested (Figure 7.2 I, J, M, N, Table 1, No GFP Expression). When integrating *spa*(-846bp) using P-element transposition (Ganshe-G1) we only observed one out of eight lines which did not have wildtype levels of GFP expression (Swanson et al., 2010). In order to further investigate this finding, we cloned *spa*(-846bp) into the pHstinger reporter vector. pHstinger contains the same reporter and selection components as Ganesh-G1, peaB, and pLeaB; however there are also gypsy insulator sequences 5' of the enhancer and 3' of GFP (Figure 7.1 D) and is integrated randomly using P-element transposition. Using this reporter method, we saw that only one line out of nine lacked proper levels of GFP expression (Figure 7.2 E-H and data not shown). Compared to 11.7% of randomly integrated *spa* reporter constructs, the 57% of Φ3C1 integrated *spa* reporter constructs which give no

Figure 7.2

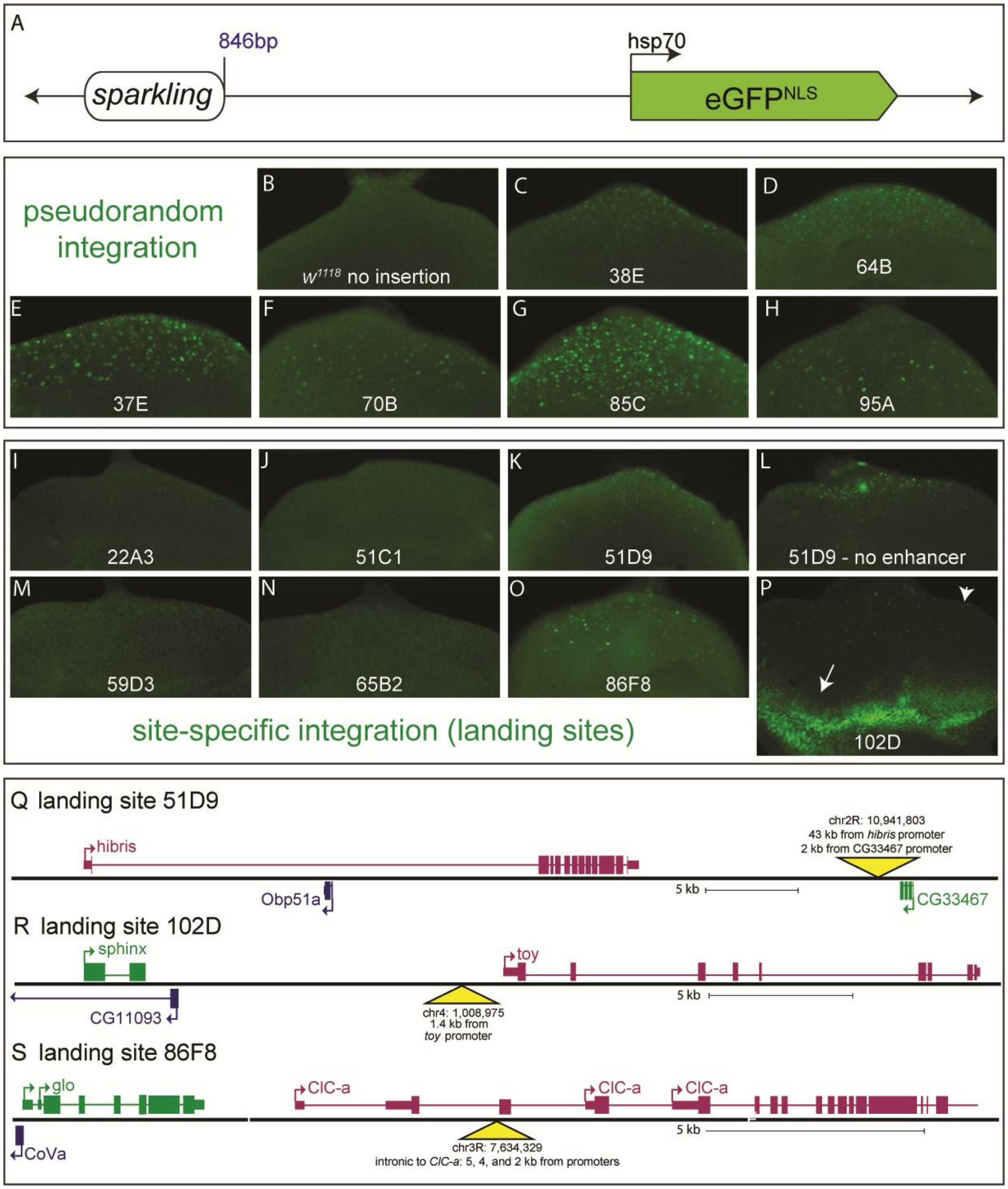


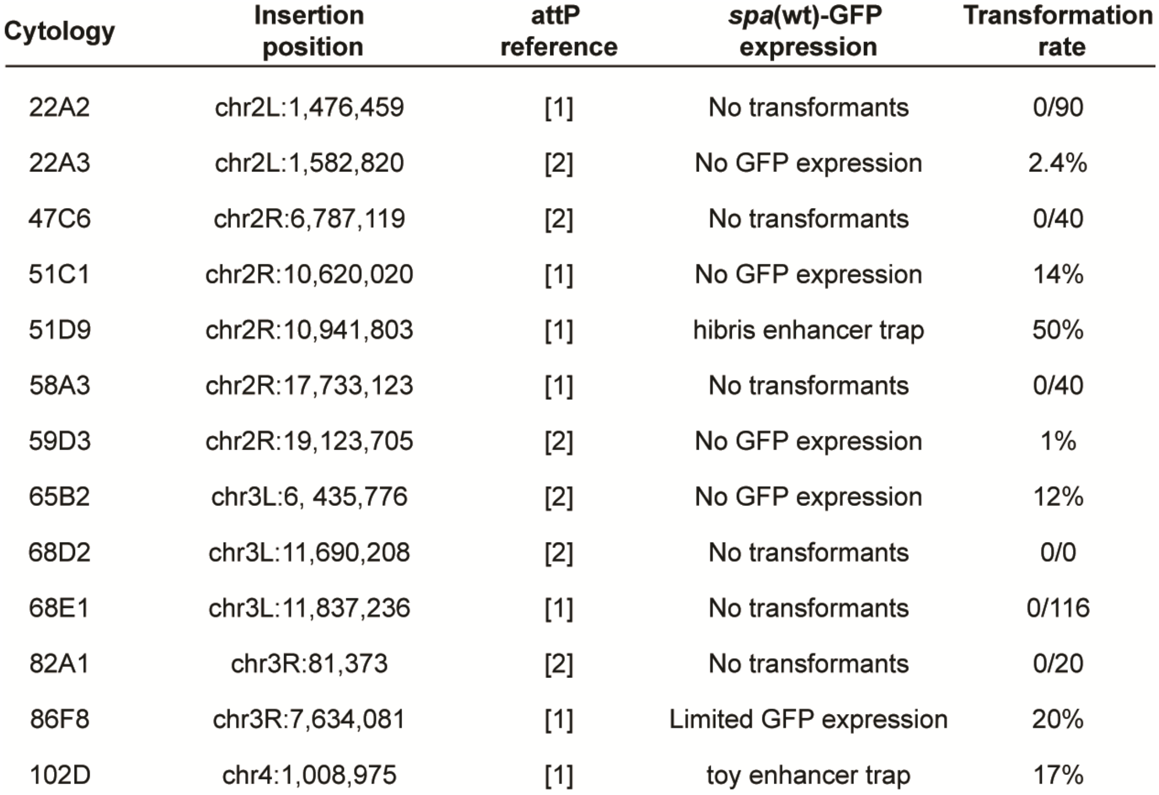
Figure 7.2 The *sparkling* enhancer behaves differently with P-element vs. site-specific integration. We generated reporter constructs containing the *dPax2 sparkling* enhancer *(spa)* placed 846bp from the GFP transcription start site (A). We established the wildtype levels of GFP expression driven by *spa* using pseudorandom integration (C-H), C-D using the Ganech-G1 cloning vector, and E-H using the pHstinger cloning vector. Integration of our reporter constructs into the genome by site-specific integration at attP landing sites posed several problems. For instance, in

the 22AC, 51C1, 59D3, and 65B2 loci *spa* does not drive GFP activity (I, J, M, N). We do see expression when *spa* is integrated at 51D9 (K). However, this expression is likely due to enhancer trapping as we also see GFP expression when an enhancerless, or empty peaB is integrated at this location (L). We also see clear enhancer trap activity and *spa* regulated GFP at the 102D locus (P). Here we see enhancer trap stimulated expression in the anterior of the eye disc (P arrow) and *spa* activity in the posterior compartment of the disc (P arrowhead). Finally, we see some limited *spa* activity when the reporter construct is integrated at 86F8 (O). Q–S depicts the genomic context of each integration locus in which *spa* drives any GFP expression. The 51D9 and 102D landing sites, in which enhancer trapping occurs, lie near gene sequences (R and S). Enhancer trapping in 51D9 likely occurs due to nearby *hibris* enhancer (Q), while enhancer trapping at 102D is likely the result of *twin of eyeless* (*toy*) regulatory sequences (R). 86F8 is the only intronic insertion in our study. It lies in the second intron of *Chlorine channel a* (*ClC-a)* (S).

GFP expression is disproportionately high. Furthermore, the GFP expression driven by *spa*(-846bp) in the remaining three landing sites, 86F8, 51D9, and 102D, is different from that observed in the randomly integrated reporters (Figure 7.2 K, O, P, and Table 7.1).

As the 51D9 integration loci has been previously associated with enhancer trapping of the *hibris* gene, we were careful to analyze expression of GFP when an enhancerless vector, that is peaB containing no *cis*-regulatory sequence, was integrated at this locus as well (Boy et al., 2010). Indeed, we see GFP expression in the third instar imaginal eye discs, partially recapitulating the known *hibris* expression pattern when both the *spa* containing vector and the enhancerless vector are integrated at 51D9 (Figure 7.2 K,L) (Dworak et al., 2001b). Somewhat surprisingly, the 51D9 attP site is 43kb downstream of the *hibris* (*hbs*) (Artero et al., 2001; Dworak et al., 2001b)promoter, and is actually closer to the *Obp51a* and CG33467 promoters. Yet, the ectopic expression patterns seen by us and others are clearly reminiscent of that of *hbs* (Artero et al., 2001; Dworak et al., 2001b). As the reporter gene activity we saw is not

Table 7.1



Cytology	Insertion position	attP reference	*spa*(wt)-GFP expression	Transformation rate
22A2	chr2L:1,476,459	[1]	No transformants	0/90
22A3	chr2L:1,582,820	[2]	No GFP expression	2.4%
47C6	chr2R:6,787,119	[2]	No transformants	0/40
51C1	chr2R:10,620,020	[1]	No GFP expression	14%
51D9	chr2R:10,941,803	[1]	hibris enhancer trap	50%
58A3	chr2R:17,733,123	[1]	No transformants	0/40
59D3	chr2R:19,123,705	[2]	No GFP expression	1%
65B2	chr3L:6, 435,776	[2]	No GFP expression	12%
68D2	chr3L:11,690,208	[2]	No transformants	0/0
68E1	chr3L:11,837,236	[1]	No transformants	0/116
82A1	chr3R:81,373	[2]	No transformants	0/20
86F8	chr3R:7,634,081	[1]	Limited GFP expression	20%
102D	chr4:1,008,975	[1]	toy enhancer trap	17%

Table 7.1 Summary of Φ3C1 mediated integration into landing sites. We found that 5 of the 12 lines we tested had extraordinarily low transformation efficiency (no transformants). In most of these lines we observed low survival after injection (surviving embryo number shown in transformation rate column. An additional 4 insertion sites were not permissive for *spa* activity (no GFP expression). *Hibris* enhancer trapping occurs in 51D9, and *twin of eyeless* (*toy*) enhancer trapping occurs in 102D. Only 86F8 shows limited expression of *spa* mediated GFP activity. [1] (Bischof et al., 2007) [2] (Venken et al., 2006).

under the control of an experiment enhancer, the GFP expression we see must be driven by regulatory information from the surrounding genomic region of the insertion locus, likely a nearby *hbs* enhancer. In fact, the entire expression pattern we see from *spa*(-846bp]) in 51D9 could be explained by the GFP driven by these additional local regulatory regions and not by *sparkling* at all (Figure 7.2K).

We also observed clear enhancer trapping in a second landing site, 102D (Figure 7.2P, Table 7.1). When *spa*(-846bp)-GFP is integrated into this landing site, GFP is active in the anterior compartment of the eye disc (Figure 7.2P, arrow). In these anterior cells, not only is *dPax2* not expressed and therefore *spa* should not be active, but the cone cell fate has not yet been specified. This anterior compartment expression is reminiscent of *twin of eyeless* (*toy*) expression (Punzo et al., 2004). As the 102D attP site is inserted 1.4kb upstream of the *toy* promoter, we suspect that this inappropriate expression is indeed due to enhancer trapping of the *toy* regulatory sequences (Figure 7.2R). Somewhat unfortunately, the best *spa* driven GFP expression is also seen in this insertion locus. The GFP expression in the posterior of the eye disc closely resembles that seen previously in randomly integrated *spa* reporters (Figure 7.2P, arrowhead). However, the overwhelming expression in the anterior compartment makes any enhancer analysis, especially quantitative analysis, difficult.

After identifying five landing site lines that were difficult to inject, four lines that had no GFP reporter activity, and two lines where enhancer trapping occurred, we were left with a single line, 86F8, in which the *spa* was capable of driving GFP expression that was not due to enhancer trap activity (Figure 7.2 O). The 86F8 attP site is integrated into the second intro of the *Chlorine channel a* (*CIC-a*) gene, where it is 5kb downstream of the primary promoter and 4kb and 2kb upstream of two secondary promoters (Figure 7.2S). While *spa* is capable of driving GFP expression in cone cells of the developing eye disc from this landing site, this expression is disperse and incomplete compared to the wildtype expression seen in randomly integrated *spa*(-846bp) reporter constructs (Figure 7.2 C-H). Taken together, only 1 of 12 (8.3%) attP landing sites gave any non-enhancer trap regulated GFP expression, and even this line, 86F8 is not suitable for studying *sparkling* activity.

7.3b *Promoter trapping is prevalent in the 86F8 landing site*

When we analyzed 86F8 transgenics containing reporter constructs, we observed that integration of different enhancers into this locus resulted in trangenic flies with differing shades of adult eye pigmentation (Figure 7.3A-E). Additionally, we also noted differential expression of 3xP3 RFP, which makes the integration of an attP site at 86F8 in the third instar larval optic nerve (Figure 1A and data not shown). Our peaB cloning vector contains the coding sequence for the *white* gene, which is critical to the generation of the red pigment that colors the adult eye (Figure 7.1A). We use this gene to mark positive integration events, and it is identical in every reporter construct. Using an empty, or

Figure 7.3

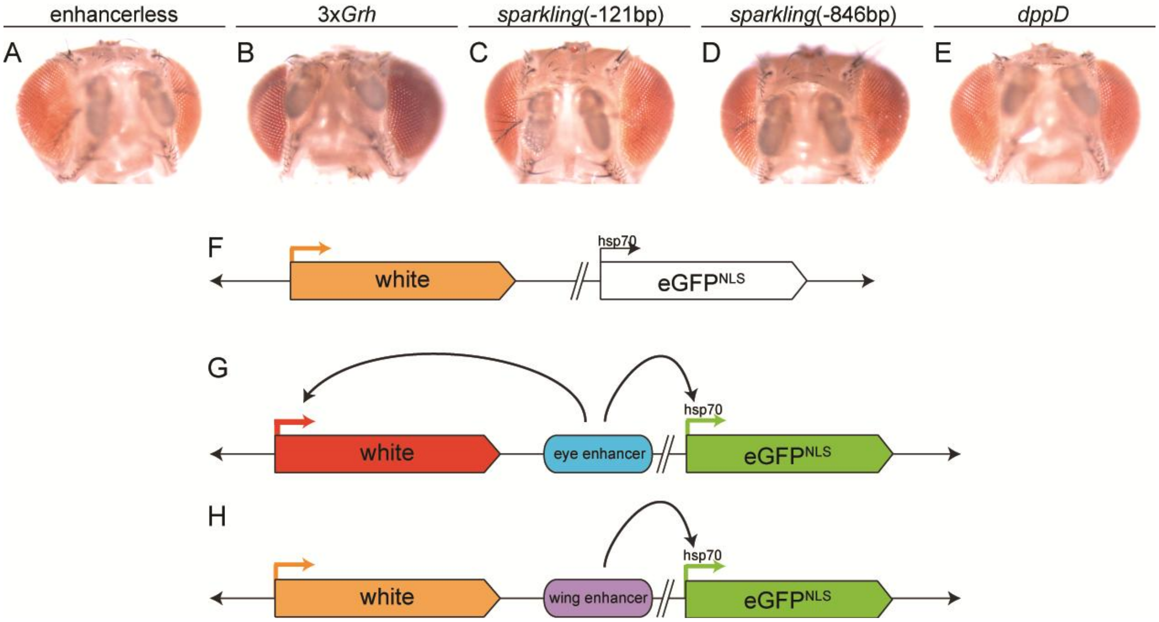


Figure 7.3 Promoter trapping occurs in the 86F8 landing site. We noticed that the level of *white* expression varied depending on which enhancer was integrated into the landing site. A synthetic enhancer consisting of 3 *Grainyhead* binding sites drives *white* expression at significantly higher levels than the expression seen when an empty, enhancerless, peaB is integrated (A and B). Promoter proximal, *sparkling*(-121 bp), and promoter distal, *spa*(-846 bp) *spa* enhancers also drive increased *white* expression, with the latter driving slightly greater levels (C and D). Meanwhile, the *dppD* wing disc enhancer does not appear to influence *white* expression (E). These observations lead to a model in which enhancerless peaB demonstrates the baseline *white* expression expected in this locus (F), an enhancer active in the eye can drive both GFP and *white* expression such that *white* levels are greater than baseline (G), and enhancers active in other tissues, such as the wing, only regulate GFP expression and do not alter *white* expression (H).

enhancerless, peaB as the baseline for expected *white* expression in the 86F8 locus, we compare these levels to the *white* expression observed when two versions of the *spa* enhancer, which are both active in the eye imaginal disc, a strong activating enhancer consisting of 3 copies of the *Grainyhead* (*Grh*) binding site that is active in all imaginal discs, and the *dppD* enhancer, which is active in the wing imaginal discs (Figure 7.3 A-E) (Uv et al., 1997). Compared to enhancerless peaB transgenic flies containing *3xGrh* synthetic enhancer, driving GFP have significantly darker adult eyes (Figure 7.3 B). Transgenics containing *spa* are also darker than integration of the enhancerless reporter (Figure 7.3C,D). Notably, when *spa* is placed more distally from the TSS, levels of *white* expression are higher (compare *spa*[-846bp] to *spa*[-121bp]). Integration of an enhancer that is not active in the eye imaginal disc, the *dppD* wing enhancer, does not result in transgenic flies with significantly darker adult eyes compared to the enhancerless peaB.

