Structure, Function, and Evolution of a Signal-Regulated Enhancer

by

Christina Ione Swanson

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
2010

Doctoral Committee:

Assistant Professor Scott E. Barolo, Chair
Professor J. Douglas Engel
Associate Professor Kenneth M. Cadigan
Associate Professor Billy Tsai
Assistant Professor Patricia J. Wittkopp

To my family, for your truly unconditional love and support.

And to Mike - the best thing that happened to me in grad school.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Almost 30 years ago, a fragment of DNA from Simian Virus 40 was shown to increase transcription of a linked rabbit β-globin gene by 200-fold (Banerji et al, 1981). This DNA element, termed an "enhancer" by the authors, could act at significant distances and independently of its orientation with respect to the promoter to drive augmented levels of transcription. Although the earliest identified enhancers were located within viral DNA, it did not take long for investigators to draw a connection between the enhancer function of the SV40 sequence and the transcription-boosting activities of cis-regulatory sequences located near endogenous genes in species as evolutionarily distant as sea urchins and humans (Khoury and Gruss, 1983). Thus began the study of enhancer elements and their role in transcription.

Enhancers were soon recognized as a critical component of transcriptional regulation in virtually every branch of life – bacteria, yeast, plants, invertebrates, and vertebrates (Guarente, 1998; Xu and Hoover, 2001; Orphanides and Reinberg, 2002). They were distinct from promoter elements in that their distance, orientation, and position relative to the transcription start site were flexible; furthermore, they were able to act on heterologous promoters to activate transcription (Khoury and Gruss, 1983). Eventually,

it became clear that enhancers do much more than simply stimulate transcription: they are capable of integrating signals from the cellular environment to control the tissue specificity, timing, and level of gene expression. Thus enhancer elements are central to our complete understanding of transcription and gene regulation in every biological system.

What do enhancers look like?

Eukaryotic enhancers are typically depicted as clusters of transcription factor binding sites located in cis to their target gene. As might be expected, studies of endogenous enhancers have revealed many variations on this basic theme. For example, although many enhancers are located in sequences 5' of the promoter, enhancers are also frequently located downstream or even within introns of the genes they regulate. They can be located very close to the promoter or tens of kilobases away. The overall size of enhancers varies as well. While most enhancers seem to range in size from several hundred base pairs to 1 kb, more extreme examples have been identified: an enhancer that drives stripes of *runt* expression in the *Drosophila* embryo is over 5 kb in size (Klingler et al, 1996), while the enhancer that drives testes-specific expression of *gonadal* spans only 53 bp (Schulz et al, 1990). Enhancers are often independent modules, in which each individual enhancer is responsible for generating a specific and unique expression pattern. The enhancers of the *even-skipped* gene are a premier example of modularity – individual enhancers are responsible for the generation of individual transverse stripes of expression in the embryo (Goto et al, 1989; Harding et al, 1989;

Small et al, 1992). However, sometimes regulatory information is dispersed over a broad region that is difficult to divide into independent modules, as in the cis-regulatory regions of the sea urchin *endo16* gene and the *Drosophila brinker* gene (Yuh and Davidson, 1996; Yao et al, 2008).

Mechanisms of enhancer function

Enhancers use multiple mechanisms to regulate transcription of their target gene. In general, regulation of transcription is accomplished by the direct DNA-binding proteins that recognize specific sequences within enhancer elements and the cofactors with which they interact. Enhancers can stimulate transcription by the direct recruitment of Pol II and the basal transcription machinery, promoting formation of the preinitiation complex (PIC) at the promoter (Szutorisz et al, 2005). They can also indirectly promote formation of the PIC via interactions with Mediator, a huge multi-unit complex that is involved in recruiting and assembling the PIC at the promoter, as well as promoting activation of transcription (Malik and Roeder, 2005; Wang et al, 2005). Enhancers also regulate transcription by influencing their local chromatin environment. Two different kinds of coregulators act to modify chromatin structure: ATP-dependent nucleosome remodeling complexes, such as those belonging to the SWI/SNF family; and enyzmes that covalently modify histones, such as histone acetyltransferases (HATs) and histone deacetylases (HDACs) (Narlikar et al, 2002; Orphanides and Reinberg, 2002). Enhancers recruit these coregulators to bring about changes in local chromatin structure that promote either activation or repression of transcription. Although chromatin

modifications and recruitment, assembly, and stimulation of the general transcription machinery are all essential for the regulation of transcription**,** it has not yet been demonstrated that these mechanisms are sufficient to activate transcription. There is growing evidence that enhancers act through additional mechanisms to regulate transcription.

For example, many enhancers are separated from their promoters by hundreds or even thousands of kilobases. The action of activating transcription at the promoter from a distance represents another important aspect of enhancer function. There are several proposed mechanisms for long-distance enhancer-promoter interactions (Figure 1.1): (1) chromatin looping between the enhancer and the promoter; (2) tracking of the transcription machinery from the enhancer to the promoter; (3) facilitated tracking, a hybrid of the looping and tracking models, in which the transcription machinery tracks along the chromatin towards the promoter, and the intervening DNA between the enhancer and the promoter is progressively looped out; and (4) linking, in which facilitator proteins are recruited in a spreading fashion from the enhancer towards the promoter (Li et al, 2006). There is good experimental evidence to support some of these models. For instance, the looping model requires physical contact between the enhancer and the promoter. Experimental techniques such as chromosome conformation capture (3C) allow us to detect these interactions *in vivo*. Using 3C, long-range physical interactions between promoters and cis-regulatory elements have been observed for several genes, including the mouse *TNF* gene and the chicken β-globin locus, providing *in vivo* evidence for looping (Tolhuls et al, 2002; Tsytsykova et al, 2007). The observation that non-coding RNAs can be transcribed between enhancers and promoters

Figure 1.1

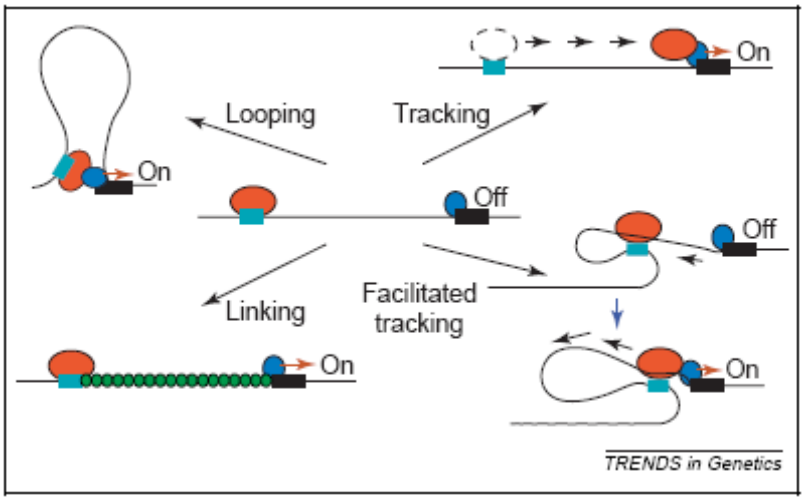


Figure 1.1. Proposed models for remote activation of gene expression by enhancer elements. In the looping model, the enhancer (blue rectangle) and the promoter (black rectangle) contact each other directly, perhaps mediated by enhancer and/or promoter binding proteins (circles), and the intervening chromatin is looped out. In the tracking model, the polymerase is recruited to the enhancer and tracks along the chromatin until it encounters the promoter. In the facilitated tracking model, the polymerase tracks as in the tracking model, and the intervening chromatin between the enhancer and the tracking polymerase is looped out. Finally, in the linking model, "linking" proteins are recruited initially at the enhancer and then spread out towards the promoter. Figure adapted from Li et al, 2006.

could be taken as evidence of a tracking mechanism; this phenomenon has been described at the human β-globin locus, as well as the *Drosophila* Antennapedia and Bithorax complexes (Gribnau et al, 2000; Bae et al, 2002; Calhoun and Levine, 2003). The tracking model would also predict that Pol II binding could be detected along the DNA between the enhancer and the promoter, an event which has been observed at the *prostate specific antigen* gene (Wang et al, 2005). Intriguingly, that study also observed direct contact between the enhancer and the promoter; taken together, these data support the facilitated tracking model of enhancer-promoter interaction. It remains unclear which mechanism most accurately describes the dynamics of distal enhancer-promoter interactions *in vivo*; perhaps there is no single "correct" model and in reality the mechanism differs from gene to gene.

Enhancer structure and organization

All enhancers are directly regulated by sequence-specific, DNA-binding transcription factors, and thus the composition and organization of transcription factor binding sites (TFBS) are a basic aspect of enhancer structure. The composition of TFBS within an enhancer often reflects the complexity of the expression pattern generated by that enhancer. In most cases, a single protein is not sufficient to specify the correct level, timing, and location of gene expression. For example, the *Drosophila even-skipped* stripe 2 enhancer (*eve* S2E), which drives expression in a single transverse stripe in the *Drosophila* embryo, contains thirteen mapped binding sites for five identified direct regulators (Bicoid, Hunchback, Giant, Kruppel, and Sloppy-paired 1) and is thought to be

regulated by at least one additional unidentified activator (Small et al, 1992; Andrioli et al, 2002). Each of these regulatory inputs is required to generate the correct level and pattern of gene expression.

The challenge of limiting expression to a particular time and place is particularly difficult for enhancers that are directly regulated by intercellular signaling pathways. Signaling pathways are active in many developmental contexts, but most enhancers do not respond universally to active signaling (Barolo and Posakony, 2002). For example, the *Drosophila prospero* enhancer, which activates gene expression in the R7 photoreceptors and cone cells of the eye, is directly regulated by EGFR signaling via direct binding of the transcription factors Pointed and Yan (Xu et al, 2000). While EGFR signaling is active in many other cell types during development, including other cell types in the eye, the *prospero* enhancer does not respond in those cells (Shilo, 2003; Voas and Rebay, 2004). The *prospero* enhancer is also directly regulated by Glass, Sine oculis, Lozenge, Seven-up, and the Notch effector Suppressor of Hairless (Hayashi et al, 2008). While each of these factors is expressed and active in cells other than R7 and cone cells, they act combinatorially on the *prospero* enhancer to direct R7- and cone cell-specific gene expression. Thus the *prospero* enhancer, like many developmentally regulated enhancers, requires a combination of activators and repressors, some signal-regulated and others locally expressed, to provide the necessary spatial and temporal information for proper gene expression. This concept – combinatorial control – is central to our understanding of enhancer function.

Spatial organization is another important aspect of enhancer structure. There are two current models of enhancer organization: the enhanceosome model and the

information display model. In the enhanceosome model, spatial organization of binding sites within the enhancer is critical for the formation of a DNA-protein complex, the enhanceosome, and stimulation of transcription (Carey, 1998). In these enhancers, changes in spacing between binding sites disrupts both DNA-protein interactions and protein-protein interactions, severely inhibiting transcriptional activation; cooperative binding and synergistic activation are key components of enhanceosome function. There are several classic examples of enhanceosome-style enhancers, including the *IFNβ* and *TCRα* enhancers, in which binding site spacing must obey strict organizational rules (Giese et al, 1995; Thanos and Maniatis, 1995). An opposing model of enhancer structure, the information display model, proposes that these enhanceosomes are the exception rather than the rule, and that spatial organization of TFBS is quite flexible in most enhancers (Arnosti and Kulkarni, 2005). In an information display enhancer, individual activators and repressors do not require cooperativity for either their DNA-binding or regulatory properties, and therefore their binding sites can be rearranged along the DNA without affecting enhancer function. This model also suggests that the transcriptional output generated by the enhancer is essentially the integrated sum of activating and repressive inputs binding to the enhancer at any given moment. The *eve* S2E is considered to be an information display enhancer. Mutated *eve* S2E constructs can be rescued by the addition of novel activator sites, suggesting that the location of activator binding sites within the enhancer is flexible (Arnosti et al, 1996). Evolutionary comparisons also suggest flexibility in TFBS position within *eve* S2E. Although the function of the *eve* S2E is conserved throughout Drosophilids and even Sepsids, alignments reveal that the arrangement of TFBS within these enhancers varies

considerably, in support of organizational flexibility (Ludwig et al, 1998; Hare et al, 2008). Furthermore, the level of gene expression generated by *eve* S2E can be manipulated *in vivo* by changing the affinity of activator binding sites or altering activator concentration, which supports the model's prediction that transcriptional output represents the sum of individual regulatory inputs (Arnosti et al, 1996).

Both the enhanceosome and information display models appear to accurately describe several characterized enhancer elements. However, in many cases the overall structure of the enhancer is not as strict as an enhanceosome, but not entirely flexible either. For example, in synthetic constructs containing binding sites for the repressor Giant and the activator Gal4, binding site organization dramatically affected the ability of Giant to repress transcriptional activation (Kulkarni and Arnosti, 2005). Although these constructs were unlikely to be behaving as enhanceosomes, organization was clearly important for interactions between Giant and Gal4, suggesting that short-range repression is sensitive to changes in spacing and organization of repressor and activator sites. In addition, many activators interact synergistically to activate transcription, implying that binding sites for those proteins would be sensitive to spatial organization as well; for example, synergistic activation of proneural gene expression by Suppressor of Hairless and Daughterless requires certain binding site configurations in *Drosophila* (Cave et al, 2005). These local interactions can be thought of as subelements within the overall enhancer structure. Refinement of the information display model to include subelements as well as individual TFBS can accurately explain the nature of many enhancers.

However, there is still uncertainty in assigning enhancers to either one model or the other. For example, a set of neuroectoderm enhancers in *Drosophila* share a common

set of regulatory motifs with characteristic spacing and organization (Erives and Levine, 2004). Furthermore, these structural characteristics appear to be conserved from *Drosophila* to *Anopheles,* indicating that the sequences, positions, and orientations of these motifs are inflexible. While the DNA-binding proteins regulating these enhancers interact synergistically, it is not clear that an enhanceosome-type complex assembles on these enhancers (Shirokawa and Courey, 1997; Bhaskar and Courey, 2002). Clearly there is significant restraint on the organization of these enhancers, but can they truly be labeled as enhanceosomes? At this point, our knowledge of enhancer structure is incomplete, making it difficult to know how to classify enhancers that fall on the spectrum between enhanceosome and information display. More examples and further experiments are required before these questions can be answered.

Unanswered questions in the field

We've come a long way towards understanding enhancer elements and their role in transcriptional regulation in the last 30 years. Yet many questions remain: How does the enhancer interact with the promoter at great distances? What is the function of noncoding transcription arising from the enhancer? How are enhancers involved in subnuclear localization of chromatin? How do enhancer sequences evolve? These are all big questions that are currently being addressed. More surprisingly, our grasp of even the most basic aspects of enhancer function is shaky: What are the basic, necessary components required to build a functional enhancer? What are the rules of enhancer organization?

As we understand it, transcription factor binding sites are the critical components of enhancers; TFBS recruit transcription factors, which go on to recruit cofactors that stimulate transcription through a variety of mechanisms. This seems very straightforward, and if true, we should be able to build synthetic versions of enhancers composed of only their TFBS. Unfortunately, this type of *in vivo* proof-of-principle experiment is rarely attempted for endogenous enhancers. In the few reported cases where this concept has been experimentally tested *in vivo* – for example, the well-characterized *Drosophila enhancer of split m4* proneural and *short gastrulation* neuroectoderm enhancers – the synthetic enhancers fail to recapitulate endogenous expression patterns (Johnson et al, 2008; Liberman and Stathopoulos, 2009). Of course, there are many reasons why these experiments might fail; these synthetic constructs may lack critical binding sites that have not yet been identified, or the synthetic constructs violate important spatial or organizational rules. However, our incomplete understanding of the basic composition and structure of even the best studied enhancers points out significant gaps in our knowledge.

Currently, no single enhancer has truly been completely characterized. Even the *eve* stripe 2 enhancer remains unsolved; deletions of any of the sequences that do not contain mapped binding sites significantly reduces enhancer function, and it is unknown whether those sequences are important because they harbor novel binding sites or because they preserve the spacing between adjacent sites (Andrioli et al, 2002). Without an enhancer in which all the essential regulatory sequences are identified, we are unable to proceed to the next steps of experimentation. Thus important questions remain unanswered, and we are hampered in our attempts to identify cis-regulatory elements *in*

silico, predict enhancer function based on TFBS composition and organization, or build enhancers that drive a desired pattern of gene expression.

The objective of this thesis research was to perform an in-depth, *in vivo* structure-function analysis of an endogenous enhancer. Our goal was to map all the critical regulatory sequences within a single enhancer, and to begin to understand the structural rules that govern enhancer function. My approaches included *in vivo* mutagenesis of enhancer sequences, construction of synthetic enhancers, evolutionary analysis of enhancer sequence and structure, and biochemical identification of novel enhancer-binding proteins, all using a signal-regulated enhancer from *Drosophila* as our model for an endogenous, tissue-specific enhancer.

The *D-Pax2 sparkling* enhancer

We selected the *D-Pax2 sparkling* enhancer for our studies. The *sparkling* (*spa*) enhancer regulates expression of *D-Pax2* in the *Drosophila* eye, an excellent system for studying intercellular signaling and cell fate specification. Furthermore, previous characterization of *spa* provided a starting point for continued investigation.

The *Drosophila* eye is composed of approximately 750 repeating units called ommatidia (Figure 1.2A, Voas and Rebay, 2004). Each ommatidium contains 8 photoreceptors (R1-8), 4 cone cells, 2 primary pigment cells, 6 secondary pigment cells, 3 tertiary pigment cells, and 3 bristle cells (Figure 1.2C). Differentiation of the eye field begins in the middle of the third larval instar stage in the eye imaginal disc, and is

Figure 1.2

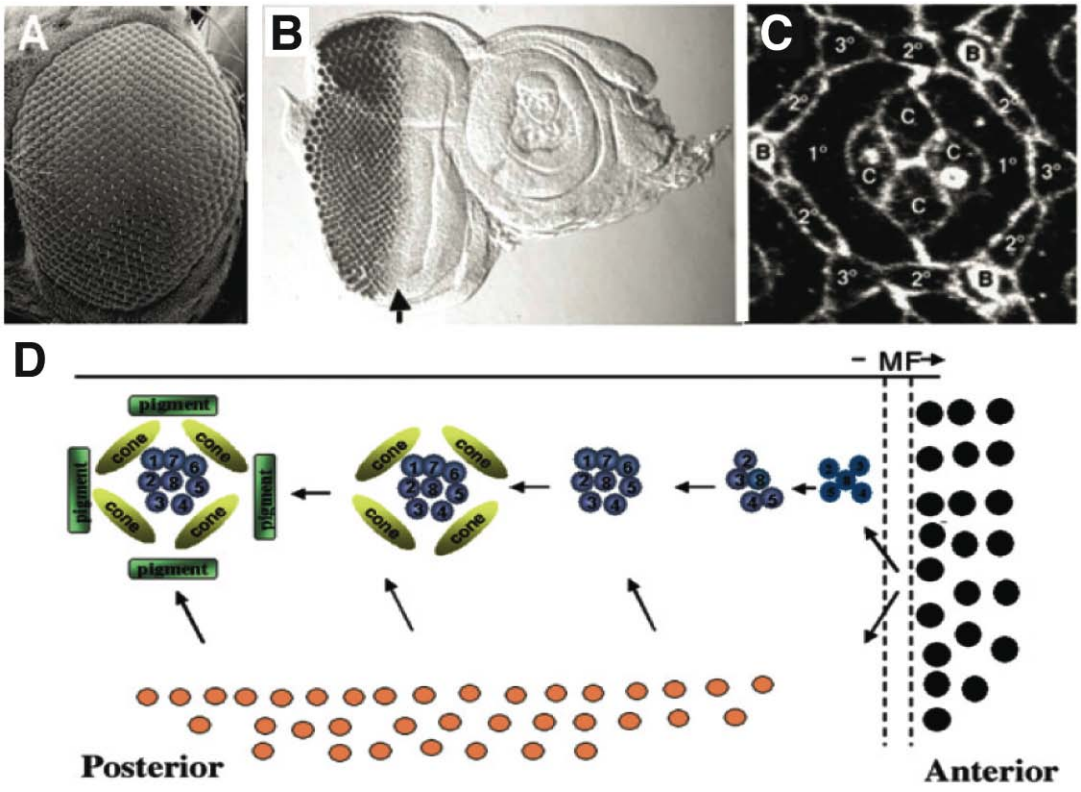


Figure 1.2. *Drosophila* eye development. (A) SEM image of a *Drosophila* eye, which is composed of repeating honeycomb-shaped units called ommatidia. (B) Third instar larval eye-antennal imaginal disc. The arrow marks the position of the morphogenetic furrow, which sweeps across the eye disc from the posterior (left) to the anterior (right). The cells posterior to the furrow are undergoing differentiation; here, those cells that have been specified as photoreceptors are stained. (C) A single pupal ommatidium, with cell types labeled: C = cone cells, 1° = primary pigment cells, 2° = secondary pigment cells, 3° = tertiary pigment cells, B = bristle cells. At this stage, photoreceptors are located in a different plane of focus, and therefore cannot be seen in this image. (D) Diagram of the sequential recruitment of cells into ommatidia behind the morphogenetic furrow (MF), which is progressing from left to right in this figure. Cells are recruited from a pool of undifferentiated precursors (orange cells at bottom) in a stereotypical order: photoreceptors R8, R2, R3, R4, and R5 are recruited first, followed by R1, R6, and R7, then the cone cells, and finally the pigment cells. Adapted from Nagaraj and Banerjee, 2004; and Voas and Rebay, 2004.

initiated by Decapentaplegic (Dpp) and Hedgehog (Hh) signaling at the morphogenetic furrow (Figure 1.2B, Silver and Rebay, 2005). As the morphogenetic furrow passes along the eye disc from the posterior to the anterior, the cells posterior to the furrow are sequentially recruited into ommatidia (Figure 1.2B, D). The order of specification and recruitment is fixed (Figure 1.2D). The R8 cells are specified first by Notch-mediated lateral inhibition, and they become the founder cells for each ommatidium. Specification of R8 is followed by recruitment and specification of R2 and R5, R3 and R4, R1 and R6, R7, the cone cells, and finally the primary, secondary, and tertiary pigment cells. The sequential specification of these cell types is regulated by the EGFR and Notch signaling pathways. Interestingly, both EGFR and Notch are required for the specification of multiple cell types during eye differentiation, and thus these two signals alone are not sufficient to specify unique cell fates in the eye. EGFR and Notch must act combinatorially with locally expressed activators and repressors to activate expression of the target genes that determine cell fate.

The *D-Pax2* gene is one of the EGFR/Notch target genes that is responsible for determining cell fate in the eye; its expression in cone cells is necessary for their proper fate specification (Fu and Noll, 1997). Cone cell-specific expression is controlled by an enhancer located in the fourth intron of the *D-Pax2* gene, the *sparkling* (*spa*) enhancer (Figure 1.3, Fu et al, 1998). Further characterization of *spa* identified a 362 bp minimal element sufficient for cone cell-specific gene expression, and furthermore mapped the location of twelve TFBS within this minimal enhancer (Flores et al, 2000). The twelve TFBS included binding sites for three distinct regulatory inputs (Figure 1.3C). First, *spa* is directly regulated by Lozenge (Lz), a RUNX-family activator that is expressed in all

Figure 1.3

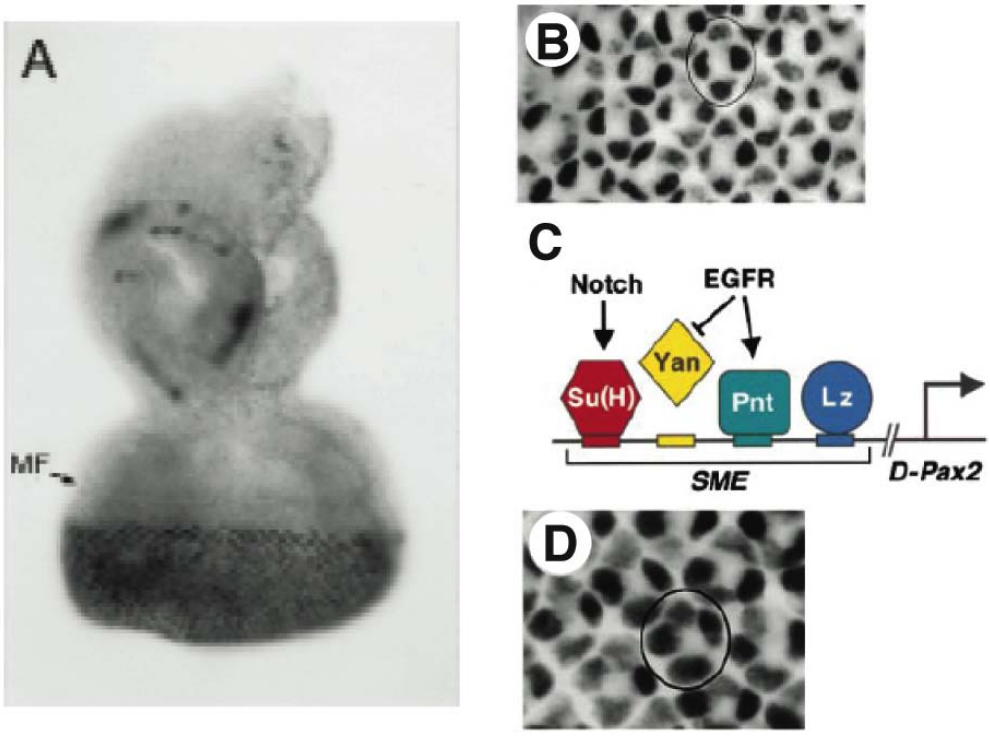


Figure 1.3. The *D-Pax2 sparkling* enhancer. (A-B) D-Pax2 is expressed in cone cells of third instar larval eye imaginal discs. (A) A third instar larval eye-antennal imaginal disc, oriented with posterior to the bottom, is stained with anti-D-Pax2 antibody; MF = morphogenetic furrow. (B) Higher magnification image shows that all four cone cells in each ommatidium express D-Pax2; circle outlines a single ommatidium. (C) Model of the *sparkling* enhancer, which regulates expression of *D-Pax2* in the cone cells. *Sparkling* is directly regulated by Notch signaling, EGFR signaling, and Lozenge. SME = *sparkling* minimal enhancer. (D) The SME, placed upstream of a LacZ reporter, is sufficient for cone cell-specific gene expression. Adapted from Fu and Noll, 1997; and Flores et al, 2000.

undifferentiated progenitor cells posterior of the morphogenetic furrow (Flores et al, 1998). Second, *spa* contains binding sites for the transcriptional effector of the Notch signaling pathway, Suppressor of Hairless (Su(H)). Su(H) acts as a transcriptional switch; in the absence of active Notch signaling, it mediates repression of target gene expression, but in the presence of active signaling, it functions as a transcriptional activator (Bray, 2006). Finally, *spa* also contains Ets binding sites, which are directly bound by both PointedP2 (Pnt) and Yan, effectors of the EGFR/MAPK signaling pathway. Pnt and Yan bind to the same sites, but Pnt mediates activation through those sites while Yan mediates repression. In the absence of EGFR signaling, Yan represses target genes, while in the presence of active EGFR signaling, Pnt is phosphorylated and activates target genes (Brunner et al, 1994; O'Neill et al, 1994; Rebay and Rubin, 1995). Both genetic manipulations and *in vivo* binding site mutations demonstrated that *spa* requires each of these regulatory inputs for proper regulation of *D-Pax2* expression.

Theoretically, it was possible that the characterized regulatory inputs, acting combinatorially, might be sufficient to dictate cone cell-specific activation of the *spa* enhancer. Furthermore, the additive effects of twelve activators bound to *spa* in putative cone cells could provide sufficient activation to generate proper levels of gene expression. Because knowledge of the regulation and TFBS composition of *spa* was potentially complete, we thought the *spa* enhancer would be a perfect candidate for this thesis project. Therefore, in my thesis, I used the *D-Pax2 sparkling* enhancer as a model to ask questions about the relationship between structure and function in endogenous enhancers.

CHAPTER II

STRUCTURAL RULES AND COMPLEX REGULATORY CIRCUITRY CONSTRAIN EXPRESSION OF A NOTCH- AND EGFR-REGULATED EYE ENHANCER

INTRODUCTION

Enhancers, or cis-regulatory elements, are the primary determinants of spatiotemporal patterns of gene expression. In order to properly regulate their target genes, enhancers must perform a number of functions, such as identifying and communicating with the promoter, sometimes over great distances, and triggering transcription in certain cells, but not in others. Many enhancers are capable of driving a heterologous promoter in the proper pattern when removed from their normal genomic context. This autonomy implies that enhancers can assemble a complete set of biochemical activities that together are sufficient for robust, patterned transcriptional activation at a remote promoter. Do different DNA-binding factors recruit distinct types of activation activities, or must the enhancer merely accumulate enough of a single, limiting activity to exceed a threshold for activation?

Different types of studies reach widely divergent conclusions about enhancer complexity. For example, Eric Davidson and colleagues, combining reporter assays with affinity purification in an extensive study of cis-regulatory logic in the sea urchin *Endo16* gene, identified 55 binding sites for 16 regulatory proteins, which form an intricate regulatory computer spanning 2300 bp of DNA (Davidson, 1999). On the other hand, most developmental genetics-based enhancer studies culminate in models requiring no more than three to five different regulators (often only one or two), binding within ~300-1000 bp of DNA, to explain the activity and specificity of a seemingly typical enhancer. In the very rare cases where the question of sufficiency is addressed in vivo, the defined regulatory sites are generally insufficient to properly reconstitute enhancer function, and an unknown activator "X" is added to the model (reviewed by Barolo and Posakony, 2002). How many cis-regulatory sites are sufficient, when combined, to recapitulate normal enhancer function, in the context of a chromosome in a normal cell?

