Stony Brook University

Interplay between distinct cis-regulatory modules mediates combinatorial regulation by Runt and other pair-rule transcription factors during Drosophila segmentation

A Dissertation Presented

by

Lisa Prazak

to

The Graduate School

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

In

Molecular and Cellular Biology

(Biochemistry and Molecular Biology)

Stony Brook University

December 2009

Stony Brook University

The Graduate School

Lisa Prazak

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, herby recommend

Acceptance of this dissertation.

Dr. J. Peter Gergen, Ph.D. – Dissertation Advisor
Professor, Department of Biochemistry and Cell Biology

Dr. John Reinitz, Ph.D. – Chairperson of Defense
Professor, Department of Applied Mathematics and Statistics

Dr. Patrick Hearing, Ph.D.
Professor, Department of Molecular Genetics and Microbiology

Dr. Stephen D. Nimer, M.D.
Professor of Molecular Pharmacology and Chemistry

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Interplay between distinct cis-regulatory modules mediates combinatorial regulation by Runt and other pair-rule transcription factors during Drosophila segmentation

by

Lisa Prazak

Doctor in Philosophy

In

Molecular and Cellular Biology

(Biochemistry and Molecular Biology)

Stony Brook University

2009

The relatively simple combinatorial rules responsible for establishing the initial metameric expression of *sloppy-paired-1 (slp1)* in two cell wide stripes in the posterior half of each parasegment in the Drosophila embryo make this system an attractive model for investigating the mechanism of regulation by Runt and other pair-rule transcription factors. I analyzed various *slp1-lacZ* reporter genes in order to identify cis-regulatory sequences responsible for this early pattern. This work identifies two distinct elements, a proximal early stripe element (PESE) and a distal early stripe element (DESE) located

from -3.8 kb to -1.8 kb and from -8.7 kb to -6.6 kb upstream of the *slp1* promoter, respectively. The distal element expresses both odd and even-numbered stripes with inappropriate expression in the anterior half of the odd-numbered parasegments due to an inability to respond to repression by Even-skipped (Eve). In contrast, the proximal element expresses only even-numbered stripes and mediates repression by Eve as well as by the combination of Runt and Fushi-tarazu (Ftz). A minimal PESE element retains Eve-dependent repression, but is expressed throughout the even-numbered parasegments due to the loss of repression by Runt and Ftz. Importantly, a composite reporter gene containing both cis-elements faithfully recapitulates *slp1* regulation in a manner beyond what is expected from combining their individual patterns. I propose a model whereby the metameric expression of *slp1* is achieved through Runt dependent regulation of interactions between the DESE and PESE enhancers and the *slp1* promoter. The role of DNA-binding by Runt in *slp1* regulation was further investigated by examining expression of *DESE-lacZ* reporters containing mutations in different Runt binding sites. These results indicate these sites are important for repression of *DESE-lacZ* but are not essential for activation. Complementary studies using a DNA-binding defective form of Runt confirm the importance of DNA-binding for *slp1* repression and also reveal a role for DNA-binding by Runt in activation of *slp1* and the *DESE-lacZ* reporter. This work reveals new insights into Runt dependent regulation and opens the door for future studies on the mechanisms underlying the developmental regulation of gene expression.

Table of Contents

List of Figures

Acknowledgments

First and foremost, I would like to thank my thesis advisor, Dr. J. Peter Gergen, for his kindness, patience, and support as he mentored me throughout my time as a graduate student. He has provided me with endless guidance and advice which has been invaluable to both my research and my life. He is truly an outstanding individual and I have been lucky to be part of his lab.

I would like to sincerely thank my committee members, Dr. John Reinitz, Dr. Patrick Hearing and Dr. Stephen Nimer for their helpful suggestions with my thesis project over the past few years, and for their comments, which have improved the contents of this dissertation. I would especially like to thank John for agreeing to serve as my chairperson, and for always providing me with thoughtful feedback.

Many thanks to Miki Fukioka for the initial identification of the distal element, a main focus of my research, as well as her helpful comments over the course of the project. I would also like to thank the Calos lab for providing me with the reagents for site-specific integration and Yelana Altshuller from Stony Brook Molecular cloning facility.