When expression of the *white* gene varies in randomly integrated p-element reporters, we understandably conclude that the regulatory information near the integration site influences the level of *white* expression, a phenomenon referred to as position effects. However, the enhancers integrated into 86F8 are all integrated into the same locus, so the position effects should be the same regardless of the enhancer used. Therefore, we would expect that the adult eye color should be the same in every transgenic we make in this landing site and indeed this is true in other landing sites such as 51D9. As the integration site is identical for each of these trangenics, as well as the vector backbone the enhancers were cloned into, the variation in *white* expression can only be due to the enhancers themselves. We conclude then that enhancers integrated in the 86F8 landing site are capable, in this genominc context, of “promoter trapping” or inappropriately interacting with and regulating the *white* and 3xP3 promoters. As the 86F8 integration site lies in the intron of a CIC-a isoform, it is possible that integration of some reporter constructs could alter expression of CIC-a proteins as well.

Based on our observations we propose of model of enhancer trapping in the 86F8 insertion locus, where insertion of an enhancerless reporter construct gives us a baseline for expected *white* expression (Figure 7.3 F). Enhancers that are not active in the eye regulate GFP in the correct tissues, but do not affect *white* expression (Figure 7.3 H). Enhancers that are active in the *Drosophila* eye regulate both GFP and *white* expression (Figure 7.3 G). In the case of *sparkling* specifically, levels of white and GFP are anti-correlated. When *spa* is placed close, to -121bp from the TSS, GFP expression is high but *white* expression is decreased compared to 3x*Grh* and *spa*(-846bp) regulated *white* expression. Distally located *spa*, 846bp from the TSS, has decreased GFP expression and increased *white* expression. In every insertion, the 3xP3 and *white* promoters are each further away from the enhancer than the hsp70 promoter driving GFP expression. The *spa* and *Grh* enhancers are still able to regulate *white* expression, suggesting these promoters are able to compete with hsp70 for interaction with these enhancers.

7.3c Enhancer trapping occurs in the Drosophila embryo at the 51D9 landing site

Due to previous reports of enhancer trapping in the 51D9 locus, we were careful to examine GFP expression driven by the hsp70 promoter in an enhancerless, or empty, peaB insertion. We have already demonstrated that enhancer trapping does indeed occur in the *Drosophila* eye imaginal disc (Figure 7.2 L). Therefore, we extended this analysis to other tissues of interest, including early embryonic development. Here, we observed significant GFP expression only when the enhancerless peaB is integrated at the 51D9 landing site, but not in the 86F8 landing site (Figure 7.4). Accordingly, this expression corresponds to known aspects of the *hibris* expression pattern (Artero et al., 2001; Dworak et al., 2001a). At embryonic stage 10 we observe expression of GFP in precursors of the visceral musculature (Figure 7.4 K, solid arrow). We also observed GFP expression in the ventral midline, another location of *hbs* activity at stage 10, along with GFP expression in non-*hibris* expressing cells (Figure 7.4 P,Q, solid arrowhead). Additional expression is seen at stage 12 in the somatic mesoderm (Figure 7.4 L, open arrow). Visceral musculature and somatic mesoderm expression continues in stages 13 and 14 (Figure 7.4 M,N). By stage 14 we also observe GFP expression in the phargeal muscles and mesoectoderm (Figure 7.4 O, curved arrow and open arrowhead). w^{1118} embryos, which have similar genetic background to our transgenics, do not have any endogenous fluorescence in these tissues (Figure 7.4A-E). Similarly, integration of the

Figure 7.4

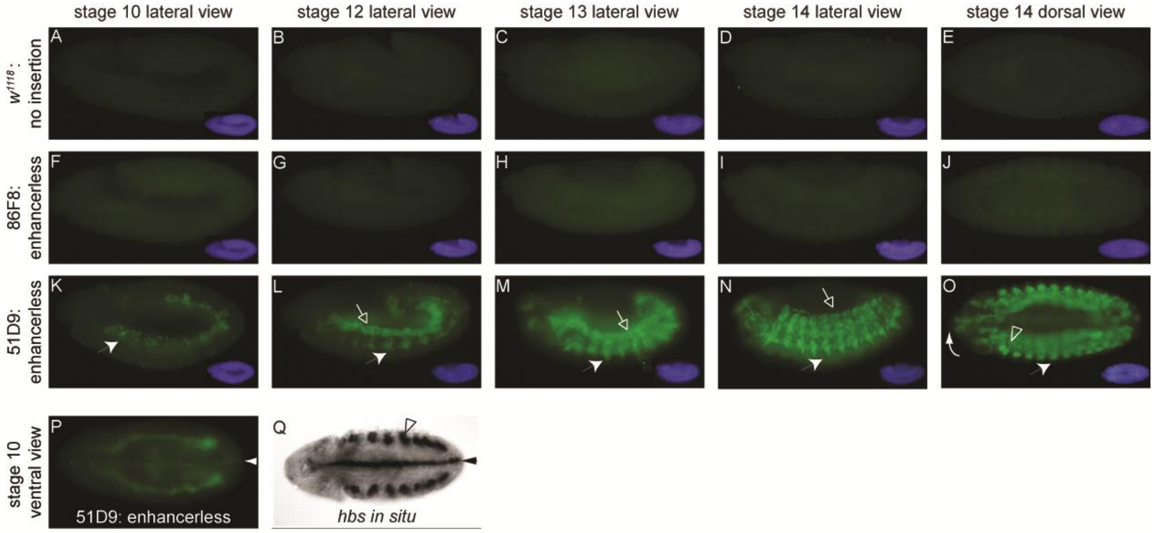


Figure 7.4 Embryonic enhancer trapping of hibris occurs in the 51D9 landing site. We analyzed GFP expression in early embryos in *w*^{1118} wildtype *Drosophila*, empty, enhancerless, peaB integrated at 86F8, and the same enhancerless peaB integrated at 51D9. We saw that *w*^{1118} embryos posess very little endogenous autoflourescence (A - E), while GFP expression is only slightly, and ubiquitously, increased over background levels (F – J) when peaB is integrated at 86F8. However, we do see GFP expression when the enhancerless vector is integrated into the 51D9 locus that is reminiscent of *hibris* expression (Q) (Dworak et al., 2001b). At embryonic stage 10 we observed GFP expression in the visceral musculature (K, solid arrow) and ventral midline (P, solid arrowhead). We see additional expression in the somatic mesoderm at stage 12 (L, open arrow), which continues in stages 13 and 14 (M and N). By stage 14 we also observe GFP expression in the pharangeal muscles (O, curved arrow) and mesoetoderm (O open arrowhead).

enhancerless peaB into the 86F8 landing site has a slight, seemingly ubiquitous increase in GFP fluorescence over w^{1118} background levels (Figure 7.4 F-J). As the GFP expression in 51D9 embryos is significantly greater than that seen in w^{1118} and 86F8 embryos, and demonstrates clear cell type specificity, we conclude that this activity is due to enhancer trapping in the 51D9 locus, likely of *hibris* enhancers. These observations render the 51D9 insertion site unsuitable for studying *cis*-regulatory sequences in the embryo and in eye imaginal disc.

7.3d Latent enhancer trapping occurs in the wing imaginal disc at the 51D9 landing site

We have already observed enhancer trapping activity in the eye and embryo at the 51D9 landing site when an enhancerless peaB is integrated, which is not seen at the 86F8 landing site (Figure 7.5A,C, Figure 4). We also analyzed potential enhancer trapping in the third instar wing imaginal disc. In the wing disc, *hbs* is expressed in the notum (Figure 7.5J, asterisks), in a stripe on either side of the wing margin (Figure 7.5 J, arrowhead), two stripes that correspond to the future L3 and L4 wing veins (Figure 7.5J, upwards arrow), as well as near the hinge region which corresponds to the future wing veins L0 and L1 (Figure 7.5 J, horizontal arrow (Dworak et al., 2001b). Trangenic 51D9 flies containing an enhancerless reporter do indeed show ectopic GFP expression, but only in the hinge (L0 and L1 wing veins) and at low levels in the notum (Figure 7.5H). Our synthetic reporter construct containing three binding sites for the transcription factor *Grainyhead*, which is expressed uniformly in all imaginal discs (Uv et al., 1997), drives GFP reporter gene expression ubiquitously in the eye and wing imaginal discs when integrated at the 86F8 landing site (Figure 7.5 B, D). Similarly, 3x*Grh* drives high levels of GFP expression across the eye disc in 51D9 as well, with *hbs* expressing cells containing even higher levels of reporter activity (Figure 7.5 D, E). Remarkably, 51D9 flies containing the *3xGrh* synthetic enhancer do not posses GFP expression in the pattern of *Grainyhead* as expected, but instead reporter expression recapitulates the *hbs* wing disc expression pattern (Figure 7.5 I, J). As the 3x*Grh* reporter construct inserted into 86F8 and 51D9 landing sites are identical, this dramatic difference in GFP expression pattern must be due to the influence of the local genomic environment, likely *hbs* regulatory sequences. Furthermore, this data suggests that any reporter construct results in 51D9 from *cis*-regulatory sequences containing activating information must be suspect for *hbs* enhancer contribution to the observed pattern.

Together our observations regarding enhancer trapping leads to the following model of gene activity in the 51D9 locus. In early embryonic tissue, one or more embryo enhancers regulate both *hibris* and GFP reporter expression in nearly the same pattern, indicating enhancer trapping at this landing site (Figure 7.6A). Meanwhile, in third instar larva, one or more disc enhancers are able to drive *hibris* expression as well as restricted aspects of that pattern from the

Figure 7.5

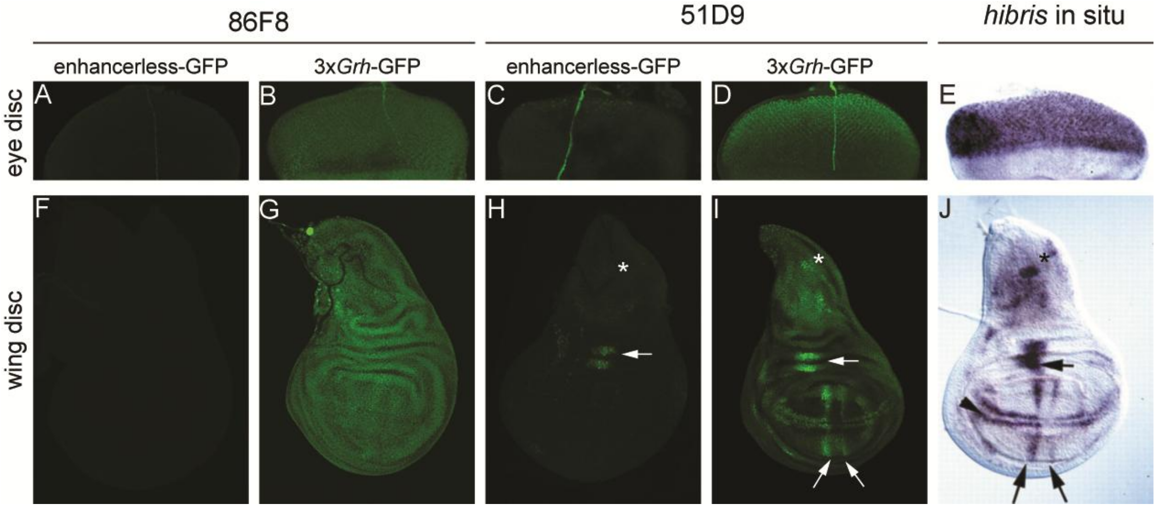


Figure 7.5 Latent enhancer trapping of *hibris* occurs in the third instar larval eye and wing imaginal discs at 51D9. *Hibris* is expressed in the eye and wing imaginal discs (E and J) (Dworak et al., 2001b). Under normal conditions, *hibris* is expressed in the posterior cells of the eye imaginal disc (E). *Hibris* is also expressed in the notum (J*), on either side of the wing margin (J, arrowhead), in two stripes that correspond to the future L3 and L4 wing veins (J, upwards arrows), and in the near hinge in the presumptive L0 and L1 wing veins (J, horizontal arrows). When an enhancerless vector is integrated at the 86F8 landing site we see no GFP expression in the eye or wing discs (A and F). When the same enhancerless vector is integrated at 51D9, GFP is expressed correlating to known aspects of the *hibris* expression pattern, including in posterior eye disc cells (C) and in the notum (H*) and hinge (H, horizontal arrow) in the wing disc. A synthetic enhancer containing three transcription factor binding sites for the strong activator, *Grainyhead* (*Grh*) drives ubiquitous GFP expression in imaginal discs from the 86F8 landing site (B and G). However, from 51D9 3x*Grh* drives both *Grh* pattern of GFP expression and increased levels of *hibris*-like expression compared to the enhancerless vector in the eye disc (D). In the wing disc 3x*Grh* –GFP expression is completely co-opted to form the *hibris* pattern of expression (I).

hsp70 promoter upstream of GFP (Figure 7.6B). However, when these disc enhancers act in conjunction with a strong activating enhancer upstream of GFP, they are able to drive a complete *hbs* pattern, revealing the “latent enhancer trapping” potential of the 51D9 landing site (Figure 7.6C).

7.4 Discussion

Drosophila transgenics generated by pseudorandom P-element transposition pose many potential difficulties and caveats for investigating *cis*-regulatory elements. For example, position effects due to regulatory information in the locus where the construct integrated can significantly affect levels of reporter gene expression, either by increasing or silencing the regulatory sequences or altering its pattern. Additionally, reporter genes can be “trapped” by local enhancers leading to expression of the reporter gene in the pattern regulated by those enhancers rather than the *cis*-regulatory sequence of interest. Also, reporter constructs frequently “home”, or integrate into a locus nearby that of the endogenous enhancer location. This phenomenon can result in altered reporter gene activity, especially when studying enhancer mutations. As such, it is the common rule in the field is that three to five independent integration lines must show the same expression levels and pattern in order for an average expression pattern to be established. This increases the time scale and workload of any project utilizing P-element transposition. Furthermore, the

Figure 7.6

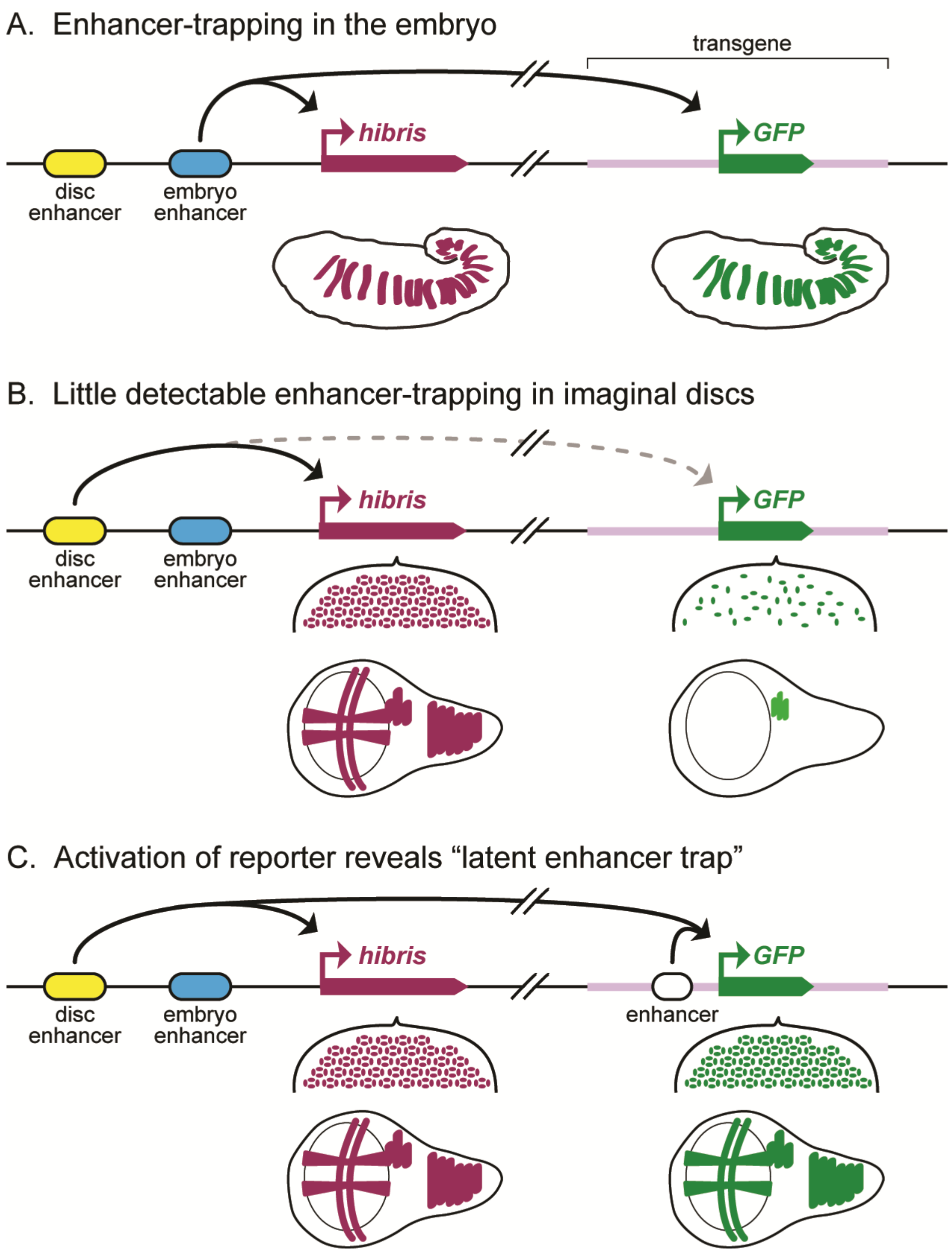

A. Enhancer-trapping in the embryo
transgene
hibris
GFP
disc
enhancer
embryo
enhancer
B. Little detectable enhancer-trapping in imaginal discs
hibris
GFP
disc
enhancer
embryo
enhancer
C. Activation of reporter reveals "latent enhancer trap"
hibris
GFP
disc
enhancer
embryo
enhancer
enhancer

Figure 7.6 Model of enhancer trapping in the 51D9 landing site. We have found that integration of an enhancerless reporter construct into 51D9 results in GFP expression in a *hibris* expression pattern due to enhancer trapping of one or more embryonic enhancers (A). The *hibris* imaginal disc enhancers are only partially able to recapitulate its expression pattern when driving GFP expression from the hsp70 promoter (B). However, a strong enhancer upstream of GFP is able to work in conjunction to latently trap the disc enhancers and drive GFP expression in a *hibris* pattern (C).

variation between lines due to position effects makes determining small changes in enhancer activity difficult and quantitative analysis tedious.