We have pursued a bottom-up approach to these questions by taking a previously well-characterized developmental enhancer and exhaustively dissecting it in vivo, both to discover the extent of its regulatory complexity and to determine whether different enhancer sub-elements perform distinct functions. We chose to study the *sparkling (spa)* enhancer of the *dPax2* gene, which is necessary and sufficient to specify the cone cell fate in certain multipotent cells in the developing *Drosophila* eye (Fu and Noll, 1997; Fu et al., 1998; Flores et al., 2000; Shi and Noll, 2009). *spa* drives cone cell-specific *dPax2* expression in response to four direct regulators, acting through twelve transcription factor binding sites (TFBSs): Suppressor of Hairless [Su(H)], under the control of Notch signaling; two Ets factors, the activator PointedP2 (Pnt) and the repressor Yan, both

controlled by EGFR/Ras/MAPK signaling; and the Runx-family protein Lozenge (Lz) (Fu et al., 1998; Flores et al., 2000; Tsuda et al., 2002) (Figure 2.1A). In their report describing the direct regulation of the *spa* enhancer by Su(H), Lz, and Ets factors, Flores et al. (2000) proposed a model in which a combinatorial code, Lz + EGFR/Pnt/Yan + Notch/Su(H), determines the cell type specificity of *spa* activity. The authors were careful to state that "the model…reflects requirements rather than sufficiency for cell fate specification." Despite this caveat, the Lz+Ets+Su(H) code is now considered to "define the combinatorial input required for cone cell specification" (Voas and Rebay, 2004; see also Pickup et al., 2009; Shi and Noll, 2009).

Because the *spa* enhancer is small (362 bp), and because the known regulatory inputs could, in theory, explain its cell-type specificity (Flores et al., 2000), we considered it an ideal test case for a comprehensive structure-function analysis. Here, we report the results of our initial tests, which reveal several surprising aspects of *spa* enhancer function in vivo.

RESULTS

For our in vivo analysis of the *spa* enhancer, we used a specially built Gateway reporter transgene vector, Ganesh-G1, in which enhancers are placed upstream of a minimal, TATA-containing promoter taken from the *Drosophila Hsp70* gene, driving an EGFP-NLS reporter (Swanson et al., 2008). A unique feature of this vector is that the enhancer is placed 846 bp upstream from the transcription start site (Figure 2.1A). Thus, in all reporter constructs (except those in Figure 2.4), the enhancer is forced to act at a

Figure 2.1

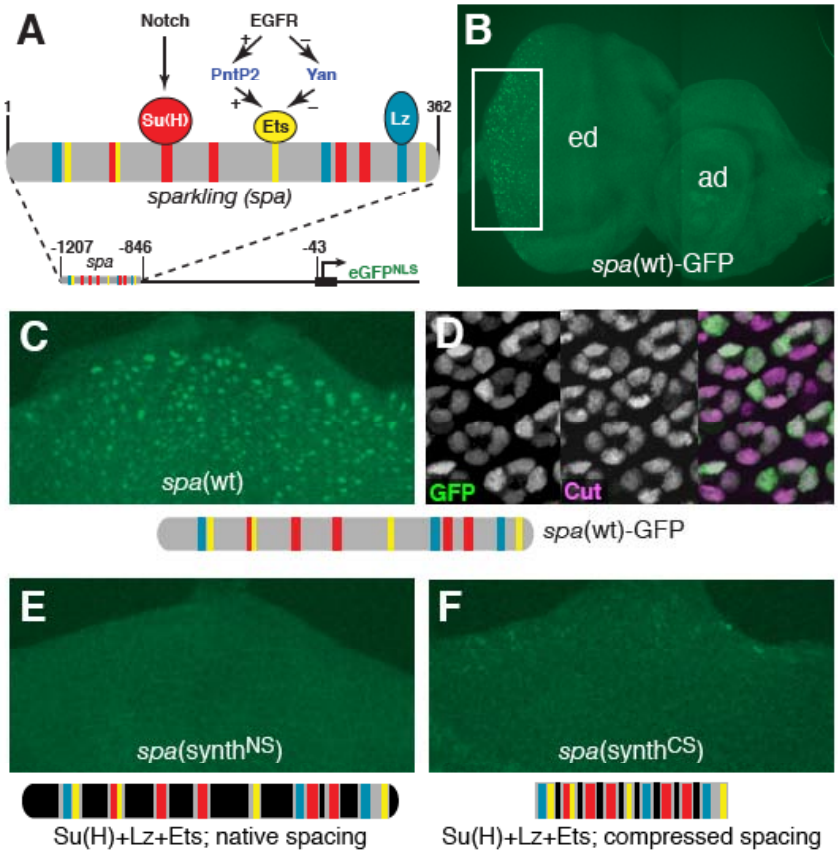


Figure 2.1. The known regulators of *spa* are insufficient for transcription in cone cells. (A) Summary of the known regulatory inputs of the *sparkling (spa)* cone cell enhancer of *D-Pax2*. Defined TF binding sites (TFBSs) are shown as colored bars; uncharacterized sequences are gray. The enhancer is placed 846 bp upstream of the transcription start site in all transgenic constructs, except those in Figure 2.4. (B-D) Expression of a GFP transgene under the control of *spa*. (B) Eye-antennal imaginal disc from a *spa*-GFP transgenic larva. (C) The posterior of an eye disc, corresponding approximately to the boxed area in (B). Posterior is to the top. (D) Eye of a 24-hour pupa carrying *spa*(wt)-GFP, stained with antibodies against GFP (green) and the cone cell nuclear marker Cut (magenta). (E) *spa*(synth^{NS}), in which the previously uncharacterized sequences have been altered (black), but the 12 defined TFBSs are present in their native arrangement and spacing, fails to drive gene expression. (F) *spa*(synth^{CS}), containing the 12 TFBSs in compressed spacing, is active in a few cells at the posterior margin, but does not recapitulate the cone cell activity of *spa*(wt).

moderate distance from the promoter. We do not consider this an unfair test of enhancer activity, given that, in its native genomic context, the *dPax2 spa* enhancer is located >7 kb from its promoter (Fu et al., 1998). We generated at least 4 independent transgenic lines for each reporter construct. Because line-to-line variability was generally low, we found that examination of 3-5 independently derived lines was sufficient for most constructs. For constructs with more variable expression (usually those with low activity), we examined additional lines (10-14) to ensure that our conclusions were not based on rare insertion effects.

When placed in Ganesh-G1, *spa* drives cone cell-specific GFP expression in developing retinas of transgenic larvae and pupae (Figures 2.1B-2.1D). This and previous work by Flores et al. (2000) demonstrate that the 362-bp *spa* enhancer contains all sequences necessary to (1) activate gene expression in vivo and (2) restrict this activation to developing cone cells.

The [Lz + Pnt + Su(H)] code is insufficient to specify cone cell expression

All three of the known positive regulators of the *spa* enhancer are required for its activity and cone cell specificity. This suggested a "combinatorial code" model for *dPax2* regulation, in which the combined activities of Lz, Pnt, and Su(H), acting through binding sites in *spa*, cooperatively activate *dPax2* expression specifically in cone cells (Flores at al., 2000; Tsuda et al., 2002; Nagaraj and Banerjee, 2007). We began our analysis by testing the simplest form of such a model, which predicts that the binding

sites within *spa* that mediate those three regulatory inputs should suffice, in combination, to drive gene expression in cone cells.

First, we built a synthetic *spa* enhancer construct in which all twelve of the defined binding sites for Lz, Su(H), and Pnt/Yan within *spa* are intact (along with 3-4 flanking base pairs to either side), and are placed in their native arrangement and spacing—but in which all other enhancer sequences are mutated by altering every second base pair. This construct, called *spa*(synth^{NS}) because of the native spacing of its TFBSs, fails to activate gene expression in vivo (Figure 2.1E). A second version of *spa*(synth^{NS}), in which the opposite set of base pairs was mutated, produced the same result (not shown). We also created *spa*(synth^{CS}), a compressed-spacing construct containing the same twelve sites, in which inter-site sequences of >12 bp have been reduced to 12 bp. *spa*(synth^{CS}) also fails to act as a cone cell enhancer, although weak GFP expression can be detected in a few non-cone cells (Figures 2.1F, 2.5H; Table S2.1). Based on these findings, we hypothesized that additional sequences, besides the twelve defined regulatory sites, are necessary for proper transcriptional regulation mediated by *spa*.

Multiple novel sites within *spa* are required for cone cell activation

In order to pinpoint the novel regulatory sequences within *spa* that make essential contributions to enhancer activity in vivo, we conducted a systematic mutational analysis of all previously uncharacterized sequences within *spa*. These sequences were divided into regions 1 through 6, and each region was deleted in turn, leaving the known TFBSs intact in all cases (Figure 2.2A). Of all segments mutated in this manner, only region 3

Figure 2.2

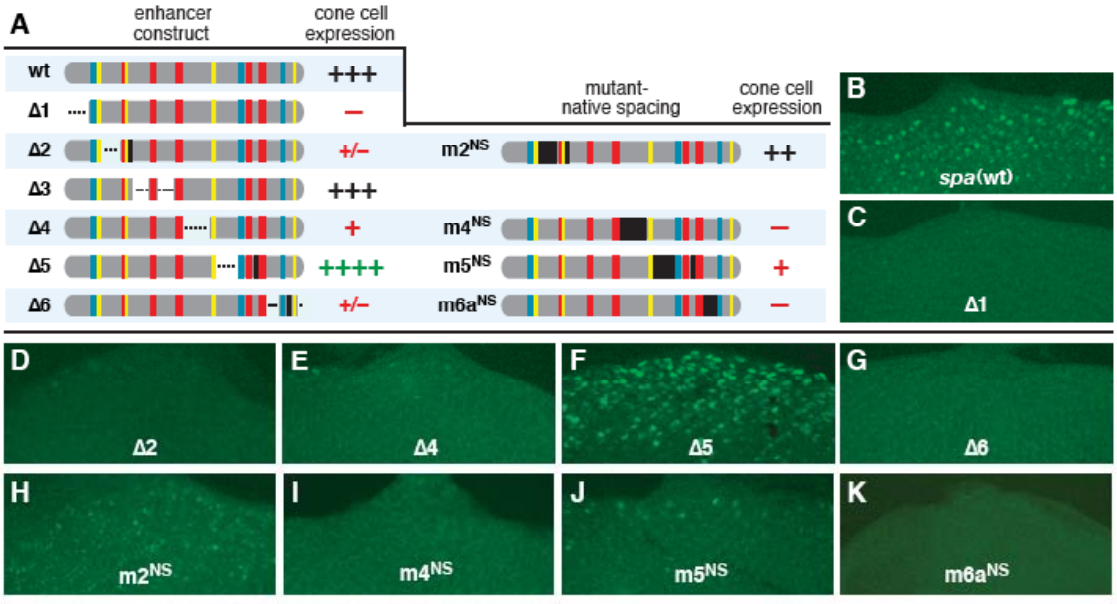


Figure 2.2. Sequence and/or spacing rules apply to multiple segments of *spa*. (A) Diagrams of *spa* enhancer constructs and summary of their cone cell activity in larval eye discs. Dotted lines indicate deletions; black bars indicated mutations that preserve native spacing (NS). In each case, the 12 known TFBSs are preserved. +++, wild-type levels and pattern of expression in cone cells; ++, moderately reduced; +, severely reduced; +/, detectable in very few cells; -, no detectable expression; ++++, augmented levels of expression. (B-K) GFP expression in eye imaginal discs driven by the wild-type *spa* enhancer (B) and mutant enhancers (C-K) carrying deletions or NS mutations in previously uncharacterized sequences, numbered 1 through 6.

makes no significant contribution to cone cell expression. Deleting regions 1, 2, 4, or 6 causes total or near-total loss of gene expression in vivo; conversely, deleting region 5 enhances expression in cone cells (Figures 2.2A-2.2G).

Spacing vs. sequence

Internal deletions of enhancer DNA cause two simultaneous changes: loss of the deleted sequence, and altered relative spacing of the sites to either side. To distinguish between these two types of effects, we made native-spacing (NS) mutations in which a specific sequence was altered, but its length was preserved. In regions 4 and 6, native-spacing alterations and deletions have similar effects, indicating that the sequence content of these regions is functionally significant (Figures 2.2D, 2.2G, 2.2H, and 2.2K). However, a native-spacing mutation in region 2 has a less severe effect than a deletion (Figure 2.2H; cf. Figure 2.2D), from which we infer that much of the regulatory contribution of region 2 can be attributed to its length, rather than its sequence.

Within region 5, deleting the DNA and altering its sequence have opposing effects. Deleting region 5 augments cone cell expression, while a native-spacing mutation causes a severe loss of activity (Figures 2.2F and 2.2J). The simplest interpretation of these results is that region 5 harbors positive regulatory sequences that are normally required, but that the deletion brings together sites on either side of region 5, increasing synergy between TFs and thus compensating for the loss of regulators normally binding to region 5. Consistent with this interpretation is the fact that Pnt and Lz, which bind to either side of region 5, physically interact and synergistically activate

transcription, as can mammalian orthologs of these factors (Flores et al., 2000; Behan et al., 2005 and references therein). The fact that multiple smaller-scale NS mutations within region 5 impair *spa* function, while none augment expression (see Figure 2.3), further supports this conclusion.

spa is densely packed with regulatory sites

The above analysis demonstrates that, in addition to the defined TFBSs, regions 1, 4, 5, and 6 of *spa* (and to a lesser extent region 2) are essential for its proper function. Each of these segments is large enough to contain several protein binding sites of typical size. To determine what proportion of these sequences has a regulatory role, we made native-spacing mutations to smaller segments (10 bp, on average) within regions 1, 4, 5, and 6. Of these 12 finer-scale mutations, ten cause severe or total loss of gene expression in cone cells (Figure 2.3). In addition, further experiments described below indicate the presence of novel repressive regulatory site(s) within *spa*, but outside of regions 1/4/5/6. Given that the consensus binding sites for the known regulators of *spa* are <9 bp in length, there is room for many regulatory sites within these regions. Together, the novel regulatory sites described here and the previously described TFBSs densely populate *spa*, with apparent "junk" or "spacer" sequences constituting a small proportion of the enhancer.

To investigate the possibility that the regulatory sites in regions 1, 4, 5, and 6 act by facilitating binding of the known activators to nearby binding sites, and the related possibility that these regions contain cryptic or non-canonical binding sites for the known

Figure 2.3

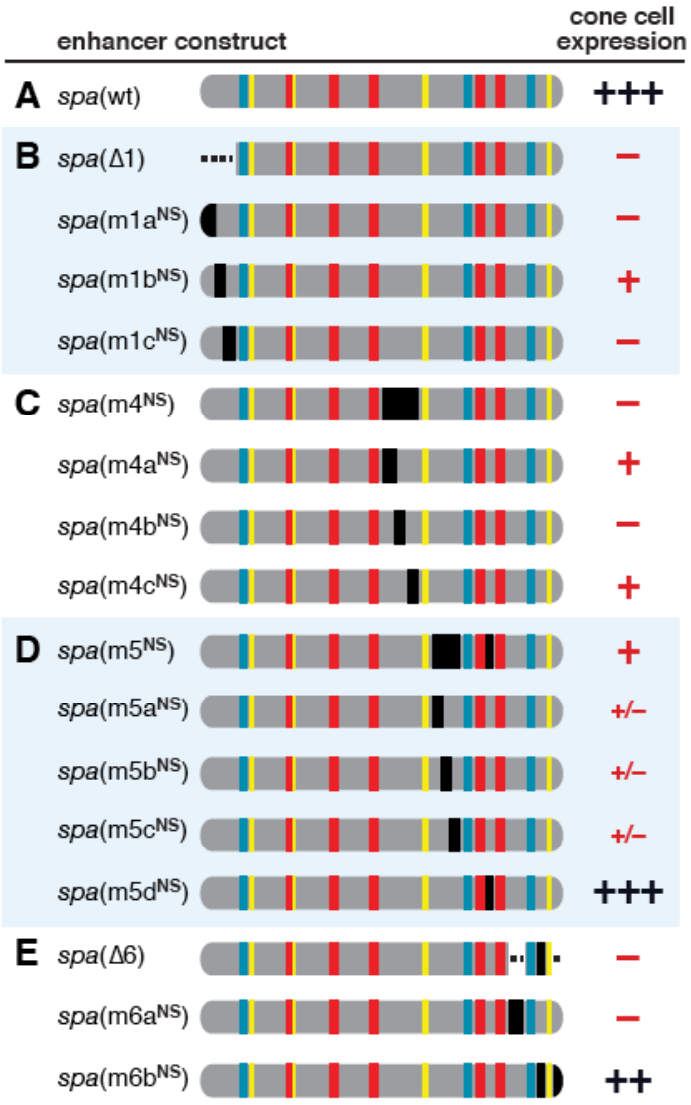


Figure 2.3. Most of *spa* is composed of critical regulatory sequences. (A-E) Diagrams of mutated *spa* enhancer constructs. Blue, yellow, and red bars indicate defined binding sites for Lz, Pnt/Yan, and Su(H), respectively. Dotted lines indicate deletions; black bars indicate mutations that preserve native spacing (NS). GFP expression in larval cone cells is summarized as in Figure 2.2.

activators, we tested the ability of Lz and Su(H) to bind to sites within *spa* in vitro. In all cases, mutating essential novel regulatory sequences did not significantly reduce the affinity of Lz or Su(H) for nearby binding sites, as determined by EMSA competition experiments (Table S2.2). Pnt does not bind in vitro to any sites flanking regions 1, 4, 5, or 6 (Flores et al., 2000). Therefore, in subsequent experiments we pursued the possibility that these regions of *spa* have novel functions, other than direct recruitment of the known activators.

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

The mutational analysis described above defined many novel regulatory sites of equal importance to the known Lz+Ets+Su(H) sites. We next attempted to isolate and study an important but poorly understood function of the enhancer: activation at a distance. As mentioned above, all of the enhancer constructs described thus far were placed 846 bp upstream of the promoter, thus forcing them to act over a moderate distance. If we could rescue the activity of a mutant enhancer by moving it close to the promoter, we reasoned, the mutated region is likely to specifically mediate remote enhancer-promoter interactions. Conversely, if a mutation cannot be rescued by promoter-proximal placement, it is likely to mediate a different step in gene activation.

The wild-type *spa* enhancer drives the same pattern from -121 bp as from -846 bp (Figure 2.4A), although activation is noticeably more robust from the more proximal position. A mutant *spa* enhancer lacking region 1 [*spa*(Δ1)], which is transcriptionally dead at -846 bp (Figure 2.2C), is completely rescued by placement at position -121,

Figure 2.4

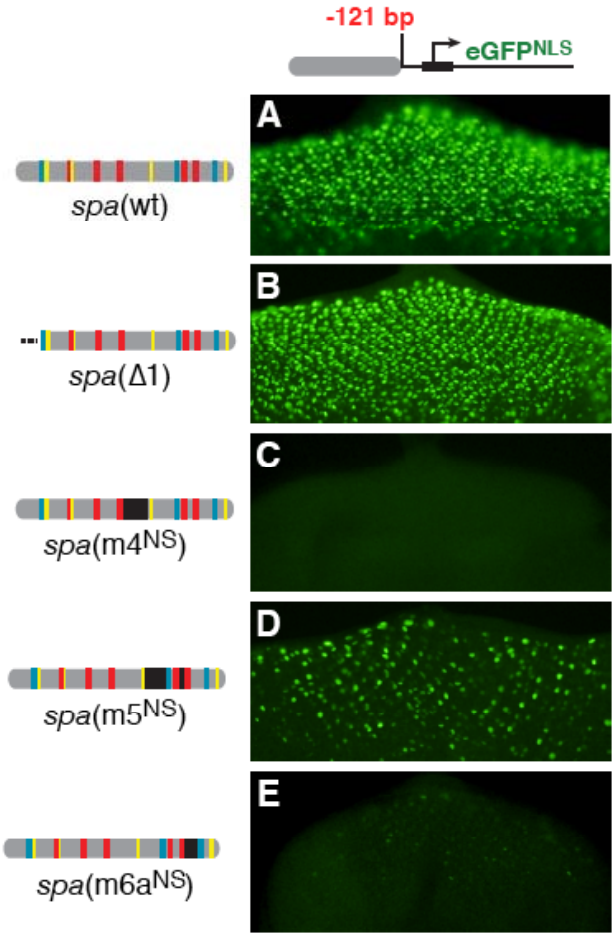


Figure 2.4. Region 1 is required for activation at a distance, but not for patterning. (A-E) Transgenic larval eye discs. In this figure, all enhancers are proximal to the minimal *Hsp70* promoter, at position -121 from the transcription start site, compared to -846 in all other figures. Because *spa* drives stronger expression from a promoter-proximal position, images are collected at a lower exposure than in other figures. (A) *spa*(wt) drives strong cone cell expression from -121 bp. (B) *spa*(Δ1), which is inactive when distant from the promoter (cf. Figure 2.2C), drives strong expression in the proper pattern from a promoter-proximal position. (C-E) In contrast to region 1, *spa* regions 4, 5, and 6a are required for wild-type levels of cone cell expression at both distal and proximal positions (cf. Figure 2.2I-K).

driving robust gene expression in the normal pattern (Figure 2.4B). 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 (Figures 2.4C-2.4E). Interestingly, each of these constructs partially recovers cone-cell activity by mid-pupal stages (not shown), suggesting that these regions may be more critical for initiation than for maintenance of gene expression. Similarly, Lz, Pnt, and Su(H) binding sites are required even when *spa* is promoter-proximal (Flores et al., 2000). Of all regulatory sites within *spa*, only region 1 is both dispensable for enhancer activity and patterning in a promoter-proximal position, and essential for activation at a distance.

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 (see Discussion). We therefore refer to region 1 as a "remote control" element to functionally distinguish it from patterning elements within *spa*, which include the defined TFBSs as well as new patterning sites to be discussed below. Future experiments will test the range, potential promoter preferences, and functional properties of this unique regulatory element.

Unlike the known TFs, Region 1 acts independently of its position within *spa*

Having mapped all essential regulatory sites within *spa*, we could then ask whether their linear organization influences gene expression in vivo. First, we tested the structural flexibility of region 1, the remote control element (RCE), by moving it from the

5’ end to the 3’ end of the enhancer. This rearranged enhancer performs normally at -846 bp (Figure 2.5G), 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.

By contrast, the Lz/Ets/Su(H) binding sites show strong position dependence. We rearranged these sites within *spa* by moving each TFBS (along with flanking sequences) to the position of another, randomly chosen, TFBS. The resultant construct, *spa*(TF scrambled), is only weakly active in cone cells (Figure 2.5G). Thus, unsurprisingly but in contrast to the RCE, the configuration of the known TFBSs within *spa* plays an important role in enhancer function in cone cells.

Cell type specificity is controlled by the structural organization of *spa*

The diminished activity of *spa*(TF scrambled), along with the altered gene expression resulting from deletions in regions 2 and 5, suggest that the spatial organization of *spa* impacts its transcriptional activity. We next took a different approach to investigate the relationship between structure and function within *spa*.

As we have demonstrated, the 12 defined TFBSs within *spa* are insufficient for cone cell enhancer activity, even when combined. Likewise, when these TFBSs are mutated, the remaining sequences are incapable of driving transcription [*spa*(KO), Figure 2.5B]. Since these two constructs, taken together, include all sequences from *spa*, we tested whether combining them would reconstitute enhancer activity. The resultant

Figure 2.5

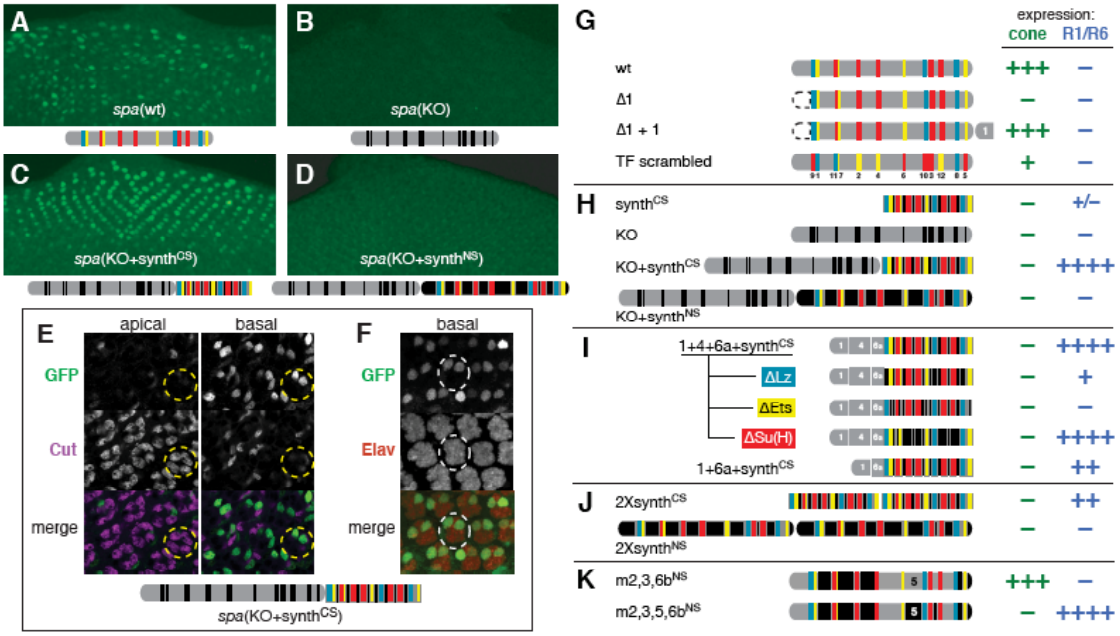


Figure 2.5. Cell-type specificity of *spa* is controlled by the arrangement of its regulatory sites. (A-D) GFP expression driven by *spa* enhancer constructs in larval eye discs. All constructs shown here are at the -846 position. (B) *spa*(KO), in which all 12 Lz/Ets/Su(H) sites are mutated, is inactive. (C) A rearranged version of *spa*, in which *spa*(KO) is placed next to the 12 TFBSs to create *spa*(KO+synth^{CS}), drives robust gene expression, but in a non-cone-cell pattern. (D) Combining *spa*(KO) with *spa*(synth^{NS}) fails to drive gene expression, though all sequences from *spa* are present.(E and F) *spa*(KO+ synth^{CS}) is expressed specifically in photoreceptors (PRs), but not in cone cells, in 24-hour pupae. (E) Confocal images at two different planes, in retinas stained with antibodies against GFP (green) and the cone cell nuclear marker Cut (magenta), show GFP in two nuclei per ommatidium, located basally to cone cells. Posterior is to the top. (F) GFP driven by *spa*(KO+ synth^{CS}) co-localizes with the PR marker Elav (red). (G-J) Organization of regulatory elements within *spa* is critical for both transcriptional activity and cell-type specificity. (G) Transferring region 1 from the 5' to the 3' end of *spa* does not impair its function in long-range transcriptional activation (Δ1+1). In contrast, randomizing the positions of the 12 Lz/Ets/Su(H) binding sites within *spa*(TF scrambled) impairs its activity in cone cells, with no ectopic activity in other cell types. (H) Rearranging the regulatory sites of *spa* converts its cell-type specificity. (I) Combining the defined TFBSs with novel regions 1, 4, and 6a drives strong R1/R6 expression. This depends on Lz and Ets sites and region 4, but not Su(H) sites. (J) 2X synth^{CS}, containing two copies of every known TFBS, is inactive in cone cells, with moderate ectopic PR activity. 2X synth^{NS} is inactive in all cell types. (K) Region 5 of *spa* mediates repression in PRs, as well as activation in cone cells. *spa*(m2,3,6b^{NS}), a native-spacing *spa* construct in which regions 2, 3, and 6v are mutated, drives normal (i.e., cone cell-specific) gene expression. Further mutating region 5 causes a switch from cone cell-specific to PR-specific gene expression.

rearranged *spa* construct, KO+synth^{CS}, drives strong gene expression in the eye (Figure 2.5C).

Three aspects of this finding are worth noting. First, the activity driven by KO+synth^{CS} is robust, exceeding *spa*(wt) in intensity (Figure 2.5C; cf. panel A). The defined TFBSs, therefore, are capable of acting synergistically with novel activator sites in *spa*, even when the enhancer is reconfigured. This, combined with our in vitro binding data mentioned above, strongly suggests that the new regulatory sequences are not merely extended binding sequences for Lz/Pnt/Su(H).

Second, when the TFBSs adjacent to *spa*(KO) are spread out to mimic their native spacing, gene expression is lost (KO+synth^{NS}, Figure 2.5D). The activity of *spa* is apparently highly dependent on close proximity, among the known TFs and/or between those TFs and novel regulators. Since KO+synth^{CS} and KO+synth^{NS} differ by only 29% in total length, and because KO+synth^{NS}, at 730 bp, is not large compared to many enhancers, this extreme dependence on short-range interactions was surprising.

Third, and most importantly, the pattern of gene expression driven by the rearranged enhancer *spa*(KO+synth^{CS}) differs from that of *spa*(wt)—in fact, the two elements drive completely non-overlapping expression patterns. Unlike *spa*(wt), whose activity co-localizes with the cone cell marker protein Cut (Figure 2.1D), KO+synth^{CS}-GFP is expressed only in nuclei located basally to Cut^{+} cells (Figure 2.5E). KO+synth^{CS} is active in a subset of basal cells expressing Elav, a marker of photoreceptor cell fate (Robinow and White, 1988). Based on the position of the two GFP^{+} cells within the Elav^{+} photoreceptor cluster, *spa*(KO+synth^{CS})'s activity is restricted to photoreceptors 1

and 6 (R1/R6) (Figure 2.5F). Thus, merely re-arranging the regulatory sites within *spa* is sufficient to cleanly switch its cell-type specificity in vivo.