Thanks to all the past and present Gergen lab members in particular John Wheeler and Deborah Swantek for the preliminary studies of the *slp1* cis-regulatory region, which laid the groundwork for these studies. Also, Pegine Walrad, Saiyu Hang and Conrad Tenenbaum. I would like to give a special thanks to Xiaoling Wang for not only providing me with scientific advise but for being a great friend throughout our time in the lab together. I would also like to thank Hilde Janssens and especially Ah-Ram Kim from the Reinitz lab for their technical advice and suggestions.

I also must thank all my friends at Stony Brook (Mary, Akua, Erin, Jamie, Lindsay, Jason, Tim, Malgosia, Grigor, Rahul, Aftab, Nusrat, John, Mike) with a special thanks to Lindsay Nelson for there support.

I especially have to give infinite thanks to my family and my husband for always having faith in me and being an endless source of support throughout my graduate career and my life.

Chapter I

Introduction to Dissertation

Chapter I

Introduction

Cis-reguatory elements in prokaryotes and eukaryotes

In 1965 Francois Jacob and Jacques Monod shared the Nobel Prize in Medicine with Andre Lwoff for their work on the *lac* operon which described transcriptional regulation of gene expression in bacteria and was the first genetic regulatory mechanism elucidated showing that genes were regulated to control the production of protein products through an mRNA intermediate (Jacob and Monod, 1961). Transcription in prokaryotic cells involves the proper recognition of a promoter by a sigma initiation factor that is responsible for specific binding of RNA polymerase to the target gene promoter. Once the preinitiation complex is established, activators and repressors act directly on RNA polymerase to regulate transcription. In eukaryotic cells the core components are similar such as the conserved core of RNA polymerase and the mechanism of transcription but the complexity differs from its prokaryotic counterpart (Sweetser et al., 1987). There are five RNA polymerase enzymes responsible for different aspects of RNA synthesis (Grummt, 1999; Herr et al., 2005; Matsui et al., 1980; Wierzbicki et al., 2009; Willis, 1993). Amongst the three nuclear polymerases, RNA polymerase II is responsible for generating messenger RNA for protein coding genes. Like in prokaryotes, proper initiation of transcription requires accessory proteins but unlike a single sigma factor there are a number of general transcription factors, TFII-B, D, E, F and H in eukaryotic cells responsible for initiation of transcription (Roeder, 1996). Although the central components are conserved, eukaryotic cells have higher levels of regulation, which have to account for the organization of DNA into

nucleosomes and higher order chromatin structures. Regulation does not involve direct interactions of activators and repressors with only promoter-proximal cis-regulatory sequences but instead involves cis-regulatory enhancers and silencers that can be thousands of base pairs removed from the site of transcription initiation. Transcription factors that bind these cis-regulatory control elements also are not thought to contact RNA polymerase directly but instead recruit non DNA-binding proteins that serve as co-activators or co-repressors. In accord with this, Mediator, a complex not present in prokaryotic cells, plays a major role in regulation of eukaryotic transcription where it is involved in relaying activation signals from gene specific enhancers (Flanagan et al., 1991; Kelleher et al., 1990; Kim et al., 1994).

The ability of enhancers to communicate a regulatory signal to a specific promoter is well established but the mechanisms by which enhancers and the DNA binding transcription factors that bind them function in regulating transcription in eukaryotic cells at distant promoters is not well understood. There are two prevalent models for the mechanism involved in enhancer communication. One model that has been proposed and experimentally supported, involves interactions of long-range enhancers with their target gene promoters resulting in the formation of chromatin loops of intervening DNA. The best evidence for this model comes from experiments on the *β-globin* locus using the Chromosome Conformation Capture technique (3C), which allows for detection of inter an intra-chromosomal interactions (Dekker et al., 2002). There are 5 human *β-globin* genes within this locus arranged in a linear fashion in the order they are developmentally expressed with the 5'-most *ε-globin* gene being expressed first in the primitive erythroid cells of the embryo (Levings and Bungert, 2002; Noordermeer and de

Laat, 2008). The locus control region (LCR), which is between 40 kb and 60 kb removed from it's target genes, is responsible for the proper regulation of this locus as deletion of the LCR results in a reduction of globin gene expression and this region has been shown to associate with the promoters of *β-globin* genes within this locus (Bender et al., 2000; Epner et al., 1998; Tolhuis et al., 2002). Along with evidence of enhancer-promoter interactions, the DNA binding transcription factors involved in *β-globin* gene expression are thought to play a role in this interaction as deletion of Erythroid Krüppel-like transcription factor and the hematopoietic transcription factor GATA-1 and it's cofactor FOG-1 are required for interactions between the LCR and β^{maj} gene promoter and are required for adult *β-globin* gene expression (Drissen et al., 2004; Vakoc et al., 2005). This suggests that the transcription factors bound to enhancers could provide a bridge for these interactions although how this may result in proper regulation of transcription is not understood.