Due to these and other limitations, the use of site specific-integration, or landing sites, in which the position effects remain constant from one reporter to the next, in principle, dramatically simplifies the process of generating and analyzing reporter flies. However, we have found that the same difficulties and caveats of random methods of integration still exist when utilizing landing sites. In fact, if we take a step back and look at integration events at the genomic level, these limitations are not only predictable, but should come as no surprise.

7.4a Analysis of a weak enhancer in multiple landing sites

We have previously analyzed the *dPax2* cone cell enhancer *sparkling* using pseudorandom integration and established a consistent pattern for the reporter expression driven by this enhancer that well exceeds the expected 3-5 lines to determine this pattern (15 of 17 lines have similar levels and patterns of expression). Using P-element mediated integration, we have never struggled to generate transgenics containing a *spa*(-846bp)-GFP reporter construct with high transformation efficiency. Conversely, we were unable to generate transgenics from 5 of the 12 of attP landing sites we used to make site-specific transgenics (Table 7.1). Notably, we are not the first to observe low transformation rates amongst the published attP lines (Laverty et al., 2011; Pfeiffer et al., 2010). With the exception of 68E1, we suspect the our difficulties generating transgenics in these landing sites is because these fly stocks appear less robust than the other attP landing site lines we utilized. Additionally, these stock lay fewer young embryos and the embryos are candidates for injection had decrease survival rates (Table 7.1). This suggests that the integration of an attP site into the 22A2, 47C6, 58A3, and 82A1 landing sites disrupt gene expression in such a way as to make the flies unhealthy. Of these only two lie near genes we could predict would affect quality of life. The attP site in 47C6 could affect the expression of Rbp5, a component of the RNA polymerase components (Aoyagi and Wassarman, 2000). Meanwhile, the attP site at 82A1 may influence the expression of *TweedleV* a protein component of the *Drosophila* cuticle (Cornman, 2010; Cornman et al., 2009).

Next we found that *spa*(-846bp) is unable to drive GFP expression from 4 of the 12 landing sites we analyzed, 22A3, 51C1, 59D3, and 65B2. This lack of GFP expression in 57% of the lines is in stark contrast to 12% of randomly integrated constructs which are unable to drive GFP expression. We postulate that that this affect may be due to the difference in where P-element transposition and piggyback transposition (used to make the attP insertion lines) typically occurs in the genome. It is well documented that P-element transposition tends to occur in highly transcribed regions such as immediately upstream of promoters and in introns, while piggyback transposase does not have a similar preference for regulator sequences (Bellen et al., 2004). Furthermore, the attP landing sites that have been propagated for wide spread research uses were specifically selected as they were mapped to intergenic DNA regions (Bischof et al., 2007). We know that the 3D space of the nucleus is organized into transcriptionaly active and inactive regions (Wilson and Berk, 2010). It is possible that simply due to the manner in which the attP sequences were integrated and selected; they are preferentially inactive regions of the genome. The four landing sites in which we observed no GFP expression for example lie 2 to 19 kb outside of gene regions. Interestingly, we were unable to find a single published study that utilizes the 51C1, 59D3, 65B2 landing sites, suggesting the lack of reporter activity in these lines is not limited to our experiences. At least three studies have successfully used the 22A3 landing site (Duncan et al., 2010; Housden et al., 2012; Schwank and Basler, 2010). Notable, this insertion lies the closest of this set to a gene promoter (2kb). It is plausible that the weak *sparkling* enhancer cannot regulate transcription in this location while other, stronger, reporters can. Only one of the 12 landing sites at which we integrated *spa*(-846) lies in the intron of a gene. The 86F8 landing site is in the 2nd intron of the CIC-a gene, with places it downstream of the primary promoter, and upstream of two secondary promoters. This is the only landing site in which we observed clear *spa* mediated gene expression (Figure 7.2). These data suggest that despite all the limitations of P-element transposition into the *Drosophila* genome, this method fortuitously integrates reporter genes into permissive regions with higher frequency than Φ3C1 mediated integration at landing sites.

7.4b Disruption of local gene expression occurs in landing sites

One of the prevalent limitations of P-element transposition is that the genomic insertions can disrupt the expression of genes in the integration location. This can result in sick, or homozygous lethal flies, making analysis of multiple insertion sites tedious. Therefore, attP landing sites provide a distinct advantage over random insertion as landing sites that result in homozygous lethality have been discarded. (Bischof et al., 2007). However, we have found evidence that integration events in landing sites can disrupt local gene expression. For example, we have already noted that several of the attP landing site fly stocks are less healthy than wildtype flies. We hypothesized that this was due to the effect of integrating a large RFP-attP cassette into these loci. It is plausible that further integration of the reporter gene into these landing sites can disrupt gene expression such that integration progeny are heterozygous lethal, resulting in the low transformation efficiency we observed.

We also observed disrupted gene expression in the 86F8 locus. Here we saw that expression of the *white* gene differed significantly in adult eyes depending on which enhancer was integrated into the landing site (Figure 7.3), suggesting the enhancers where somehow interacting and regulating the other local promoters in this integration locus. Significantly, the slight change in sequence between the *spa*(-846bp) and *spa*(-121bp) constructs results in differential regulation of the *white* gene. It is plausible then that the interaction between an enhancer and these local promoters could affect the reporter gene readout, and this affect could be altered by changes within the enhancer. In this scenario reporter gene outcome would misrepresent the actually enhancer activity

7.4c Position affects occur in landing sites

We also saw numerous instances of position affects, or enhancer trapping, when using landing sites. For example, reporter genes in 102D trap *toy* expression while those in 51D9 trap *hibris* expression (Figure 7.2). Interestingly, enhancer trapping in 51D9 is not a simple story. Integration of an enhancerless vector into this locus results in GFP expression in both the eye imaginal disc and *Drosophila* embryo (Figure 7.2 and Figure7.4). Given its expression pattern we could have easily incorrectly identified this expression in the eye disc as *spa*(-846bp) reporter activity if we had not also examined an enhancerless vector into this landing site. Conversely, insertion of the enhancerless peaB in to the 51D9 landing sites drives only a small part of the *hibris* expression pattern in the wing imaginal disc. However, upon addition of a simple activator binding sequence (*3xGrh*) we observed GFP expression that recapitulates the *hibris* expression pattern rather than the expected *Grainyhead* pattern (Figure 7.5). This "latent enhancer" trapping reveals the importance of understanding any integration locus and the potential for neighboring regulatory sequences to affect reporter construct results. Contribution from the *hibris* enhancers to a reporter constructs expression pattern could easily lead to a mis-interpretation of an enhancer's activity. Furthermore, simply looking at the genomic sequence around an integration site is not sufficient to predict potential difficulties. The 51D9 integration locus is 43kb from the *hibris* promoter and actually lies only 2kb from the CG33467 promoter, yet reporter genes integrated into this locus clearly demonstrate enhancer trapping for *hbs* expression (Artero et al., 2001; Dworak et al., 2001b).

It is worth noting that 86F8 and 51D9 are the two most widely used landing sites, excluding attP2 which is used for the RNAi insertion project (Ni et al., 2008). Of published studies using landing sites, 24% utilize 86F8 and and 41% utilize 51D9 (Boy et al., 2010; Cande et al., 2009; Frankel et al., 2010; Haley et al., 2010; Housden et al., 2012; Joshi et al., 2010; Perry et al., 2011; Potter and Luo, 2010; Rebeiz et al., 2011; Sayal et al., 2011; Schwank and Basler, 2010). Every one of these studies using 51D9 analyzes tissues in which we have observed enhancer trapping (embryo, eye disc, and wing disc). These two landing sites are preferred as they are known throughout the field to have high transformation rates and strong reporter gene expression in multiple tissues. We would argue that the beneficial aspect of these landing sites is because, like most P-element integration events, these attP sites are near active regulatory sequences. However, based on our observations, the very nature of these sites that makes them useful for studying reporter constructs can also severely confound analysis of gene expression.

7.4d Perspectives on integration methods

The results of this study demonstrate that the many of the caveats and limitations that exist for random methods integration still exist when using site-specific integration. Furthermore, these caveats and quirks are not limited to those identified here but likely exist for every integration locus as we are not the first, nor will we be the last, to observe differential reporter gene expression based on attP landing site location (Markstein et al., 2008; Pfeiffer et al., 2010). We would suggest then that landing site integration be held to the same standards as have been long held for research performed using random methods of integration. New patterns of expression should be established using at least three independent insertion locations. As attP landing sites are traditionally thought to be free from position effects, multi-line analysis is rarely performed. Additionally, this standard need not be limited to *Drosophila* research, but expanded to any model organism study that utilizes genomic integration, including mouse. The potential for position affects exists regardless of organism. For example, the ROSA26 locus in mouse has long been thought to drive ubiquitous expression in all tissues; however, more recent expression and functional assays suggest that many transgenes are not expressed in high enough levels from this locus in every tissue (Yu and McMahon, 2006).

In order to obtain clear reliable picture of gene expression it is important that any position affects from an integration location are understood and accounted for. Site-specific integration is an incredibly powerful tool to eliminate time and variation that makes quantitative analysis difficult. However, we cannot dismiss random integration as useful tool for analyzing expression patterns in its own right. Ultimately, it is most important in any scientific study to choose the method of analysis most appropriate to the questions that are being asked.

7.5 Experimental methods

7.5a Vector and reporter gene construction and transgenesis

peGFPattB (peaB) was first constructed by digesting the pUASattB vector (GenBank EF362409) (Bischof et al., 2007) with BamHI to remove the UAS-MCS-SV40 cassette which was replaced with annealed oligos containing BglII, HindIII, SphI, XhoI, XbaI, EcoNI, and a second BglII restriction enzyme sites. The BglII sites were lost upon ligation to the BamHI overhangs in the vector. Oligo sequences are listed below:

Top: 5’ gatctaagcttgctagcatgcatctcgagattctagacctacgtaagga 3’

Bottom: 5’ gatctccttacgtaggtctagaatctcgagatgcatgctagcaagctta 3’

Subsequently, the hsp70 promoter and eGFP-NLS sequence was digested from pHstinger (Genbank AF242365) (Barolo, 2000) using SphI and SpeI and ligated to the above plasmid after digestion with SphI and XbaI to generate peaB cloning vector. pLeaB is an additional vector consisting of the peaB backbone and an additional LoxP site 5’ of the *white* gene. The 5’ LoxP site was added to the peaB vector using annealed oligos containing the LoxP sequence and NaeI overhangs. Oligo sequences are listed below:

Top: 5’ggccagctgacgcgtataacttcgtataatgtatgctatacgaagttatacgcgtccc 3’

Bottom: 5’gggacgcgtataacttcgtatagcatacattatacgaagttatacgcgtcagctggcc 3’

Sparkling enhancer sequence with either an 846bp or 121bp spacer was generated by sewing PCR and ligated into peaB, pLeaB, or pHstinger multiple cloning sites following EcoRI and BamHI digestion. The Ganesh-G1 integration vector (GenBank EF420135) containing *spa* was generated as described previously (Swanson et al., 2010; Swanson et al., 2008). The *3xGrainyhead* synthetic enhancer was generated by assembly PCR and ligated to peaB following digest with EcoRI and BamHI. Similarly, the *dppD* wing disc enhancer was amplified from genomic DNA by PCR and cloned to peaB via EcoRI and BamHI digestion.

Reporter vectors containing attB sites were integrated into the *Drosophila* genome using vasa-intDm (Φ3C1) as previously described using the appropriate integration landing site fly stock obtained from the Bloomington Stock Center (Bischof et al., 2007; Venken et al., 2006). Vasa-intDm (Φ3c1 integrase) and its associated 3xP3 GFP marker, integrated on the X chromosome, were subsequently removed by crossing transgenic males to the *w^{1118}* females for at least two generations. P-element transformation was performed using *w^{1118}* flies as described previously (Rubin and Spradling, 1982).

7.5b Tissue preparation and Immunohistochemistry

Eye and wing disc tissues were dissected from third instar larvae. Disc tissues were then fixed in 4% paraformaldehyde for 30 minutes at room temperature, washed three times with 1xPBS, and mounted in ProLong Gold with 4’, 6’-diamidino-2 phenylidole (DAPI) (Invitrogen).

Embryos were prepared by collecting one 6-12 hour and two 0-6 hour embryo and dechorinated with bleach for one minute and then washing with embryo wash (1xPBS+0.002% TritonX-100) and water. Embryos were fixed by shaking for 25 minutes at 600 rpm at room temperature in 5ml heptane, 4.5ml 4% paraformaldehyde in PBS, and 0.5ml 0.5M EGTA pH8.0). After removing the bottom phase (containing formaldehyde) the embryos were devitillinized in cold methanol twice. After removing heptane (top layer), and rinsing three times with methanol, the embryos were stored at -20°C in methanol. Prior to antibody staining, embryos were rehydrated and then blocked for one hour with rocking in 1xPBS containing 10% BSA and 0.01% TritonX-100. Primary antibody staining against GFP was performed overnight at 4°C with rocking using rabbit anti-GFP (Invitrogen A11122) diluted 1:100. Secondary antibody staining was performed for 2 hours at room temperature with rocking using goat anti-rabbit 488nm (Invitrogen A11008) diluted 1:2000. After staining, embryos were mounted in ProLong Gold +DAPI.

Three day old adult heads were prepared by first crossing homozygous transgenic males to w^{1118} females to generate heterozygous progeny. The first 24 hours of flies to eclose were discarded. The second 24 hours of flies to eclose were removed to a fresh vial and allowed to age for an additional 48 hours. Adult heads were detached prior to imaging using a razor blade.

7.5c Microscopy

For third instar larval eye discs and embryos, GFP fluorescence and bright field imaging was performed with an Olympus BX5I microscope and an

Olympus DP70 digital camera. Confocal images of third instar larval eye and wing discs were captured on an Olympus FluoView 500 Laser Scanning Confocal Microscope mounted on an Olympus 1X71 inverted microscope. Adult heads were imaged using a LiecaMZ12.5 dissecting microscope equipped with LeicaFireCam software.

7.6 Acknowledgments

We would like to thank Andrea Ramos for her contributions to this work. The research was performed in part with support of NIH grant GM076509.

7.7 References

Aoyagi, N., and Wassarman, D.A. (2000). Genes encoding Drosophila melanogaster RNA polymerase II general transcription factors: diversity in TFIIA and TFIID components contributes to gene-specific transcriptional regulation. J Cell Biol *150*, F45-50.

Artero, R.D., Castanon, I., and Baylies, M.K. (2001). The immunoglobulin-like protein Hibris functions as a dose-dependent regulator of myoblast fusion and is differentially controlled by Ras and Notch signaling. Development *128*, 4251-4264.

Barolo, S., Carver, L.A., Posakony, J.W. (2000). GFP and beta-galactosidase transformation vectors for promoter/enhancer analysis in Drosophila. BioTechniques *29*, 726-732.

Bateman, J.R., Lee, A.M., and Wu, C.T. (2006). Site-specific transformation of Drosophila via phiC31 integrase-mediated cassette exchange. Genetics *173*, 769-777.

Bellen, H.J., Levis, R.W., Liao, G., He, Y., Carlson, J.W., Tsang, G., Evans-Holm, M., Hiesinger, P.R., Schulze, K.L., Rubin, G.M.*, et al.* (2004). The BDGP gene disruption project: single transposon insertions associated with 40% of Drosophila genes. Genetics *167*, 761-781.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. Proceedings of the National Academy of Sciences of the United States of America *104*, 3312-3317.

Boy, A.L., Zhai, Z., Habring-Muller, A., Kussler-Schneider, Y., Kaspar, P., and Lohmann, I. (2010). Vectors for efficient and high-throughput construction of fluorescent *Drosophila* reporters using the PhiC31 site-specific integration system. Genesis *48*, 452-456.

Cande, J., Goltsev, Y., and Levine, M.S. (2009). Conservation of enhancer location in divergent insects. Proc Natl Acad Sci U S A *106*, 14414-14419.

Cornman, R.S. (2010). The distribution of GYR- and YLP-like motifs in Drosophila suggests a general role in cuticle assembly and other protein-protein interactions. PLoS One *5*.

Cornman, R.S., Chen, Y.P., Schatz, M.C., Street, C., Zhao, Y., Desany, B., Egholm, M., Hutchison, S., Pettis, J.S., Lipkin, W.I.*, et al.* (2009). Genomic analyses of the microsporidian Nosema ceranae, an emergent pathogen of honey bees. PLoS Pathog *5*, e1000466.

Duncan, D., Kiefel, P., and Duncan, I. (2010). Control of the spineless antennal enhancer: direct repression of antennal target genes by Antennapedia. Dev Biol *347*, 82-91.

Dworak, D., Loskiewicz, J., and Janik, M. (2001a). Asymptotic solutions of neutron transport equation and the limits of correct use of diffusion approximation for rocks. Appl Radiat Isot *54*, 845-848.