Ectopic photoreceptor-specific transcription depends on Lz and Ets binding sites, additional novel regulatory sequences, and tight clustering of regulatory sites

We next attempted to identify the regulatory sites responsible for ectopic activity of *spa* in photoreceptors (PRs). Combining regions 1, 4, and 6a with the known TFBSs (1+4+6a+synth^{CS}) results in strong R1/R6 expression; removing region 4 from this construct weakens its activity (Figure 2.5I). By selectively mutating TFBSs, we found that R1/R6 expression requires Lz and Ets sites, but not Su(H) sites (Figure 2.5I). This is consistent with the fact that R1/R6 receive MAPK signaling and express Lz at high levels, but do not respond to Notch signaling (reviewed by Voas and Rebay, 2004).

Based on our remote vs. proximal enhancer analysis (Figure 2.4), we hypothesized that different regulatory sequences within *spa* contribute distinct activities to gene activation. If this is so, one type of activity may not be able to functionally substitute for another. We tested this idea by creating tandem repeats of the synth^{CS} and synth^{NS} constructs, which contain two copies of each known TFBS, in compressed or native spacing, respectively. 2Xsynth^{CS} is inactive in cone cells and relatively weakly active in PRs, while 2Xsynth^{NS} is inactive in all cell types (Figure 2.5J). We therefore conclude that the Lz+Ets+Su(H) combination is insufficient for cone cell activation. Further, the fact that additional Lz/Ets/Su(H) sites fail to compensate for the missing novel regulatory sites adds support to the idea that some parts of the enhancer perform

functions in transcriptional activation that are qualitatively distinct from those of the known regulators.

Interestingly, when *spa*(synth^{NS}) is placed at -121 bp, we observe occasional position-effect-dependent activity in cone cells (1 out of 7 lines) or PRs (1 of 7 lines) (Figure S3). The pattern of gene expression in these two lines depends on the site of transgene insertion, which is consistent with the conclusion that Lz+Ets+Su(H) can contribute to gene expression in multiple cell types, but only in combination with additional regulatory inputs.

A novel, short-range, cell type-specific repressor activity prevents *spa* activation in photoreceptors

In both *spa* constructs driving strong ectopic R1/R6 activity, *spa*(KO+synth^{CS}) and *spa* (1+4+6a+synth^{CS}), the configuration of defined TFBSs differs from wild-type in two respects: their spacing relative to one another is reduced, and their linear order and position relative to the novel regulatory sequences is altered. Ectopic photoreceptor expression, then, could have three possible (non-exclusive) causes: (1) tight TF clustering may increase synergy by Lz and Pnt in R1/R6, or altered spacing between TFs and novel sites may cause (2) inappropriate synergistic activation and/or (3) weakened repressive interactions in PRs. In order to test these models, and to further explore the role of enhancer structure, we generated compound mutations in multiple regions of *spa*, while keeping the spacing/arrangement of the remaining sequences intact.

First, we simultaneously mutated regions 2, 3, and 6b of *spa*, none of which are essential for cone cell expression. This construct, *spa*(m2,3,6b^{NS}), is comparable to *spa*(wt) in its pattern and levels of expression (Figure 2.5K). Next, we additionally mutated region 5 in this construct to create *spa*(m2,3,5,6b^{NS}). Remember that when region 5 alone is mutated, cone cell expression is severely reduced, and no ectopic expression is seen (Figures 2.2J and 2.3D). However, when region 5 is mutated simultaneously with 2/3/6b, a discrete switch from cone cell- to R1/R6-specific expression occurs (Figure 2.5K). Therefore, region 5 mediates repression in PRs, in addition to activation in cone cells. This repressive activity must be redundant with additional repressor site(s) in regions 2/3/6b. It must also have a very limited range of action, since moving Lz and Ets sites to the 3' end of the enhancer, without altering the repressor sites (KO+synth^{CS}), de-represses *spa* in R1/R6.

spa enhancer evolution: function is conserved despite rapid turnover of regulatory sequences

Taking this study and previous work into account, *spa* is among the most finely mapped enhancers with respect to regulatory sites essential for function in vivo. We made use of the recent sequencing of multiple *Drosophila* species genomes (*Drosophila* 12 Genomes Consortium, 2007) to investigate the evolutionary history of *spa*. We will focus on the *D. melanogaster-D. pseudoobscura* (*mel-pse*) comparison, which is commonly used to study cis-regulatory sequence evolution; the two populations diverged ~25 million years ago (e.g., Ludwig et al., 2005 and references therein). As we will

discuss below, blocks of sequence conservation between *melanogaster spa* and *pseudoobscura* are relatively few and short, and many TFBSs and novel regulatory sites were not alignable (Figure 2.6A). We were therefore surprised to find that a 409-bp *pseudoobscura* sequence we identified as the putative ortholog of *spa* was able to drive cone cell-specific reporter gene expression in transgenic *D. melanogaster*, indistinguishably in pattern and intensity from *melanogaster spa* (Figure 2.6B).

We wish to point out several notable aspects of *spa* sequence evolution. First, its distribution of sequence conservation appears to be unusual among developmental enhancers. When total *mel-pse* sequence identity is considered, *spa* (65% identity) falls only slightly below the range of six well-studied *Drosophila* enhancers we analyzed for comparison (70% to 88%**;** Table S2.4). However, *spa* is relatively poor in extended blocks of conserved sequence; it contains only one block of 100% conservation of ≥10 bp in length (in region 1, the RCE), constituting 3.9% of total enhancer sequence (Figure 2.6C). By contrast, in the six reference enhancers, an average of 52% of sequence lies in perfectly conserved blocks of ≥10 bp (range is 37% to 75%). If the cutoff is lowered to 9/10 identity, 29% of *spa* is in conserved blocks, while in the six reference enhancers, the average is 67% (range is 48% to 89%; Figure 2.6C; Table S2.4). Even more strikingly, in the six reference enhancers, an average of one-third of the sequence is in perfectly conserved blocks of ≥20 bp (range is 20% to 58%), while *spa* has no conserved blocks of this size (Figure 2.6C; Table S2.4). The paucity of extended blocks of conservation within *spa* is exemplified by the fact that only 24% of *spa* can be aligned with its *pseudoobscura* ortholog by the "BLAST 2 Sequences" algorithm (Tatusova and Madden, 1999), compared to an average of 66% for the six reference enhancers (range is

Figure 2.6

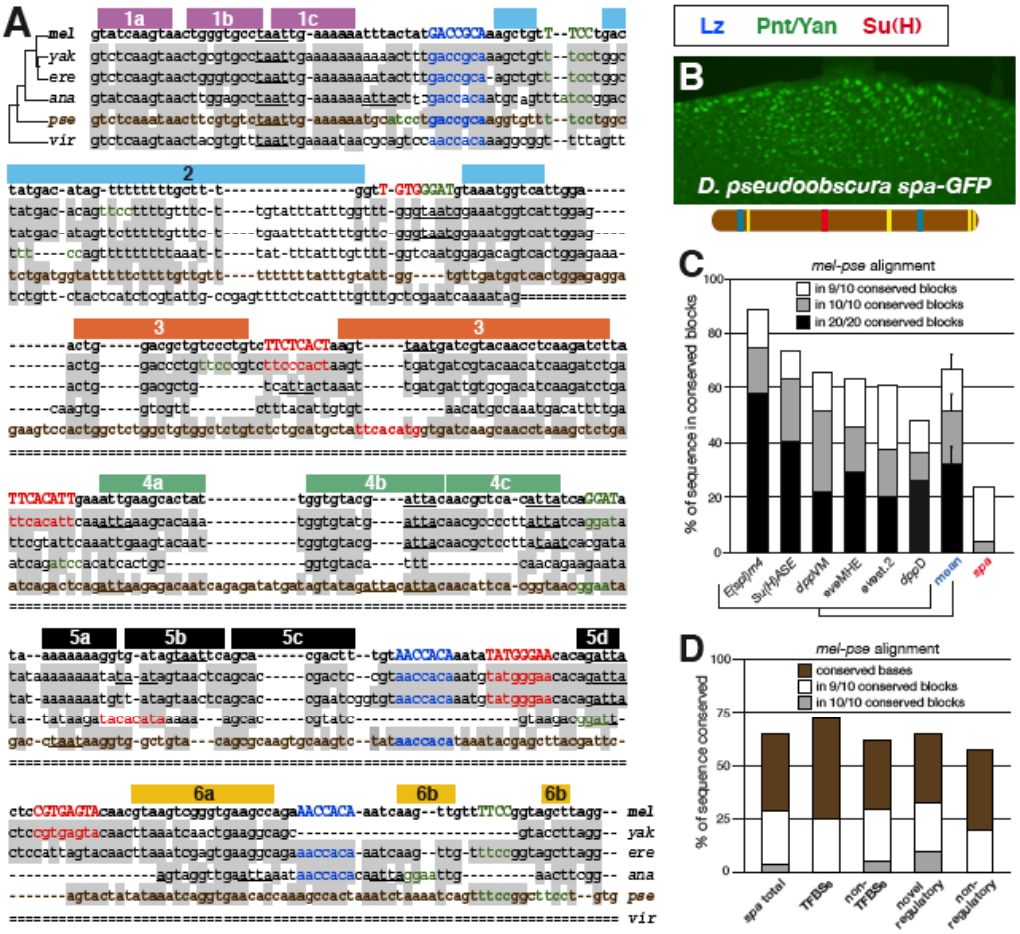


Figure 2.6. *spa* function is conserved in two *Drosophila* species, despite rapid turnover of known and novel regulatory sequences. (A) Alignment of the *spa* enhancer of *D. melanogaster* (*mel)* and orthologous sequences from *D. yakuba (yak), D. erecta (ere), D. ananassae* (*ana*), and *D. pseudoobscura* (*pse*). Known binding sites for Lz, Pnt/Yan, and Su(H), and predicted orthologous sites, are colored blue, green, and red, respectively. Regions 1 through 6 are labeled with colored bars. TAAT motifs are underlined. Conserved bases are shaded gray. (B) The 409-bp *D. pse* sequence shown in panel A drives robust cone cell-specific gene expression in eye discs of transgenic *D. mel*. (C) *spa* has few blocks of conserved sequence, based on a pairwise *mel-pse* enhancer alignment. 90-99% and 100% conserved sequence blocks are defined as contiguous ≥10 bp sequences with 90-99% or perfect conservation, respectively, counting gaps as mismatches. Results for six other well-studied developmental enhancers are shown, along with mean values. Error bars indicate SEM. (D) Conservation is a poor indicator of the regulatory significance of *spa* enhancer sequences. "Novel regulatory" sequence is regions 1, 4, 5abc, and 6a; "non-regulatory" sequence is regions 2, 3, 5d, and 6b. Sequence identity is calculated from a pairwise *mel-pse* BLASTZ alignment as ((# of identical bases)/(total bases residing in those sequences)) X 100%.

42% to 99%, Table S2.5). The genomic BLAST program at DroSpeGe (Gilbert, 2007) is unable to align any part of *spa* with *D. pseudoobscura* genomic sequence, whereas an average of 52% of the sequence of the six reference enhancers can be aligned by this method (range is 30% to 79%; Table S2.5). Lack of sequence conservation does not appear to result from a reshuffling of regulatory sequences, as *melanogaster* vs. *pseudoobscura* dot-plot analysis does not detect any rearrangements within *spa* (Figure S2.6).

Second, of the 12 identified binding sites for Lz, Pnt/Yan, and Su(H), only three can be unambiguously aligned with orthologous predicted binding sites in *pseudoobscura*. Four other predicted binding sites for these TFs were found in the *pseudoobscura* enhancer, but had no definitive orthologs in *melanogaster spa*, due to significant differences in sequence and/or position (Figure 2.6A). The average *mel-pse* sequence identity of TFBSs in *spa* is 72.6%, lower than that of all six reference enhancers (Figures 2.6D and S3**;** Table S2.4). Overall, *pseudoobscura spa* contains fewer predicted TFBSs than *melanogaster spa*: 1 vs. 5 predicted Su(H) sites, 2 vs. 3 predicted Lz sites, and 4 GGAW consensus Ets sites vs. 6 in *melanogaster*.

Third, with respect to the novel (that is, previously uncharacterized) sequences within *spa*, we do not observe a strong correlation between functional significance and sequence conservation. Of those functionally essential novel sequences identified in this report (regions 1/4/5abc/6a), the total *mel-pse* sequence identity is not greatly higher than that of sequences making little or no contribution to activation (regions 2/3/5d/6b) (65% vs. 58% identity; Figure 2.6D). Thus, in the context of the *spa* enhancer, we find

evolutionary sequence conservation to be a poor indicator of functional importance in transcriptional regulation.

DISCUSSION

The goal of this study was to use a well-characterized, signal-regulated developmental enhancer to examine, in fine detail, the regulatory interactions and structural rules governing transcriptional activation in vivo. Taking the elegant work of Flores and colleagues (2000) as a starting point, we have used functional in vivo assays to test the power of the proposed combinatorial code of "Notch/Su(H) + Lz + MAPK/Ets" to explain the activity and cell-type specificity of the *spa* cone cell enhancer of *dPax2*. In the course of this work, we have discovered several surprising properties of *spa* that are not accounted for in current models of enhancer function.

The *spa* patterning code is massively combinatorial

We chose the *spa* enhancer for our fine-scale analysis because (1) the known direct regulators and their binding sites are well defined, (2) they could in theory constitute the sum total of the patterning information received by the enhancer, and (3) the enhancer, at 362 bp, is relatively small, simplifying mutational analyses. To our surprise, a large proportion of the previously uncharacterized sequence within *spa* is vital for normal enhancer activity in vivo, and of that subset, a large proportion directly influences cell-type specificity. These findings are summarized in Figure 2.7A.

Figure 2.7

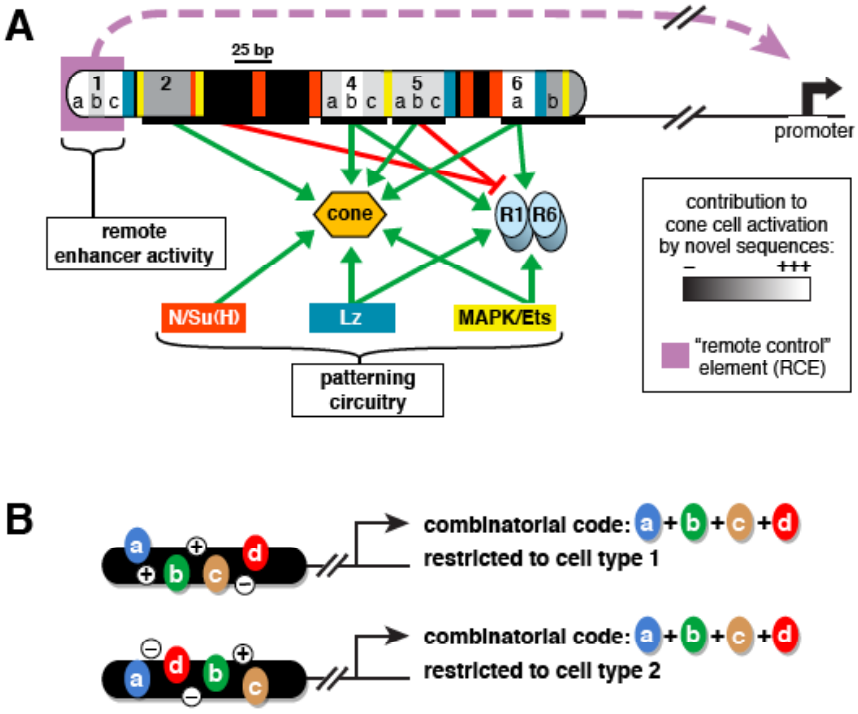


Figure 2.7. Short-range positive and negative interactions shape the cell-type specificity of enhancer function. (A) Summary of *spa* regulation: two functional classes of regulatory sites govern the enhancer activity of *spa* in vivo. The *spa* enhancer requires the presence and proper arrangement of many regulatory sub-elements for its transcriptional activity and cell-type specificity. Region 1 appears to be required for remote enhancer activity, but dispensable for patterning. In addition, proper cell-type patterning of *spa* in the developing eye is considerably more complex than previously thought, and depends on short-range interactions among many regulatory sites. Green arrows indicate activation mediated by sites within *spa*; red bars indicate cell-type-specific repression activities. (B) A simple "combinatorial code" model is insufficient to explain the cell type specificity of *spa*, as the same regulatory elements (represented here as factors a, b, c, and d) can be rearranged to generate transcription in either cone cells or photoreceptors. Thus any model describing cone cell-specific transcriptional activation by *spa* must also incorporate rules of spatial organization.

Activation in cone cells: In addition to necessary inputs from Lz, Pnt, and Su(H), we have identified three segments of *spa*, regions 4, 5, and 6, that make essential contributions to gene expression in cone cells. In addition, region 2 makes a relatively minor contribution. (Region 1, another essential domain, will be discussed separately.) Fine-scale mutagenesis reveals that within regions 4/5/6, very little DNA is dispensable for cone cell activation. The novel regulatory sites in *spa* are very likely bound by factors other than Lz/Pnt/Su(H), for the following reasons: no sequences resembling Pnt/Lz/Su(H) binding sites reside in these regions; mutations in the novel sites have different effects than removing the defined TFBSs or the proteins that bind them; doubling the known TFBSs fails to compensate for the loss of the novel sequences; and most importantly, mutating the novel regulatory regions does not significantly affect binding of the known activators to nearby binding sites in vitro (Table S2.2). As for the expression pattern of the novel regulators, we cannot tell from a loss-of-function analysis whether they are cone cell-specific, eye-specific, or ubiquitous—we only know that multiple novel sites are necessary both for normal cone cell expression and ectopic PR expression. Besides Lz, Pnt, and Su(H), we know of no transcriptional activators present in cone cells; Cut, Prospero, and Tramtrack are expressed in cone cells, but are thought to act as transcriptional repressors (e.g., Lai and Li, 1999; Cook et al., 2003; Sato et al., 2006). The transcription factor Hindsight is required for *dPax2* expression and cone cell induction, but acts indirectly, activating *Delta* in R1/R6 to induce Notch signaling in cone cells (Pickup et al., 2009).

Unsurprisingly, placing the enhancer closer to the promoter boosts expression of *spa*(wt), as well as some of the impaired mutants (Figures 2.4 and S2.1). Remember that

spa is located at +7 kb in its native locus, and that nearly all mutational studies place the enhancer immediately upstream of the promoter. If our entire analysis had been performed at -121 bp, we would have underrated the functional significance of several critical regulatory sequences, and would have eliminated region 1 entirely as a regulatory site. Other well-characterized enhancers, which have been analyzed in a promoter-proximal position only, may therefore contain more critical regulatory sites than is currently realized.

Like many transcriptional activators, all three known direct activators of *spa* (or their orthologs) recruit p300/CBP histone acetyltransferase coactivator complexes (e.g., Kitabayashi et al., 1998; Barolo and Posakony, 2002). Doubling the number of binding sites for these TFs (to 6 Lz, 8 Ets, and 10 Su[H] sites) does not suffice to drive cone cell expression in the absence of the novel regulatory regions (Figure 2.5). It may be, then, that the novel activators of *spa* employ mechanisms that are distinct from those of the known activators. The remote activity of *spa*, mediated by region 1, may be an example of such a mechanism.

Cell-type specificity: We were able to convert *spa* into a photoreceptor R1/R6-specific enhancer in three ways: (1) by moving the defined TFBSs to one side of the enhancer in a tight cluster; (2) by placing Lz and Ets sites next to regions 1/4/6a; and (3) by mutating regions 2/3/5/6b within *spa* while maintaining the native spacing of all other sites. From these experiments, we conclude that *spa* contains short-range repressor sites that prevent ectopic activation in PRs by Lz + Pnt + regions 4 + 6a. *spa* contains at least two redundant repressor sites, since both region 5 and regions 2/3/6b must be mutated to attain ectopic R1/R6 expression.

klumpfuss, which encodes a putative transcriptional repressor, is directly activated by Lz in R1/R6/R7, but is also present in cone cells (Wildonger et al., 2005, and references therein), making it an unlikely repressor of *spa*. *seven-up*, another known transcriptional repressor, is expressed in R3/R4/R1/R6 and could therefore act to repress *spa* in photoreceptors (Mlodzik et al., 1990; Cooney et al., 1993). However, we did not identify putative *seven-up* binding sites within *spa*. Phyllopod, an E3 ubiquitin ligase component, represses *dPax2* and the cone cell fate in R1/R6/R7, but the transcription factor mediating this effect is not yet known (Shi and Noll, 2009). Perhaps the best candidate for a photoreceptor-specific direct repressor of *spa* is *Bar*, which encodes the closely related and redundant homeodomain TFs BarH1 and BarH2. *Bar* expression is activated by Lz in R1/R6 and is required for R1/R6 cell fates (Higashijima et al., 1992; Crew et al., 1997). Furthermore, misexpression of *BarH1* in presumptive cone cells can transform them into photoreceptors (Hayashi et al., 1998). It is unclear whether Bar-family proteins act as repressors, activators, or both. BarH1/2 can bind sequences containing the homeodomain-binding core consensus TAAT (Noyes et al., 2008), and region 5 of *spa* contains two TAAT motifs (underlined in Figure 2.6A). Future studies will explore the possibility that *Bar* directly represses *spa* in photoreceptors.

The combinatorial code of *spa*, then, requires multiple inputs in addition to Lz, MAPK/Ets, and Notch/Su(H). Indeed, our data suggest that the known regulators can contribute to expression in multiple cell types, depending on context. The novel control elements we have identified within *spa* are necessary not only to facilitate transcriptional activation, but also to steer the Lz/Ets/Su(H) code toward cone cell-specific gene expression.

Functional evidence for a new 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 (e.g., Yoshida et al., 1999; Carter et al., 2002; Song et al., 2007). 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, without requiring an LCR or other large-scale genomic regulatory apparatus. 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 (e.g., Calhoun et al., 2002; Chen et al., 2005, Akbari et al., 2008). 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 ER, AR, and Sp1 (Wang et al., 2005; Williams et al., 2007; Pan et al., 2008). Second, studies of distant enhancers of the *cut* and *Ultrabithorax* genes have revealed a role for the cohesin-associated factor Nipped-B, especially with respect to bypassing insulators (Misulovin et al., 2008, and references therein), but it has not been demonstrated that Nipped-B, or any other enhancer-binding regulator, is required *only* when the enhancer is remote.

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.

Enhancer structure: shaped and constrained by short-range patterning interactions

As discussed above, it is fairly easy to switch *spa* from cone cell expression to R1/R6 expression (though, curiously, we have yet to generate a construct that is active in both cell types). Our results show that multiple regions of *spa* mediate a repression activity in R1/R6 but not in cone cells. We further conclude that these *spa*-binding repressor(s) act in a short-range manner; that is, they must be located very near to relevant activator binding sites, since moving Lz and Pnt sites to one side of *spa*, without removing the repressor sites (KO+synth^{CS}), abolishes repression. Despite this failure of repression, synergistic interactions among Lz + Ets sites and novel sequences still occur in this re-organized enhancer—at least in R1/R6 cells. Cone cell-specific expression is lost, however, revealing (along with other experiments) that transcriptional activation in cone cells is highly sensitive to the organization of regulatory sites within *spa*. Slightly wider spacing of regulatory sites (KO+synth^{NS}) kills the enhancer altogether, suggesting that synergistic positive interactions within *spa*, though apparently longer in range than repressive interactions, are severely limited in their range. The structural organization of *spa*, then, appears to be constrained by a complex network of short-range positive and negative interactions (Figure 2.7B). Activator sites must be spaced closely enough to trigger synergistic activation in cone cells; at the same time, repressor sites must be positioned to disrupt this synergy in non-cone cells, preventing ectopic activation.

Recent work by Crocker et al. (2008) has shown that changes to enhancer organization can "fine-tune" the output of a combinatorial code, subtly changing the sensitivity of the enhancer to a morphogen. Given the importance of the structure of the *spa* enhancer for its proper function, we propose that *any* combinatorial code model, no matter how complex, is insufficient to describe the regulation of *spa*, since the same components can be rearranged to produce drastically different patterns.

Conservation of *spa* function despite lack of sequence conservation: insights into enhancer structure

One might expect that the regulatory and organizational complexity of the *spa* enhancer, and its extreme sensitivity to mutation, would be reflected in strict evolutionary constraints upon enhancer sequence and structure. Yet we observe very poor conservation of *spa* sequence, both in the known TFBSs and in novel regulatory elements shown to be critical for the function of *melanogaster spa*. The reduced presence of Lz/Ets/Su(H) sites in *D. pse* could potentially be attributed to redundancy of those sites in *D. mel*, or to compensatory gain of binding sites for alternate factors in the *D. pse* enhancer. Perhaps more difficult to understand is the apparent loss of critical regulatory sequences in regions 4, 5, and 6a in *D. pse*; our experiments in *D. mel* suggest that the absence of those inputs would result in loss of cone cell expression and/or ectopic activation. It remains possible that many of these inputs are in fact conserved, but that conservation is not obvious due to binding site degeneracy and/or rearrangement of elements within the enhancer. Fine-scale comparative studies are ongoing.

spa is by no means the first example of an enhancer that is functionally maintained despite a lack of sequence conservation (for a review of this topic, see Wittkopp, 2006). The most thoroughly characterized example of this phenomenon is the *eve* stripe 2 enhancer; its function is conserved despite changes in binding site composition and organization (Ludwig et al., 2000; Ludwig et al., 2005; Hare et al., 2008). Note, however, that the sequence of *spa* has changed much more rapidly than that of *eve* stripe 2 (Figures 2.6 and S2.4-S2.5), with no apparent change in function. In general, the ability of an enhancer to maintain its function in the face of rapid sequence evolution suggests that enhancer structure must be quite flexible. These observations support the "billboard" model of enhancer structure, which proposes that as long as individual regulatory units within an enhancer remain intact, the organization of those units within the enhancer is flexible (Arnosti and Kulkarni, 2005). Yet our findings concerning the importance of local interactions among densely clustered, precisely positioned transcription factors are more consistent with the tightly structured "enhanceosome" model (Thanos and Maniatis, 1995). Further structure-function analysis will be necessary to fully understand the players and rules governing this regulatory element.

EXPERIMENTAL PROCEDURES

Generation of mutant and synthetic enhancers

The 362-bp *sparkling* enhancer was amplified from w^{1118} genomic DNA with the following primers: 5'-CACCGGATCCgtatcaagtaactgggtgcctaattg-3'; 5'-GGGTCTAGAcctaagctaccggaaaacaacttg-3'. The 409-bp *D. pseudoobscura spa* enhancer was PCR-amplified from genomic DNA with the following primers: 5'-CACCGGATCCgtctcaaataacttcgtgtc-3'; 5'-GGGTCTAGAcacaggaagccggaaactg-3'. Lower-case 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.

In mutating novel 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. In mutating TFBSs, we converted Lz sites from RACCRCA to RA**AA**RCA; Ets sites from GGAW to **TT**AW; and Su(H) sites from YGTGDGAA (or related sequence) to YGTGD**C**AA; these changes eliminate TF binding in vitro (Barolo et al., 2000; Flores et al., 2000; references therein).

Enhancer cloning, vectors, and transgenesis

PCR-amplified enhancer constructs were TOPO-cloned into the pENTR/D-TOPO vector (Invitrogen). *Spa*(synth^{CS}) was created by annealing two complementary oligonucleotides and ligation into the Gateway donor vector pBS-ENTR-TOPO (Swanson et al., 2008). 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 (Figure 2.4) 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.

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.

DNA sequence alignment

The *sparkling* multi-species alignment is based on BLASTZ alignments and was taken from the UCSC genome browser (http://genome.ucsc.edu). Pairwise *mel-pse* alignments were performed using zPicture (Ovcharenko et al., 2004; http://zpicture.dcode.org). Pairwise BLAST searches were performed with the "BLAST 2 Sequences" algorithm at NCBI (Tatusova and Madden, 1999; www.ncbi.nlm.nih.gov/BLAST/bl2seq/wblast2.cgi) and at DroSpeGe (Gilbert, 2007; http://insects.eugenes.org/species).

Protein expression and purification

Lozenge protein was expressed using the TNT T7 Coupled Reticulocyte Lysate System (Promega) from pET3c-Lz (kindly provided by Richard Mann, Columbia University). His-tagged Su(H) was expressed from pRSET-6XHis-Su(H) (kindly provided by Jim Posakony, University of California, San Diego). RosettaBlue (DE3) Competent Cells (Novagen) transformed with pRSET-6XHis-Su(H) were grown overnight in 250 ml of LB plus 30 µg/ml carbenicillin in a 37°C shaking incubator. The next morning, cells were spun down for 10 min at 6000 rpm, and the pellet was resuspended in 10 ml LB. Two 500 ml LB+carb cultures were each inoculated with 2.5 ml of resuspended cells and grown in a 30°C shaking incubator until they reached an OD_{600} of 0.6-0.7. Each 500 ml culture was then induced with 500 µl 1M IPTG and

grown at 30°C for an additional 2 hours. Cultures were spun down at 4°C for 10 min at 6000 rpm. Each pellet was resuspended on ice in 10 ml of lysis buffer (0.1 M NaCl, 0.1% Tween-20, 10 mM Tris-Cl (pH 8.0), 5 mM Imidazole, 1 mM DTT, 1 mM PMSF, 1 tablet Complete Mini EDTA-free (Roche)). Cells were lysed at 4°C by sonication (60% power, 5 X 30s, at 30s intervals), then centrifuged for 20 min at 10000 rpm at 4°C. All following purification steps took place in a 4°C cold room. Two 0.8 X 4 cm Poly-Prep Chromatography Columns (Bio-Rad) were packed with 0.15 ml of Ni-NTA agarose (Qiagen) and cleared with 10 ml Wash Buffer 1 (0.1 M NaCl, 0.1% Tween-20, 10 mM Tris-Cl (pH 8.0), 5 mM Imidazole). Ten ml of supernatant was then applied to each column. Each column was washed with 1 ml Wash Buffer 1, followed by 1 ml Wash Buffer 2 (0.1 M NaCl, 10 mM Tris-Cl (pH 8.0), 15 mM Imidazole). Protein was eluted with 6 X 100 µl elution buffer (0.1 M NaCl, 10 mM Tris-Cl (pH 8.0), 200 mM Imidazole, pH adjusted to 8.0). Elutions containing protein were pooled and glycerol was added to 10% final concentration before being aliquoted, flash frozen with dry ice and ethanol, and stored at -80°C.