Dworak, H.A., Charles, M.A., Pellerano, L.B., and Sink, H. (2001b). Characterization of Drosophila hibris, a gene related to human nephrin. Development *128*, 4265-4276.

Evans, N.C., Swanson, C.I., and Barolo, S. (2012). Sparkling insights into enhancer structure, function, and evolution. Curr Top Dev Biol *98*, 97-120.

Evans, N.C., Swanson, C.I., Barolo,S (2012). sparkling insights to enhancer sparkling Insights into Enhancer Structure, Function, and Evolution. Transcriptional Switches During Development *98*.

Flores, G.V., Duan, H., Yan, H., Nagaraj, R., Fu, W., Zou, Y., Noll, M., and Banerjee, U. (2000). Combinatorial signaling in the specification of unique cell fates. Cell *103*, 75-85.

Frankel, N., Davis, G.K., Vargas, D., Wang, S., Payre, F., and Stern, D.L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature *466*, 490-493.

Fu, W., Duan, H., Frei, E., and Noll, M. (1998). *shaven* and *sparkling* are mutations in separate enhancers of the *Drosophila Pax2* homolog. Development *125*, 2943-2950.

Fu, W., and Noll, M. (1997). The *Pax2* homolog *sparkling* is required for development of cone and pigment cells in the *Drosophila* eye. Genes Dev *11*, 2066-2078.

Groth, A.C., Fish, M., Nusse, R., and Calos, M.P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics *166*, 1775-1782.

Haley, B., Foys, B., and Levine, M. (2010). Vectors and parameters that enhance the efficacy of RNAi-mediated gene disruption in transgenic Drosophila. Proc Natl Acad Sci U S A *107*, 11435-11440.

Housden, B.E., Millen, K., and Bray, S.J. (2012). Drosophila Reporter Vectors Compatible with PhiC31 Integrase Transgenesis Techniques and Their Use to Generate New Notch Reporter Fly Lines. G3 (Bethesda) *2*, 79-82.

Jagle, U., Gasser, J.A., Muller, M., and Kinzel, B. (2007). Conditional transgene expression mediated by the mouse beta-actin locus. Genesis *45*, 659-666.

Johnson, L.A., Zhao, Y., Golden, K., and Barolo, S. (2008). Reverse-engineering a transcriptional enhancer: a case study in Drosophila. Tissue Eng Part A *14*, 1549-1559.

Joshi, R., Sun, L., and Mann, R. (2010). Dissecting the functional specificities of two Hox proteins. Genes Dev *24*, 1533-1545.

Kalay, G., and Wittkopp, P.J. (2010). Nomadic enhancers: tissue-specific *cis*-regulatory elements of *yellow* have divergent genomic positions among *Drosophila* species. PLoS Genet *6*, e1001222.

Laverty, C., Li, F., Belikoff, E.J., and Scott, M.J. (2011). Abnormal dosage compensation of reporter genes driven by the Drosophila glass multiple reporter (GMR) enhancer-promoter. PLoS One *6*, e20455.

Levis, R., Hazelrigg, T., and Rubin, G.M. (1985). Effects of genomic position on the expression of transduced copies of the white gene of Drosophila. Science *229*, 558-561.

Liu, W.Y., Wang, Y., Qin, Y., Wang, Y.P., and Zhu, Z.Y. (2007). Site-directed gene integration in transgenic zebrafish mediated by cre recombinase using a combination of mutant lox sites. Mar Biotechnol (NY) *9*, 420-428.

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

Mello, C.C., Kramer, J.M., Stinchcomb, D., and Ambros, V. (1991). Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. Embo J *10*, 3959-3970.

Ni, J.Q., Markstein, M., Binari, R., Pfeiffer, B., Liu, L.P., Villalta, C., Booker, M., Perkins, L., and Perrimon, N. (2008). Vector and parameters for targeted transgenic RNA interference in Drosophila melanogaster. Nat Methods *5*, 49-51.

Parker, D.S., White, M.A., Ramos, A.I., Cohen, B.A., and Barolo, S. (2011). The cis-Regulatory Logic of Hedgehog Gradient Responses: Key Roles for Gli Binding Affinity, Competition, and Cooperativity. Sci Signal *4*, ra38.

Perry, M.W., Boettiger, A.N., and Levine, M. (2011). Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proc Natl Acad Sci U S A *108*, 13570-13575.

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

Potter, C.J., and Luo, L. (2010). Splinkerette PCR for mapping transposable elements in Drosophila. PLoS One *5*, e10168.

Punzo, C., Plaza, S., Seimiya, M., Schnupf, P., Kurata, S., Jaeger, J., and Gehring, W.J. (2004). Functional divergence between eyeless and twin of eyeless in Drosophila melanogaster. Development *131*, 3943-3953.

Rebeiz, M., Jikomes, N., Kassner, V.A., and Carroll, S.B. (2011). Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. Proc Natl Acad Sci U S A *108*, 10036-10043.

Robert, V., and Bessereau, J.L. (2007). Targeted engineering of the Caenorhabditis elegans genome following Mos1-triggered chromosomal breaks. Embo J *26*, 170-183.

Rubin, G.M., and Spradling, A.C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. Science *218*, 348-353.

Sato, Y., Kasai, T., Nakagawa, S., Tanabe, K., Watanabe, T., Kawakami, K., and Takahashi, Y. (2007). Stable integration and conditional expression of electroporated transgenes in chicken embryos. Dev Biol *305*, 616-624.

Sayal, R., Ryu, S.M., and Arnosti, D.N. (2011). Optimization of reporter gene architecture for quantitative measurements of gene expression in the Drosophila embryo. Fly (Austin) *5*, 47-52.

Schwank, G., and Basler, K. (2010). Regulation of organ growth by morphogen gradients. Cold Spring Harb Perspect Biol *2*, a001669.

Shi, Y., and Noll, M. (2009). Determination of cell fates in the R7 equivalence group of the *Drosophila* eye by the concerted regulation of D-Pax2 and TTK88. Developmental Biology *331*, 68-77.

Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A., and Kucherlapati, R.S. (1985). Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature *317*, 230-234.

Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat Genet *21*, 70-71.

Spradling, A.C., and Rubin, G.M. (1982). Transposition of cloned P elements into Drosophila germ line chromosomes. Science *218*, 341-347.

Swanson, C.I., Evans, N.C., and Barolo, S. (2010). Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. Dev Cell *18*, 359-370.

Swanson, C.I., Hinrichs, T., Johnson, L.A., Zhao, Y., and Barolo, S. (2008). A directional recombination cloning system for restriction- and ligation-free construction of GFP, DsRed, and lacZ transgenic *Drosophila* reporters. Gene *408*, 180-186.

Swanson, C.I., Schwimmer, D.B., and Barolo, S. (2011). Rapid evolutionary rewiring of a structurally constrained eye enhancer. Curr Biol *21*, 1186-1196.

Thomas, K.R., and Capecchi, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell *51*, 503-512.

Thorpe, H.M., and Smith, M.C. (1998). In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. Proc Natl Acad Sci U S A *95*, 5505-5510.

Uv, A.E., Harrison, E.J., and Bray, S.J. (1997). Tissue-specific splicing and functions of the Drosophila transcription factor Grainyhead. Mol Cell Biol *17*, 6727-6735.

Venken, K.J., He, Y., Hoskins, R.A., and Bellen, H.J. (2006). P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. Science *314*, 1747-1751.

Voas, M.G., and Rebay, I. (2004). Signal integration during development: insights from the *Drosophila* eye. Dev Dyn *229*, 162-175.

Wilson, K.L., and Berk, J.M. (2010). The nuclear envelope at a glance. J Cell Sci *123*, 1973-1978.

Yu, J., and McMahon, A.P. (2006). Reproducible and inducible knockdown of gene expression in mice. Genesis *44*, 252-261.

Zhu, Z.Y., and Sun, Y.H. (2000). Embryonic and genetic manipulation in fish. Cell Res *10*, 17-27.

CHAPTER 8

CONCLUSIONS AND FUTURE DIRECTIONS

Despite more than three decades of research since the discovery of the Simian Virus 40 (SV40) enhancer, we still do not completely understand the mechanisms by which enhancers act to regulate gene transcription or the structural rules that govern their location and organization in the genome. The objective of this dissertation was to investigate the long-range gene regulation at the level of the transcriptional enhancer as well as to delve deeper into the basic, fundamental aspects of enhancer regulation and organization. Using the *dPax2* cone cell enhancer *sparkling* as an example distantly located enhancer we undertook a mutational and biochemical analysis of the enhancer and also examined the additional regulatory sequences of the *dPax2* genomic locus.

8.1 A current model of *sparkling* action

The *dPax2* cone cell enhancer now represents one of the most extensively characterized developmental enhancers in any organism (Evans et al., 2012; Flores et al., 2000; Fu et al., 1998; Fu and Noll, 1997; Johnson et al., 2008; Swanson et al., 2010; Swanson et al., 2011). We know that the enhancer is regulated through interaction with the transcription factors Lozenge, PntP2, Yan, and Suppressor of Hairless. Additionally, *sparkling* requires at least four distinct inputs from the remainder of the enhancer, which lie in *spa* regions 1, 4, 5, and 6. This dissertation, along with previous work from the laboratory, has characterized the role of each of these regions in *spa* function. Region 1, or the RCE, is responsible for mediating the long-range function of the enhancer and can contribute cone cell pattering information to the enhancer. Region 4 of the enhancer is critical for both the initiation and maintenance of gene expression in cone cells. Regions 5 and 6 appear to be required more for proper initiation of gene expression than for maintenance during later stages of development. Region 5 is also required to repress enhancer activity in photoreceptors (Figure 8.1). We also found that region 4 is capable of binding protein specifically through the HD site in the 4b subelement, the same site that is critical for an interaction with Sine oculis *in vitro*. Meanwhile, region 5 also interacts with protein through the 5b subelement, but this interaction does not require an intact HD site. Given these unique roles of each of these sequences we originally thought the sequences with each enhancer encode unique protein interactions and contribute distinct actions the total of *sparkling* function. However, we found that some of these sequences have overlapping functions. For example, regions 4, 5, and 6 can all act from the position of the RCE to stimulate distal gene transcription, although only region 4 does so at wildtype levels. The RCE, which we long thought contained no patterning information, can also act as a copy of region 4 to promote gene expression in cone cell. In fact, if the RCE is placed in the wildtype region 4 location within the enhancer, the distally placed sparking

Figure 8.1

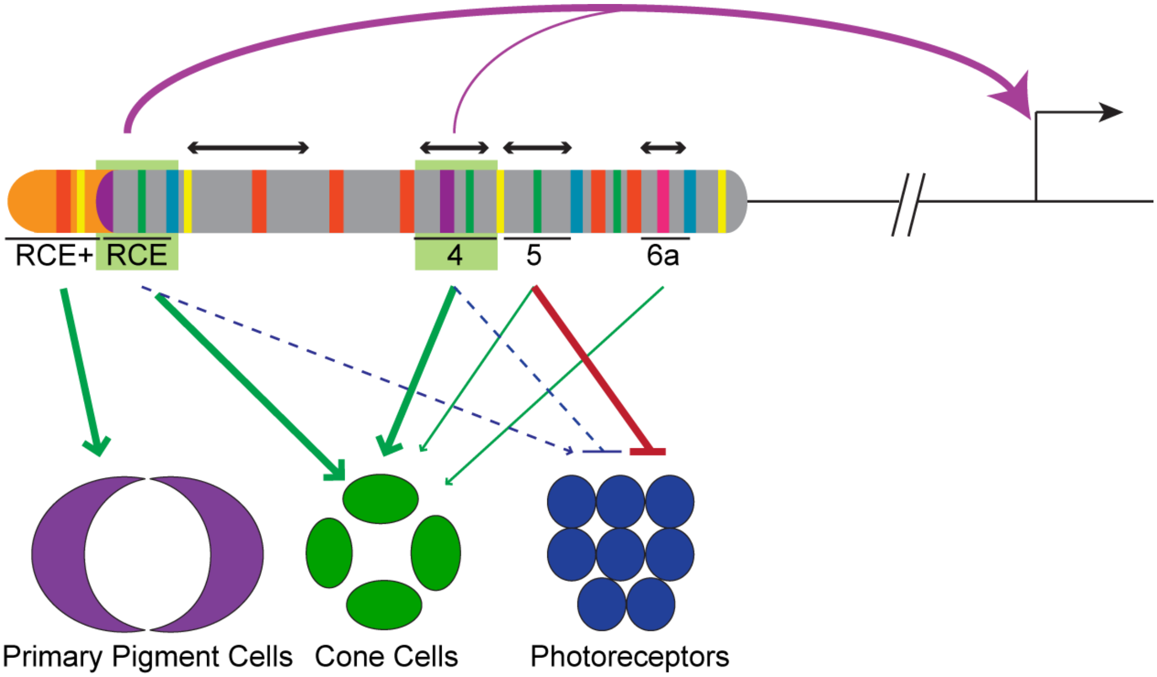


Figure 8.1 Current model of *sparkling* activity. This work and previous studies have shown that the RCE, region 4, region 5, and region 6 all contain critical inputs to *spa* function. Furthermore we know that sequences within 4, 5, and 6 require specific spatial organization (black arrows). The RCE is required to facilitate the distal gene transcription, and *spa* region 4 can contribute to this action when present in two copies (purple arrows). We also know that the RCE, 4, 5, and 5 all contain patterning information in cone cells (green arrows). Meanwhile, region 5 contains information required to repress *spa* activity in photoreceptors (red line). Furthermore, either region 4 can also repress expression in photoreceptors, or the RCE can promote photoreceptor expression when in the location of region 4 (blue lines). Interestingly, the RCE and region 4 sequences can compensate for one another, and the RCE can perform both sequences roles from the position of region 4 (green boxes). Finally the addition of 60bp of upstream conserved sequence (orange) is enables the enhancer to drive gene expression in primary pigment cells in addition to cone cells. Putative transcription factor binding sites: Su(H) red, Ets yellow, Lz blue, So purple, HD green, unknown region 6 site pink.

enhancer does not require a copy of region 4 at all. Suggesting the inputs within the RCE can regulate distal gene activity and provide proper patterning information, as well as synergize with the sequences around the region 4 position to enable enhancer activity (Figure 8.1). Despite finding that some regions of the enhancer are flexible in both their location and precise identify of the input, we also found that some inputs are more rigid. For example, *spa* region 4 cannot substitute for region 5 and *spa* region 5 cannot substitute for region 4, suggesting each of these regions contributes a unique input to the enhancer.

8.2 The *dPax2* 4th intron contains additional regulatory sequences

In addition to cone cells, *dPax2* is also expressed in primary pigment cells and the mechanosensory bristles (Fu et al., 1998). The regulatory sequences responsible for bristle cell development lie in two enhancers upstream of the gene's promoter (Johnson et al., 2011). We know that the *sparkling* enhancer is in the 4th intron of the gene, and that the primary pigment cell enhancer is likely also in this intron (Fu et al., 1998; Fu and Noll, 1997). We found that the addition of at minimum, 60bp of DNA sequence upstream of the *spa* enhancer enables the enhancer to drive reporter gene expression primary pigment cells (Figure 8.1). As specification of the primary pigment cells also requires the known regulators of *sparkling,* it is possible the addition of a small amount of additional information is able to convert these inputs into a primary pigment cell enhancer, as we also saw in Chapter 2 with photoreceptors. Alternatively, the *dPax2* primary pigment cell enhancer could also overlap with the *sparkling* enhancer. As we have yet to find a different sequence within the 4th intron that drives primary pigment cell reporter expression, this possibility is becoming more and more likely. Our experimental analysis shows that gene expression in primary pigment cells requires not only the upstream intronic sequence, but all or almost all, of the *sparkling* enhancer sequence. The activity of *spa* homolgous sequences in *D. erecta, D. ana,* and *D. virilis* as well as *sparking* enhancer mutations altering the affinity of Su(H) binding sites indicates that increased Notch input may play a critical role in the primary pigment cell enhancer action (Swanson et al., 2010). We can test this possibility by mutating, or changing the affinity of, the Su(H) sites within these enhancers and assessing both cone cell and primary pigment cell gene expression. Furthermore, we can decrease Notch signaling in these discs and assess the effect on enhancer action.

We also found a second sequence within the *dPax2* 4th intron that is capable of driving GFP expression in the cone cells of the developing *Drosophila* eye imaginal disc. This sequence contains binding sites for the known regulators of *spa* activity Lz, Su(H), and Ets factors. As such we would like to examine the contribute of these sites, as well as any potential Sine Oculis binding sites, to enhancer function. We are very interested in the location of this enhancer within the 4th intron of the gene, as it lies directly upstream of a secondary promoter of *dPax2.* This secondary promoter is capable activating transcription from a small exon which splices to the 5th exon in the same frame as *dPax2.* The upstream exons are all small and very poorly conserved compared to the downstream exons. We would like to establish the expression of *dPax2* from both the primary and secondary promoters in the developing eye disc using transcript specific *in situ,* as well as to assess the ability of both the *sparkling* and this new cone cell enhancer to regulate both promoters. It is possible that each enhancer regulates a specific promoter, or that they can both regulate either promoter, or that our new enhancer does not regulate gene expression in the endogenous location at all.

In order to continue our analysis of both the *sparkling* enhancer, the primary pigment cell enhancer, and the second cone cell enhancer, it is essential we study enhancer action in the endogenous context. To do so we will generate *dPax2* BAC reporter constructs and rescue constructs. Using this method, we can remove each enhancer, or parts of each enhancer, in order to study their activity in their wildtype genomic context. The will aid us in not only studying long-range gene regulation, but to examine the ability of these enhancers to work together to stimulate gene transcription. Our second cone cell enhancer sequence is present in the sparkling mutant flies used to discover the *sparkling* enhancer (Fu et al., 1998). Therefore, it is likely that this enhancer is not able to regulate *dPax2* expression alone, unlike a traditional shadow enhancer (Barolo, 2012; Hong et al., 2008; Perry et al., 2010). However, it is possible that the breakpoint of the spa^{pol} mutation, less than 100bp upstream does affect the action of this enhancer. As such, we could recreate the mutation in our BAC analysis and examine the activity of this enhancer in the altered genomic context.