In vitro binding assays

Gel shift probes were made using custom oligos from either Invitrogen or IDT. Gel shift probe labeling reactions contained 37 µl dH_2O, 5 µl 10X PNK Buffer, 1 µl top strand oligo (2 µM), 1 µl bottom strand oligo (2 µM), 5 µl $\gamma^{32}P$, and 1 µl T4 PNK (NEB). Reactions were incubated at 37°C for one hour, boiled at 80°C for 5 min, then allowed to cool slowly to room temperature. Labeled probes were purified twice using Illustra

ProbeQuant G-50 Micro Columns (GE). The Lz probe contained the first Lz site from *spa* along with flanking sequence. The Su(H) probe contained the third Su(H) site from *spa* along with flanking sequence. The sequences of labeled probes are below, with Lz sites in blue, Su(H) sites in red, and binding site mutations in bold and underlined:

Lz: 5′-aaaatttactatGACCGCAaagctgtttcc-3′

MutLz: 5′-aaaatttactatGA**AA**GCAaagctgtttcc-3′

Su(H): 5′-tcaagatcttaTTCACATTgaaattgaagc-3′

MutSu(H): 5′-tcaagatcttaTT**GGG**ATTgaaattgaagc-3′

Cold competitors were assembled as follows: 10X - 25 µl dH_2O, 5 µl 10X PNK Buffer, 10 µl top strand oligo (2 µM), 10 µl bottom strand oligo (2 µM); 100X - 43 µl dH_2O, 5 µl 10X PNK Buffer, 1 µl top strand oligo (200 µM), 1 µl bottom strand oligo (200 µM); 1000X - 25 µl dH_2O, 5 µl 10X PNK Buffer, 10 µl top strand oligo (200 µM), 10 µl bottom strand oligo (200 µM). Cold competitors were incubated at 37°C for one hour, boiled at 80°C for 5 min, then allowed to cool slowly to room temperature. Sequences of cold competitor probes match the sequences of *in vivo* reporter constructs, and include region of interest along with adjacent binding sites and additional flanking sequence on either side. Sequences of cold competitors are below, with Lz sites in blue, Su(H) sites in red, Ets sites in green, binding site mutations in bold and underlined, and mutations outside binding sites in alternating caps and lower case, where the bases in caps have been mutated by non-complementary transversion:

1WT: 5’-gtatcaagtaactgggtgcctaattgaaaaaatttactatGACCGCAaagctgtttc-3′

m1A^{NS}: 5′-gGaGcCaTtCactgggtgcctaattgaaaaaatttactatGACCGCAaagctgtttc-3′

m1B^{NS}: 5′-gtatcaagtaacGgTgGgAcGaattgaaaaaatttactatGACCGCAaagctgtttc-3′

m1C^{NS}: 5′-gtatcaagtaactgggtgcctaCtGgCaCaCatttactatGACCGCAaagctgtttc-3′

1BSmut: 5′-gtatcaagtaactgggtgcctaattgaaaaaatttactatGA**AA**GCAaagctgtttc-3′

4WT:

5'caagatcttaTTCACATTgaaattgaagcactattggtgtacgattacaacgctcacattatcaGGATat aaaaaaaa-3′

m4^{NS}:

5′-caagatcttaTTCACATTgaaaGtTaCgAaAtCtGgTtTtCcTaGtCcCaAgAtAaAaGtCtcaGGAT ataaaaaaaa-3′

m4A^{NS}:

5′-caagatcttaTTCACATTgaaCtGgCaTcCcGaGtggtgtacgattacaacgctcacattatcaGGAT ataaaaaaaa-3′

m4B^{NS}:

5′-caagatcttaTTCACATTgaaattgaagcactattTgGgGaAgCtGaAaacgctcacattatcaGGAT ataaaaaaaa-3′

m4C^{NS}:

5′-caagatcttaTTCACATTgaaattgaagcactattggtgtacgattacCaAgAtAaAaGtCtcaGGAT ataaaaaaaa-3′

4BSmut:

5′-caagatcttaTT**G**ACATTgaaattgaagcactattggtgtacgattacaacgctcacattatca**TT**AT ataaaaaaaa-3′

5WT: 5′-ttatcaGGATataaaaaaaaggtgatagtaattcagcacgactttgtAACCACAaatata-3′

m5^{NS}: 5′-ttatcaGGATataaCaCaCaTgGgCtCgGaCtGcCgAaAgCcAttgtAACCACAaatata-3′

m5A^{NS}:

5′-ttatcaGGATataCaCaCaCgTtgatagtaattcagcacgactttgtAACCACAaatata-3′

m5B^{NS}:

5′-ttatcaGGATataaaaaaaaggtTaGaTtCaGtAagcacgactttgtAACCACAaatata-3′

m5C^{NS}:

5′-ttatcaGGATataaaaaaaaggtgatagtaattcaTcCcTaAtGtgtAACCACAaatata-3′

5BSmut:

5′-ttatca**TT**ATataaaaaaaaggtgatagtaattcagcacgactttgtAA**AA**ACAaatata-3′

6AWT: 5′-cagattactcCGTGAGTAcaacgtaagtcgggtgaagccagaAACCACAaatcaagttg-3′

m6A^{NS}: 5′-cagattactcCGTGAGTAcaacTtCaTtAgTgGgCaTcAagaAACCACAaatcaagttg-3′

6BSmut:

5′-cagattactcCGTGA**C**TAcaacgtaagtcgggtgaagccagaAA**AA**ACAaatcaagttg-3′

Gel shifts reactions were assembled on ice and contained 1 µl 10X Gel Shift 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), 1 µl poly d(I-C) (1 mg/ml), 1 µl DTT (100 µM), 1 µl labeled probe, 1 µl competitor (if included), and protein and dH_2O to a final volume of 10 µl. Lozenge gel shift reactions contained 3 µl of TNT reaction, while Su(H) gel shift reactions contained 5 µl of purified 6XHis-Su(H). Gel shift reactions were incubated on ice for 15 min before being loaded into 5% or 6% acrylamide gels that had been pre-run for 30 min; gels were then run at 120-140V in 0.5X TBE. Gels were dried for one hour at 80°C, exposed overnight to a storage phosphor screen (GE), and scanned using a Typhoon 9400 Variable Mode Imager. Quantification was performed using ImageJ. Calculations of competition were based on measurements of mean signal from which background had been subtracted.

ACKNOWLEDGMENTS

This manuscript was written by Christina I. Swanson, Nicole C. Evans, and Scott Barolo. This research was supported in part by a Center for Organogenesis Training Grant (5T32HD007505) to C.I.S. and by NIH grant GM076509 to S.B. We thank Ying Zhao, Zeeshaan Bhatti, and Trish Hinrichs for research support, and Ben Novitch and Ken Cadigan for sharing reagents. We are grateful to the following people for helpful discussions: Ken Cadigan, Tim Blauwkamp, Deneen Wellik, Ben Novitch, Doug Engel,

Tom Glaser, Trisha Wittkopp, Jim Posakony, Billy Tsai, Albert Erives, Robert Drewell, and many members of the lab.

Table S2.1

Supplemental Table S2.1. Summary of transgenic reporter expression data. All constructs are placed at -846 bp from promoter, except those labeled "-121," which are placed at -121 bp. CC, cone cell expression; PR, photoreceptor 1+6 expression. 3 to 14 independently derived transgenic lines were examined for each construct. See Figure 2.2 for scoring key. Asterisks on lines of *spa*(synth^{CS}) and 2x*spa*(synth^{CS}) denote lines with weak and incomplete R1+R6 expression, plus weak-to-moderate expression at the posterior margin of the eye disc, which, based on z-position and nuclear size, is likely to include peripodial cells.

figure	*spa* construct	TG line	CC	PR	figure	*spa* construct	TG line	CC	PR
1	wt	3a	++	-	2	Δ1	1	-	-
		4	+++	-			2b	-	-
		5	-	-			3a	-	-
		6	+++	-			3b	-	-
		7b	-	-			3c	-	-
		8b	+++	-			4a	-	-
		9a	+++	-			4b	-	-
		9b	+++	-			5b	-	-
		call	+++	-			5c	-	-
	synth^{NS}	1a	-	-			6b	-	-
		1b	-	-			call	-	-
		2a	-	-		Δ2	1	+	-
		2b	-	-			4	-	-
		3a	-	-			5a	+	-
		4a	-	-			5b	+	-
		5a	-	-			5c	-	-
		5b	-	-			6	+	-
		call	-	-			avg	+/-	-
	synth^{CS}	1a*	-	+/-		Δ3	1a	+++	-
		2a	-	-			1c	+++	-
		2b	-	-			2	++	-
		2c	-	-			3	+++	-
		3*	-	+			call	+++	-
		4*	-	+		Δ4	1a	+	-
		5a*	-	+			1b	+++	-
		5b	-	-			2a	+	-
		7a	-	-			2b	-	-
		7b*	-	+/-			2c	-	-
		7c*	-	+/-			2e	+	-
		10b*	-	+/-			call	+	-
		10c	-	-		Δ5	1a	++++	-
		call	-	+/-			1b	++++	-
							2b	++++	-
							2c	++++	-
							3a	++++	-
							3b	++++	-
							call	++++	-

figure	*spa* construct	TG line	CC	PR	figure	*spa* construct	TG line	CC	PR
2	Δ6	1a	-	-	3	m1a^{NS}	1a	+/-	-
(cont.)		1c	-	-			1b	-	-
		2	-	-			3a	-	-
		3a	-	-			3c	+	-
		3c	-	-			4a	-	-
		3d	+	-			4b	+/-	-
		3e	+	-			5	-	-
		3f	-	-			call	-	-
		4a	+	-		m1b^{NS}	1	+/-	-
		4b	+	-			2a	-	-
		call	+/-	-			2b	-	-
	m2^{NS}	2	+++	-			3	++	-
		4a	+++	-			4	+	-
		4b	+	-			6a	+	-
		5b	++	-			7a	+	-
		5c	++	-			8	++	-
		6a	-	-			9a	+++	-
		8	++	-			9b	++	-
		9a	++	-			call	+	-
		9b	++	-		m1c^{NS}	1	-	-
		9c	++	-			2a	+/-	-
		call	++	-			2b	+/-	-
	m4^{NS}	1	-	-			2c	+/-	-
		2a	-	-			3a	-	-
		2c	-	-			3b	-	-
		3	-	-			4a	-	-
		call	-	-			4b	-	-
	m5^{NS}	1a	+	-			call	-	-
		1b	-	-		m4a^{NS}	3	+/-	-
		1c	+	-			5b	+	-
		2b	-	-			7b	+	-
		3b	+	-			call	+	-
		4	+/-	-		m4b^{NS}	1b	-	-
		5a	+	-			3a	-	-
		5b	+	-			4b	-	-
		call	+	-			call	-	-
	m6a^{NS}	1a	+/-	-		m4c^{NS}	2a	+	-
		2b	+/-	-			3a	+/-	-
		3b	-	-			3b	++	-
		4	-	-			4a	-	-
		5a	-	-			call	+	-
		5b	-	-		m5a^{NS}	1a	-	-
		8b	-	-			1c	+	-
		8c	-	-			2	-	-
		call	-	-			3a1	+/-	-
							3b	+/-	-
							call	+/-	-

figure	*spa* construct	TG line	CC	PR	figure	*spa* construct	TG line	CC	PR
3	m5b^{NS}	2a	+/-	-	4	wt	1a	++++	-
(cont.)		3c	-	-		-121	2c	++++	-
		4a	+/-	-			3b	++++	-
		6b	+/-	-			3c	++++	-
		ca	+/-	-			4a	++++	-
	m5c^{NS}	1a	+	-			4b	++++	-
		1b	+/-	-			4c	++++	-
		2b	+	-			ca	++++	-
		4b	-	-		Δ1	1a	++++	-
		ca	+/-	-		-121	1b	++++	-
	m5d^{NS}	1a	-	-			1c	++++	-
		1b	+++	-			1d	++++	-
		1c	+++	-			2	++++	-
		1d	-	-			4a	++++	-
		3a	++	-			4b	++++	-
		3c	++++	-			5a	++++	-
		ca	+++	-			5b	-	-
	m6a^{NS}	1a	+/-	-			5c	++++	-
		2b	+/-	-			6b	++++	-
		3b	-	-			ca	++++	-
		4	-	-		m4^{NS}	1	-	-
		5a	-	-		-121	2	-	-
		5b	-	-			3	-	-
		8b	-	-			3b	-	-
		8c	-	-			3c	-	-
		ca	-	-			4	-	-
	m6b^{NS}	1a	+++	-			7	-	-
		3a	++	-			ca	-	-
		3b	++	-		m5^{NS}	1a	++++	+
		ca	++	-		-121	1b	+++	+
							1c	+++	-
							2	++	-
							2c	+++	-
							ca	+++	-
						m6a^{NS}	1a	++	-
						-121	1b	++	-
							2	+++	-
							4	++	-
							5	++	-
							6	-	-
							7	+++	-
							11	++	-
							14	++	-
							18	++	-
							ca	++	-

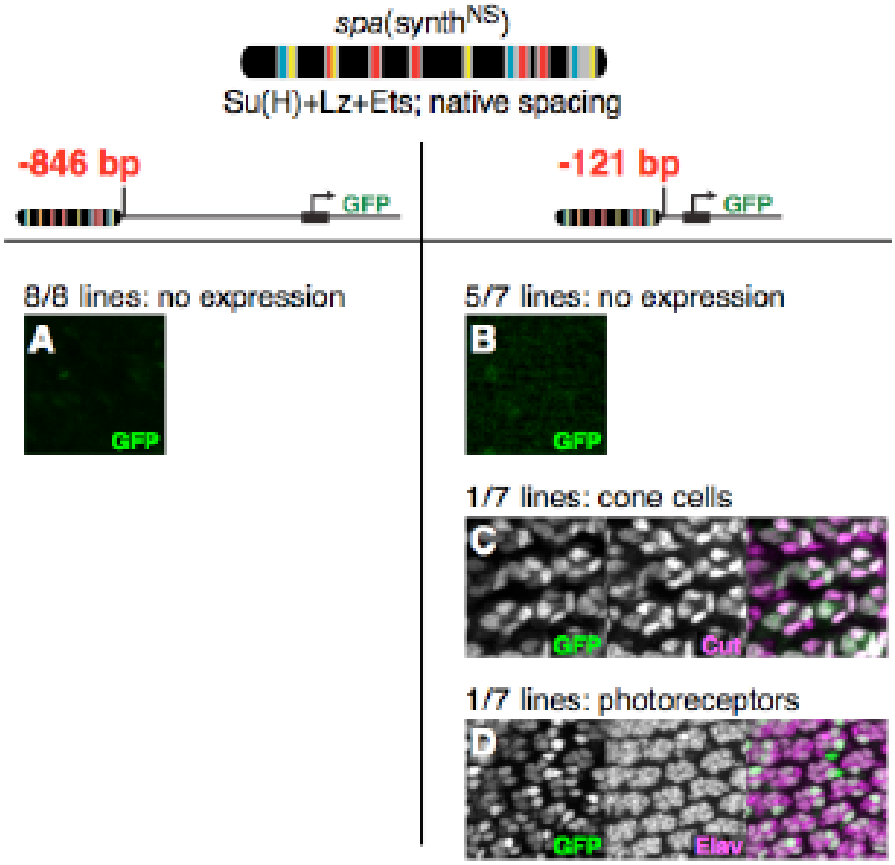
figure	*spa* construct	TG line	CC	PR	figure	*spa* construct	TG line	CC	PR
5	KO	2a	-	-	5	1+4+6a+	1	-	-
		2b	-	-	(cont.)	synth^{CS},	2	-	-
		2c	-	-		ΔEts	3b	-	-
		3	-	-			4	-	-
		4a	-	-			call	-	-
		5	-	-		1+4+6a+	1	-	++++
		call	-	-		synth^{CS},	2	-	+++
	KO+	1	-	+++		ΔSuH	4	-	++++
	synth^{CS}	2a	-	+++			5b	-	++++
		2b	-	++++			call	-	++++
		3a	-	++++		1+6a+	4	-	++
		3b	-	+++		synth^{CS}	8b	-	+
		4	-	+++			9a	-	+++
		5	-	++++			call	-	++
		6a	-	++++		2x	1'	-	+
		6b	-	++++		synth^{CS}	2a'	-	++
		7a	-	++++			2b'	-	++
		9a	-	+			5'	-	++
		10b	-	++++			call	-	++
		11a	-	-		2x	1a	-	-
		13a	-	-		synth^{NS}	2b	-	-
		call	-	++++			4	-	-
	KO+	1a	-	-			5b	-	-
	synth^{NS}	3b	-	-			6	-	-
		6a	-	-			call	-	-
		8a	-	-		m2,3,6b^{NS}	1b	–	-
		call	-	-			3	+++	-
	Δ1+1	1a	+++	-			4	+++	-
		1b	+++	-			6	+++	-
		1c	+++	-			7	++++	-
		call	+++	-			call	+++	-
	TF	6	+	-		m2,3,5,	1a	-	+
	scram-	10	+	-		6b^{NS}	1b	-	+++
	bled	11a	+	-			3	-	++++
		call	+	-			4c	-	++++
	1+4+6a+	1	-	+++			call	-	++++
	synth^{CS}	6	-	++++	6	*pse spa*	1	++	-
		10	-	++++			4a	-	-
		13b	-	+++			5	+++	-
		call	-	++++			6	+++	-
	1+4+6a+	5	-	+			8	+++	-
	synth^{CS},	6	-	-			call	+++	-
	ΔLz	9	-	+					
		13	-	+					
		call	-	+					

Table S2.2

Lz binding in vitro (given as % competition)				Su(H) binding in vitro (given as % competition)		
cold competitor fold excess:	10x	100x		cold competitor fold excess:	100x	1000x
Region 1:				Region 4:		
1 wt	32.8	74.2		4 wt	33.8	69.8
1 mBS	1.1	1.2		4 mBS	21.0	63.0
m1a	36.8	72.7		m4a	32.3	76.5
m1b	34.2	74.6		m4b	25.3	73.7
m1c	39.6	71.2		m4c	51.1	84.3
Region 5:				Region 6a:		
5 wt	37.7	79.4		6a wt	10.3	52.1
5 mBS	5.8	8.9		6a mBS	8.8	36.0
m5a	39.5	85.4		m6a	15.7	55.4
m5b	44.7	81.1				
m5c	39.7	78.4				
Region 6a:						
6a wt	35.1	75.1				
6a mBS	5.6	11.5				
m6a	44.4	72.6				

Supplemental Table S2.2. Novel regulatory regions of *spa* do not significantly contribute to in vitro binding of the known regulatory TFs. Experimental details and probe sequences are described in Experimental Procedures. Probes include novel regulatory regions 1, 4, 5, or 6a, plus any immediately flanking Lz, Su(H), or Ets binding sites (see Figure 2.6A for annotated sequence). Left, Lz binding to sites flanking regions 1, 5, and 6a; right, Su(H) binding to sites flanking regions 4 and 6a. Pnt binding was not examined, because the predicted Ets sites flanking these enhancer regions do not bind Pnt in vitro (Flores et al, 2000). In all cases, mutations to novel regulatory regions of *spa* (all of which reduce enhancer activity in vivo) do not significantly affect binding of Lz or Su(H) to nearby sites. By contrast, binding is reduced in all cases by mutating the Lz or Su(H) sites themselves (red figures). Note that mutating Su(H) sites has a milder effect on binding than mutating Lz sites, and that a greater excess of competitor is required for Su(H) than for Lz. This apparent difference in relative binding strength is consistent with the fact that the Lz sites perfectly match the optimal binding consensus, while the Su(H) sites deviate from the optimal consensus and are predicted to be lower-affinity sites (Flores et al, 2000; Nellesen et al, 1999).

Figure S2.3



Supplemental Figure S2.3. A synthetic version of *spa*, containing only Su(H)+Lz+Ets binding sites, is poised for activation in multiple cell types of the developing eye. *spa*(synth^{NS}), a synthetic enhancer containing the Su(H)+Lz+Ets sites from *spa* in their native arrangement, placed at -846 bp upstream of a promoter, is inactive in all cell types of the eye, in all 8 lines examine (A). When this construct was placed closer to the promoter (-121 bp), no expression was observed in 5 of 7 independent transgenic lines (B). However, 2 of 7 lines show insertion site-dependent activity in the eye: one line is active in cone cells, as shown by co-expression with Cut (C), while another line is active in multiple photoreceptors, as shown by co-expression with Elva (D).

Table S2.4

Enhancer	Total bp	Conserved bp (%)[a]	TFBS total bp[b]	TFBS conserved bp (%)[c]	9/10 conserved blocks (%)[d]	10/10 conserved blocks (%)[e]	20/20 conserved blocks (%)[e]
E(spl)m4[f]	279	87.8%	38	97.4%	88.9	74.6	58.1
*Su(H)*ASE[g]	372	82.8%	40	100.0%	73.4	63.2	40.3
*dpp*VM[h]	419	70.9%	89	100.0%	65.6	51.3	21.7
*eve*MHE[i]	311	74.3%	130	71.5%	63.0	45.7	29.3
*eve*st.2[j]	483	73.1%	163	84.7%	60.7	37.7	19.9
*dpp*D[k]	372	69.6%	24	100.0%	48.3	36.5	26.1
mean±SEM (excluding *spa*)		76.4% ±3.0%		92.3% ±4.8%	66.6 ±5.6	51.5 ±6.1	32.6±5.9
spa[l]	362	64.6%	73	72.6%	29.3	3.9	0
mean±SEM (including *spa*)		74.7% ±3.0%		89.5% ±4.9%	61.3 ±7.1	44.7 ±8.5	27.9 ±6.8

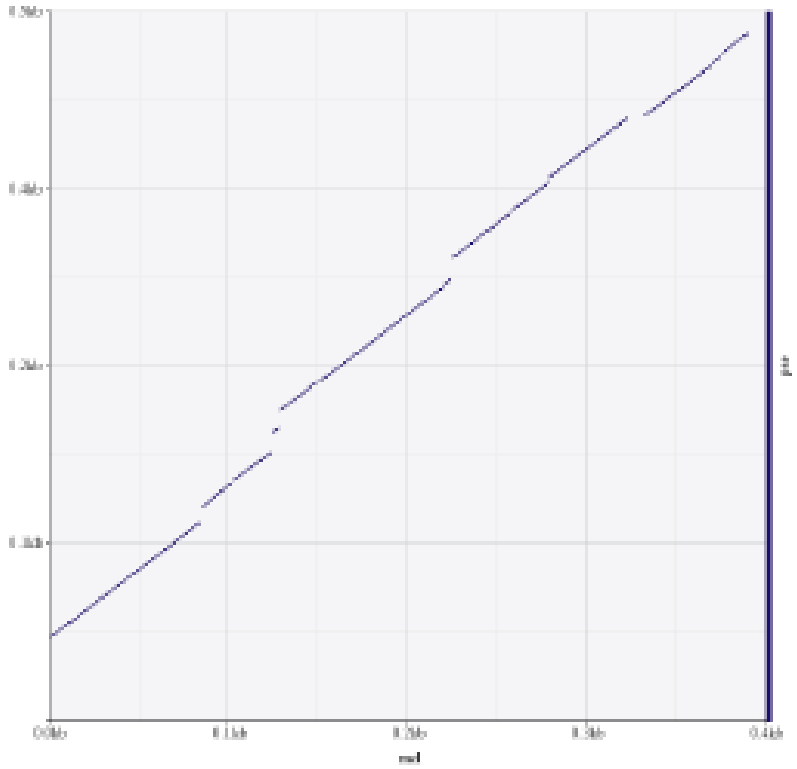
Supplemental Table S2.4. *D. melanogaster-D. pseudoobscura* sequence identity and blocks of conservation within *spa* and six other developmental enhancers. (a) Derived from BLASTZ alignments of orthologous enhancer sequences. Calculated as ((# matched bases in a BLASTZ alignment)/(*mel* sequence length)) X 100%. (b) TFBSs, transcription factor binding sites. Binding site definitions taken from previous reports, references below. (c) *mel-pse* sequence identity of all defined or predicted transcription factor binding sites (TFBSs) in seven enhancers. Orthologous sequences are taken from BLASTZ alignments. (d) Derived from BLASTZ alignments of orthologous enhancer sequences. Defined as the total length of contiguous sequences of ≥10 bp in which ≥90% of bp are conserved (counting gaps as mismatched bases), as a percentage of total enhancer length. (e) Defined as the total length of contiguous conserved sequences of ≥10 bp or ≥20bp (counting gaps as mismatched bases), as a percentage of total enhancer length. (f) Nellesen et al, 1999; (g) Barolo et al, 2000; (h) Sun et al, 1995; Yang et al, 2000; Zaffran et al, 2001; Stultz et al, 2006; (i) Halfon et al, 2000; (j) Small et al, 1992; Arnosti et al, 1996; Andrioli et al, 2002; (k) Muller and Basler, 2000; (l) Flores et al, 2000.

Table S2.5

Enhancer	bl2seq alignable (%)[a]	DroSpeGe BLAST alignable (%)[b]
E(spl)m4	53.8	42.8
*Su(H)*ASE	41.5	54.2
*dpp*VM	98.9	79.2
*eve*MHE	64.6	29.6
*eve*st.2	48.0	31.1
*dpp*D	91.1	72.0
mean±SEM (excluding *spa*)	66.3 ±9.6	51.5±8.5
spa	24.3	0
mean±SEM (including *spa*)	60.3±10.1	44.1±10.3

Supplemental Figure S2.5. Alignability of *D. melanogaster* and *D. pseudoobscura* orthologs of *spa* and six other developmental enhancers. (a) *mel* enhancer sequence alignable with orthologous *pse* sequence by the BLAST 2 Sequences program at NCBI (www.ncbi.nlm.nih.gov/BLAST/bl2seq/wblast2.cgi; Tatusova and Madden, 1999). Default settings were used, with the following exceptions: the "open gap" penalty was set to 2, the "extension gap" penalty was set to 1, and the low-complexity filter was disabled. (b) *mel* enhancer sequence alignable with total *pse* genomic sequence by the BLAST program at DroSpeGe (insects.eugenes.org/species/blast; Gilbert, 2007). Default settings were used, except that the low-complexity filter was disabled.

Table S2.6



Supplemental Table S2.6. Dot-plot alignment of *D. melanogaster spa* with orthologous sequences from *D. pseudoobscura*. The plot was created in zPicture (Ovcharenko et al, 2004; http://zpicture.dcode.org) from a BLASTZ alignment. Co-linear sequence conservation indicates that no large-scale rearrangements of enhancer structure have occurred since the divergence of *D. melanogaster* and *D. pseudoobscura*.

CHAPTER III

CONSERVATION OF ENHANCER FUNCTION DESPITE EVOLUTION OF SEQUENCE AND ORGANIZATION

INTRODUCTION

Enhancers are cis-regulatory elements (CREs) that control the level, timing, and cell-type specificity of gene expression. Enhancers are composed of transcription factor binding sites (TFBSs) which recruit sequence-specific transcription factors and cofactors in order to regulate the transcriptional state of the target gene (Orphanides and Reinberg, 2002). In some enhancers, the organization of TFBSs is also critical for enhancer function (Thanos and Maniatis, 1995; Erives and Levine, 2004; Senger et al, 2004). Therefore, both the DNA sequence and the structure are important for enhancer function. However, evolutionary analyses of CREs reveal rapid evolution of enhancer sequences that could alter both TFBS sequence and overall enhancer structure, potentially affecting enhancer function (Richards et al, 2005; Moses et al, 2006; Li et al, 2007).

Several studies have shown that evolution of enhancer sequences can lead to changes in gene expression; in some cases, enhancer evolution is responsible for morphological differences between two species (Shapiro et al, 2004; Gompel et al, 2005;

Marcellini and Simpson, 2006, Jeong et al, 2008; Williams et al, 2008). However, it seems likely that selective pressure acts to preserve the function of most CREs. Indeed, there are many examples of enhancers that have retained their function despite sequence divergence (Takahashi et al, 1999; Fisher et al, 2006; Piano et al, 1999; Yamamoto et al, 2007; Wratten et al, 2006). In some of these examples, the sequence is so degenerate that typical alignment tools are unable to align functionally conserved elements (Fisher et al, 2006; Wratten et al, 2006). In most cases, the direct regulatory inputs governing enhancer activity are conserved, yet the number and organization of transcription factor binding sites (TFBS) has changed, suggesting rapid binding site turnover has occurred (Piano et al, 1999; Wratten et al, 2006). It has been suggested that enhancer activity is maintained despite mutation of important regulatory sequences due to binding site redundancy, compensatory mutations, and organizational flexibility (Ludwig, 2002; Wittkopp, 2007). However, these hypotheses have rarely been tested experimentally.

The best *in vivo* evidence for the structure-function implications of enhancer evolution comes from analyses of the *Drosophila even-skipped* stripe 2 enhancer (*eve* S2E). This enhancer drives expression of *eve* in a transverse stripe of cells in the syncytial *Drosophila* embryo (Goto et al, 1989; Harding et al, 1989). The function of *eve* S2E is conserved not only among Drosophilids, but also among a more distantly related group of flies, the sepsids (Ludwig et al, 1998; Hare et al, 2008). However, S2E sequence comparisons reveal significant divergence, including poor conservation of binding sites and changes in spacing between TFBS (Ludwig and Kreitman, 1995; Ludwig et al, 1998; Hare et al, 2008). It was proposed that the *eve* S2E is evolving under stabilizing selection, and that compensatory mutations have occurred in each lineage to

preserve enhancer function despite turnover of functionally important binding sites. This hypothesis was directly tested for the *Drosophila melanogaster* (*D. mel*) and *Drosophila pseudoobscura* (*D. pse*) stripe 2 enhancers. Chimeric enhancers containing sequences from both the *D. mel* and *D. pse* stripe 2 enhancers cannot recapitulate the endogenous stripe 2 expression pattern, presumably because compensatory mutations in both lineages have rendered *D. mel* and *D. pse* enhancer sequences incompatible (Ludwig et al, 2000). However, beyond this elegant work with the *eve* S2E, very little has been done to examine the nature of enhancer evolution *in vivo*.