8.3 A potential role for Sine oculis in *sparkling* activity

The Six family protein, Sine oculis (So) is a critical regulator of *Drosophila* eye development and cell type specification ((Blanco et al., 2010; Bovolenta et al., 1998; Cheyette et al., 1994; Daniel et al., 1999; Serikaku and O'Tousa, 1994). As transcription factor, it has been shown to interact with several eye specific enhancers and activate gene transcription, likely through its cofactor, eyes absent (Blanco et al., 2010; Halder et al., 1998; Pauli et al., 2005; Pignoni et al., 1997; Yan et al., 2003). Therefore, it would be not be surprising if Sine oculis (So) is a critical regulator of the *sparkling* enhancer as well. As such, we identified two putative So binding motifs within the enhancer and have see that these sites are able to interact with these sites *in vitro* (Figure 8.1)*.* However, targeted mutation of one of these sites within the enhancer did not affect enhancer activity, leaving us unsure as to the potential role for So in regulating *spa.* To further examine the role of So to *spa* activity within the enhancer we will mutate all of the So binding sites within the wildtype enhancer, possibly in conjunction with all of the HD binding sites. We also plan to examine reporter gene expression, *dPax2* expression, and cone cell specification upon knockdown of So in the developing eye disc. We already know that decreased So expression during larval development results in loss of cone cell specification in third instar larva. However, as the adult eyes are phenotypicaly normal, specification must recover at some point during development (N. Evans, unpublished observation).

We observed a clear association between Sine oculis binding sites and the classic homeodomain binding site (TAAT) in the *sparkling* enhancer. Given the relationship between these sites, we would like to assess the ability of So to interact cooperatively with one of the many retinal homeodomain proteins using *in vitro* gel shift assays as well as altering the spacing between the So and HD sites *in vitro* and *in vivo.* We would also like to test the possibility that Sine oculis itself is able to interact with both the So consensus site and the HD site through either the formation of So dimmers or through two DNA binding domains in a single protein (Czerny et al., 1999).

The putative So binding sites within the enhancer lie within the RCE and region 4, which can both convey long-range transcriptional activity. Given the importance of So in *Drosophila* eye specification, and the essential nature of the Six proteins in vertebrate development, a potential role for these proteins in distal gene regulation is very interesting. However, there is no known function of So that would directly implicate it for this role. Therefore, we will test the ability of both the RCE and So to enable reporter gene expression driven by distally placed synthetic and eye specific enhancers.

8.4 *sparkling* enhancer promiscuity

The observation that the promoter proximal wildtype enhancer drives slightly decreased levels of gene expression compared to the enhancer lacking the RCE began as a moderately interesting, but likely coincidal piece of data.

However, as we dug deeper into the role of the RCE in facilitating *spa* action this observation has become more and more interesting. We have found that the wildtype enhancer drives decreased expression whenever the reporter construct is not insulated. However, upon addition of insulator sequences flanking the enhancer and reporter gene, the constructs drive similar levels of expression. Furthermore, we have seen that placement of the *spa* enhancer into known promoter rich genomic locations such as the 86F8 and attP2 landing sites significantly affects wildtype enhancer action (Bischof et al., 2007; Groth et al., 2004). Therefore we hypothesize that the wildtype enhancer, potentially through the RCE, is capable of interacting with multiple local promoters and therefore, spending less "time" activating GFP expression than an enhancer lacking the RCE. Addition of the insulators sequences would inhibit these interactions such that *spa*(wt) can only interact with a single promoter. Given the lack of reporter expression from *spa*(RCE+)-846bp, it is also possible that the extra 60bp enable the enhancer to interact so strongly with other local promoters that no GFP expression is seen. We would like to test the ability of the wildtype enhancers to regulate multiple promoters using promoter competition assays in which *spa* is allowed to activate gene transcription from different core promoters driving the expression of unique reporter genes. We are especially interested in whether the *spa* enhancer, the primary pigment cell enhancer, and our newly identified cone cell enhancer activate transcription preferentially from the primary or secondary *dPax2* promoters. We can also test the ability of *spa*(RCE+)-846bp to regulate transcription from different core promoters utilizing promoter competition assays in which two promoters drive different reporter genes.

8.5 Potential mechanisms of RCE activity

Throughout this work we have identified a significant number of protein candidates for both interaction with the RCE and for regulating the *sparkling* enhancer using candidate analysis, motif analysis, and affinity purification. The identity of these proteins combined with our *in vivo* observations; provide evidence for several of the models of long-range enhancer activity.

8.5a Looping

One of the prominent proposed mechanisms of distal enhancer action is through looping of DNA to bring the enhancer and its target promoter into close proximity and stimulate gene transcription. This is thought to occur by the formation of protein complexes between the enhancer and promoter that alter the 3D structure of the chromatin. As such we assessed the ability Zeste, a protein capable of binding specific sequences within an enhancer and a promoter and induce chromatin loops, to interact with the RCE, but did not see an interaction *in vitro* (Kostyuchenko et al., 2009; Laney and Biggin, 1997; Mohrmann et al., 2002; Qian et al., 1992). Accordingly, we found that addition of the second copy of the RCE had no affect on *spa* activity making it unlikely that if the RCE functions to form loops through protein homo-oligomerization. This is not unexpected, as we know that *spa* can activate both the endogenous *dPax2* promoter and the hsp70 promoter in our reporter vector, which is separated from the enhancer by a completely independent DNA sequence.

In terms of a looping mechanism then, it would be more likely that *spa* would bind proteins known to form large proteins complexes that force looping of DNA such as those that interact with insulators. Our affinity purification indentified two proteins known to interact with insulators and promote DNA looping, iswi and Mi2 (Li et al., 2010; Mutskov et al., 2002). However, we have been unable to detect interaction of these proteins, or of CTGF another protein known to interact with insulators and cohesion to promoter looping, with the RCE *in vito.* Another canididate, Dichaete, is an HMG family member (High Mobility Group) which has been shown to bind to and bend DNA (Pil et al., 1993). This protein is not expressed in the correct cells to be involved in *spa* activity; however, it is possible a family member with similar function and binding affinity is (Mukherjee et al., 2000). Ideally, to test looping as a mechanism of *spa* activity, we would prefer to utilize chromatin capture assays. However, we are limited by both the small amount of tissue we have to work with (cone cells of the developing *Drosophila* eye) and the distance between *spa* and the promoter; even in its endogenous location, looping between the *spa* enhancer and the *dPax2* promoter would be difficult to detect above background levels (Dekker, 2006).

8.5b Linking

In this model the enhancer and promoter remain separated from each other spatially, and the enhancer sets up the formation of a protein complex that spreads across the chromatin between the enhancer and promoter, ultimately activating target gene transcription. We identified motifs for the Lim domain family of proteins within the RCE. As Lim proteins contain homeodomain binding sites, the importance of HD sites in the *spa* enhancer provides further support the potential of Lim proteins to interact with *spa.* Lim proteins are thought to form a complex with the protein Chip, and in turn form a linking protein complex between the distal enhancer and its target promoter (Morcillo et al., 1996). This is an unlikely mechanism for *spa* action however as we know the enhancer can function from both the *dPax2* 4th intron, and in our reporter construct we contain very different intervening DNA sequences. However, as HD sites are prevalent in almost any DNA sequence, we could test this model by changing the spacer DNA sequence in our reporter constructs.

8.5c Tracking

We found that the RCE can act to promote distal *spa* activity even when it is separated from the enhancer by a significant distance. This observation supports a tracking mechanism for *spa* activity. As *spa* can act from downstream to activate *dPax2* transcription and upstream to activate GFP transcription the RCE would have to recruit the basal transcription machinery and stimulate transcription in both directions in order tracking to be a plausible mechanism -

unless, or course, *spa* actually activates *dPax2* from the secondary, downstream, promoter. Consistent with this mode of enhancer action, we have identified several members of the basal transcription machinery as potential protein interacting partners of the RCE including TATA binding proteins (Tbp) and TBP-associated factors 6 and 9 (Taf6 and 9) (Blackwood and Kadonaga, 1998; Thomas and Chiang, 2006; Zhu et al., 2007). Accordingly, Taf6 is the only one of our candidate proteins we have been able to demonstrate interacts with the RCE *in vitro.* In order to further test tracking as a mechanism RCE activity, we could analyze transcription of *spa* and the surrounding genomic sequences in both the endogenous locations and from our reporter constructs. We would also like to perform Chromatin Immunoprecipitation (ChIP) for Pol ll at the *spa* enhancer and intervening sequences; however we are again limited by the amount of tissue available to work with.

8.5d Non-coding RNAs

The ability of Taf6 to interact with the RCE could also support the production of enhancer-like non coding RNAs from the *sparkling* genomic region (Orom et al., 2010). Additionally the identification of RNA binding proteins such as no on or off transient A (nonA), polyA-binding protein (pABp), and penguin (pen) could all support this mechanism of *spa* action (Derry et al., 2006; Kozlova et al., 2006; Maleszka et al., 1996). If RNA transcriptions are detected from the *spa* enhancer or nearby, we could decrease levels of this RNA using siRNA, and assess the effect on endogenous or reporter gene transcription. Again, this mechanism is unlikely as *spa* stimulates transcription from both the *dPax2* and hsp70 heterologous promoter over different intervening sequences. However, long non coding RNAs have previously been shown to stimulate transcription from both endogenous and heterologous promoters (Orom et al., 2010).

In general, the entire study of enhancer biology warrents more attention to long-range gene regulation. Almost every enhancer in the genome requires some mechanism interacting with its target promoter from a distal location. However, most enhancer studies only analyze activity from a promoter proximal position and disregard this important component of enhancer action. We hope that in time, more long-range elements will be identified within enhancers, and through extensive study the mechanisms by which these elements act to regulate gene transcription will be better understood.

8.6 Refrences

Barolo, S. (2012). Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. Bioessays *34*, 135-141.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. Proceedings of the National Academy of Sciences of the United States of America *104*, 3312-3317.

Blackwood, E.M., and Kadonaga, J.T. (1998). Going the distance: a current view of enhancer action. Science *281*, 60-63.

Blanco, J., Pauli, T., Seimiya, M., Udolph, G., and Gehring, W.J. (2010). Genetic interactions of eyes absent, twin of eyeless and orthodenticle regulate sine oculis expression during ocellar development in Drosophila. Dev Biol *344*, 1088-1099.

Bovolenta, P., Mallamaci, A., Puelles, L., and Boncinelli, E. (1998). Expression pattern of cSix3, a member of the Six/sine oculis family of transcription factors. Mech Dev *70*, 201-203.

Cheyette, B.N.R., Green, P.J., Martin, K., Garren, H., Hartenstein, V., and Zipursky, S.L. (1994). The drosophila sine oculis locus encodes a homeodomain-containing protein required for the development of the entire visual system. Neuron *12*, 977-996.

Czerny, T., Halder, G., Kloter, U., Souabni, A., Gehring, W.J., and Busslinger, M. (1999). twin of eyeless, a second Pax-6 gene of Drosophila, acts upstream of eyeless in the control of eye development. Mol Cell *3*, 297-307.

Daniel, A., Dumstrei, K., Lengyel, J.A., and Hartenstein, V. (1999). The control of cell fate in the embryonic visual system by atonal, tailless and EGFR signaling. Development *126*, 2945-2954.

Dekker, J. (2006). The three 'C' s of chromosome conformation capture: controls, controls, controls. Nat Methods *3*, 17-21.

Derry, M.C., Yanagiya, A., Martineau, Y., and Sonenberg, N. (2006). Regulation of poly(A)-binding protein through PABP-interacting proteins. Cold Spring Harb Symp Quant Biol *71*, 537-543.

Evans, N.C., Swanson, C.I., and Barolo, S. (2012). Sparkling insights into enhancer structure, function, and evolution. Curr Top Dev Biol *98*, 97-120.

Flores, G.V., Duan, H., Yan, H., Nagaraj, R., Fu, W., Zou, Y., Noll, M., and Banerjee, U. (2000). Combinatorial signaling in the specification of unique cell fates. Cell *103*, 75-85.

Fu, W., Duan, H., Frei, E., and Noll, M. (1998). *shaven* and *sparkling* are mutations in separate enhancers of the *Drosophila Pax2* homolog. Development *125*, 2943-2950.

Fu, W., and Noll, M. (1997). The *Pax2* homolog *sparkling* is required for development of cone and pigment cells in the *Drosophila* eye. Genes Dev *11*, 2066-2078.

Groth, A.C., Fish, M., Nusse, R., and Calos, M.P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics *166*, 1775-1782.

Halder, G., Callaerts, P., Flister, S., Walldorf, U., Kloter, U., and Gehring, W.J. (1998). Eyeless initiates the expression of both sine oculis and eyes absent during Drosophila compound eye development. Development *125*, 2181-2191.

Hong, J.W., Hendrix, D.A., and Levine, M.S. (2008). Shadow enhancers as a source of evolutionary novelty. Science *321*, 1314.

Johnson, L.A., Zhao, Y., Golden, K., and Barolo, S. (2008). Reverse-engineering a transcriptional enhancer: a case study in Drosophila. Tissue Eng Part A *14*, 1549-1559.

Johnson, S.A., Harmon, K.J., Smiley, S.G., Still, F.M., and Kavaler, J. (2011). Discrete regulatory regions control early and late expression of D-Pax2 during external sensory organ development. Dev Dyn *240*, 1769-1778.

Kostyuchenko, M., Savitskaya, E., Koryagina, E., Melnikova, L., Karakozova, M., and Georgiev, P. (2009). Zeste can facilitate long-range enhancer-promoter communication and insulator bypass in Drosophila melanogaster. Chromosoma *118*, 665-674.

Kozlova, N., Braga, J., Lundgren, J., Rino, J., Young, P., Carmo-Fonseca, M., and Visa, N. (2006). Studies on the role of NonA in mRNA biogenesis. Exp Cell Res *312*, 2619-2630.

Laney, J.D., and Biggin, M.D. (1997). Zeste-mediated activation by an enhancer is independent of cooperative DNA binding in vivo. Proc Natl Acad Sci U S A *94*, 3602-3604.

Li, M., Belozerov, V.E., and Cai, H.N. (2010). Modulation of chromatin boundary activities by nucleosome-remodeling activities in Drosophila melanogaster. Mol Cell Biol *30*, 1067-1076.

Maleszka, R., Hanes, S.D., Hackett, R.L., de Couet, H.G., and Miklos, G.L. (1996). The Drosophila melanogaster dodo (dod) gene, conserved in humans, is functionally interchangeable with the ESS1 cell division gene of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *93*, 447-451.

Mohrmann, L., Kal, A.J., and Verrijzer, C.P. (2002). Characterization of the extended Myb-like DNA-binding domain of trithorax group protein Zeste. J Biol Chem *277*, 47385-47392.

Morcillo, P., Rosen, C., and Dorsett, D. (1996). Genes regulating the remote wing margin enhancer in the Drosophila cut locus. Genetics *144*, 1143-1154.

Mukherjee, A., Shan, X., Mutsuddi, M., Ma, Y., and Nambu, J.R. (2000). The Drosophila sox gene, fish-hook, is required for postembryonic development. Dev Biol *217*, 91-106.

Mutskov, V.J., Farrell, C.M., Wade, P.A., Wolffe, A.P., and Felsenfeld, G. (2002). The barrier function of an insulator couples high histone acetylation levels with specific protection of promoter DNA from methylation. Genes Dev *16*, 1540-1554.

Orom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q.*, et al.* (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell *143*, 46-58.

Pauli, T., Seimiya, M., Blanco, J., and Gehring, W.J. (2005). Identification of functional sine oculis motifs in the autoregulatory element of its own gene, in the eyeless enhancer and in the signalling gene hedgehog. Development *132*, 2771-2782.

Perry, M.W., Boettiger, A.N., Bothma, J.P., and Levine, M. (2010). Shadow enhancers foster robustness of Drosophila gastrulation. Curr Biol *20*, 1562-1567.

Pignoni, F., Hu, B., Zavitz, K.H., Xiao, J., Garrity, P.A., and Zipursky, S.L. (1997). The eye-specification proteins So and Eya form a complex and regulate multiple steps in Drosophila eye development. Cell *91*, 881-891.

Pil, P.M., Chow, C.S., and Lippard, S.J. (1993). High-mobility-group 1 protein mediates DNA bending as determined by ring closures. Proc Natl Acad Sci U S A *90*, 9465-9469.

Qian, S., Varjavand, B., and Pirrotta, V. (1992). Molecular analysis of the zeste-white interaction reveals a promoter-proximal element essential for distant enhancer-promoter communication. Genetics *131*, 79-90.

Serikaku, M.A., and O'Tousa, J.E. (1994). sine oculis is a homeobox gene required for Drosophila visual system development. Genetics *138*, 1137-1150.

Swanson, C.I., Evans, N.C., and Barolo, S. (2010). Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. Dev Cell *18*, 359-370.

Swanson, C.I., Schwimmer, D.B., and Barolo, S. (2011). Rapid evolutionary rewiring of a structurally constrained eye enhancer. Curr Biol *21*, 1186-1196.

Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol *41*, 105-178.

Yan, H., Canon, J., and Banerjee, U. (2003). A transcriptional chain linking eye specification to terminal determination of cone cells in the Drosophila eye. Dev Biol *263*, 323-329.

Zhu, X., Ling, J., Zhang, L., Pi, W., Wu, M., and Tuan, D. (2007). A facilitated tracking and transcription mechanism of long-range enhancer function. Nucleic Acids Res *35*, 5532-5544.

APENDIX 1

PROMOTER PROXIMAL AND DISTAL ANALYSIS OF THE *DPPD* WING IMAGINAL DISC ENHANCER

1.1 Abstract

The *dppD* enhancer is responsible for regulating expression of decapentaplegic (dpp) in the *Drosophila* wing imaginal disc, specifically along the A/P boundary of the disc. This minimal enhancer lies at 26 kb form the nearest *dpp* promoter;therefore, it is likely this enhancer requires a mechanism for long-range regulation to allow the enhancer to regulate gene expression in a temporal and cell type specific manner. As this enhancer has been analyzed at the subregion level, has known inputs, and must act at a distance from the gene it regulates, this enhancer is a good candidate for searching for a “remote control” element in addition to that discovered in the *sparkling dPax2* enhancer. Therefore we undertook a mutational analysis of the enhancer at both a promoter proximal and promoter distal position. Unfortunately, we were unable to conclusively identify a region of the *dppD* specifically required for long-range enhancer function. The experimental results of this study further elucidate the role of the essential DNA sequences within the *dppD* enhancer as well as identify

putative transcription factor interaction partners based on motif analysis and known *in vivo* function.