Evolutionary comparisons have the potential to provide valuable insight into the rules that govern structure and function of enhancer elements. Therefore, we decided to analyze the evolution of the *Drosophila D-Pax2 sparkling* enhancer (*spa*). *Spa* is an excellent candidate for evolutionary analysis for several reasons: the regulation of *spa* has been well characterized; all the regulatory sequences within a minimal version of *spa* have been finely mapped; unlike *eve* S2E, it is signal-regulated enhancer; previous studies have identified strict constraints on enhancer structure; and sequence alignments reveal rapid evolution of enhancer sequences (Flores et al, 2000; Swanson et al, 2010).

The *D. mel spa* enhancer drives cone cell-specific expression of *D-Pax2* during *Drosophila* eye morphogenesis; expression of *D-Pax2* in the cone cells is essential for their proper fate specification (Fu and Noll, 1997; Fu et al, 1998). The *spa* enhancer contains twelve mapped TFBS which mediate three direct regulatory inputs: Notch signaling via binding of Suppressor of Hairless (Su(H)); EGFR/MAPK signaling via binding of the Ets proteins Pointed P2 (PntP2, an activator) and Yan (a repressor); and Lozenge (Lz), a RUNX-family activator (Figure 3.1A, Flores et al, 2000). In addition to

Figure 3.1

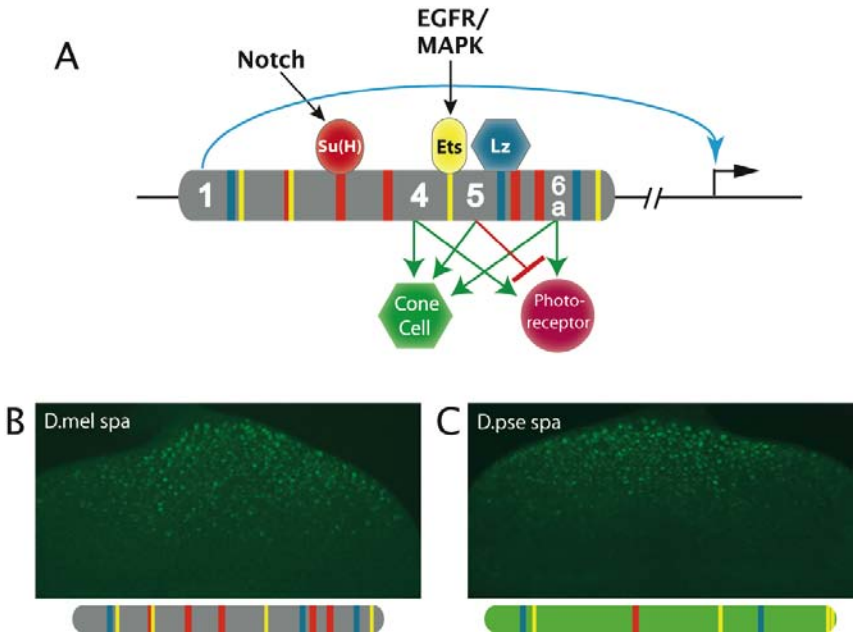


Figure 3.1. Functional conservation of the *D. melanogaster* and *D. pseudoobscura sparkling* enhancers. (A) The regulation of *D. mel spa* is complex; in addition to twelve characterized binding sites for Lozenge (blue), the Ets proteins PointedP2 and Yan (yellow), and Suppressor of Hairless (red), *spa* contains essential regulatory elements in the regions labeled 1, 4, 5, and 6A (Flores et al, 2000; Swanson et al, 2010). The regulatory sequences in Regions 4 and 6A contribute to transcriptional activation in multiple cell types, whereas regulatory sequences in Region 5 contribute to cone cell-specific activation. Region 5 also contains a position-dependent element that represses gene expression in the photoreceptors. Region 1 is not required for proper transcriptional activation or patterning when the enhancer is proximal to the promoter, but is required when the enhancer is located at a distance from the promoter, suggesting that Region 1 mediates long-distance enhancer-promoter interactions (blue arrow). The organization of these regulatory elements within *D. mel spa* is also critical for proper transcriptional activation and gene patterning. (B) The *D. mel spa* enhancer, placed 846 bp upstream of a GFP reporter, drives cone cell-specific gene expression in late third instar larval imaginal eye discs. (C) The *D. pse spa* enhancer drives cone cell-specific gene expression at levels indistinguishable from the *D. mel spa* enhancer.

the mapped binding sites for Su(H), PntP2/Yan, and Lz, there are critical regulatory elements in several other regions of the enhancer, although the proteins that may bind those elements have not yet been identified (Figure 3.1A, Swanson et al, 2010). Alignments reveal an overall lack of sequence conservation within *spa*, and more specifically poor conservation of critical regulatory elements. Nevertheless, the orthologous *D. pse spa* enhancer is capable of driving reporter gene expression in a pattern that is indistinguishable from that driven by the *D. mel spa* enhancer (Figure 3.1B-C, Swanson et al, 2010). We have characterized the differences in regulation and organization of the *D. mel* and *D. pse spa* enhancers in order to understand how two enhancers with such different sequences and structures can generate identical transcriptional outputs. By building chimeric enhancers containing sequences from both *mel spa* and *pse spa*, we discovered that the organization of regulatory elements within these two enhancers is surprisingly divergent. In addition, while both enhancers respond to the same direct regulatory inputs, binding site turnover events have led to changes in TFBS composition. Our findings provide insight into the mechanisms by which *spa* function was maintained despite mutation of important regulatory sequences and remodeling of enhancer organization.

RESULTS

Conservation of function, but not sequence, in the *spa* enhancer

Previous studies have revealed complexity in both the regulation and organization of the *D. mel spa* enhancer. As stated above, *spa* is directly bound by the transcription factors Lz, PntP2/Yan, and Su(H); each of these inputs is required for enhancer function (Figure 3.1A, Flores et al, 2000). The sequences outside these characterized binding sites were subdivided into smaller regions (1A-1C, 2, 3, 4A-C, 5A-D, 6A-B) that were individually tested for regulatory function (Figure 3.1A, Figure 3.2; Swanson et al, 2010). Regions 1A through 1C, 4A through 4C, 5A through 5C, and 6A were all found to make essential contributions to enhancer function; for the duration of this report these sequences will be referred to as Regions 1, 4, 5, and 6A. The regulatory functions of these novel regions are distinct from one another; Region 1 is required only when the enhancer is located at a distance from the promoter, Regions 4 and 6A can contribute to transcriptional activation in multiple cell types, and Region 5 is necessary for both activation in cone cells and repression in photoreceptors (Figure 3.1A). Thus *D. mel spa* is combinatorially regulated by Lz, Pnt, Yan, Su(H), and an unknown number of additional regulatory factors binding within Regions 1, 4, 5, and 6A.

The organization of these regulatory elements within *D. mel spa* is also a critical component of enhancer function. Changes in spacing between TFBS can alter the transcriptional output of *spa*, sometimes resulting in impaired enhancer activity, and

Figure 3.2

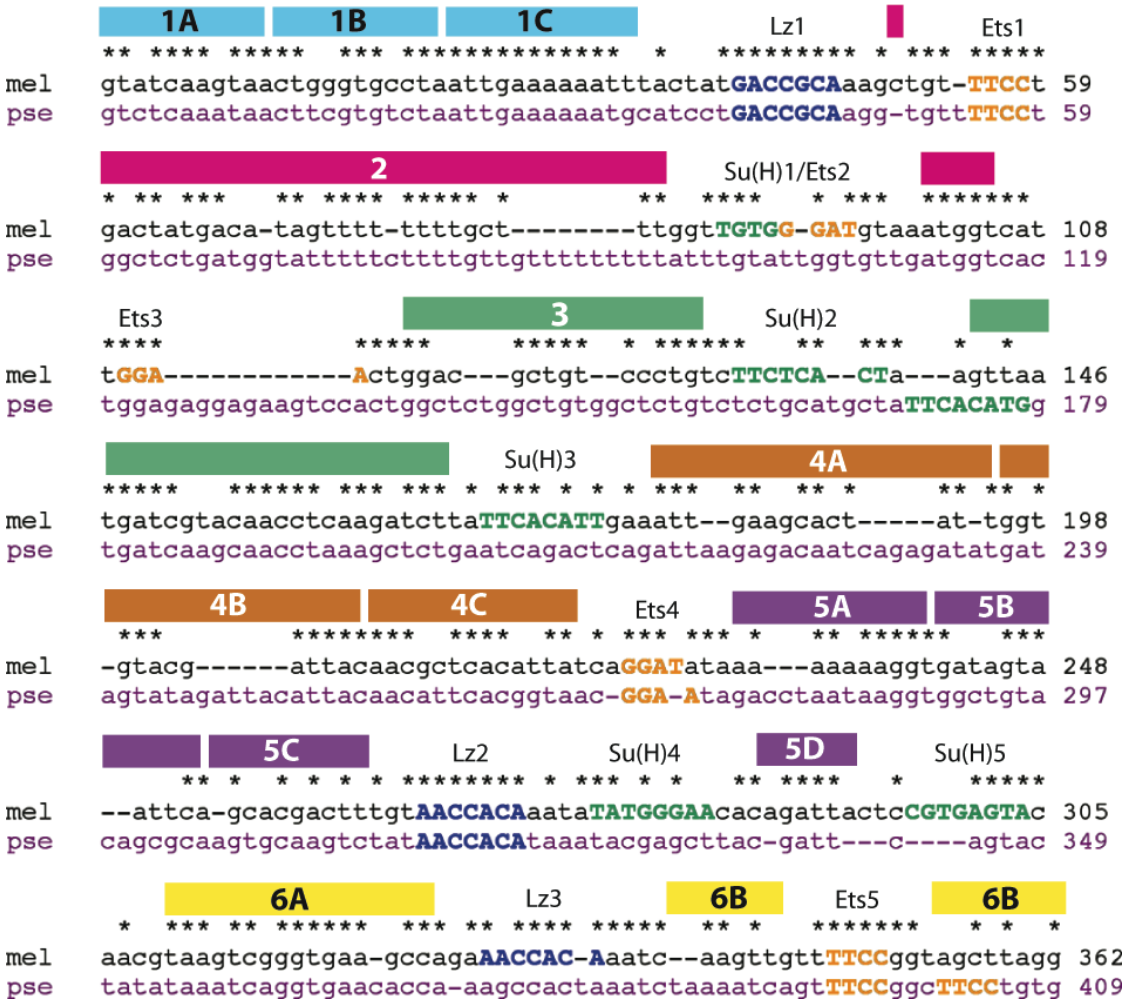


Figure 3.2. Alignment of *D. mel* and *D. pse spa* enhancer sequences. The alignment was generated by ClustalW using the default settings (Chenna et al, 2003); see supplemental figures for additional alignments. Lz binding sites are blue, Ets binding sites are orange (rather than yellow as in other figures, so that text is readable), and Su(H) sites are green. Binding sites in *D. mel spa* have been confirmed experimentally (Flores et al, 2000); binding sites in *D. pse spa* are predicted and have not yet been tested for *in vitro* binding. Colored boxes above sequence designate *D. mel* regions as defined in our previous analysis of *spa*. Stars marked conserved bases. This alignment was used to design regional swap experiments described in Figure 3.3.

other times resulting in increased transcriptional activation (Swanson et al, 2010). Rearrangement of regulatory elements within *spa* can also impair enhancer function; in some cases, rearrangements can switch the cell-type specificity of the enhancer, activating transcription in photoreceptors but not in cone cells (Swanson et al, 2010). Therefore, spatial relationships among regulatory elements in *D. mel spa* are crucial for proper transcriptional activation and patterning of gene expression. Because of the complex nature of *spa* structure and function, we expected the evolution of *spa* sequence to be highly constrained. Surprisingly, that is not the case; *spa* sequences appear to be evolving rapidly based on an overall lack of sequence conservation (Figures 3.2, S3.1, S3.2).

We tested the functional conservation of the *D. mel* and *D. pse spa* enhancers by comparing their ability to drive reporter gene expression in transgenic *Drosophila* imaginal eye discs. For all reporter gene constructs, we examined reporter gene expression in at least five independent transgenic lines. Although the sequences of the *D. mel* and *D. pse spa* enhancers are divergent, they drive reporter gene expression in patterns that are indistinguishable from one another, both in level and cell-type specificity (Figure 3.1B-C). We concluded that the *spa* enhancer is functionally conserved between *D. mel* and *D. pse*.

Sequence alignments reveal that many of the regulatory elements critical to proper function of *D. mel spa* are poorly conserved (Figures 3.2, S3.2, S3.2). Only two regulatory elements from *D. mel spa* are clearly conserved among all twelve sequenced *Drosophila* species: the Lz1 site and the Region 1 regulatory element (Figure 3.2, Figure S3.1). *D. mel* and *D. pse spa* also share the Ets1 site (Figure 3.2). Beyond those three

regulatory elements, it is difficult to say with confidence whether any of the remaining regulatory elements in *D. mel* spa have direct orthologs in *D. pse spa*; it is possible that the single Su(H) site, the second Lz site, and one or two of the Ets sites in *D. pse* are orthologous to sites in *D. mel spa*, but it seems equally likely that these sites are not orthologous to *D. mel* sites based on sequence and position within the *D. pse* enhancer. Overall, there are only seven putative sites for Lz, Ets, and Su(H) proteins in *D. pse*, as opposed to twelve binding sites for those proteins in *D. mel* (Figure 3.2). Further experimentation is necessary to characterize putative binding sites in *D. pse spa*, both *in vitro* and *in vivo*. Finally, although the specific, essential regulatory sequences within Regions 4, 5, and 6A have not been identified, we do not detect sequence conservation in those regions. In conclusion, sequence comparisons reveal a lack of conservation of essential regulatory elements and a significant change in TFBS composition in the *spa* enhancer despite a conservation of overall enhancer function.

The *sparkling* enhancer is evolving under stabilizing selection

The composition of binding sites within *spa* has been rapidly evolving; although the conservation of unmapped TFBS from 4, 5, and 6 is unclear, most binding sites for Lz, PntP2/Yan, and Su(H) are not strictly conserved. Such rapid turnover of binding site composition suggests that compensatory mutations may have occurred within *D. pse spa* to preserve its function. To test this hypothesis, we built chimeric constructs combining sequences from both the *D. mel spa* and *D. pse spa* enhancers and tested whether regions from these enhancers were functionally interchangeable.

Figure 3.3

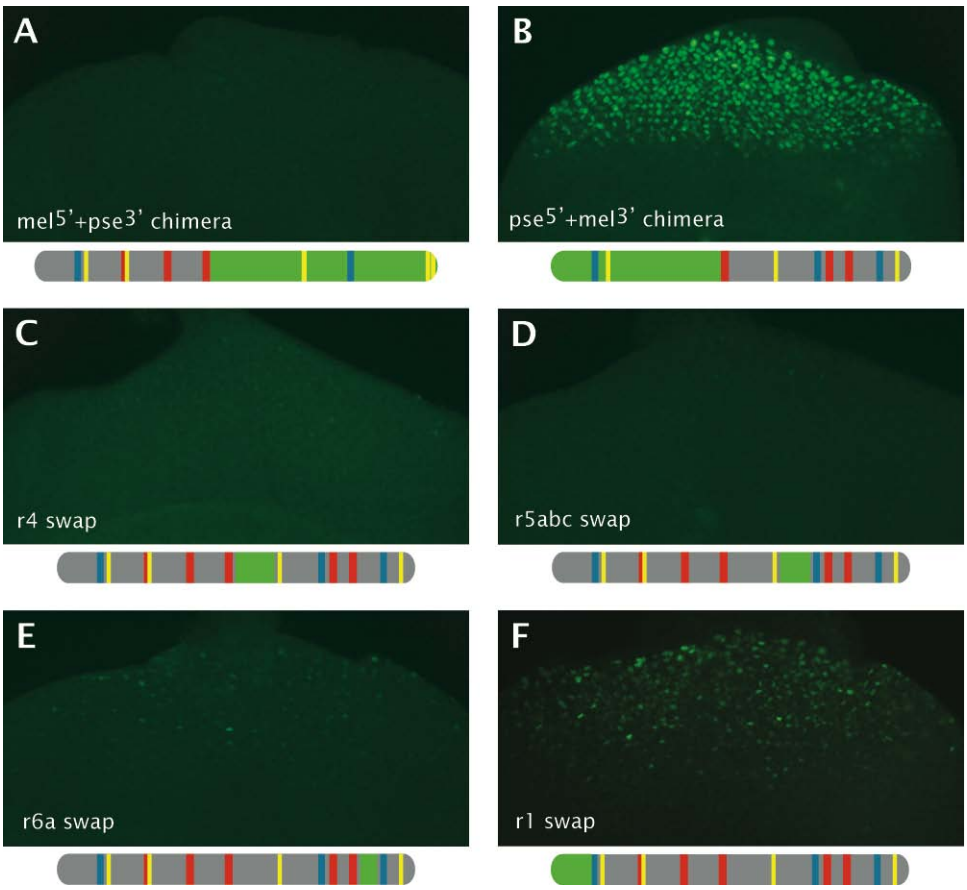


Figure 3.3. *D. mel spa* and *D. pse spa* sequences are not functionally equivalent. (A-B) Chimeric constructs combining sequences from *D. mel* and *D. pse spa* are unable to exactly recapitulate the wild-type expression of either endogenous enhancer; gray sequences are from *D. mel* and green sequences are from *D. pse*. The $mel^{5'}+pse^{3'}$ chimera is unable to activate transcription (A), while the $pse^{5'}+mel^{3'}$ chimera activates cone cell-specific expression, but at elevated levels compared to either wild-type enhancer (compare to Figure 3.1 B-C). Furthermore, *D. mel* Region 4, 5, and 6A sequences cannot be replaced by orthologous sequences from *D. pse spa* (C-E). The function of Region 1 is conserved, as orthologous sequence from *D. pse spa* can functionally replace *D. mel spa* Region 1 (F).

First, we tested the ability of the 5' half of *D. mel spa* and the 3' half of *D. pse spa* to cooperatively activate gene expression. Surprisingly, this chimeric construct was unable to drive reporter gene expression *in vivo* (Figure 3.3A), indicating essential regulatory sites are absent from this construct. In contrast, the reciprocal chimera, 5' *D. pse spa* plus 3' *D. mel spa*, drives robust cone-cell specific gene expression, but at levels that exceed the expression of either intact endogenous enhancer (Figure 3.3B). The ability of this construct, $pse^{5'}$+$mel^{3'}$ chimera, to drive gene expression at elevated levels suggests an increased number of activator sites in this chimeric enhancer compared to either wild-type enhancer; interestingly, this chimera only contains 9 Lz/Ets/Su(H) sites, fewer than *D. mel spa*, indicating that extra activation may come from unmapped binding sites. Neither chimeric enhancer can recapitulate wild-type *spa* function; one enhancer is transcriptionally inactive, while the other activates cone cell-specific expression, but not at wild-type levels. The lack of compatibility between the *mel* and *pse* 5' and 3' halves suggests that the differences in TFBS composition and organization between these two enhancers are functionally relevant..

The lack of activity from $mel^{5'}$+ $pse^{3'}$ indicates that the 3' end of *D. pse spa* lacks critical regulatory elements that are located at the 3' end of *D. mel spa*. *D. mel spa* contains essential regulatory sequences in Regions 4, 5, and 6A that do not appear to be conserved based on sequence alignments. However, it seemed possible that some or all of these regions are functionally conserved, but that conservation is not apparent, perhaps because of degeneracy in binding site sequence. We tested the functional conservation of these regions by replacing sequence in *D. mel spa* with orthologous sequence from *D. pse spa*. Orthologous sequence was defined by ClustalW alignment (Figure 3.2), although

the lack of sequence conservation made it difficult to be certain we were truly swapping sequence that was orthologous in any way. Mutation of any of these regions in *D. mel spa* severely impairs enhancer function, and the replacement of that sequence with *D. pse* sequence was unable to rescue any of these enhancers (Figure 3.3C-E). A similar experiment in which Region 1 of *D. mel spa* was replaced with the orthologous sequence from *D. pse spa*, on the other hand, did not affect enhancer function (Figure 3.3F). This result was expected based on the high conservation of Region 1 sequence. We conclude that regulatory elements in Region 1 of *D. mel spa* and *D. pse spa* are conserved, but regulatory elements in *D. mel* Regions 4, 5, and 6A are not. It is possible that the *D. pse* enhancer is not directly regulated by the Region 4, 5, and 6A inputs; alternatively, compensatory sites for those inputs may have been gained elsewhere in *D. pse spa* to maintain overall enhancer function.

Compensatory evolution and reorganization of elements within *D. pse spa*

Our chimeric constructs revealed that many of the regulatory elements within the 3' portion of *D. mel spa* (Regions 4, 5, and 6A) appear to be absent from the 3' portion of *D. pse spa*. In order to understand how stabilizing selection has acted to preserve the function of *D. pse spa*, we attempted to identify compensatory elements in that enhancer. We performed these experiments in the context of $pse^{5'}+mel^{3'}$, enabling us to individually assess potential redundancy of regulatory elements that reside in Regions 4, 5, and 6A of *D. mel spa* with sequences in the 5' half of *D. pse spa*.

Figure 3.4

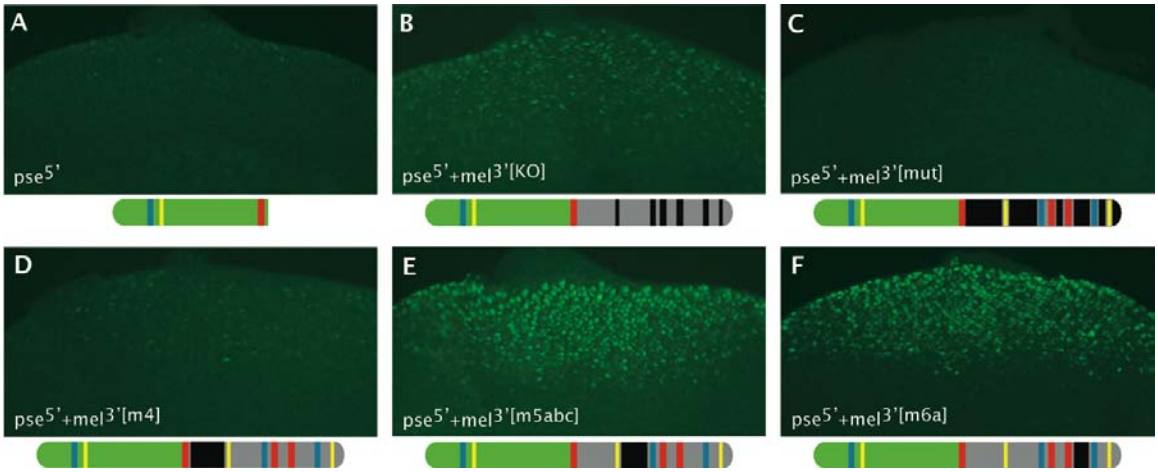


Figure 3.4. Differences in organization of regulatory elements between *D. mel* and *D. pse spa*. (A) The 5' half of *D. pse spa* is insufficient to activate robust gene expression on its own. (B) *pse*$^{5'}$ does not require additional input from Lz/Ets/Su(H) to drive cone cell expression at wild-type levels; black boxes represent mutated Lz/Ets/Su(H) sites. (C) However, *pse*$^{5'}$ does require additional regulatory input from *D. mel* Regions 4, 5, and 6A; black regions have been mutated. (D) Mutation of *D. mel* Region 4 results in a severe loss of activity in the context of the *pse*$^{5'}$*+mel*$^{3'}$ chimera (compare to Figure 3.3B), whereas mutation of either Region 5 or Region 6A in that context does not affect the activity of the chimeric enhancer (E-F).

Our previous set of experiments suggested that much of the regulatory function of *D. pse spa* was contained within its 5' half, particularly since the 3' half of this enhancer was not able to complement the 5' half of *D. mel spa* at all (Figure 3.3A). However, the 5' half of *D. pse spa* is not sufficient on its own to activate wild-type levels of gene expression (Figure 3.4A). We then asked whether $pse^{5'}$ requires contributions from the Lz/Ets/Su(H) sites, or from the regulatory elements within Regions 4, 5, and 6A. Surprisingly, the 5' half of *D. pse spa* requires no additional contribution from the 3' *mel* Lz/Ets/Su(H) sites to drive wild-type levels of gene expression, despite the fact that 5' *D. pse spa* contains only a single binding site for each of these proteins (Figure 3.4B). However, mutation of the *mel* sequences outside the Lz, Ets, and Su(H) sites resulted in a complete loss of enhancer activity in the context of $pse^{5'}$+$mel^{3'}$ (Figure 3.4C). Clearly, $pse^{5'}$ requires some or all of the sequences in Regions 4, 5, and 6A for proper function. Therefore, it seemed likely that these regulatory inputs are indeed conserved in *D. pse*, and compensatory sites for these inputs have evolved in novel locations in *D. pse spa*. Our next set of experiments sought to map the presence of Region 4, 5, and 6A regulatory elements in the *D. pse spa* enhancer.

Again in the context of $pse^{5'}$+$mel^{3'}$, we made targeted mutations in regions 4, 5, and 6A of the *mel* 3'sequence to determine whether the activities in those regions were complemented by regulatory elements in the 5' half of *D. pse spa*. We found that mutation of regions 5 or 6A had no effect on gene expression in this chimeric construct, indicating that regulatory elements that are functionally equivalent to elements in 5 and 6A reside in the 5' half of *D. pse spa* (Figure 3.4E-F). Mutation of Region 4, however, resulted in a significant reduction in reporter gene expression (Figure 3.4D). Therefore,

the 5' half of *D. pse spa* does not contain sequences that can functionally replace the regulatory element within *D. mel spa* Region 4. We suspect that this regulatory element is in fact conserved in *D. pse*, but may be located more 3' to the sequence we considered to be the orthologous to *D. mel* Region 4 in the r4 swap construct; future experiments will test this hypothesis. These data suggest a model in which the TFBS composition of *D. pse* and *D. mel spa* are conserved overall, but individual TFBS are not conserved. Instead, multiple binding site loss and gain events have resulted in two enhancers with conserved function but very different TFBS organizations.

Conservation of Regulation by Lz, Ets, and Su(H)

While it appears that the regulatory inputs that control *spa* are conserved, the relative contribution of each of these inputs to enhancer function may not be entirely static. In particular, the relative lack of Lz, Ets, and Su(H) sites in *D. pse spa* is surprising. We wanted to explore the possible reasons behind differences in the number of binding sites for these proteins, in hopes of better understanding the regulatory circuitry that governs these two enhancers.

The individual requirement for Lz, Notch, and EGFR inputs in regulation of *D. mel spa* has been previously demonstrated (Flores et al, 2000). However, the sites in *D. pse spa* are putative and have not been tested *in vivo*. We mutated these sites in reporter constructs to test the direct regulation of *D. pse spa* by Lz, PntP2/Yan, and Su(H). Because our reporter constructs differ in several respects from the constructs used to test the *mel spa* sites *in vivo* in the past (they drive expression of GFP rather than LacZ, and

Figure 3.5

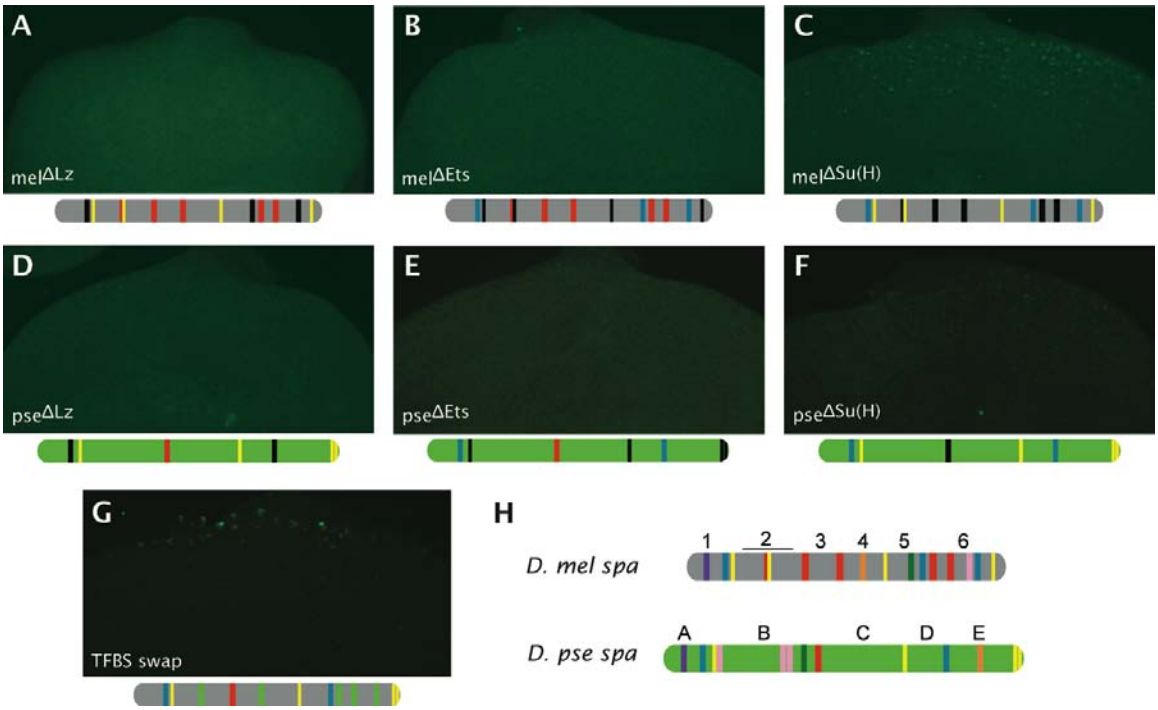


Figure 3.5. Both *D. mel* and *D. pse spa* require Lz, Ets, and Su(H) binding sites for proper transcriptional activation. (A-C) Binding site mutations in the *D. mel spa* enhancer. Black boxes represent mutated binding sites. Mutation of all Lz sites (A) or all Ets sites (B) results in a complete loss of enhancer activity. Mutation of Su(H) sites in *D. mel spa* results in a significant loss of gene expression (C). (D-F) Binding site mutations in the *D. pse spa* enhancer. As in *D. mel spa*, mutation of all Lz sites (D) or all Ets sites (E) results in a loss of enhancer activity. Mutation of the single putative Su(H) site in *D. pse spa* also significantly disrupts enhancer function (F). A version of *D. mel spa* that contains only the Lz/Ets/Su(H) sites present in *D. pse spa* fails to activate robust gene expression (G). In this construct, red, blue, and yellow sites are intact binding sites of the same sequence as corresponding sites in *D. pse spa*; green sequences represent sites that were mutated to orthologous sequence in *D. pse* (not predicted to be functional binding sites). (H) Model of *D. mel* and *D. pse spa* enhancers. Both enhancers require binding sites for Lz, PntP2/Yan, and Su(H) (blue, yellow, and red), as well as binding sites for the regulatory elements that bind in *D. mel* Regions 1, 4, 5, and 6A (purple, orange, green, and pink). The location of the orange, green, and pink sites has changed in *D. pse*, and there may also be more binding sites for these proteins in *D. pse* than in *D. mel spa* (for example, three pink sites in *D. pse* as opposed to only one in *D. mel*).

are located at -846 relative to the promoter rather than directly upstream), we also made binding site mutations in our *D. mel spa* construct for direct comparison with our *D. pse spa* results.

We found that, for both species' enhancers, mutation of either all Lz sites or all Ets sites severely diminished enhancer activity, confirming the necessity of direct regulatory input from Lz and PntP2/Yan (Figure 3.5A, B, D, and E). Mutation of all the Su(H) sites in either *D. mel spa* or *D. pse spa* significantly reduced enhancer function, but we observed a higher level of residual enhancer activity than expected (Figure 3.5C and F). In a previous report, eight putative Su(H) sites were mutated, whereas only five were mutated in our construct; we chose not to mutate the additional three sites because they did not bind Su(H) protein *in vitro* (Flores et al, 2000). However, it is possible that the weak activation we observe from our mutant enhancers represents residual Su(H) input from these weak binding sites. Alternatively, because Su(H) functions as a repressor in the absence of Notch signaling, the weak expression observed may result from derepression in cell types that lack active Notch signaling. Additional *in vivo* and *in vitro* experiments will be performed to explore the role of Su(H) regulation of *spa*. However, it seems clear that Su(H) binding sites contribute to transcriptional activation by *spa* in both *D. mel* and *D. pse*.

These results confirm that both the *D. mel* and *D. pse spa* enhancers require binding sites for Lz, PntP2/Yan, and Su(H). Because *D. pse spa* has only about half as many Lz/Ets/Su(H) sites as *D. mel spa*, we postulated that perhaps there is redundancy in the *D. mel spa* sites. We tested this hypothesis by building a version of *D. mel spa* that contains only the Lz/Ets/Su(H) sites from *D. pse spa*. Because it is impossible to identify

truly orthologous sites in the two species, we chose the sites most similar in sequence and position for preservation in this *D. mel* TF swap experiment; we also included the additional Ets site present at the very 3' end of *D. pse spa* in the *D. mel* TF swap construct (a previous report showed the *D. mel spa* sequence that was replaced with the novel *D. pse spa* Ets site was not necessary for *D. mel spa* function; Swanson et al, 2010). This construct, which is essentially a version of *D. mel spa* in which half the Lz/Ets/Su(H) sites have been mutated, drove gene expression in a weak and diffuse pattern (Figure 3.5G) . This result suggests that additional TFBS in *D. mel spa* are not simply redundant.

DISCUSSION

It is becoming evident that stabilizing selection acts on many cis-regulatory elements to maintain their function despite rapid sequence evolution and binding site turnover. The *Drosophila D-Pax2 sparkling* enhancer is an excellent example of an enhancer in which overall enhancer activity has been preserved despite lack of sequence conservation, probably due to compensatory mutations that have occurred in each lineage to offset binding site turnover. We have taken the evolutionary analysis of *spa* a step further by mapping the location of compensatory mutations between *D. mel spa* and *D. pse spa*. The characterization of two functionally equivalent yet very divergent enhancers provides insight into the evolution of enhancer sequence and organization, as well as the rules of TFBS composition and organization that govern the *spa* enhancer.

TFBS Composition

Alignment of the *spa* enhancer reveals a striking lack of conservation of critical regulatory sequences. While the *D. pse* and *D. mel spa* enhancers are regulated by the same transcriptional network, the number and organization of individual TFBS has changed due to binding site turnover and compensatory mutations. As a result, *D. mel spa* contains more sites for Lz, PntP2/Yan, and Su(H). *D. pse spa*, on the other hand, may have gained more binding sites for the unidentified proteins that bind in Regions 4, 5, and 6A, perhaps to compensate for the absence of additional Lz, Ets, and Su(H) sites.

It is possible that certain Lz, PntP2/Yan, and Su(H) sites in the *D. mel spa* enhancer are redundant, and that specific sites could be lost without any impact on enhancer function. Indeed, binding site redundancy is an important aspect of enhancer evolution, allowing individual binding site mutations to become fixed because their effects are not deleterious (Wittkopp, 2006). At least one site for each of these TFs is present in every species, and it has been shown they are important for both transcriptional activation and proper patterning of gene expression in *D. mel* and *D. pse* (Flores et al, 2000; Swanson et al, 2010). Yet the melanogaster subgroup has a significantly greater number of sites for these factors, and our data suggest that those extra sites are not all simply redundant (Figures 3.5G and S3.1). The increased presence of Lz/Ets/Su(H) sites seems to be specific to the melanogaster subgroup, and the most parsimonious explanation for this observation would be gain of those sites in that lineage. The presence of extra sites in the melanogaster subgroup could reflect changes in the signaling pathways or transcription factors themselves in those species. However, we

think that is unlikely as the *D. pse spa* enhancer behaves very similarly to the *D. mel spa* enhancer in a *melanogaster* background, which strongly suggests conservation of the *trans* environment; also, the protein sequence of all four *spa*-binding proteins is highly conserved between *D. mel* and *D. pse*, with very high conservation (≥96%) of DNA-binding domains (not shown). An intriguing possibility is that the gain of sites for Lz, PntP2/Yan, and Su(H) was necessary to offset absence of other binding sites.

Indeed, some of our data suggest that *D. pse spa* contains more sites for as-yet unidentified activators of the *sparkling* enhancer than *D. mel spa*. Some of our chimeric constructs drive significantly higher levels of gene expression compared to either endogenous enhancer, suggesting an overall increase in activator binding sites in those chimeras. Since this occurs in constructs with fewer characterized Lz/Ets/Su(H) binding sites than *D. mel spa*, the additional activation most likely originates from unmapped binding sites. Importantly, the 5' half of *D. pse* harbors binding sites that compensate for the loss of *mel* Regions 5 and 6A, and mutation of those regions in the context of $pse^{5'}+mel^{3'}$ chimera does not even diminish the elevated transcriptional activity of that construct (Figures 3.3B and 3.4E-F). A model in which *D. pse* 5' contains more binding sites for the Region 5 and Region 6A activators than *D. mel* Regions 5 and 6A would explain the elevated transcriptional activity of the constructs $pse^{5'}+mel^{3'}$ chimera, $pse^{5'}+mel^{3'\ (m5abc)}$, and $pse^{5'}+mel^{3'\ (m6A)}$(Figure 3.5H). In fact, motif analysis has identified several conserved motifs in 5' *D. pse* that may represent novel binding sites; intriguingly, one of these motifs that is present in one copy in *D. mel* Region 6A is present in 3 copies in *D. pse* 5' (Figure 3.5H). The presence of additional, novel sites for

the Region 4, 5, and/or 6A activators in *D. pse* may represent binding site gains that compensate for the relative lack of activation from Lz, PntP2, and Su(H).

Taken together, multiple binding site gain and loss events have occurred throughout the evolution of *spa* enhancer, resulting in differences in the TFBS composition of *D. mel spa* and *D. pse spa*. Although the previously described direct regulatory inputs are conserved and that conservation is likely critical for cone-cell specific gene expression, variation in the numbers of individual TFBS suggests some flexibility in how those proteins regulate transcriptional activation. Stabilizing selection has preserved the function of *spa* despite binding site turnover by selecting for compensatory mutations that maintain overall activation by *spa*; in *D. pseudoobscura*, this may have been achieved by gain of additional sites for the Region 4/5/6A activators, whereas activation in *D. melanogaster* and the melanogaster subgroup may be maintained by an increased number of Lz/Ets/Su(H) sites.

Enhancer structure and organization

Previous characterization of the *D. mel spa* enhancer revealed a surprising degree of inflexibility; reorganizations of the enhancer often impaired enhancer function or even resulted in a change in the cell-type specificity of enhancer activity from cone cells to photoreceptors (Swanson et al, 2010). Therefore the evolutionary changes in *spa* organization came as a surprise to us, suggesting that the architecture is actually more flexible than we thought.

The overall size of *spa* has not changed significantly between *mel* and *pse*; the *D. mel spa* enhancer is 362 bp total, and the *D. pse spa* enhancer is 409 bp total. While there is a slight expansion in enhancer size in *D. pse*, consistent with the overall slightly larger genome size of *D. pse*, there is likely to be evolutionary pressure to maintain a compressed enhancer structure (Richards et al 2005),. The compressed enhancer size allows for the conservation of local activating and repressive interactions we suspect occur within *spa* (Swanson et al, 2010). Examination of those aspects of *spa* organization that are shared between *D. mel* and *D. pse* may provide insights into important features of *spa* enhancer structure.

For example, Lozenge and PntP2 have been shown to physically interact, and can act synergistically in the regulation of target genes (Behan et al, 2005). Perhaps the conservation of the closely spaced Lz/Ets sites in the 5' portion of the enhancer is significant. This 5' binding site pair is conserved throughout the Sophophora lineage; only *D. vir* and *D. moj* lack a closely spaced Lz/Ets binding site pair. Thus closely spaced Lz/Ets sites may be an important feature for gene activation in cone cells.

Our previous work also identified the presence of an element in *D. mel* Region 5 that mediated repression in photoreceptors; that repressive function appeared to be a short-range activity, suggesting that its placement within the enhancer is critical for proper patterning (Swanson et al, 2010). Despite the profound organizational differences between *D. mel* and *D. pse*, *D. pse spa* does not drive ectopic expression in cell types other than cone cells. Furthermore, none of our chimeric constructs were ectopically activated in photoreceptors. Therefore, none of our constructs disrupt Region 5 repressor function. While we do not know the exact motif to which the repressor binds in *D. mel*

spa Region 5, it is worth noting that that region is flanked by an Ets site and a Lz site. The mutation of Region 5 in the context of the $pse^{5'}$+$mel^{3'}$ chimeric enhancer did not result in ectopic gene expression, suggesting that the repressive function is intact in that construct, either somewhere within the 5' half of *pse* or elsewhere in the 3' half of *mel*. Our previous study did note that the repressor site in Region 5 was redundant with a site or sites in Regions 2/3/6B; since only region 6B is present in the chimeric context, it is possible that a redundant site in 6B is still able to mediate repression in the absence of Region 5. Alternatively, the 5' half of *pse* may harbor a site for this repressor, likely in *pse* Region B (Figure 3.5H). Interestingly, *pse* Region B is flanked by an Ets site and a Su(H) site (Figure 3.5H). The common presence of a flanking Ets site in both *mel* Region 5 and *pse* Region B may suggest that the unknown repressor interacts with the Ets proteins PntP2 and Yan; this possibility is intriguing because the unknown repressor could be influencing either activation or repression by blocking activation by PntP2, or by cooperating with Yan to mediate repression.

Conserved structural features are difficult to identify in *spa* because not all critical regulatory sequences have been narrowed down to the level of a TFBS. We expect that the identification of novel binding sites and the proteins that bind those sites, perhaps via the motif analyses these data have helped us perform, will aid in elucidation of the rules that govern *spa* structure.

EXPERIMENTAL PROCEDURES

Generation of chimeric enhancers and transgenic reporter flies

Enhancer constructs were built using PCR-based techniques. Constructs that combine *D. melanogaster* and *D. pseudoobscura* sequences were generated by overlap extension (sewing) PCR using previously described *D. mel spa* and *D. pse spa* enhancers as templates (Swanson et al, 2010). Enhancers that contained multiple internal mutations were generated by assembly PCR (Swanson et al, 2010). Binding sites were mutated as follows: Lz sites were converted from RACCRCA to RA**AA**RCA; Ets sites from GGAW to **TT**AW; and Su(H) sites from YGTGDGAA to YGTGD**C**AA. These mutations have been shown to eliminate *in vitro* binding (Bailey and Posakony, 1995; Flores et al, 2000). PCR products were cloned into pENTR/D-TOPO (Invitrogen) and then subcloned via Gateway cloning into Ganesh-G1 for injection (Invitrogen, Swanson et al, 2008). Embryos from w^{1118} flies were injected as described by Rubin and Spradling (1982).

Dissections and Microscopy

Eye imaginal discs were dissected from transgenic late third-instar larvae and fixed in 4% paraformaldehyde for 30 minutes at room temperature. Endogenous GFP expression was visualized using an Olympus BX51 microscope and Olympus DP70

digital camera. At least 5 independent transgenic lines were dissected and analyzed for each reporter construct.

DNA Sequence Alignment

Orthologous *Drosophila sparkling* sequences were identified by BLAST searches on the DroSpeGe website (Gilbert, 2007; http://insects.eugenes.org/species). The alignments in Figures 3.2 and S3.1 were generated on ClustalW using the default settings (Chenna et al, 2003; www.ebi.ac.uk/tools/ClustalW). The alignment in Figure S3.2 was taken from the UCSC genome browser (http://genome.ucsc.edu).

Figure S3.1

Supplemental Figure 3.1. ClustalW alignment of the *sparkling* enhancer across all twelve sequenced *Drosophila* species (Chenna et al, 2003). Putative Lozenge sites are in blue, putative Ets sites are in orange (rather than yellow, as in other figures, so that sequence is readable), and Suppressor of Hairless sites are in red. Sequence shaded light blue is conserved with *D. melanogaster* sequence (shaded gray). Stars mark bases that are conserved across all twelve genomes.

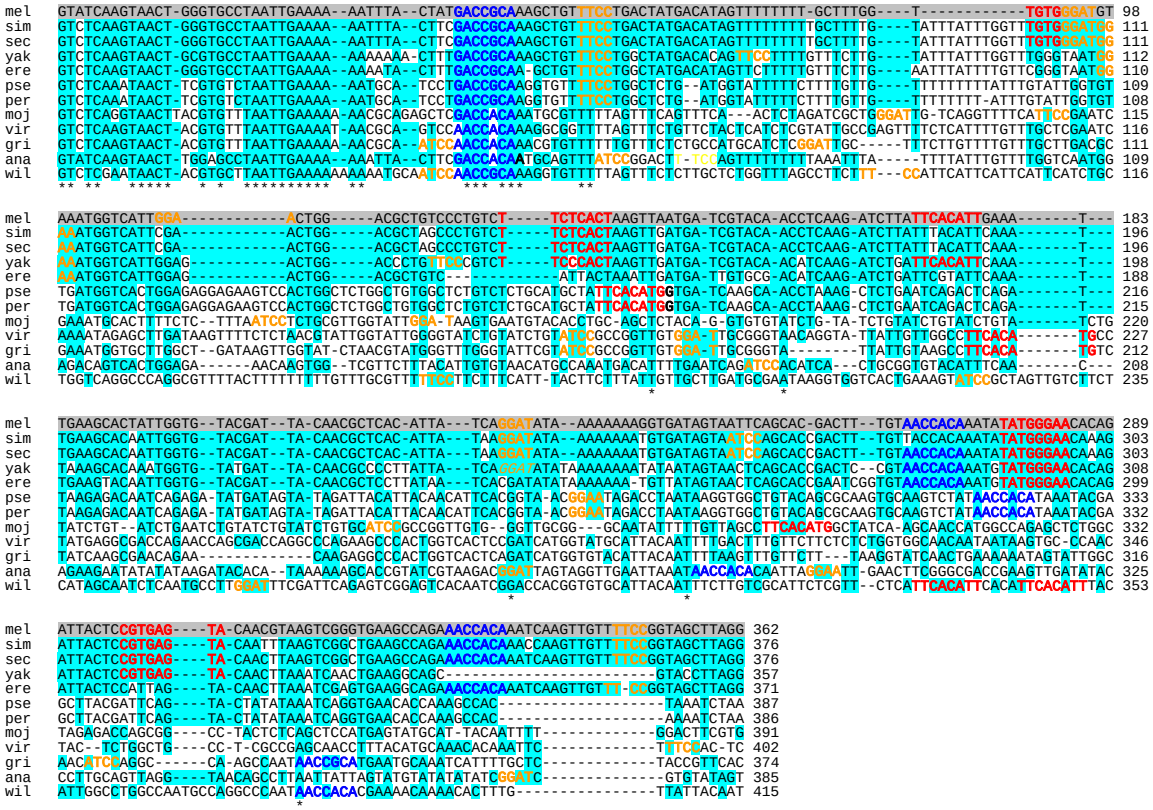


Figure S3.2

Supplemental Figure 3.2. BLASTZ alignment of *sparkling* enhancer sequence taken from the UCSC Genome Browser. Note that *D. willistoni*, *D. virilis*, *D. mojavensis*, and *D. grimshawi* sequences cannot be fully aligned with *D. melanogaster sparkling* sequence. Putative Lozenge sites are in blue, putative Ets sites are in orange (rather then yellow as in other figures, to make sequence readable), and putative Suppressor of Hairless sites are in red. Stars mark bases conserved across all twelve genomes.

```
D. mel  gtatcaagt-aactgggtgcctaattg----aaaaaatttactatgaccgcaaagctgtt--tcctgactatgac-atag
D. sim  gtctcaagt-aactgggtgcctaattg----aaaaaatttacttcgaccgcaaagctgtt--tcctgactatgac-atag
D. sec  gtctcaagt-aactgggtgcctaattg----aaaaaatttacttcgaccgcaaagctgtt--tcctgactatgac-atag
D. yak  gtctcaagt-aactgcgtgcctaattg---aaaaaaaaaaactttgaccgcaaagctgtt--tcctggctatgac-acag
D. ere  gtctcaagt-aactgggtgcctaattg----aaaaaaaatactttgaccgc-aagctgtt--tcctggctatgac-atag
D. ana  gtatcaagt-aacttggagcctaattg----aaaaaaattacttcgaccacaatgcagtttatccggacttt----ccag
D. pse  gtctcaaat-aacttcgtgtctaattg----aaaaaatgcatcctgaccgcaaggtgttt--tcctggctctgatggtat
D. per  gtctcaaat-aacttcgtgtctaattg----aaaaaatgcatcctgaccgcaaggtgttt--tcctggctctgatggtat
D. wil  gtctcgaat-aactacgtgcttaattgaaaaaaaaaatgcaatccaaccgcaaaggtgtt--t------------ttag
D. vir  gtctcaagt-aactacgtgtttaattg---aaaataacgcagtccaaccacaaaggcggt--tttagtttctgtt-ctac
D. moj  gtctcaggtaacttacgtgtttaattgaaaaaaa---cgcagagctcgaccacaaatgcgtttttagtttcagtttcaac
D. gri  gtctcaagt-aactacgtgtttaattg---aaaaaaacgcaatccaaccacaaacgtgtt--ttttgtttctctg-ccat
        ** **   * *  *  * *  ******    ***      *         *          *               *

D. mel  -ttttttttgctt-t----------------ggtt-gtgggatgtaaatggtcattgg-a-----------actg----
D. sim  ttttttttttgctt-t----tgtatttatttggttt-gtgggatggaaatggtcattcg-a-----------actg----
D. sec  ttttttttttgctt-t----tgtatttatttggttt-gtgggatggaaatggtcattcg-a-----------actg----
D. yak  ttcctttttgtttc-t----tgtatttatttggttt-gggtaatggaaatggtcattgg-ag----------actg----
D. ere  ttctttttgtttc-t----tgaatttattttgttc-gggtaatggaaatggtcattgg-ag----------actg----
D. ana  ttttttttttaaat-t----tat-tttatttgtttt-ggtcaatggagacagtcactgg-agaaa------caagtg----
D. pse  ttttcttttgttgtt----ttttttttatttgtatt-gg----tgttgatggtcactgg-agaggagaagtccactggctc
D. per  ttttcttttgttg-t----ttttttttatttgtatt-gg----tgttgatggtcactgg-agaggagaagtccactggctc
D. wil  -tttctcttgctc-t----ggtt-------------------------==============================
D. vir  tcatctcgtattg-ccgagttttctcattttgtttgctcgaatcaaaatag=============================
D. moj  tctagatcgctggatgtcaggttttcattccgaatcgaaatgcacttttctctttaatcctctgcgttggt=======
D. gri  gcatctcggattg-c-----tttcttgttttgtttgcttgacgcgaaatggtgcttgg=====================

D. mel  -gacgctgtccctgtcttctcactaagt-------taatgatcgtacaacctcaagatctta--ttcacattgaaattga
D. sim  -gacgctagccctgtcttctcactaagt-------tgatgatcgtacaacctcaagatctta--tttacattcaaattga
D. sec  -gacgctagccctgtcttctcactaagt-------tgatgatcgtacaacctcaagatctta--tttacattcaaattga
D. yak  -gaccctgttccgtctttcccactaagt-------tgatgatcgtacaacatcaagatctga--ttcacattcaaattaa
D. ere  -gacgctg--------tcattactaaat-------tgatgattgtgcgacatcaagatctga--ttcgtattcaaattga
D. ana  -gtcgtt--------ctttacattgtgt------------aacatgccaaatgacattttga--atcagatccacatcac
D. pse  tggctgtggctctgtctctgcatgctattc--acatggtgatcaagcaacctaaagctctga--atcagactcagattaa
D. per  tggctgtggctctgtctctgcatgctattc--acatggtgatcaagcaacctaaagctctga--atcagactcagattaa
D. wil  ================================================================================
D. vir  ================================================================================
D. moj  ================================================================================
D. gri  ================================================================================

D. mel  agcactat-----------tggtgtacg----attacaacgctca-cattatcaggatata--aaaaaaaggtg-at-ag
D. sim  agcacaat-----------tggtgtacg----attacaacgctca-cattataaggatata--aaaaaaatgtg-at-ag
D. sec  agcacaat-----------tggtgtacg----attacaacgctca-cattataaggatata--aaaaaaatgtg-at-ag
D. yak  agcacaaa-----------tggtgtatg----attacaacgccccttattatcaggatataaaaaaaaatata-at-ag
D. ere  agtacaat-----------tggtgtacg----attacaacgctccttataatcacgatatat-aaaaaaatgtt-at-ag
D. ana  tgc----------------ggtgtaca----ttt----------caacagaagaatata--tataagatacacat-aa
D. pse  gagacaatcagagatatgatagtatagattacattacaacattca-cggtaacggaatagac-ctaataaggtg-gc-tg
D. per  gagacaatcagagatatgatagtatagattacattacaacattca-cggtaacggaatagac-ctaataaggtg-gc-tg
D. wil  ==================================================tataagagtataa-aaacagaggaa-ataaa
D. vir  ================================================================================
D. moj  ================================================================================
D. gri  ================================================================================
```

```
D. mel  taattcagca------cgactt--tgtaaccacaaatatatgggaacacagattact-ccgtgagtac------aacgta
D. sim  taatccagcac-----cgactt--tgttaccacaaatatatgggaacaaagattact-ccgtgagtac------aattta
D. sec  taatccagcac-----cgactt--tgtaaccacaaatatatgggaacaaagattact-ccgtgagtac------aactta
D. yak  taactcagcac-----cgactc--cgtaaccacaaatgtatgggaacacagattact-ccgtgagtac------aactta
D. ere  taactcagcac-----cgaatcggtgtaaccacaaatgtatgggaacacagattact-ccattagtac------aactta
D. ana  aaa---agcac-----cgtatc-------------------gtaagacggatt------------------------
D. pse  ta---cagcgcaagtgcaagtc--tataaccacataaatacgagcttacgattc---------agtac------tatata
D. per  ta---cagcgcaagtgcaagtc--tataaccacataaatacgagcttacgattc---------agtac------tatata
D. wil  tagataagctc-----aaa-----tgtaattg-----atatgaacttgtagattcttgcaatgaatacctttttaaacta
D. vir  ================================================================================
D. moj  ================================================================================
D. gri  ================================================================================


D. mel  agtcgggtgaagccagaaaccaca-aatcaag--ttgtttttccggtagcttagg
D. sim  agtcggctgaagccagaaaccaca-aaccaag--ttgtttttccggtagcttagg
D. sec  agtcggctgaagccagaaaccaca-aatcaag--ttgtttttccggtagcttagg
D. yak  aatcaactgaaggcagc-------------------------gtaccttagg
D. ere  aatcgagtgaaggcagaaaccaca-aatcaag--ttg-tttccggtagcttagg
D. ana  agtaggttgaattaaataaccacacaattaggaattg---------aacttcgg
D. pse  aatcaggtgaacaccaaagccactaaatctaaaatcagttttccggcttcct--g
D. per  aatcaggtgaacaccaaagccacaaaatctaaaatcagttttccggcttcct--g
D. wil  aataagttgaagactcagactcaagagttaag------------gaaataaag
D. vir  ================================================
D. moj  ============================tcaaa--ttttaccatagatagtaaaag
D. gri  ================================================
```

CHAPTER IV

ATTEMPTED IDENTIFICATION OF NOVEL REGULATORS OF THE *sparkling* ENHANCER

INTRODUCTION

During *Drosophila* eye development, the Notch and EGFR signaling pathways are required for the fate specification of all twelve of the unique cells types that comprise each individual ommatidium (Voas and Rebay, 2004). These two signaling pathways are used repeatedly to determine individual cell fates, and are directly involved in regulating expression of many cell-type specific transcription factors that control cell fate specification. Because regulation by Notch and EGFR alone is not sufficient to define cell-type specific expression patterns, additional regulatory inputs must be required to generate specific outputs.

It has been proposed that combinatorial regulation by Notch, EGFR, and locally expressed activators and repressors provides sufficient spatial and temporal information for cell-type specific expression (Voas and Rebay, 2004). For example, expression of *prospero* in R7 and the four cone cells is combinatorially regulated by Glass, Sine Oculis, Lozenge, the Notch effector Suppressor of Hairless (Su(H)), and the EGFR effectors Yan

and PointedP2 (PntP2) (Xu et al, 2000; Hayashi et al, 2008). This combination of regulatory inputs, each of which acts directly on an eye-specific *prospero* enhancer, is suggested to provide the specificity required to activate expression of *prospero* in R7 and the cone cells, but repress its expression in other cells of the eye that experience active Notch and EGFR signaling.

Similarly, expression of *D-Pax2* in the cone cells is regulated by the *sparkling* (*spa*) enhancer. The *spa* enhancer has been shown to contain binding sites for Lz, Su(H), and the Ets factors PntP2 and Yan (Flores et al, 2000). Direct regulation by these three inputs – Lz, Notch, and EGFR – was proposed to be sufficient to generate cone-cell specific expression of *D-Pax2* (Voas and Rebay, 2004). Additional characterization of *spa*, however, revealed that sequences outside the twelve mapped binding sites for Lz, Su(H), and PntP2/Yan are required for enhancer function (Swanson et al, 2010). Therefore, Lz, Notch, and EGFR are not sufficient to regulate *D-Pax2* expression in the cone cells.

Mutagenesis experiments identified four regions of *spa* outside the Lz/Ets/Su(H) binding sites that are critical for enhancer function (Swanson et al, 2010). These regions, designated as Regions 1, 4, 5, and 6A, range in size from ~20-40 base pairs. Targeted mutations made within these regions indicated that most of that sequence is functionally significant, and given the size of these regions, it is possible that each region harbors one or more transcription factor binding sites. Data suggest that Regions 4 and 6A are important for robust transcriptional activation, Region 5 is important for activation in cone cells as well as repression in photoreceptors R1 and R6, and Region 1 is dispensable for proper patterning but is required for transcriptional activation when the enhancer is

located at a distance from the promoter (Swanson et al, 2010). Because the regulatory functions of these regions differ from one another, it seems likely that they are recruiting distinct activities to the enhancer. Therefore, we suspect that Regions 1, 4, 5, and 6A are bound by multiple unidentified proteins that directly regulate *spa* function.

Although multiple transcription factors are known to be expressed in the eye during this stage of development, only a few proteins are good candidates to regulate *spa* based on expression, activator or repressor function, and binding site sequence. Two promising candidates are BarH1/H2 and Sine Oculis. BarH1 and BarH2 are functionally redundant homeodomain proteins expressed in the photoreceptors R1 and R6 and primary pigment cells that may act as transcriptional repressors (Higashijima et al, 1992; Reig et al, 2007). There are several putative homeobox binding sites within *spa*, including two in Region 5. Bar could potentially bind Region 5 to repress *spa* activity in R1/R6; loss of Bar binding to those sites might explain why mutation of Region 5 can, in some contexts, lead to derepression of *spa* in R1 and R6. Sine Oculis (So) is an activator expressed in cells ahead of and behind the morphogenetic furrow in the eye (Serikaku and O'Tousa, 1994; Pauli et al, 2005). So may be acting through a putative binding site in Region 1 to regulate enhancer activity, although it is unclear whether So could be responsible for the long-distance function of Region 1. We are currently pursuing these and other candidates in our lab. However, we have yet to identify candidates that bind in Regions 4 or 6A, and we have not yet determined whether either of our candidates truly regulates *D-Pax2* expression *in vivo*.