PROMOTER PROXIMAL AND DISTAL ANALYSIS OF THE *DPPD* WING IMAGINAL DISC ENHANCER

1.2 Introduction

The *Drosophila* morphogen protein *decapentaplegic* (*dpp*) is necessary for the proper patterning and growth of numerous tissues including the embryo, heart and the fifteen imaginal discs. Imaginal discs are the larval tissues that will undergo morphogenesis during pupation to become the adult appendages and other organs (Affolter and Basler, 2007). As the name suggests, *decapenta* (fifteen) – *plegic* (paralysis), loss of *dpp* expression results in failure of the imaginal discs to form correctly. *dpp* is a homolog of the vertebrate bone morphogenic proteins (specifically BMP 2 and 4) and a member of the TGF-β superfamily of single cascade proteins(Entchev et al., 2000). *dpp's* function in regulating tissue growth and patterning has been best demonstrated in the *Drosophila* wing imaginal disc during larval development. Here, *dpp* ligand binds to receptors *Thickveins* (Tkv) and *Punt* to activate *mother against dpp* (*mad*) by phosphorylation (Kim et al., 1997; Ruberte et al., 1995). Phosphorylated mads can act as transcription factors which repress the gene *brinker*, a repressor of *dpp* target genes, and activates wing patterning genes such as *vestigal*, *optomoter blind* (omb), and *splat* and growth regulators such as the microRNA *bantam (Campbell and Tomlinson, 1999; Kim et al., 1997; Marty et al., 2000; Nellen et al., 1996; Oh and Irvine, 2011).*

The pattern of *dpp* expression itself in the wing is tightly controlled through the intersection of Hedgehog (Hh), via the transcription factor Cubitus interuptus (Ci), and Engrailed (En) activity. In the wing imaginal disc *En* and *Hh* are expressed only in the posterior compartment while *Ci* is expressed only in the anterior compartment (Lee et al., 1992; Morata and Lawrence, 1975; Schwartz et al., 1995) (Figure 1.1A). Hh ligand secreted from the posterior compartment generates a gradient of Hh across the anterior compartment, with the highest levels at the anterior/posterior boundary (A/P) (Figure 1.1A). In the absence of Hh signaling the Ci transcription factor is cleaved to a 75kDa form, which acts as a transcriptional repressor (Ci^{REP}) (Aza-Blanc et al., 1997). In cells that receive Hh signal, Ci remains uncleaved in a 155kDa form and is converted to a transcriptional activator (Ci^{ACT}) (Chen et al., 1999; Methot and Basler, 1999). Therefore, the gradient of Hh across the anterior compartment results in opposing gradients of Ci^{REP} vs. Ci^{ACT}, where Ci^{ACT} is present in highest at the A/P boundary (Figure 1.1A). Only were Ci^{ACT} is present in great enough levels is transcription of *dpp* stimulated (Methot and Basler, 1999; Tabata and Kornberg, 1994). Expression of *dpp* is simultaneously repressed in the posterior compartment by En, generating a precise stripe of *dpp* along the A/P boundary (Zecca et al., 1995) (Figure 1.1A, top).

In a 2000 *Development* paper, Müller and Basler set out to identify the enhancer responsible for integrating the Hh and En signals and regulating the precise stripe of *dpp* expression in the wing disc (Muller and Basler, 2000). The *dpp* genomic locus contains at least 25kb of potential regulator sequences 3' of the gene (Figure 1.1A bottom), evidenced by the approximately 30 mutant alleles spanning this region which all result in diminished *dpp* expression (Blackman et al., 1987; St Johnston et al., 1990). Therefore, it is likely this region contains an enhancer, or enhancers, capable of regulating *dpp* expression. Furthermore, a specific 4kb sequence from this region had previously been shown to complement *dpp* mutant alleles, and in a reporter construct was capable of driving LacZ expression which recapitulates the *dpp* expression pattern in the third instar larval wing disc (Masucci et al., 1990). Starting from this fragment, Müller and Basler subsequently identified an 800bp DNA fragment that is sufficient to direct LacZ expression in a stripe along the A/P boundary in the wing disc (Figure 1.1B). Next they systematically analyzed the contributions of each part of this enhancer by making eight 100bp deletions within the enhancer (ΔA-ΔH, Figure 1.1 C-I)(Muller and Basler, 2000).

Recall that *dpp* is regulated by Hh signaling and En. Müller and Basler demonstrated that this action is direct. Deletion of region F resulted in expansion of reporter gene expression into the posterior wing compartment, suggesting this region may contain an En binding site (Figure 1.1H). Indeed, they identified and mutated a potential En site (TAATCA) and saw even greater derepression of LacZ in the posterior compartment (Muller and Basler, 2000). It should be noted,

Figure 1.1

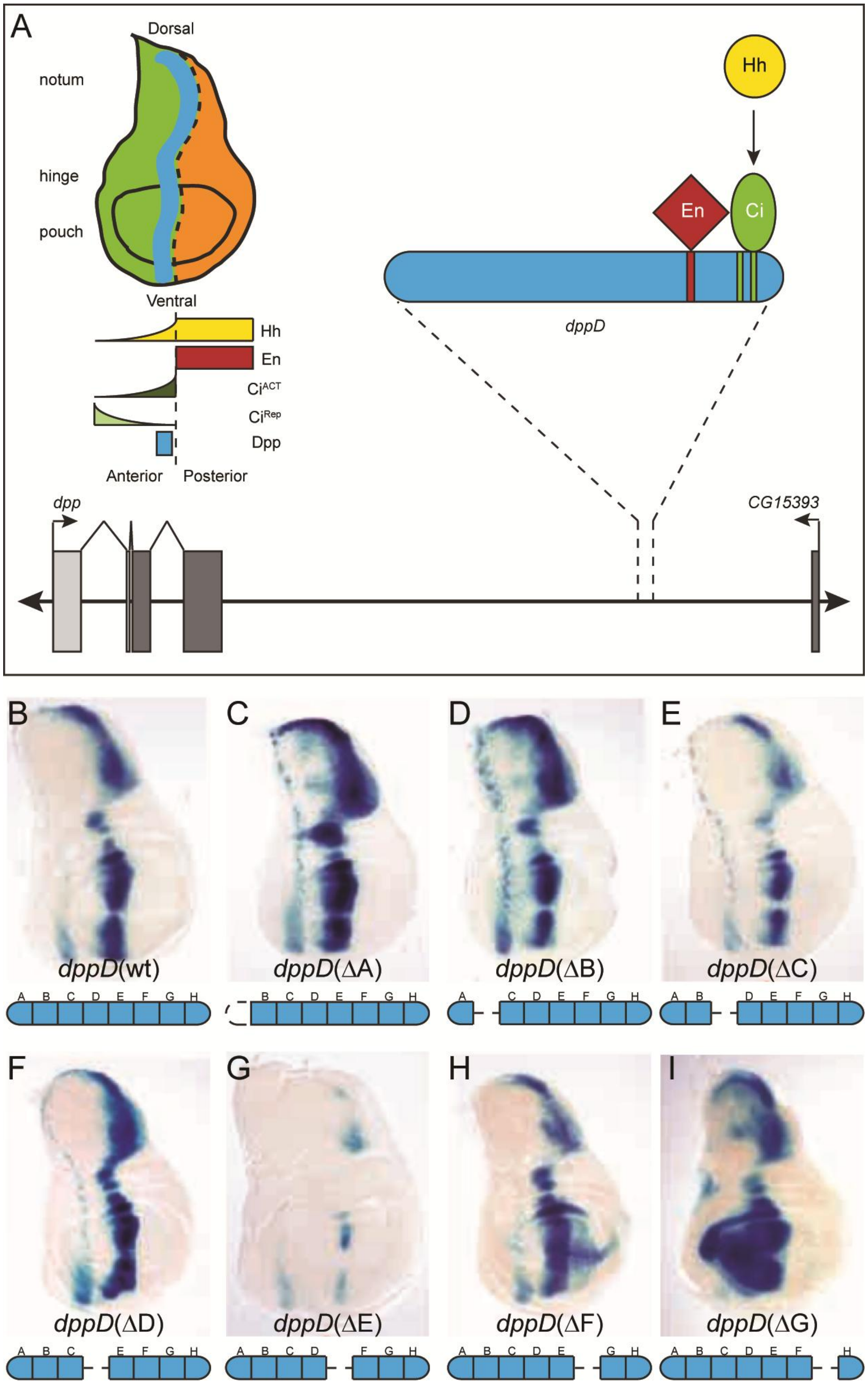

A
Dorsal
notum
hinge
pouch
Ventral
Hh
En
Ci^{ACT}
Ci^{Rep}
Dpp
Anterior
Posterior
Hh
En
Ci
dppD
dpp
CG15393
B
C
D
E
dppD(wt)
dppD(ΔA)
dppD(ΔB)
dppD(ΔC)
A B C D E F G H
B C D E F G H
A C D E F G H
A B D E F G H
F
G
H
I
dppD(ΔD)
dppD(ΔE)
dppD(ΔF)
dppD(ΔG)
A B C E F G H
A B C D F G H
A B C D E G H
A B C D E F H

Figure 1.1 *Decapentaplegic* (*dpp*) expression in the third instar larval wing disc is restricted to a stripe at the anterior/posterior boundary that extends from the notum, through the hinge and to the wing pouch. This specific pattern of gene expression is set up through the integration of Hedgehog (Hh) and Engrailed (En) signaling events. En is expressed only in the posterior compartment of the wing disc where it represses *dpp* expression. Hh is also expressed exclusively in the posterior compartment. As Hh is a secreted protein, it moves across the A/P boundary and sets up a morphogen gradient. This gradient results in high levels of the activator from of Cubitus interuptus (Ci) at the A/P boundary and high levels of the repressor form of Ci at the margins. Together these signals activate *dpp* at the A/P boundary and repress it elsewhere (A top). One of the enhancers that integrates these signals to drive *dpp* expression in the wing disc is *dppD*. The *dppD* enhancer lies 27kb downstream of the gene (A bottom). The *dppD* enhancer contains binding sites for En and Ci (A). Previous studies have shown that loss of the En sequence in region F results in posterior derepression of *dppD* activity (H). Conversely, deletion of the Ci sites in region G result in anterior derepression (I). Region E likely contains a critical activation sequence as deletion of this region results in significantly decreased levels of reporter gene activity (G). Meanwhile, loss of regions A – D individually does not dramatically alter gene expression (B – F) (Muller and Basler, 2000).

loss of region F does not result in complete derepression, suggesting an additional En might be present elsewhere in the enhancer. Conversely, loss of region G showed expansion of LacZ expression into the anterior wing pouch and notum of the wing disc (Figure 1.1I). One explanation of this result is loss of a repressor binding site. As Ci acts as a repressor in the most anterior portions of the disc, it is a likely candidate for binding to region G. This anterior derepression was subsequently attributed to two 10bp subregions within G, that each contain a sequence that deviates by two basepairs from the Gli consensus binding site (TGGGT/AGGTC) which when mutated at a single basepair also result in anterior reporter gene expression (Muller and Basler, 2000). Interestingly, LacZ is still expressed at the A/P boundary when region G is deleted. As the authors demonstrated, Ci^{ACT} works through these same binding sites; it is likely additional activator binding sites such as another Hh target exist elsewhere in the enhancer. Together, loss of enhancer regions F and G demonstrate this enhancer is directly regulated by En and Ci (Figure 1.1 A) (Muller and Basler, 2000).

Loss of other regions of the enhancer demonstrate that region E contains a binding site, or binding sites, for an unknown transcriptional activator, as deletion of this sequence results in a severe decrease in reporter gene activity (Figure 1.1 G). Conversely, region H does not seem to contain any patterning information. Deletions of regions A-D individually do not significantly affect gene expression, with only loss of region C resulting in a slight decrease in enhancer activity (Figure 1.1 C-F). However, part of the A-D sequences must be necessary for enhancer activity as a fragment containing regions E-G is insufficient for proper reporter gene activation (Muller and Basler, 2000).

The *dpp* minimal enhancer identified by Müller and Basler in 2000 is between 52.7kb and 26.4kb 3' of the *dpp* promoter, depending on *dpp* isoform (Figure 1.1 A bottom); therefore, it is likely this enhancer requires a mechanism for long-range regulation to allow the enhancer to regulate gene expression in a temporal and cell type specific manner. As this enhancer has been analyzed at the subregion level, has known inputs, and must act at a distance from the gene it regulates, this enhancer – hereafter referred to as *dppD* – is a good candidate for searching for a "remote control" element in a second enhancer. Furthermore, the Müller and Basler experiments demonstrated that additional inputs other than Ci must be involved in controlling this enhancer's function. As such, we originally set out to understand the *dpp* enhancer inputs outside the contribution of Ci and En. As such the Barolo lab began its analysis of *dppD*, making three major changes to the enhancer compared to the previous study. First, region H was discarded as it did not appear to contribute to enhancer activity. Second, the single En site in region F, and two Ci sites in region G were mutated concurrently. Third, enhancer activity was analyzed at both a promoter proximal position (121bp from the LacZ transcription start site) and at a moderate distance (846bp) from the transcription start site (TSS).

1.3 Results

1.3a dppD activity at in the promoter distal position

As expected, region H was not essential and this enhancer hereafter referred to as wildtype [*dppD*(wt)], drives LacZ expression in a *dpp* expression pattern from this promoter distal position (Figure 1.2 B). The *dppD* enhancer lacking En and Ci binding sites [*dppD*(mut)] drives weak gene expression that is derepressed in both the anterior and posterior compartments due to loss of Ci^{REP} and En mediated repression (Figure1.2 C). Loss of regions A,B,C, and D individually did not significantly affect gene expression previously, yet A –D sequence is necessary for complete patterning (Muller and Basler, 2000). Reasoning that regions A-D might contain redundant regulatory information which would allow for compensation of each region’s function, we decided to delete these regions in tandem, all in the context the En and Ci sites mutated. Deletion of Region A alone results in decreased expression at this margins of the wing and slightly higher levels of expression at the A/P boundary (Figure 1.2 D). Loss of A and B together demonstrates decreased levels of expression that does not extend as far toward the wing margins (Figure 1.2 E). When regions A, B, and C are deleted, enhancer activity is extremely weak (Figure 1.2 F). Finally, when A, B, C, and D are lost together enhancer activity is completely lost (Figure 1.2 L). Together, these observations indicate that activator inputs, independent of Ci, lie in *dppD* regions A, B, C, and D. Furthermore, region A likely contains a repressor binding site.

Given these results we decided to mutate region D in addition to region E, the sequence previously demonstrated to contain an activation input (Muller and Basler, 2000). Instead of deleting these regions, we mutated the sequence while retaining the native spacing of the enhancer by changing every other basepair by non-complementary transversion. Interestingly, in our hands, loss of BOTH regions D and E independently result in complete loss of LacZ expression (Figure 1.2 M, N), suggesting each of these regions contains essential activator inputs. Notably, these inputs are not sufficient for *dppD*(mut) activity, as constructs that contain both these inputs do not give correct reporter gene activity [ie *dppD*(ΔA) and *dppD*(ΔAB)]. These results differ from those of Müller and Basler as they saw no change in LacZ expression upon deletion of region D (Figure 2.1F) (Muller and Basler, 2000). The difference in our results could be explained by the change in spacing in the previous experiments, which could juxtapose two activator binding sites that work synergistically to compensate for loss of region D, or the sequences within region D are sensitive to the change in

Figure 1.2

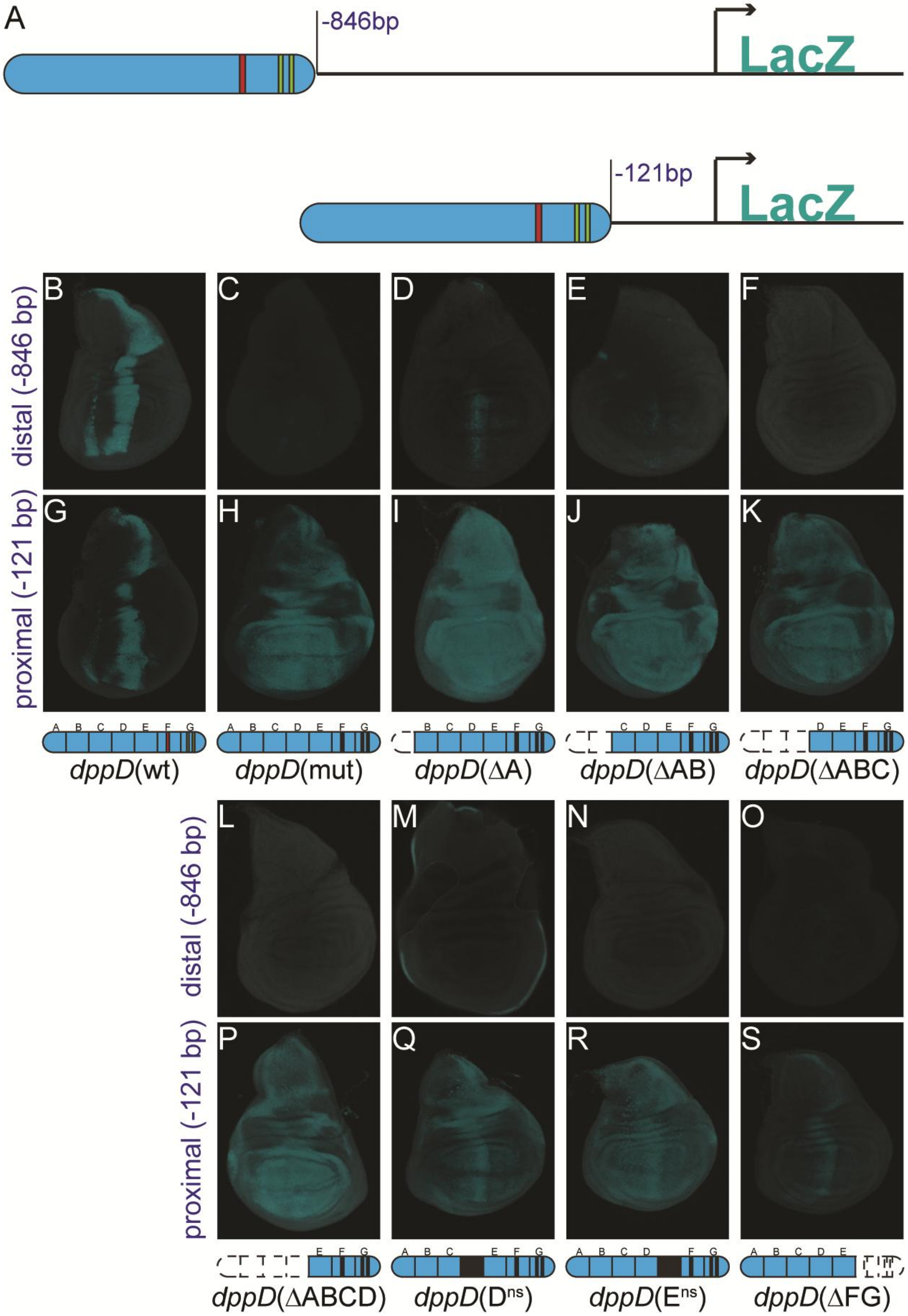

A
-846bp
LacZ
-121bp
LacZ
distal (-846 bp)
B
C
D
E
F
proximal (-121 bp)
G
H
I
J
K
A B C D E F G
dppD(wt)
A B C D E F G
dppD(mut)
B C D E F G
dppD(ΔA)
C D E F G
dppD(ΔAB)
D E F G
dppD(ΔABC)
distal (-846 bp)
L
M
N
O
proximal (-121 bp)
P
Q
R
S
E F G
dppD(ΔABCD)
A B C E F G
dppD(Dns)
A B C D F G
dppD(Ens)
A B C D E
dppD(ΔFG)