There are several possible methods to identify novel *spa*-binding proteins other than the candidate approach. A yeast one-hybrid screen is a classic experimental

approach to identify proteins that bind specific DNA sequences. There are also multiple biochemical approaches to isolating and characterizing DNA-binding proteins. We decided to use biochemistry to purify and identify proteins that bind the *spa* enhancer. Our initial biochemistry experiments focused on Region 1 because of its functional importance, very high sequence conservation, and possible role in mediating long distance enhancer-promoter interactions.

RESULTS

Region 1-specific binding activity in embryonic nuclear extract

Because we suspect that Region 1 contains at least one novel transcription factor binding site, we wanted to test whether nuclear proteins specifically recognized Region 1 sequences *in vitro*. Therefore, we performed gel shift experiments with Region 1 sequences and embryonic nuclear extract to visualize protein-DNA interactions. We chose to use embryonic nuclear extract for our gel shift experiments because it is feasible to obtain enough protein from embryos for biochemical applications, whereas it is challenging to isolate large quantities of protein from tissues such as the imaginal discs. Furthermore, many proteins expressed in the eye are also expressed in the embryo, including Su(H), Lz, Pnt, and Yan (Schweisguth and Posakony, 1992; Scholz et al, 1993; Lebestky et al, 2000; Price and Lai, 2000).

We found that a wild-type Region 1 probe was consistently bound by a high-molecular weight activity in the embryonic nuclear extract; based on its size, this shift

Figure 4.1

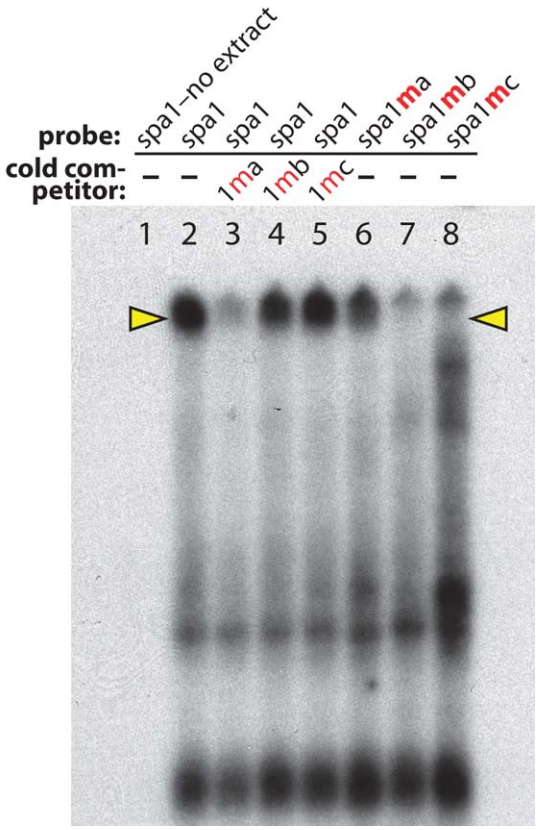


Figure 4.1. Embryonic nuclear extract contains a sequence-specific Region 1-binding activity. The wild-type *spa* Region 1 sequence is shifted by a high molecular-weight activity (Lane 2, yellow arrowheads). Binding of this activity is lost when mutations B and C are made to Region 1 sequence (Lanes 7 and 8). Mutant mB and mC probes also fail to compete with the wild-type probe for binding of this activity, suggesting that Region 1-binding protein(s) interact specifically with sequences within Regions 1B and 1C (Lanes 4 and 5).

most likely represents a complex of multiple proteins bound to the Region 1 probe (Figure 4.1). This DNA-protein interaction was sequence-specific, because mutations in the Region 1 probe could disrupt binding (Figure 4.1). In particular, mutation of Region 1B resulted in a complete loss of binding activity, and unlabeled 1B mutant probe was unable to compete with the wild-type Region 1 probe for binding of this complex (Figure 4.1). These data suggest that a protein complex in embryonic nuclear extract recognizes specific sequences in Region 1. The proteins within this complex may be novel regulators of the *spa* enhancer.

Biochemical purification of potential *spa*-binding proteins

We next designed a biochemical approach to purify proteins from embryonic nuclear extract that bound specifically to sequences within Region 1. We used biotinylated oligos bound to streptavidin-coated magnetic beads to pull down DNA-binding proteins from embryonic nuclear extract. The double-stranded oligos were comprised of either the wild-type Region 1 sequence or the 1B mutant sequence; the 1B mutant sequence was an appropriate negative control since it was not shifted by embryonic nuclear extract in our gel shift experiments. Comparisons with our negative control allowed us to identify proteins that recognized specific sequences within Region 1.

Both the wild-type Region 1 and mutant 1B sequences were bound by multiple proteins in embryonic nuclear extract. Separation of these proteins by SDS-PAGE revealed that most proteins bound both sequences, suggesting that those DNA-protein

Figure 4.2

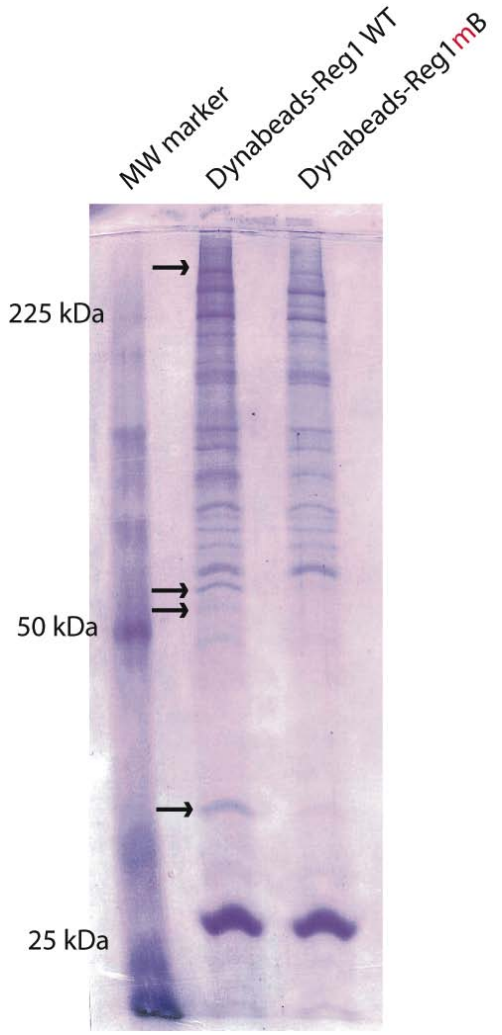


Figure 4.2. Purification of Region 1-binding proteins. Region 1 wild-type sequence and mutated sequence were used to pull down proteins from embryonic nuclear extract; proteins were then eluted from the DNA (which was bound to streptavidin-coated Dynabeads) and loaded onto an SDS-PAGE gel. Examination of the stained gel revealed that several bands pulled down by the wild-type probe were unique and were not pulled down by the 1mB probe (arrows; top-most band not visible in this image). These four bands were cut directly from the gel and submitted for analysis by mass spectrometry.

interactions are non-specific (Figure 4.2). However, four protein bands were uniquely pulled down by the wild-type Region 1 sequence, indicating that those proteins recognize specific sequences within Region 1 (Figure 4.2). These protein bands were cut out from the SDS-PAGE gel and identified by mass spectrometry.

Characterization of Region 1-binding proteins

Mass spectrometry identified six proteins within those unique bands, each of which may be a novel regulator of *spa* based on its ability to bind Region 1 in a sequence-specific manner. These candidates were identified as Aly, *Adh* Distal Factor-1 (Adf-1), Pep, Splicing factor 2 (SF2), CG3800, and CG5787. Two of these proteins, CG3800 and CG5787, have not been characterized in *Drosophila*. Two others, Aly and Adf-1, have identified roles in regulating gene expression. Adf-1 was originally identified as a sequence-specific DNA-binding transcriptional activator of the *Alcohol dehydrogenase* gene in *Drosophila*; however, no putative Adf-1 binding sites can be identified in Region 1 (Heberlein et al, 1985). The mammalian homolog of Aly has been shown to function as a coactivator, while yeast homologs have been implicated in both transcription and mRNA export (Bruhn et al, 1997; Strasser and Hurt, 2000; Strasser et al, 2002). The role of Aly in *Drosophila* has not been fully characterized, although it is not absolutely required for mRNA export (Gatfield and Izaurralde, 2002). The last two proteins, Pep and SF2, have no known role in regulating transcription. Pep is a zinc-finger RNA-binding protein that has been shown to associate with specific mRNAs in hnRNP complexes (Amero et al, 1991; Amero et al, 1993; Hamann and Stratling, 1998).

SF2 has been primarily characterized as an essential splicing factor or an alternative splicing factor, although intriguingly, overexpression of SF2 in the *Drosophila* eye results in a cone cell phenotype (Ge and Manley, 1990; Krainer et al, 1990; Gabut et al, 2007).

Although none of these proteins is an obvious candidate for regulation of *D-Pax2* expression in the *Drosophila* eye, we decided to perform preliminary characterization of each of these candidates. First we asked whether any of these proteins were able to bind Region 1 sequence specifically *in vitro*. All six candidate proteins were expressed from full-length cDNAs in an *in vitro* transcription/translation system. Two different full-length cDNAs exist for Pep, and we tested both isoforms. Expressed proteins were tested for their ability to bind a wild-type Region 1 probe and a 1B mutant probe in gel shift experiments. We found that none of these proteins is able to bind specifically to Region 1 sequence *in vitro* (Figure 4.3). However, the lack of ability to directly bind Region 1 sequence does not immediately rule out any of these candidates. If these proteins bind DNA as part of a complex, individual proteins may not be able to bind Region 1 without other members of the complex present. We did test the ability of the candidate proteins to bind Region 1 together, in case the candidates themselves comprise a DNA-binding complex, but did not see any shift in the combined presence of all six candidates and Region 1 DNA (not shown). We concluded that none of our candidates can independently bind to sequences within Region 1.

Because at this point we could not rule out any of the candidates, we proceeded to test the requirement for these proteins *in vivo*. We obtained *UAS-RNAi* transgenic flies for each of the candidate proteins. RNAi knockdown was performed *in vivo* by crossing

Figure 4.3

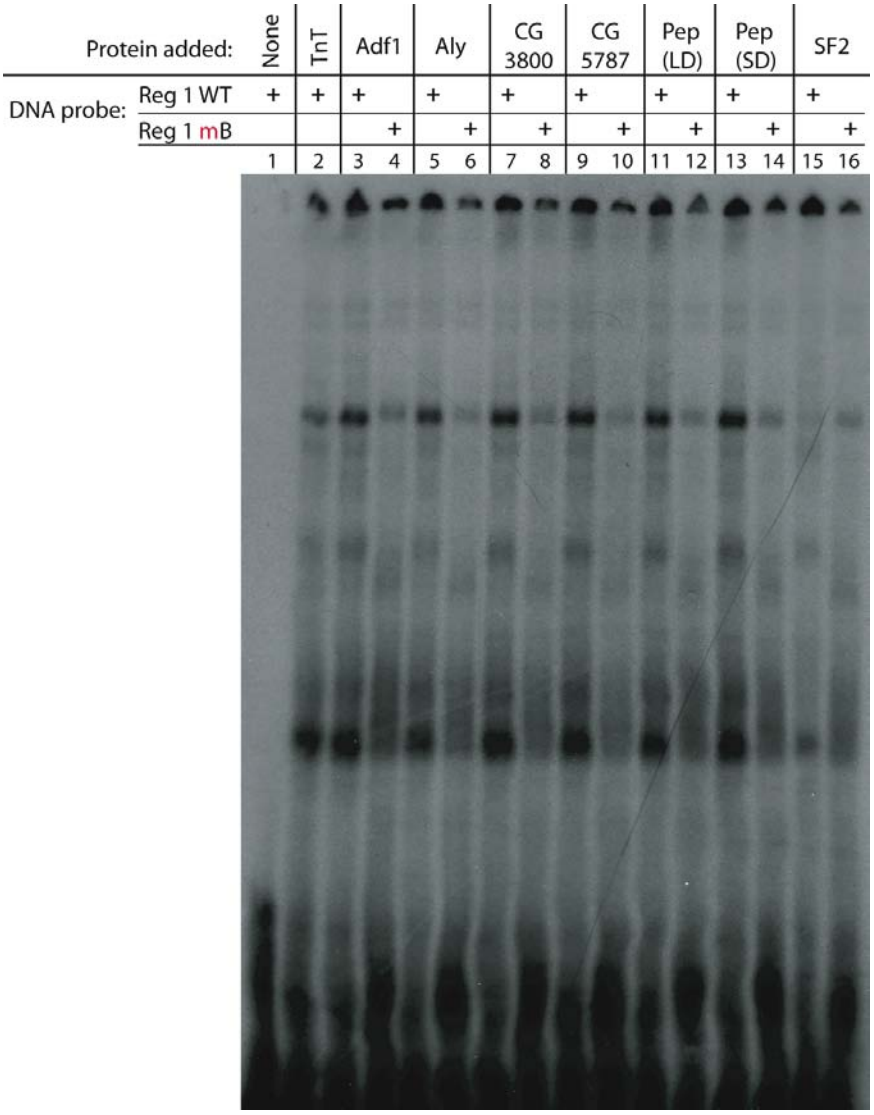


Figure 4.3. Candidate regulators of *spa* do not bind Region 1 sequence *in vitro*. Candidate proteins were expressed from full-length cDNAs using a TnT *in vitro* transcription/translation system. The Region 1 wild-type sequence was bound by multiple proteins present in the TnT lysate, presumably non-specifically (Lane 1). Presence of expressed proteins did not result in unique shifts of the Region 1 sequence, suggesting that none of these proteins directly binds Region 1 *in vitro* (Lanes 3, 5, 7, 9, 11, 13, and 15). Pep LD and SD lanes represent the two different isoforms of Pep that were tested. Mutant 1mB probes were not specifically shifted by any of the candidate proteins either (Lanes 4, 6, 8, 10, 12, 14, and 16).

GMR-Gal4 transgenic flies with *UAS-RNAi* transgenic flies; the *GMR* driver is active in all cells behind the morphogenetic furrow of the imaginal eye disc, and therefore knockdown of candidate proteins should occur in all cells behind the furrow (Freeman, 1996). We assayed knockdown phenotypes in two ways: by antibody staining for the cone cell marker Cut, and by examining the eyes of adult flies. Loss of *D-Pax2* expression would result in loss of cone cell specification, resulting in both loss of Cut expression and in a visible eye phenotype in adult flies. As a positive control, we also knocked down expression of *D-Pax2*. In *D-Pax2* knockdown flies, we observed a loss of expression of the cone cell marker Cut in late larval imaginal eye discs, as well as a *sparkling* eye phenotype in the adults (Figure 4.4H, data not shown). However, none of the candidate gene knockdown flies displayed either phenotype (Figure 4.4B-G, data not shown). The presence of *GMR-Gal4* in the absence of an RNAi transgene had no effect on either Cut expression or eye development (Figure 4.4A, data not shown). We concluded that none of the candidate genes is essential for *D-Pax2* expression, and thus none of our candidates is likely to be an important regulator of the *spa* enhancer.

DISCUSSION

The *spa* enhancer is most likely regulated by additional proteins other than Lz, Su(H), PntP2, and Yan, but the identity of those novel regulators remains unknown. We attempted to purify and characterize novel *spa*-binding proteins using a biochemical approach. Although we were able to isolate proteins based on binding to wild-type Region 1 sequence *in vitro*, none of these proteins was required for *spa* function *in vivo*.

Figure 4.4

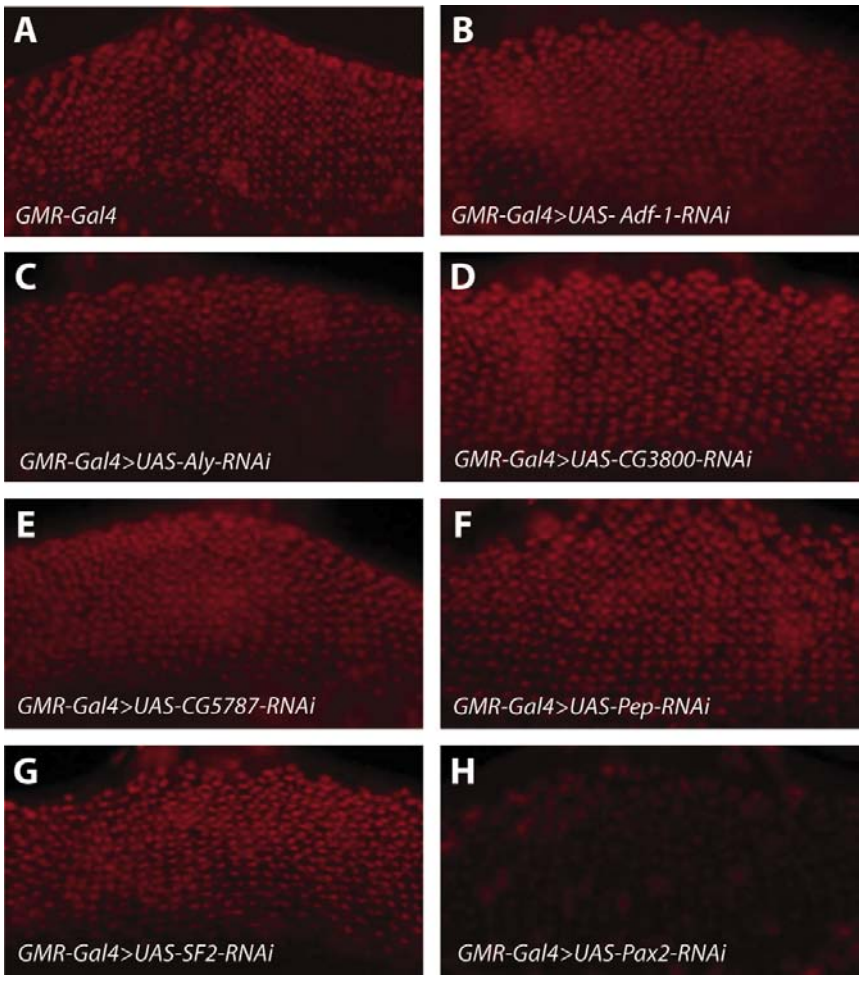


Figure 4.4. RNAi knockdown of candidate proteins in the *Drosophila* eye. Knockdown of candidate proteins was accomplished by crossing *GMR-Gal4* flies with *UAS-RNAi* flies; knockdown phenotype was assessed by staining for expression of Cut, a cone cell marker. The presence of the *GMR-Gal4* transgene alone does not affect cone cell development as assessed by Cut expression (A). RNAi knockdown of candidate proteins did not appear to affect cone cell development (B-G), whereas RNAi knockdown of *D-Pax2*, the gene controlled by the *spa* enhancer, resulted in a loss of Cut staining, suggesting a loss of cone cells in the eye (H).

Therefore, it seems unlikely that any of these proteins is an important novel regulator of *D-Pax2* expression in the *Drosophila* eye.

Identification of additional regulators of *D-Pax2* expression is still a primary objective of our lab. The biochemistry strategy described here could be improved upon before reattempting protein pulldowns with Region 1. For example, rather than using the 1B mutant probe, it may be more effective to use a mutant probe in which the entire Region 1 sequence has been mutated. We are also performing pulldown experiments with other important regulatory sequences within *spa*. In addition, we are pursuing a yeast one-hybrid approach to identify *spa*-binding proteins. Finally, we continue to pursue a candidate approach to identify novel regulators of *spa* function.

Through multiple approaches, we believe we will discover additional direct regulators of the *spa* enhancer. Identification of additional regulators of *spa* will not only help us understand the specification of cone cell fate in the *Drosophila* eye, but will also allow us to examine protein-protein interactions that occur among regulators of *spa* and the role those interactions have in enhancer function.

EXPERIMENTAL PROCEDURES

Gel shifts with Region 1 sequences

We purified embryonic nuclear extract from 0-12 hour embryos collected over three days. Purification of embryonic nuclear extract was performed as described (Sullivan et al, 2000). Gel shift probes were simultaneously annealed and labeled by

combining 37 µl sdH_2O, 5 µl 10X PNK buffer, 1 µl top strand oligo (2 µM), 1 µl bottom strand oligo (2 µM), 5 µl γ^{32}P-ATP, and 1 µl T4 PNK (NEB); incubating at 37ºC for one hour; boiling at 80ºC for 5 minutes; and allowing to cool slowly to room temperature. Labeled probes were purified twice using GE ProbeQuant G-50 spin columns. Cold competitor probes were made by combining 39 µl sdH_2O, 5 µl 10X PNK Buffer, 1 µl top strand oligo (200 µM), and 1 µl bottom strand oligo (200 µM). From there they were prepared just as labeled probes with the exception that they were not purified. Probe sequences were as follows, with mutated bases in uppercase:

Region 1 wild-type: gtatcaagtaactgggtgcctaattgaaaaaatttactat

Region 1mA: gGaGcCaTtCactgggtgcctaattgaaaaaatttactat

Region 1mB: gtatcaagtaacGgTgGgAcGaattgaaaaaatttactat

Region 1mC: gtatcaagtaactgggtgcctaCtGgCaCaCatttactat

Gel shift reactions included 1 µl labeled probe, 1 µl nuclear extract (approximately 8 µg total protein), 1 µl 10X Gel Shift 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), 1 µl poly d(I-C) (1 mg/ml), 1 µl DTT (100 µM), and 5 µl dH_2O. In reactions that included cold competitor, 1 µl dH_2O was replaced with 1 µl cold competitor. Gel shift reactions were incubated on ice for 15 minutes, then were loaded onto 6% polyacrylamide gels. Gels were run for 4-5 hours in 0.5X TBE at 120 V, vacuum dried for one hour at 80ºC, then exposed to film.

Biochemical purification of Region 1-binding proteins

The Region 1 wild-type and mB probes used for our pulldown experiments had the same sequences as the corresponding gel shift probes. For pulldowns, top strand oligos were biotinylated at the 5' end. Double stranded probes were made by combining 500 pmol biotinylated top strand oligo with 500 pmol bottom strand oligo and adding dH_2O to final volume of 100 µl, boiling for 5 minutes at 80ºC, then allowing to cool slowly to room temperature.

100 µl of magnetic, streptavidin-coated Dynabeads (10 mg/ml, Invitrogen) were added to a fresh epitube. Using a magnet, the beads were washed twice with 100 µl 2X BW buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl) and then resuspended in 100 µl 2X BW buffer. 100 µl probe (500 pmol) was added, and beads and probe were incubated for 15 minutes at room temperature on a rotating wheel. Using a magnet, the supernatant was removed and beads, now bound to probe, were washed 3 times in 1X BW buffer before resuspension in 50 µl 1X gel shift buffer. Separately, 35 µl 10X gel shift buffer, 35 µl poly d(I-C) (1 mg/ml), 5 µl 0.5 M DTT, 50 µl 7X Roche Complete protease inhibitor, 25 µl glycerol, and 150 µl nuclear embryonic extract (approximately 1.2 mg total protein) were combined. This mixture was incubated at room temperature for 10 minutes, mixing gently every few minutes. The mixture was then added to the resuspended beads and incubated at room temperature for 20 minutes on a rotating wheel. Using a magnet, the beads were washed once in 1X gel shift buffer containing 0.1 mg/ml poly d(I-C), then washed an additional 3 times in 1X gel shift 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), then boiled for 5 minutes at 100ºC to elute proteins. Using a magnet, supernatant (including eluted protein) was removed from the beads and loaded onto an SDS-PAGE gel (12% resolving gel, 5% stacking gel). Pulldowns with Region 1 wild-type and Region 1mB probes were performed in parallel, with eluted proteins loaded into adjacent lanes on the same gel for comparison. SDS-PAGE gels were run in 1X Tris-Glycine buffer at 120 V for 5-6 hours, then stained overnight with Coomassie blue as described (Wong et al, 2000). Bands of interest were cut directly from an SDS-PAGE gel and analyzed using LC-MS/MS by the University of Michigan Protein Structure Facility.

In vitro expression of candidate proteins

Full-length cDNAs for each candidate protein were obtained from the Drosophila Genomics Resource Center. Full-length proteins were expressed from PCR products using the TNT ® SP6 High-Yield Wheat Germ Protein Expression System according to their protocol (Promega). PCR primer sequences are listed below:

SP6+Adf1for: tagcattatttaggtgacactatagaacagaccaccatggacaagctggatgcc

Adf1rev: ttttttttttttttttttttttttttttttttcatcattagtacttgttgtgctgtg

SP6+Alyfor: tagcattatttaggtgacactatagaacagaccaccatggtggacaaaattgaaatg

Alyrev: ttttttttttttttttttttttttttttttttcatcattagatcttcatgtcgttg

SP6+CG3800for: tagcattatttaggtgacactatagaacagaccaccatgtcgatgtctgccacgtg

CG3800rev: ttttttttttttttttttttttttttttttttcatcactagttccgtccgcccttc

SP6+CG5787for: tagcattatttaggtgacactatagaacagaccaccatggactaccgccaaatgaag

CG5787rev: ttttttttttttttttttttttttttttttttcatcattactctatagagggaaag

SP6+PepLDfor: tagcattatttaggtgacactatagaacagaccaccatggctttccgtggtaaccag

PepLDrev: ttttttttttttttttttttttttttttttttcatcattagtagcgattgtagcg

SP6+PepSDfor: tagcattatttaggtgacactatagaacagaccaccatgttctactgtcatctgtg

PepSDrev: ttttttttttttttttttttttttttttttttcatcattagtagcgattgtagcg

SP6+SF2for: tagcattatttaggtgacactatagaacagaccaccatgggatcacgcaacgagtg

SF2rev: ttttttttttttttttttttttttttttttttcatcattaatagttagaacgtgag

Gel shifts with candidate proteins were performed as described above, with the exception that these shifts included 3 µl of expressed protein and as a result the water added was reduced to 3 µl. In the shift that included all six candidate proteins, 1 µl of each protein and no water was included for that reaction.

RNAi knockdown of candidate protein expression

Transgenic RNAi fly lines were obtained from the Vienna Drosophila RNAi Center. Gal4 driver lines were obtained from the Bloomington Stock Center. The *GMR-Gal4* line used here is stock number 1104. *GMR-Gal4* flies were crossed to *UAS-RNAi* flies for *in vivo* knockdown of protein expression. Antibody staining was performed using a mouse anti-Cut primary antibody from the Developmental Studies Hybridoma Bank, diluted 1:50. Staining and imaging were performed as described previously (Swanson et al, 2010).

CHAPTER V

CONCLUSION

Summary of findings

Even after thirty years of research, we still do not completely understand the rules that govern the function of enhancer elements *in vivo*. Our inability to draw general conclusions about enhancer structure and function is at least partially due to a lack of studies that examine the basic, fundamental aspects of enhancer regulation and organization. Therefore, we performed an in-depth analysis of the *Drosophila D-Pax2 sparkling* enhancer to gain further insight into the relationship between structure and function in enhancer elements *in vivo*.

The objective of this thesis was to attempt to understand the regulation and organization of the *spa* enhancer. Our data suggest that both the regulation and organization of *spa* are surprisingly complex. Previous research had identified three direct inputs that regulate *spa* activity: the Notch-regulated transcription factor Suppressor of Hairless, the EGFR-regulated transcription factors PointedP2 and Yan, and the transcriptional activator Lozenge (Flores et al, 2000). We discovered that in addition to the twelve identified binding sites for those proteins, *spa* harbors essential regulatory

elements in multiple sequences that lie outside its characterized binding sites. It is unclear how many additional transcription factor binding sites these novel regulatory regions may contain. However, we have reason to believe that these sequences are bound by at least three distinct, unidentified regulators of *spa* that make individual contributions to enhancer function: a regulatory element within Region 1 may be important for long-range enhancer-promoter interactions, regulatory elements within Regions 4 and 6A are important for transcriptional activation, and a regulatory element within Region 5 is important for repressing gene expression in photoreceptors. Taken together, there are at least six direct regulatory inputs that bind *spa* to generate cone cell-specific gene expression, although the identity of additional *spa*-binding proteins remains unknown. Our findings reveal an unexpected level of complexity in the combinatorial control of this enhancer.

We have found that the order and spacing of these regulatory sequences in *spa* is critical to enhancer function. Changes in spacing between transcription factor binding sites can have significant impacts on enhancer activity; we have observed both reduced and increased transcriptional responses brought about by altered spacing within *spa*. Therefore, spatial relationships within this enhancer directly impact transcriptional output. The order of transcription factor binding sites within *spa* also affects enhancer function. Rearrangements of regulatory elements can result in a change in the cell-type specificity of gene expression, converting *spa* from a cone cell enhancer to an R1/R6 enhancer. Our data indicate that spatial relationships and local interactions within *spa* are essential not only for activation of transcription, but also for proper patterning of gene expression.

However, there is also evidence for flexibility in enhancer structure. Although rearrangements can alter the pattern of gene expression generated by *spa*, the reorganized enhancer is still capable of generating a robust transcriptional response. Evidence of flexibility also comes from our analysis of the evolution of the *spa* enhancer. Despite a lack of sequence conservation, the function of the *spa* enhancer is preserved in *D. pseudoobscura*. Our data suggest that while the overall regulation of *D. mel spa* and *D. pse spa* is conserved, the exact composition and organization of transcription factor binding sites has diverged without changing the function of the enhancer. We conclude that some aspects of *spa* structure are likely to be quite flexible, while other regulatory sequences within *spa* are spatially constrained; thus the organization of *spa* is also quite complex.

Implications

Our findings suggest that both the regulation and organization of enhancer elements may be more complicated than suspected. Few enhancer studies go beyond identifying obvious transcriptional regulators, mapping their direct binding sites, and perhaps testing that those sites are necessary *in vivo*. Our data show that this approach is likely to overlook important regulatory elements that are critical to enhancer function. While it is impractical for every enhancer to be characterized to the extent of our analysis of *spa*, it is sobering to consider that our knowledge of the regulation of most enhancers is likely to be incomplete.

Even our understanding of *spa* is incomplete; we have yet to identify all the direct regulators of *spa*, and we still cannot define the precise combination of activating and repressive inputs sufficient to activate cone cell-specific gene expression. Yet this research has brought renewed attention to the complicated nature of transcriptional regulation. At least for *spa*, and perhaps more generally, many regulatory inputs are required to generate an appropriate transcriptional response. In addition, our data suggest that a variety of activities must be recruited to properly regulate transcription; in *spa*, for example, simply doubling the number of Lz, Ets, and Su(H) sites cannot compensate for the absence of the other regulatory inputs. Furthermore, some of these regulatory inputs may be playing unique roles in regulating gene expression, as does the remote control element in Region 1. In summary, regulation of enhancer function is highly combinatorial, in every sense of the word.

This analysis has also provided insights into the rules of enhancer organization. We have shown that enhancer structure is to some extent flexible, and even a drastically rearranged version of *spa* is able to activate robust gene expression. Yet there must be organizational rules as well, since structural changes can affect the level of transcriptional activation and even the patterning of gene expression. It seems that *spa* is not as strictly structured as an enhanceosome, particularly since over evolutionary time much of the sequence and organization of the enhancer has changed without altering function. However, *spa*'s structure is not as flexible as a straightforward information display either. Most likely, further characterization of *spa* will reveal specific synergistic or repressive interactions that occur among *spa*-binding proteins that require particular binding site configurations. We predict that the binding sites for those proteins will be structurally

constrained, but binding sites for other proteins can be flexibly rearranged without affecting enhancer function. The identification of novel *spa*-binding proteins and accurate mapping of their binding sites will allow future exploration of these types of subelements within *spa*. We think it is very likely that this type of organization – overall flexibility as long as specific subelement structure is maintained – is common among endogenous enhancer elements.

A final point to take away from our analysis of *spa* is that enhancer sequences evolve rapidly compared to coding sequences. Evolution is often used *in silico* to identify cis-regulatory elements, and is sometimes used to draw conclusions about the functional relevance of particular sequences within cis-regulatory elements. It is unclear whether *spa* would have been identified based on sequence conservation alone, but it is clear that many essential sequences within *spa* are not strictly conserved due to sequence evolution, compensatory mutation, and binding site turnover. While evolution is a valuable tool for identifying cis-regulatory elements, our data show that conclusions about functional importance are best drawn from *in vivo* analyses.

Future Directions

The work described here has raised as many questions as it has answered. There is much left to be learned about *spa* function, as well as about enhancer elements in general.

This thesis project has led to the identification of several important regulatory elements within the *spa* enhancer, some of which have potentially novel and exciting

roles in enhancer function. The identification of additional proteins that regulate *spa*, as well as elucidation of the specific sequences they recognize within *spa*, is an important step in understanding this enhancer. The characterization of novel regulators of *spa* will not only provide insight into cone cell fate specification and combinatorial control of gene expression, but also allow us to ask many more questions about enhancer structure and function. Once these proteins are identified, we can investigate the physical interactions that occur among *spa*-binding proteins, which will lead to better understanding of the organizational rules within *spa*. It will also be interesting to characterize the function of novel *spa*-binding proteins; we can identify the cofactors each protein recruits to the enhancer, as well as follow up on potentially novel mechanisms of transcriptional regulation, such as the function of the remote control element in Region 1. Finally, we can try to understand the roles that individual proteins play in *spa* function: Does *spa* function strictly require this unique combination of proteins? Or are some of these activities replaceable? How many ways are there to build the *sparkling* enhancer?

It is also important to remember that this analysis, although performed entirely *in vivo*, is still artificial in that we used transgenic reporter constructs with a heterologous promoter to assay *spa* function. This approach is very useful to understand the structure and function of the minimal *spa* enhancer, which contains all the necessary information to activate a cone cell-specific expression pattern. However, transcriptional activation at the endogenous *D-Pax2* locus is likely to be even more complicated. For example, although we tested *spa* function at a greater distance from the promoter than is typically used in reporter assays, at the endogenous locus *spa* acts at a distance of 7 kilobases to

direct transcriptional activation at the promoter. We also have yet to make a version of *spa* that drives gene expression from the endogenous *D-Pax2* promoter; initial attempts using a minimal version of the *D-Pax2* promoter were unsuccessful, hinting at another level of complexity in *spa* function. Finally, the endogenous *D-Pax2* locus most likely contains other cis-regulatory elements that affect *spa* function, such as insulators, silencers, and other enhancers. It is even possible that the minimal version of *spa* used for this study does not include all the regulatory sequences that contribute to cone cell-specific gene expression. The endogenous *spa* enhancer may be larger than the fragment used here, or additional cone cell-specific regulatory sequences may be distributed elsewhere in the *D-Pax2* locus. Our lab is actively exploring these other aspects of *spa* function.

In conclusion, our analysis of the *spa* enhancer has provided valuable insight into the complex relationship between structure and function in enhancer elements. Our data suggest that the typical enhancer may be much more crowded with regulatory elements than suspected. In addition, those regulatory elements recruit a variety of activities to the enhancer, some of which may be playing unique roles in regulating transcription. Finally, the organization of those regulatory elements is critical not only for activating the correct level of transcription, but also for generating the proper pattern of gene expression.

COMPLETE REFERENCES

Akbari, OS; Bae, E; Johnsen, H; Villaluz, A; Wong, D; and Drewel, RA. (2008). A novel promoter-tethering element regulates enhancer-driven gene expresion at the bithorax complex in the *Drosophila* embryo. *Development* **135**, 123-131.

Alexandre, C; Jacinto, A; Ingham, PW. (1996). Transcriptional activation of *hedgehog* target genes in *Drosophila* is mediated directly by the Cubitus interruptus protein, a member of the GLI family of zinc finger DNA-binding proteins. *Genes and Development* **10**, 2003-2013.

Amero, SA; Elgin, SC; and Beyer, AL. (1991). A unique zinc finger protein is associated preferentially with active ecdysone-responsive loci in *Drosophila*. *Genes and Development* **5**, 188-200.

Amero, SA; Matunis, MJ; Matunis, EL; Hockensmith, JW; Raychaudhuri, G; and Beyer, AL. (1993). A unique ribonucleoprotein complex assembles preferentially on ecdysone-reponsive sites in *Drosophila melanogaster*. *Molecular and Cellular Biology* **13(9)**, 5323-5330.

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

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

Arnosti, DN; and Kulkarni, MM. (2005). Transcriptional enhancers: intelligent enhanceosomes or flexible billboards? *Journal of Cellular Biochemistry* **94**, 890-898.

Bae, E; Calhoun, VC; Levine, M; Lewis, EB; and Drewell, RA. (2002). Characterization of the intergenic RNA profile at *Abdominal-A* and *Abdominal-B* in the *Drosophila* bithorax complex. *PNAS* **99(26)**, 16847-16852.

Bailey, AM; and Posakony, JW. (1995). Suppressor of hairless directly activates transcription of enhancer of split complex genes in response to Notch receptor activity. *Genes and Development* **9(21)**, 2609-2622.

Banerji, J; Sandro, R; Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299-308.

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

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

Behan, KJ; Fair, J; Singh, S; Bogwitz, M; Perry, T; Grubor, V; Cunningham, F; Nichols, CD; Cheung, TL; Batterham, P; and Pollock, JA. (2005). Alternative splicing removes an Ets interaction domain from Lozenge during *Drosophila* eye development. *Developmental Genes and Evolution* **215**, 423-435.

Bruhn, L; Munnerlyn, A; and Grosschedl, R. (1997). ALY, a context-dependent coactivator of LEF-1 and AML-1, is required for TCRalpha enhancer function. *Genes and Development* **11**, 640-653.

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.

Bhaskar, V; and Courey, AJ. (2002). The MADF-BESS domain factor Dip3 potentiates synergistic activation by Dorsal and Twist. *Gene* **299**, 173-184.

Bray, SJ. (2006). Notch signalling: a simple pathway becomes complex. *Nature Reviews* **7**, 678-689.

Calhoun, VC; Stathopoulos, A; and Levine, M. (2002). Promoter-proximal tethering elements regulate enhancer-promoter specificity in the *Drosophila Antennapedia* complex. *PNAS* **99**, 9243-9247.

Calhoun, VC; and Levine, M. (2003). Long-range enhancer-promoter interactions in the *Scr-Antp* interval of the *Drosophila* Antennapedia complex. *PNAS* **100(17)**, 9878-9883.

Carey, M. (1998). The enhanceosome and transcriptional synergy. *Cell* **92**, 5-8.

Carroll, SB. (2008). Evo-Devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25-36.

Carter, D; Chakalova, L; Osborne, CS; Dai, YF; and Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. *Nature Genetics* **32**, 623-626.

Cave, JW; Loh, F; Surpris, JW; Xia, L; and Caudy, MA. (2005). A DNA transcription code for cell-specific gene activation by Notch signaling. *Current Biology* **15**, 94-104.

Chen, Q; Lin, L; Smith, S; Lin, Q; and Zhou, J. (2005). Multiple promoter targeting sequences exist in *Abdominal-B* to regulate long-range gene activation. *Developmental Biology* **286**, 629-636.

Chenna, R; Sugawara, H; Koike, T; Lopez, R; Gibson, TJ; Higgins, DG; Thompson, JD. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Resources* **31(13)**, 3497-3500.

Cook, T; Pichaud, F; Sonneville, R; Papatsenko, D; and Desplan, C. (2003). Distinction between color photoreceptor cell fates is controlled by Prospero in *Drosophila*. *Developmental Cell* **4**, 853-864.

Cooney, AJ; Leng, X; Tsai, SY; O'Malley, BW; and Tsai, M. (1993). Multiple mechanisms of Chicken Ovalbumin Upstream Promoter Transcription Factor-dependent repression of transactivation by the vitamin D, thyroid hormone, and retinoic acid receptors. *Journal of Biological Chemistry* **268**, 4152-4160.

Crew, JR; Batterham, P; and Pollock, JA. (1997). Developing compound eye in *lozenge* mutants of *Drosophila: lozenge* expression in the R7 equivalence group. *Dev. Genes Evol* **206**, 481-493.

Crocker, J; Tamori, Y; and Erives, A. (2008). Evolution acts on enhancer organization to fine-tune gradient threshold readouts. *PloS Biology* **6**, e263.

Davidson, EH. (1999). A view from the genome: spatial control of transcription in sea urchin development. *Curr Opin Genet Dev* **9**, 530-541.

Dillon, PJ; and Rosen, CA. (1990). A rapid method for the construction of synthetic genes using the polymerase chain reaction. *BioTechniques* **9**, 298-300.

Drosophila 12 Genomes Consortium (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203-218.

Erives, A; and Levine, M. (2004). Coordinate enhancers share common organizational features in the *Drosophila* genome. *PNAS* **101(11)**, 3851-3856.

Fisher, S; Grice, EA; Vinton, RM; Bessling, SL; and McCallion, AS. (2006). Conservation of *RET* regulatory function from human to zebrafish without sequence similarity. *Science* **312**, 276-279.

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

Flores, GV; 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.

Freeman, M. (1996). Reiterative use of the EGF receptor triggers differentiation of all cell types in the *Drosophila* eye. *Cell* **87**, 651-660.

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

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.

Gabut, M; Dejardin, J; Tazi, J; and Soret, J. (2007). The SR family proteins B52 and dASF/SF2 modulate development of the *Drosophila* visual system by regulating specific RNA targets. *Molecular and Cellular Biology* **27(8)**, 3087-3097.

Gatfield, D; and Izaurralde, E. (2002). REF1/Aly and the additional exon junction complex proteins are dispensable for nuclear mRNA export. *Journal of Cell Biology* **159 (4)**, 579-588.

Ge, H; and Manley, JL. (1990). A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA *in vitro*. *Cell* **62**, 25-34.

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

Gilbert, DG. (2007). DroSpeGe: rapid access database for new *Drosophila* species genomes. *Nucleic Acids Res* **35**, D480-485.

Gompel, N; Prud'homme, B; Wittkopp, PJ; Kassner, VA; and Carroll, SB. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* **433**, 481-487.

Goto, T; Macdonal, 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.

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

Guarente, L. (1988). UASs and enhancers: common mechanism of transcriptional activation in yeast and mammals. *Cell* **52**, 303-305.

Halfon, MS; Carmena, A; Gisselbrecth, S; Sackerson, CM; Jimenez, F; Baylies, MK; and Michelson, AM. (2000). Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. *Cell* **103**, 63-74.

Hamann, S; and Stratling, WH. (1998). Specific binding of *Drosophila* nuclear protein PEP (protein on ecdysone puffs) to hsp70 DNA and RNA. *Nucleic Acids Research* **26(18)**, 4108-4115.

Harding, K; Hoey, T; Warrior, R; and Levine, M. (1989). Autoregulatory and gap gene response elements of the *even-skipped* promoter of *Drosophila*. *The EMBO Journal* **8(4)**, 1205-1212.

Hare, EE; Peterson, BK; Iyer, VN; Meier, R; and Eisen, MB. (2008). Sepsid *even-skipped* enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLoS Genetics* **4(6)**, e1000106.

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. *Developmental Biology* **200**, 131-145.

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

Heberlein, U; England, B; and Tijan, R. (1985). Characterization of *Drosophila* transcription factors that activate the tandem promoters of the Alcohol Dehydrogenase gene. *Cell* **41**, 965-977.

Higashijima, S; Kojima, T; Michiue, T; Ishimaru, S; Emori, Y; and Saigo, K. (1992). Dual *Bar* homeobox genes of *Drosophila* required in two photoreceptor cells, R1 and R6, and primary pigment cells for normal eye development. *Genes and Development* **6**, 50-60.

Jeong, S; Rebeiz, M; Andolfatto, P; Werner, T; True, J; and Carroll, SB. (2008). The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. *Cell* **132**, 783-793.

Johnson, LA; Zhao, Y; Golden, K; and Barolo, S. (2008). Reverse-engineering a transcriptional enhancer: a case study in *Drosophila*. *Tissue Engineering: Part A* **14(9)**, 1549-1559.

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

Kitabayashi, I; Yokoyama, A; Shimizu, K; and Ohki, M. (1998). Interaction and functional cooperation of the leukemia-associated factors AML1 and p300 in myeloid cell differentiation. *EMBO J* **17**, 2994-3004.

Klingler, M; Soong, J; Butler, B; Gergen, JP. (1996). Disperse versus compact elements for the regulation of *runt* stripes in *Drosophila*. *Developmental Biology*. **177**, 73-84.

Krainer, AR; Conway, GC; and Kozak, D. (1990). The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell* **62**, 35-42.

Kuhn, RM; Karolchik, D; Zweig, AS; Wang, T; Smith, KE; Rosenbloom, KR; Rhead, B; Raney, BJ; Pohl, A; Pheasant, M; Meyer, L; Hsu, F; Hinrichs, AS; Harte, RA; Giardine, B; Fujita, P; Diekhans, M; Dreszer, T; Clawson, H; Barber, GP; Haussler, D; and Kent, WJ. (2009). The UCSC Genome Browser Database: update 2009. *Nucleic Acids Res* **37**, D755-761.

Kulkarni, MM; and Arnosti, DN. (2005). Cis-regulatory logic of short-range transcriptional repression in *Drosophila melanogaster*. *Molecular and Cellular Biology* **25(9)**, 3411-3420.

Lai, ZC; and Li, Y. (1999). Tramtrack69 is positively and autonomously required for *Drosophila* photoreceptor development. *Genetics* **152**, 299-305.

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

Li, L; Zhu, Q; He, X; Sinha, S; and Halfon, MS. (2007). Large-scale analysis of transcriptional cis-regulatory modules reveals both common features and distinct subclasses. *Genome Biology* **8(6)**, R101.

Li, Q; Barkess, G; and Qian, H. (2006). Chromatin looping and the probability of transcription. *TRENDS in Genetics* **22(4)**, 197-202.

Liberman, LM; and Stathopoulos, A. (2009). Design flexibility in cis-regulatory control of gene expression: synthetic and comparative evidence. *Developmental Biology* **327**, 578-589.

Ludwig, MZ; and Kreitman, M. (1995). Evolutionary dynamics of the enhancer region of *even-skipped* in *Drosophila*. *Molecular Biology and Evolution* **12(6)**, 1002-1011.

Ludwig, MZ; Patel, NH; and Kreitman, M. (1998). Functional analysis of *eve* stripe 2 enhancer evolution in *Drosophila*: rules governing conservatoin and change. *Development* **125**, 949-958.

Ludwig, MZ; Bergman, C; Patel, NH; and Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* **403**, 564-567.

Ludwig, MZ. (2002). Functional evolution of noncoding DNA. *Current Opinion in Genetics and Development* **12**, 634-639.

Ludwig, MZ; Palsson, A; Alekseeva, E; Bergman, CM; Nathan, J; and Kreitman, M. (2005). Functional evolution of a cis-regulatory module. *PloS Biology* **3**, e93.

Malik, S; and Roeder, RG. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. *TRENDS in Biochemical Science* **30(5),** 256-263.

Marcellini, S; and Simpson, P. Two or Four Bristles: Functional evolution of an enhancer of *scute* in Drosophilidae. *PloS Biology* **4(12)**, 2252-2261.

Misulovin, Z; Schwartz, YB; Li, XY; Kahn, TG; Gause, M; Macarthur, S; Fay, JC; Eisen, MB; Pirrotta, V; Biggin, MD; and Dorsett, D. (2008). Association of cohesin

and Nipped-B with transcriptionally active regions of the *Drosophila melanogaster* genome. *Chromosoma* **117**, 89-102.

Mlodzik, M; Hiromi, Y; Weber, U; Goodman, CS; and Rubin, GM. (1990). The *Drosophila seven-up* gene, a member of the Steroid Receptor gene superfamily, controls photoreceptor cell fates. *Cell* **60**, 211-224.

Moses, AM; Pollard, DA; Nix, DA; Iyer, VN; Li, X; Biggin, MD; and Eisen, MB. (2006). Large-scale turnover of functional transcription factor binding sites in *Drosophila*. *PloS Computational Biology* **2(10)**, 1219-1231.

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.

Nagaraj, R; and Banerjee, U. (2004). The little R cell that could. *Int J Dev Biol* **48**, 755-760.

Nagaraj, R; and Banerjee, U. (2007). Combinatorial signaling in the specification of primary pigment cells in the *Drosophila* eye. *Development* **134**, 825-831.

Narlikar, GJ; Fan, H; and Kingston, RE. (2002). Cooperation between complexes that regulate chromatin structure and transcription. *Cell* **108**, 475-487.

Nellesen, DT; Lai, EC; and Posakony, JW. (1999). Discrete enhancer elements mediate selective responsiveness of *Enhancer of split* complex genes to common transcriptional activators. *Developmental Biology* **213**, 33-53.

Noyes, MB; Christensen, RG; Wakabayashi, A; Stormo, GD; Brodsky, MH; and Wolfe, SA. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* **133**, 1277-1289.

O'Neill, EM; Rebay, I; Tjian, R; and Rubin, GM. (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.

Orphanides, G; and Reinberg, D. (2002). A unified theory of gene expression. *Cell* **108**, 439-451.

Ovcharenko, I; Loots, GG; Hardison, RC; Miller, W; and Stubbs, L. (2004). zPicture: dynamic alignment and visualization tool for analyzing conservation profiles. *Genome Resources* **14(3)**, 472-477.

Pan, YF; Wansa, KDSA; Liu, MH; Zhao, B; Hong, SZ; Tan, PY; Lim, KS; Borque, G; Liu, ET; and Cheung, E. (2008). Regulation of Estrogen-Receptor-mediated long range transcription via evolutionarily conserved distal response elements. *Journal of Biological Chemistry* **283**, 32977-32988.

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

Piano, F; Parisi, MJ; Karess, R; and Kambysellis, MP. (1999). Evidence for redundancy but not *trans* factor-*cis* element coevolution in the regulation of *Drosophila Yp* genes. *Genetics* **152**, 605-616.

Pickup, AT; Ming, L; and Lipshitz, HD. (2009). Hindsight modulates Delta expression during *Drosophila* cone cell induction. *Development* **136**, 975-982.

Price, MD; and Lai, Z-C. (1999). The *yan* gene is highly conserved in *Drosophila* and its expression suggests a complex role throughout development. *Development Genes and Evolution* **209**, 207-217.

Rebay, I; and Rubin, GM. (1995). Yan functions as a general inhibitor of differentiation and is negatively regulated by activation of the Ras1/MAPK pathway. *Cell* **81**, 857-866.

Reig, G; Cabrejos, ME; Concha, ML. (2007). Functions of BarH transcription factors during embryonic development. *Developmental Biology* **302**, 367-375.

Richards, S; Liu, Y; Bettencourt, BR; Hradecky, P; Letovsky, S; Nielsen, R; Thornton, K; Hubisz, MJ; Chen, R; Meisel, RP; Couronne, O; Hua, S; Smith, MA; Zhang, P; Liu, J; Bussemaker, HJ; van Batenburg, MF; Howells, SL; Schere, SE; Sodergren, E; Matthews, BB; Crosby, MA; Schroeder, AJ; Ortiz-Barrientos, D; Rives, CM; Metzker, ML; Muzny, DM; Scott, G; Steffen, D; Wheeler, DA; Worley, KC; Havlak, P; Durbin, KJ; Egan, A; Gill, R; Hume, J; Morgan, MB; Miner, G; Hamilton, Y; Huang, Y; Waldron, L; Verduzco, D; Clerc-Blankenburg, KP; Dubchak, I; Noor, MAF; Anderson, W; White, KP; Clardk, AG; Schaeffer, SW; Gelbart, W; Weinstock, GM; and Gibbs, RA. (2005). Comparative genome sequencing of *Drosophila pseudoobscura*: Chromosomal, gene, and cis-element evolution. *Genome Resources* **15**, 1-18.

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

Robinow, S; and White, K. (1988). The locus *elav* of *Drosophila melanogaster* is expressed in neurons at all developmental stages. *Developmental Biology* **126**, 294-303.

Rubin, GM; and Spradling, AC. (1982). Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218**, 348-353.

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.

Schulz, RA; Xie, XL; Miksch, JL. (1990). *Cis*-acting sequences required for the germ lines expression of the *Drosophila gonadal* gene. *Developmental. Biology*. **140**, 455-458.

Schwartz, S; Kent, WJ; Smit, A; Zhang, Z; Baertsch, R; Hardison, RC; Haussler, D; and Miller, W. (2003). Human-mouse alignments with BLASTZ. *Genome Res* **13**, 103-107.

Schweisguth, F; and Posakony, JW. (1992). *Suppressor of Hairless*, the *Drosophila* homolog of the mouse *recombination signal-binding protein* gene, controls sensory organ cell fates. *Cell* **69**, 1199-1212.

Senger, K; Armstrong, GW; Rowell, WJ; Kwan, JM; Markstein, M; and Levine, M. (2004). Immunity regulatory DNAs share common organization features in *Drosophila*. *Molecular Cell* **13**, 19-32.

Serikaku, MA; and O'Tousa, JE. (1994). *Sine oculis* is a homeobox gene required for *Drosophila* visual system development. *Genetics* **138**, 1137-1150.

Seto, H; Hayashi, Y; Kwon, E; Taguchi, O; and Yamguchi, M. (2006). Antagonistic regulation of the *Drosophila* PCNA gene promoter by DREF and Cut. *Genes Cells* **11**, 499-512.

Shapiro, MD; Marks, ME; Pelchel, CL; Blackman, BK; Nereng, KS; Jonsson, B; Schluter, D; and Kingsley, DM. (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717-723.

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. (2003). Signaling by the *Drosophila* epidermal growth factor receptor pathway during development. *Experimental Cell Research* **284**, 140-149.

Shirokawa, JM; and Courey, AJ. (1997). A direct contact between the Dorsal rel homology domain and Twist may mediate transcriptional synergy. *Molecular and Cellular Biology* **17(6)**, 3345-3355.

Silver, SJ; and Rebay, I. (2005). Signaling circuitries in development: insights from the retinal determination gene network. *Development* **132(1)**, 3-13.

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

Song, SH; Hou, C; and Dean, A. (2007). A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. *Molecular Cell* **28**, 810-822.

Strasser, K; and Hurt, E. (2000). Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. *The EMBO Journal* **19(3)**, 410-420.

Strasser, K; Masuda, S; Mason, P; Pfannstiel, J; Oppizzi, M; Rodriguez-Navarro, S; Rondon, AG; Aguilera, A; Struhl, K; Reed, R; and Hurt, E. (2002). TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417**, 304-308.

Stultz, BG; Jackson, DG; Mortin, MA; Yang, X; Beachy, PA; and Hursh, DA. (2006). Transcriptional activation by Extradenticle in the *Drosophila* visceral mesoderm. *Developmental Biology* **290**, 482-494.

Sullivan, W; Ashburner, M; and Hawley, RS. (2000). *Drosophila* protocols. *Cold Spring Harbor Laboratory Press*; Cold Spring Harbor, New York.

Sun, B; Hursh, DA; Jackson, D; and Beachy, PA. (1995). Ultrabithorax protein is necessary but not sufficient for full activation of *decapentaplegic* expression in the visceral mesoderm. *EMBO* **14**, 520-535.

Swanson, CI; Hinrichs, T; Johnson, LA; 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, CI; Evans, NC; and Barolo, S. (2010). Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. *Developmental Cell*, accepted.

Szutorisz, H; Dillon, N; and Tora, L. (2005). The role of enhancers as centres for general transcription factor recruitment. *TRENDS in Biochemical Science* **30(11)**, 593-599.

Takahashi, H; Mitani, Y; Satoh, G; and Satoh, N. (1999). Evolutionary alterations of the minimal promoter for notochord-specific *Brachyury* expression in ascidian embryos. *Development* **126**, 3725-3734.

Tatusova, T.A., and Madden, T.L. (1999). BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* **174**, 247-

250.
Thanos, D; and Maniatis, T. (1995). Virus induction of human IFNβ gene expression requires the assembly of an enhanceosome. *Cell* **83**, 1091-1100.
Tolhuls, B; Palstra, R; Splinter, E; Grosveld, F; and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active β-globin locus. *Molecular Cell* **10**, 1453-1465.
Tsuda, L; Nagaraj, R; Zipursky, SL; and Banerjee, U. (2002). An EGFR/Ebi/Sno pathway promotes delta expression by inactivation Su(H)/SMRTER repression during inductive Notch signaling. *Cell* **110**, 625-637.
Tsytsykova, AV; Rajsbaum, R; Falvo, JV; Ligeiro, F; Neely, SR; and Goldfeld, AE. (2007). Activation-dependent intrachromosomal interactions formed by the *TNF* gene promoter and two distal enhancers. *PNAS* **104(43)**, 16850-16855.
Voas, MG; and Rebay, I. (2004). Signal integration during development: insights from the *Drosophila* eye. *Developmental Dynamics* **229**, 162-175.
Wang, G; Balamotis, MA; Stevens, JL; Yamaguchi, Y; Handa, H; and Berk, AJ. (2005). Mediator requirement for both recruitment and postrecruitment steps in transcription initiation. *Molecular Cell* **17**, 683-694.
Wang, Q; Carroll, JS; and Brown, M. (2005). Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Molecular Cell* **19**, 631-642.
Wildonger, J; Sosinsky, A; Honig, B; and Mann, RS. (2005). Lozenge directly activates *argos* and *klumpfuss* to regulate programmed cell death. *Genes and Development* **19**, 1034-1039.
Williams, AO; Isaacs, RJ; and Stowell, KM. (2007). Down-regulation of human topoisomerase IIα expression correlates with relative amounts of specificity factors Sp1 and Sp3 bound at proximal and distal promoter regions. *BMC Molecular Biology* **8(36)**.
Williams, TM; Selegue, JE; Werner, T; Gompel, N; Kopp, A; and Carroll, SB. (2008). The regulation and evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*. *Cell* **134**, 610-623.
Wittkopp, P. (2006). Evolution of cis-regulatory sequence and function in Diptera. *Heredity* **97**, 139-147.
Wong, C; Sridhara, S; Bardwell, JCA; and Jakob, U. (2000). Heating greatly speeds Coomassie blue staining and destaining. *Biotechniques* **28**, 426-432.
Wratten, NS; McGregor, AP; Shaw, PJ; and Dover, GA. (2006). Evolutionary and functional analysis of the *tailless* enhancer in *Musca domestica* and *Drosophila melanogaster*. *Evolution and Development* **8(1)**, 6-15.
Xu, C; Kauffmann, RC; Zhang, J; Kladny, S; and Carthew, RW. (2000). Overlapping activators and repressors delimit transcriptional response to receptor tyrosine kinase signals in the *Drosophila* eye. *Cell* **103**, 87-97.
Xu, H; and Hoover, TR. (2001). Transcriptional regulation at a distance in bacteria. *Current Opinion in Microbiology* **4**, 138-144.
Yamamoto, T; Kawamoto, R; Fujii, T; Sakamoto, N; and Shibata, T. (2007). DNA variations within the sea urchin *Otx* gene enhancer. *FEBS Letters* **581**, 5234-5240.

Yao, L; 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.

Yang, X; van Beest, M; Clevers, H; Jones, T; Hursh, DA, and Mortin, MA. (2000). *Decapentaplegic* is a direct target of dTcf repression in the *Drosophila* visceral mesoderm. *Development* **127** 3695-3702.

Yoshida, C; Tokumasu, F; Hohmura, KI; Bungert, J; Hayashi, N; Nagasawa, T; Engel, JD; Yamamoto, M; Takeyasu, K; and Igarashi, K. (1999). Long range interaction of cis-DNA elements mediated by architectural transcription factor Bach1. *Genes Cells* **4**, 643-655.

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

Zaffran, S; Kuchler, A; Lee, HH; and Frasch, M. (2001). *Biniou* (*FoxF*), a central component in a regulatory network controlling visceral mesoderm development and midgut morphogenesis in *Drosophila*. *Genes and Development* **15**, 2900-2915.