Figure 1.2 We analyzed *dppD* enhancer activity at both promoter proximal (-121 bp) and promoter distal (-846 bp) positions (A). The LacZ expression driven by the wildtype *dppD* is not significantly different at the two distances (B and G). We next mutated the En and Ci binding sites and observed both anterior and posterior derepression that is significantly increased in the promoter proximal position (C and H). In the context of no En or Ci input, we then mutated or deleted the other regions of *dppD.* At -846 bp mutations of regions D, E, and F/G completely abolish enhancer activity (L – O). Loss of regions A, B, and C sequentially decrease gene expression with no expression observed when A, B, C, and D are removed together (D – F). At -121 bp loss of no region abolished enhancer activity. Deletion of region A results in derepression in the wing pouch, hinge, and notum as well as in the D/V boundary (I). Subsequent deletion of region B results in loss of this derepression in the pouch, hinge, and notum but not in the D/V boundary (J). Loss of A, B, and C together results in loss of enhancer activity in the D/V boundary and in the A/P stripe region (K). Loss of A, B, C, and D together only slightly decrease gene expression (P). Mutation of regions D, E and F/G all decrease enhancer activity in the pouch and hinge while deletion of F/G also decreases activity in the notum (Q – S). Cyan is LacZ antibody staining and gray is DAPI.

distance from the transcription start site (TSS) to -846bp, which will be addressed below. Finally, loss of regions F and G together also result in loss of reporter gene expression (Figure 1.2 O)

1.3b dppD activity in the promoter proximal position

Having established the contribution of each region to *dppD* enhancer function, A, B, and C contribute to total levels of expression and D, E, and F/G are essential for enhancer activity, we proceeded to address the question of long-range gene regulation within the *dppD* enhancer, employing the same strategy we used to identify the RCE in the *sparkling* enhancer (Chapter 2). In this vein the *dppD*(wt) and *dppD*(mut) constructs were placed in the promoter proximal position of -121bp from the TSS (Figure 1.2A). Interestingly, the wildtype enhancer is not appreciatively affected by this change in distance; however, the reporter gene expression driven by *dpp* lacking the Ci and En sites [*dppD*(m)] is significantly increased in this position (Figure 1.2 G-H). This finding suggests that the Ci and En inputs are “strong enough” to overcome a distance from the promoter, but the other enhancer inputs are not. From this starting point, we moved the mutant enhancer constructs described above to the same promoter proximal position of -121bp from the TSS. The expression patterns driven by these enhancer constructs is then compared to *dppD*(mut) at both distances, looking for a sequence within the enhancer that is essential at the distal position and dispensable in the promoter proximal position.

Loss of region A at -121 bp has expanded reporter gene expression in the posterior compartment compared to *dppD*(mut) in the wing pouch, hinge, and notum, suggesting derepression due to loss of a repressor input beyond that of En and Ci^{REP}. LacZ expression also expands to cells in the dorsal/ventral (D/V) boundary which is absent in *dppD*(mut) (Figure 1.2 H,I). Deletion of regions A and B together results in loss of the posterior derepression seen with deletion of A alone, suggesting the presence of a binding site for an activator expressed in the wing pouch within the region B sequence. LacZ is still expressed in the cells at the D/V boundary (Figure 1.2 J). However, this reporter gene expression at the D/V boundary is lost upon deletion of region C with regions A and B. In fact loss of region C in combination with A and B at -121bp most resembles the *dppD*(mut)-121bp expression pattern, although with slightly less reporter gene activity overall (Figure 1.2K). Finally, the most significant change in expression was seen when region D was deleted along with A, B, and C. Here, we observed a decrease in expression in the anterior and posterior wing pouch, and a severe loss of expression in the notum and hinge regions (Figure 1.2 P). In *dppD*(mut) there is a small area of LacZ expression outside the wing pouch in the posterior part of the disc, just ventral to the hinge (Figure 1.2 H). Interestingly, this expression remains with loss of regions A, B, and C, but is lost upon enhancer truncation through region D. It is also notable that loss of region D results in failure of reporter gene activity in the cells immediately flanking the D/V boundary, suggesting *dpp*D region D contains an activator sequences for each of these cell types as well as in the pouch, hinge, and notum.

Recall that mutation of regions D and E in the distal position resulted in loss LacZ activity indicating sequences within these regions are essential for enhancer activity. It was therefore surprising that we observed reporter gene expression from the both *dppD*(mD^{ns}) and *dppD*(mE^{ns}) enhancers in the promoter proximal position (Figure 1.2 Q, R). Unlike *dppD*(mut)-121bp which has uniform levels of expression across the disc, mutation of D and E results in decreased expression in the wing pouch, while expression in a stripe at the A/P boundary and in the notum remain high, suggesting that the regions contain binding sites for activator proteins expressed in the wing pouch. As seen in *dppD*(ΔABCD)-121bp, *dppD*(mD^{ns})-121bp also diminished expression in the posterior margin close to the hinge; however we do not see disrupted expression in the cells flanking the D/V boundary suggesting the sequences responsible for this activity in region D are redundant with another part of this enhancer, most likely in A, B, or C (Figure 1.2 Q). Similar to *dppD*(mD^{ns})-121bp and *dpp*(mE^{ns})-121bp deletion of regions F and G in this position results in an even greater decrease in LacZ expression in the pouch, while activity in the A/P stripe remains (Figure 1.2 S). Regions F and G must also contain information for activation in the notum as we only see reporter gene expression at the A/P boundary in this region of the disc. The presence of a stripe of LacZ expression at the A/P boundary when D, E, or F/G are lost, suggests the existence of an activator binding site for expression EITHER redundantly in these regions, OR an activator site in regions regions A-C. Indeed, upon careful examination of many insertion sites with varying exposure levels, a stripe can be seen in *dppD*(mut)-121bp, *dppD*(ΔA)-121bp, and *dppD*(ΔAB)-121bp, above the significant levels of pouch expression, which is then lost in *dpp*(ΔABC) suggesting such an activation sequence lies within region C.

1.3c Motif analysis identifies potential dppD interacting proteins

Our laboratory's work with the *dppD* enhancer has demonstrated that each 100bp region (A-G) contains regulator information. As such, we analyzed the enhancer for additional protein binding sites. We looked both for binding sites of interest using genepallete (Rebeiz and Posakony, 2004) and took an unbiased approach to motive analysis using evoprinter (Odenwald et al., 2005). Potential protein interaction partners were identified with the help of TomTom (Gupta et al., 2007) and the list of candidates was subsequently refined based on known expression patters and the contributions of each region (A-G) to enhancer function. Interestingly, the majority of the putative binding sites we identified are

Figure 1.3

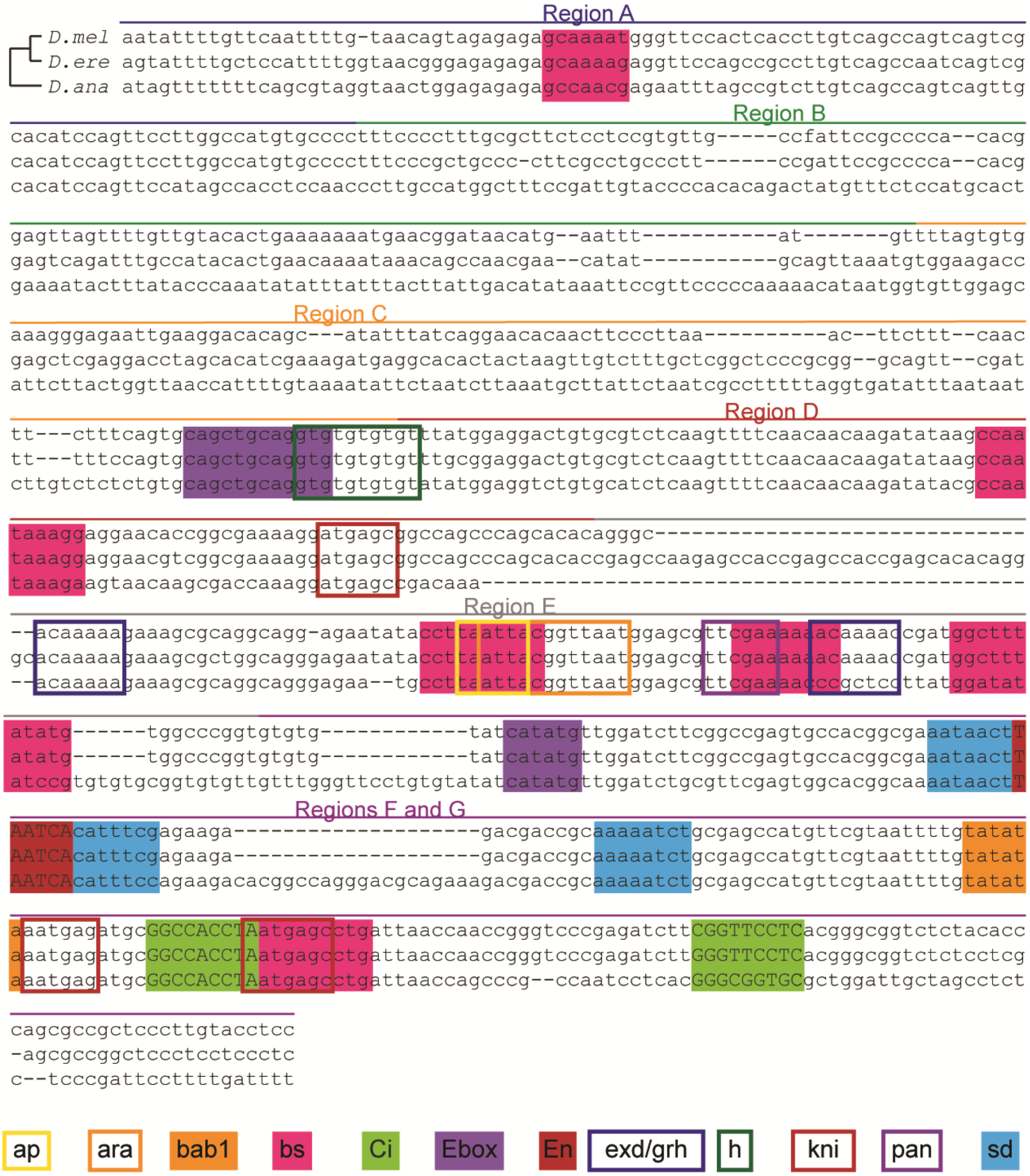

Region A
D.mel aatattttgttcaattttg-taacagtagagagagcaaaatgggttccactcaccttgtcagccagtcagtcg
D.ere agtattttgctccattttggtaacgggagagagagcaaaagaggttccagccgccttgtcagccaatcagtcg
D.ana atagtttttttcagcgtaggtaactggagagagagccaacgagaatttagccgtcttgtcagccagtcagttg
Region B
cacatccagttccttggccatgtgcccctttcccctttgcgcttctcctccgtgttg-----ccfattccgcccca--cacg
cacatccagttccttggccatgtgcccctttcccgctgccc-cttcgcctgccctt------ccgattccgcccca--cacg
cacatccagttccatagccacctccaacccttgccatggctttccgattgtaccccacacagactatgtttctccatgcact
gagttagttttgttgtacactgaaaaaaatgaacggataacatg--aattt-----------at-------gttttagtgtg
gagtcagatttgccatacactgaacaaaataaacagccaacgaa--catat-----------gcagttaaatgtggaagacc
gaaaatactttatacccaaatatatttatttacttattgacatataaattccgttcccccaaaaacataatggtgttggagc
Region C
aaagggagaattgaaggacacagc---atatttatcaggaacacaacttcccttaa----------ac--ttcttt--caac
gagctcgaggacctagcacatcgaaagatgaggcacactactaagttgtctttgctcggctcccgcgg--gcagtt--cgat
attcttactggttaaccattttgtaaaatattctaatcttaaatgcttattctaatcgccttttttaggtgatatttaataat
Region D
tt---ctttcagtgcagctgcaggtgtgtgtgtttatggaggactgtgcgtctcaagttttcaacaacaagatataagccaa
tt---tttccagtgcagctgcaggtgtgtgtgtttgcggaggactgtgcgtctcaagttttcaacaacaagatataagccaa
cttgtctctctgtgcagctgcaggtgtgtgtgtatatggaggtctgtgcatctcaagttttcaacaacaagatatacgccaa
taaaggaggaacaccggcgaaaaggatgagcggccagcccagcacacagggc--------------------------------
taaaggaggaacgtcggcgaaaaggatgagcggccagcccagcacaccgagccaagagccaccgagccaccgagcacacagg
taaagaagtaacaagcgaccaaaggatgagccgacaaa--
Region E
--acaaaaagaaagcgcaggcagg-agaatataccttaattacggttaatggagcgttcgaaaaaacaaaaacgatggcttt
gcacaaaaagaaagcgctggcagggagaatataccttaattacggttaatggagcgttcgaaaaaacaaaaacgatggcttt
--acaaaaagaaagcgcaggcagggagaa--tgccttaattacggttaatggagcgttcgaaaaccgctccttatggatat
atatg------tggcccggtgtgtg------------tatcatatgttggatcttcggccgagtgccacggcgaaataactT
atatg------tggcccggtgtgtg------------tatcatatgttggatcttcggccgagtgccacggcgaaataactT
atccgtgtgtgcggtgtgttgtttgggttcctgtgtatatcatatgttggatctgcgttcgagtggcacggcaaaataactT
Regions F and G
AATCAcatttcgagaaga--------------------gacgaccgcaaaaatctgcgagccatgttcgtaattttgtatat
AATCAcatttcgagaaga--------------------gacgaccgcaaaaatctgcgagccatgttcgtaattttgtatat
AATCAcatttccagaagacacggccagggacgcagaaagacgaccgcaaaaatctgcgagccatgttcgtaattttgtatat
aaatgagatgcGGCCACCTAatgagcctgattaaccaaccgggtcccgagatcttCGGTTCCTCacgggcggtctctacacc
aaatgagatgcGGCCACCTAatgagcctgattaaccaaccgggtcccgagatcttGGGTTCCTCacgggcggtctctcctcg
aaatgagatgcGGCCACCTAatgagcctgattaaccagcccg--ccaatcctcacGGGCGGTGCgctggattgctagcctct
cagcgccgctcccttgtacctcc
-agcgccggctccctcctccctc
c--tcccgattccttttgatttt
ap
ara
bab1
bs
Ci
Ebox
En
exd/grh
h
kni
pan
sd

Figure 1.3 Evolutionary comparison of *dppD* enhancer sequence. Using candidate protein binding sites and evoprinter motif analysis we identified several potential protein binding partners for *dppD*. The *D. melanagaster* (*D.mel*), *D.erecta* (*D.ere*), and *D.ananasse* (*D.ana*) sequences are shown with the regions A – G highlighted above the sequence. Known DNA binding sites for Cubitis interuptis (Ci) and Engrailed (En) are depicted in solid green and red boxes respectively. Other potential binding sites include apterous (ap), araucan (ara), bric á brac (bab1), blistered (bs), Ebox domain, extradenticle (ext), grainyhead (grh), hairy (h), knirps (kni), pangolin (pan), and scalloped (sd).

located in regions E, F, and G of the enhancer. Using these methods we identified putative binding sites for Ebox binding proteins (Figure 1.3). This class of transcription factors has been shown to interact with DNA and stimulate transcription (Chaudhary and Skinner, 1999). We also identified binding sites for numerous proteins that are expressed in the wing pouch including knirps (kni), bric á brac (bab), blistered (bs), apterous (ap), araucan (ara) (Fristrom et al., 1994; Gomez-Skarmeta et al., 1996; Klein and Arias, 1998b; Lunde et al., 2003). ara and ap are also expressed in the hinge region of the wing imaginal disc. Furthermore, we identified a binding site for the downstream affector of the Wingless (Wg) signaling pathway, pangolin (pan) (Prasad et al., 2003). We also found binding sites for the more ubiquitously expressed proteins grainyhead (grh), extradenticle (exd), scalloped (sd), and hairy (h) (Carroll, 1989; Gonzalez-Crespo and Morata, 1995; Uv et al., 1997).

1.4 Discussion

In this study we have performed a promoter proximal and distal analysis of the *dppD* enhancer, the regulatory sequence responsible for directing *dpp* expression in a stripe along the A/P boundary in the *Drosophia* third instar wing imaginal disc. It is important to acknowledge the weaknesses and limitations of this analysis. First and foremost, we based our enhancer mutations on an analysis in which the *dppD* enhancer was placed immediately adjacent to the hsp70 promoter driving LacZ expression (Muller and Basler, 2000). However, we performed our initial analysis of *dppD* with the enhancer placed at a moderate distance from the promoter (846bp upstream of the TSS). While Müller and Basler found that regions A-D did not individually affect *dppD* activity (Muller and Basler, 2000), we found that region D is essential for enhancer activity in the promoter proximal position (Figure 1.2). Based on these contradictory results we cannot be certain of the affect of loss of A-C individually on the distally placed enhancer's activity. Furthermore, the massive derepression we observed when the Ci and En inputs were lost in the promoter proximal position made it difficult to assess the contribution of regions A-C to enhancer activity. In order to elucidate the role of these regions in *dppD* activity, mutations to these regions need to be made individually, and in the context of wildtype and En input. Additionally, in the promoter proximal position, we observed a dramatic decrease in gene expression when regions F and G are lost. Unfortunately this mutation spans 200bp, which contain numerous putative binding sites (Figure 1.3). Smaller mutations are necessary to further analyze the action of these enhancer regions.

1.4a The affect of promoter proximity on dppD enhancer activity

The nature of our results, and the inherent limitations of this work make it difficult analyze long-range gene regulation by the *dppD* enhancer. We were surprised to find that the wildtype *dppD* enhancer is relatively unaffected by change in enhancer position with regards to the hsp70 promoter. Conversely, *dppD*(mut) is extraordinarily affected by the move to a more promoter proximal position. At -846bp, the *dppD*(mut) enhancer drives weak expression that is derepressed in the anterior and posterior wing pouch; whereas at -121bp *dppD*(mut) drives surprisingly strong expression that extends further towards wing margins and into the hinge and notum regions of the wing disc that the distal enhancer (Figure 1.2). These results indicate that the Ci and En sites contribute strongly to enhancer activation and repression compared to the other inputs into the enhancer. Furthermore, the Ci and En binding sites, either through their binding or another overlapping site, seem to play a role in the long-range activity of the enhancer, as we see low levels of expression when the *dppD*(mut) enhancer is placed distally, but high levels of expression occur at the proximal position. Alternatively, the Ci and En activation and repression inputs are strong enough to overcome the 846 spacer and activation transcription, and moving the enhancer father away from the promoter would be a better test of long range gene regulation for *dppD.*

The contradiction between the Müller and Basler results, and our results may indicate a role for region D in long-range gene activity as well. The authors previously demonstrated that DELETION of region D did not have any effect on reporter gene transcription at the promoter proximal position (Muller and Basler, 2000). Conversely, in our hands MUTATION of region D abolished gene expression at the promoter distal location (Figure 1.2). These results would suggest that region D is not necessary in the promoter proximal position, but is critical in the promoter distal position; however, when we mutate region D in the promoter proximal position we find that it is required for a *dppD*(mut) expression pattern. The experimental conditions between the Müller and Basler reporter constructs and ours are too different to make any conclusions from these results pertaining to distal gene regulation. Nevertheless, region D warrants further experimental analysis in the search for remote control element within the *dppD* enhancer.

Loss of region C in conjunction with regions A and B most resembles promoter proximal expression of *dppD*(mut). *dppD*(ΔA)-121bp and *dppD*(ΔAB)-121bp both drive LacZ expression at higher levels that *dppD*(mut)-121bp. *dppD*(ΔABC)-121bp however, drives lower levels of expression, which are only slightly diminished compared to those regulated by *dppD*(mut)-121bp (Figure 1.2). It is possible that this large, 100bp mutation, removes both a long-range enhancer element and a weak activating sequence. As with region D, *dppD* region C warrants follow-up experiments regarding the enhancer's ability to regulate distal gene regulation.

1.4b The contribution of each enhancer region to dppD function.

Our, and others, observations indicate that Ci and En are the primary inputs into the *dppD* enhancer. However, numerous reporter constructs have demonstrated that these are not the only inputs required for *dppD* activity. Loss of the Ci and En sites at the 3' end of the enhancer results in derepression of enhancer activity in the wing disc, hinge, and notum, clearly demonstrating that additional inputs can regulate this enhancer. The extent of expression in these constructs indicates that at least some of these inputs are broadly expressed, and that it is Ci and En act to position the stripe of expression. The increase in expression seen when our enhancer constructs are placed in the promoter proximal position can help determine the location of these inputs within the enhancer.

For example, loss of region A results in derepression in the pouch, hinge, and notum such that expression reaches the very edges of the disc. This observation suggests that region A has at least one binding site for a transcriptional repressor. Müller and Basler's work demonstrated that at least one transcriptional repressor lies outside the En binding site in region F, and we have shown that repressor input is likely in region A (Muller and Basler, 2000). We identified a putative *blistered* binding site within this region. *Blistered,* or serum response factor (*dsrf*). Interestingly, *bs* is expressed in the wing pouch, and has been shown to act as a transcriptional repressor of *rhomboid*. Accordingly, loss of blistered protein results in an expansion of *rhomboid* expression (Fristrom et al., 1994). Loss of region A also results in an expansion of LacZ expression into the cells at the D/V boundary, where the enhancer is normally inactive. *Wingless* is one of the signaling molecules expressed in these cells (Prasad et al., 2003). We only identified a single pangolin (dTCF) binding site in region E; however, we can rule out the role of a Wingless target in this expression.

Our results suggest that region B contains at least one binding site for a transcriptional activator that is expressed in the wing pouch. Our motif analysis did not identify any putative binding sites in this region. Similarly, region C likely contains at least one binding site for a transcriptional activator expressed in the wing pouch, hinge, and notum. Furthermore, this sequence contains information necessary to promote enhancer activity in a stripe at the A/P boundary and at the very posterior margin of the wing pouch. Interestingly, we identified a putative hairy binding site in region C. The expression pattern of hairy in the third instar imaginal disc can account for all of these locations of *dppD* activity (Carroll, 1989*).* As such, the ability of hairy to regulate the *dppD* enhancer is specifically interesting for further investigation can be further assessed using electron mobility shift assays, targeted enhancer mutations, and analysis of *dpp* expression with loss if *hairy* expression.

Loss of regions D, E, and F/G all result in a decrease in reporter gene activity specifically in the wing pouch and notum, but not dramatically in the A/P boundary stripe. Therefore, it is likely that each of these regions contains at least one binding site for a transcriptional activator expressed broadly. Indeed we identified a putative binding site for both extradenticle and grainyhead in region E. As both of these proteins are expressed ubiquitously in the wing disc, loss of either input could explain the decreased expression seen with loss of region E (Gonzalez-Crespo and Morata, 1995; Uv et al., 1997). Region E also contains a binding site for the downstream target of *Dpp, araucan.* As this protein is expressed in the wing pouch, it could act in a feedback manner to regulation the *dppD* enhancer (Gomez-Skarmeta et al., 1996). Region E also contains a binding site for the dorsal pouch protein apterous (Klein and Arias, 1998a). *dppD* region D contains a binding site for knirps, which is expressed and varying levels throughout the wing pouch *(Lunde et al., 2003).* Like region E, regions F/G contains a binding site for a broadly expressed transcriptional activator, scalloped, as well as a putative binding site for bric á brac which is expressed in the wing pouch. (Campbell et al., 1992).

Each of these putative protein interactions requires further investigation using electron mobility shift assays, targeted enhancer mutations, and analysis of *dpp* expression with mutant or RNAi mediated gene knockdown. Furthermore, additional precisely designed experiments are required to study long-range enhancer elements in the *dppD* enhancer. Our observations regarding the RCE in *sparking* indicated that we may have to look outside the minimal *dppD* enhancer in order to identify the DNA elements responsible for facilitating distal gene regulation in the *dpp* locus.

1.5 Experimental methods

1.5a Generation of enhancer constructs and transgensis

The wildtype *dppD* enhancer was amplified from w^{1118} genomic DNA. Enhancer mutations were produced by standard PCR to generate truncation mutants, or assembly PCR to generate enhancers with internal mutations. All enhancers were tagged with a 5’ EcoRI and 3’ XhoI restriction sites during PCR amplification. Enhancer constructs were subsequently TOPO-cloned into the pENTR/D-Topo vector (Invitrogen) and then Gateway-cloned into the Ganesh-Z1 LacZ reporter vector (Swanson et al., 2008) via LR recombination (Invitrogen). To generate the *dppD* enhancer fragments with a 121bp spacer, the wildtype enhancer was again amplified from w^{1118} genomic DNA and *dppD*(mut) was re-generated by assembly PCR. These enhancers were then TOPO and Gateway cloned as described above, except into the final reporter vector Ganesh-Z2, which lacks the 0.7kb spacer present in Ganesh-Z1 (Swanson et al., 2008). The additional enhancer mutations were then constructed by swapping the *dppD*(wt) enhancer in Ganesh-Z2 with the enhancers from Ganesh-Z1 after EcoRI and XhoI restriction enzyme digest and standard ligation methods. P-element transformation was performed using *w*^{1118} flies as described previously (Rubin and Spradling, 1982).

1.5b Tissue preparation, immunohistochemistry, and microscopy

Wing disc tissues were dissected from third instar larvae. Disc tissues were then fixed in 4% paraformaldehyde for 30 minutes at room temperature and then washed 3x10 minutes with PBS-Tx (1xPBS+0.1% TritonX-100). Fixed wing discs were then incubated in PBX-Tx+2%BSA for 1-3 hours and then incubated overnight in with primary antibody staining against LacZ (*Drosophila* Studies Hydrodoma Bank) diluted 1:100. After washing 3x10 minutes with PBS-Tx, secondary antibody staining was performed for two hours at room temperature with rocking using goat anti-mouse 568nm secondary antibody (Invitrogen) diluted 1:1000. After staining, wing discs were washed 3x20 minutes with PBS-Tx and mounted in ProLong Gold with 4', 6'-diamidino-2 phenylidole (DAPI) (Invitrogen). LacZ antibody staining was then visualized and imaged using an Olympus BX5I microscope and an Olympus DP70 digital camera.

1.5c DNA motif analysis

Potential transcription factor binding sites for candidate proteins were identified using genepallete (Rebeiz and Posakony, 2004). Novel DNA motifs were discovered using evoprinterHD to first identify clusters of conserved sequences within the *dppD* enhancer (Odenwald et al., 2005). This was followed using *cis*-decoder to determine which sequences from these conserved clusters are likely to be transcription factor binding sites (Brody et al., 2007). TomTom from the Meme Suite was then used to find transcription factors whose binding sites resemble the motifs these programs identified (Gupta et al., 2007).

1.6 Acknowledgments

We would like to thank Paulette Florence, Dave Parker, and Andrea Ramos for contributions to the cloning and generation of transgenic flies and Amy Strom for her help with disc dissection and staining.

1.7 References

Affolter, M., and Basler, K. (2007). The Decapentaplegic morphogen gradient: from pattern formation to growth regulation. Nat Rev Genet *8*, 663-674.

Aza-Blanc, P., Ramirez-Weber, F.A., Laget, M.P., Schwartz, C., and Kornberg, T.B. (1997). Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. Cell *89*, 1043-1053.

Blackman, R.K., Grimaila, R., Koehler, M.M., and Gelbart, W.M. (1987). Mobilization of hobo elements residing within the decapentaplegic gene complex: suggestion of a new hybrid dysgenesis system in Drosophila melanogaster. Cell *49*, 497-505.

Brody, T., Rasband, W., Baler, K., Kuzin, A., Kundu, M., and Odenwald, W.F. (2007). cis-Decoder discovers constellations of conserved DNA sequences shared among tissue-specific enhancers. Genome Biol *8*, R75.
Campbell, G., and Tomlinson, A. (1999). Transducing the Dpp morphogen gradient in the wing of Drosophila: regulation of Dpp targets by brinker. Cell *96*, 553-562.
Campbell, S., Inamdar, M., Rodrigues, V., Raghavan, V., Palazzolo, M., and Chovnick, A. (1992). The scalloped gene encodes a novel, evolutionarily conserved transcription factor required for sensory organ differentiation in Drosophila. Genes Dev *6*, 367-379.
Carroll, S.B.W., J.S. (1989). The role of the hairy gene during Drosophila morphagenesis: stripes in imaginal discs. Genes and Development *3*, 905-916.
Chaudhary, J., and Skinner, M.K. (1999). Basic helix-loop-helix proteins can act at the E-box within the serum response element of the c-fos promoter to influence hormone-induced promoter activation in Sertoli cells. Mol Endocrinol *13*, 774-786.
Chen, C.H., von Kessler, D.P., Park, W., Wang, B., Ma, Y., and Beachy, P.A. (1999). Nuclear trafficking of Cubitus interruptus in the transcriptional regulation of Hedgehog target gene expression. Cell *98*, 305-316.
Entchev, E.V., Schwabedissen, A., and Gonzalez-Gaitan, M. (2000). Gradient formation of the TGF-beta homolog Dpp. Cell *103*, 981-991.
Fristrom, D., Gotwals, P., Eaton, S., Kornberg, T.B., Sturtevant, M., Bier, E., and Fristrom, J.W. (1994). Blistered: a gene required for vein/intervein formation in wings of Drosophila. Development *120*, 2661-2671.
Gomez-Skarmeta, J.L., Diez del Corral, R., de la Calle-Mustienes, E., Ferre-Marco, D., and Modolell, J. (1996). Araucan and caupolican, two members of the novel iroquois complex, encode homeoproteins that control proneural and vein-forming genes. Cell *85*, 95-105.
Gonzalez-Crespo, S., and Morata, G. (1995). Control of Drosophila adult pattern by extradenticle. Development *121*, 2117-2125.
Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol *8*, R24.
Kim, J., Johnson, K., Chen, H.J., Carroll, S., and Laughon, A. (1997). Drosophila Mad binds to DNA and directly mediates activation of vestigial by Decapentaplegic. Nature *388*, 304-308.
Klein, T., and Arias, A.M. (1998a). Different spatial and temporal interactions between Notch, wingless, and vestigial specify proximal and distal pattern elements of the wing in Drosophila. Dev Biol *194*, 196-212.
Klein, T., and Arias, A.M. (1998b). Interactions among Delta, Serrate and Fringe modulate Notch activity during Drosophila wing development. Development *125*, 2951-2962.
Lee, J.J., von Kessler, D.P., Parks, S., and Beachy, P.A. (1992). Secretion and localized transcription suggest a role in positional signaling for products of the segmentation gene hedgehog. Cell *71*, 33-50.

Lunde, K., Trimble, J.L., Guichard, A., Guss, K.A., Nauber, U., and Bier, E. (2003). Activation of the knirps locus links patterning to morphogenesis of the second wing vein in Drosophila. Development *130*, 235-248.

Marty, T., Muller, B., Basler, K., and Affolter, M. (2000). Schnurri mediates Dpp-dependent repression of brinker transcription. Nat Cell Biol *2*, 745-749.

Masucci, J.D., Miltenberger, R.J., and Hoffmann, F.M. (1990). Pattern-specific expression of the Drosophila decapentaplegic gene in imaginal disks is regulated by 3' cis-regulatory elements. Genes Dev *4*, 2011-2023.

Methot, N., and Basler, K. (1999). Hedgehog controls limb development by regulating the activities of distinct transcriptional activator and repressor forms of Cubitus interruptus. Cell *96*, 819-831.

Morata, G., and Lawrence, P.A. (1975). Control of compartment development by the engrailed gene in Drosophila. Nature *255*, 614-617.

Muller, B., and Basler, K. (2000). The repressor and activator forms of Cubitus interruptus control Hedgehog target genes through common generic gli-binding sites. Development *127*, 2999-3007.

Nellen, D., Burke, R., Struhl, G., and Basler, K. (1996). Direct and long-range action of a DPP morphogen gradient. Cell *85*, 357-368.

Odenwald, W.F., Rasband, W., Kuzin, A., and Brody, T. (2005). EVOPRINTER, a multigenomic comparative tool for rapid identification of functionally important DNA. Proc Natl Acad Sci U S A *102*, 14700-14705.

Oh, H., and Irvine, K.D. (2011). Cooperative regulation of growth by Yorkie and Mad through bantam. Dev Cell *20*, 109-122.

Prasad, M., Bajpai, R., and Shashidhara, L.S. (2003). Regulation of Wingless and Vestigial expression in wing and haltere discs of Drosophila. Development *130*, 1537-1547.

Rebeiz, M., and Posakony, J.W. (2004). GenePalette: a universal software tool for genome sequence visualization and analysis. Developmental Biology *271*, 431-438.

Ruberte, E., Marty, T., Nellen, D., Affolter, M., and Basler, K. (1995). An absolute requirement for both the type II and type I receptors, punt and thick veins, for dpp signaling in vivo. Cell *80*, 889-897.

Rubin, G.M., and Spradling, A.C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. Science *218*, 348-353.

Schwartz, C., Locke, J., Nishida, C., and Kornberg, T.B. (1995). Analysis of cubitus interruptus regulation in Drosophila embryos and imaginal disks. Development *121*, 1625-1635.

St Johnston, R.D., Hoffmann, F.M., Blackman, R.K., Segal, D., Grimaila, R., Padgett, R.W., Irick, H.A., and Gelbart, W.M. (1990). Molecular organization of the decapentaplegic gene in Drosophila melanogaster. Genes Dev *4*, 1114-1127.

Swanson, C.I., Hinrichs, T., Johnson, L.A., Zhao, Y., and Barolo, S. (2008). A directional recombination cloning system for restriction- and ligation-free construction of GFP, DsRed, and lacZ transgenic *Drosophila* reporters. Gene *408*, 180-186.

Tabata, T., and Kornberg, T.B. (1994). Hedgehog is a signaling protein with a key role in patterning Drosophila imaginal discs. Cell *76*, 89-102.

Uv, A.E., Harrison, E.J., and Bray, S.J. (1997). Tissue-specific splicing and functions of the Drosophila transcription factor Grainyhead. Mol Cell Biol *17*, 6727-6735.

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