A question that is raised by the looping model described above is the ability of an enhancer to selectively act on a specific target promoter without affecting other genes within the vicinity of a particular *cis*-regulatory element. Although the idea of chromatin looping seems to be a well established model for transcriptional regulation by distant enhancers, it does not explain the properties of boundary or insulator elements that have been identified in regulating specific communications between an enhancer and a target promoter (Chung et al., 1993; Kellum and Schedl, 1992). The mechanisms by which insulator elements function in eukaryotic cells to prevent particular interactions between an enhancer and a non target promoter element is hard to comprehend if a looping mechanism is involved as this chromatin configuration would be thought to be able to

bypass such an element. Another model that could explain the properties of insulators in regulating transcription would be by linking or tracking where a signal is propagated along DNA explaining the ability of insulators to function. The ability of an insulator to block communication when placed in between an enhancer and a promoter seems to correlate better with a tracking model for transcriptional regulation (Barges et al., 2000; Hagstrom et al., 1996).

The transcriptional control of gene expression described above is essential for development and the proper regulation of eukaryotic transcription relies on specific interactions between *cis*-regulatory elements and *trans*-acting proteins that act through sequence specific contacts on DNA. These interactions result in the assembly of multi-protein regulatory complexes and enhancer/promoter interactions that promote the proper regulation of target gene expression. Although many of the general factors have been identified, the mechanism by which this regulation occurs is not well understood. Further studies on transcriptional regulation would benefit from studies on well-defined enhancers in a system amenable to experimental manipulation.

Segmentation in Drosophila

The segmentation pathway in Drosophila has been subject to extensive genetic and molecular analysis, making this an excellent model system for investigating the mechanisms of transcriptional regulation. The regulation of *hunchback* (*hb*) by the maternal Bicoid gradient is just one example of the extent to which this system has been examined (Tautz, 1988). The research involving the regulation of *hb* revealed that three Bicoid binding sites within the promoter region from -300 bp to -50 bp are necessary and sufficient for activation of *hb* and the affinity of these sites defines the domain of zygotic

gene expression (Driever and Nusslein-Volhard, 1989; Driever et al., 1989). Another example of this well studied system is the regulation of *even-skipped* stripe 2 which is activated in response to Bicoid and Hunchback and which has its anterior and posterior borders defined by repression from Giant and Kruppel, respectively (Small et al., 1992). This regulation is dependent on a 480 bp region from -1070 bp to -1550 bp upstream of *eve* that is regulated by combinatorial cues resulting from multiple binding sites for each of these four different transcription factors (Ingham et al., 1988; Small et al., 1992). These are just two examples that indicate the amount of information present in this system.

One of the key transcriptional regulators of segmentation in the early Drosophila embryo is the primary pair-rule gene *runt*. Runt is the founding member of a family of heteromeric DNA-binding transcription factors that share a 128 amino acid motif termed the Runt domain due to sequence homology to the Runt protein. The Runx family of transcription factors are able to both activate and repress there targets depending on the developmental context. An intriguing aspect of these proteins that is not understood is the ability to carry out both of these regulatory functions in the same cell (Canon and Banerjee, 2003; Fu and Noll, 1997; Stein et al., 2004).

The Runx family of transcription factors is conserved from see urchins to humans where it functions during hematopoiesis, osteogenesis and neurogenesis and alterations affecting the function of this protein or resulting in misexpression can result in leukemias, cleidocranial dysplasia, defective neuronal connectivity and stomach cancer (Fig. 1.1) (Cohen, 2001; Nimmo and Woollard, 2008). The multitude of developmental processes regulated by Runx transcription factors highlights the importance of this family to normal

development and the need to investigate how these proteins function to better understand why things go awry when function is disrupted. In Drosophila there are four Runx genes, two of which, *lozenge* and *runt* are involved in visual system development, sex determination, neurogenesis and segmentation (Canon and Banerjee, 2000; Duffy and Gergen, 1991; Duffy et al., 1991; Wheeler et al., 2000). Given the well conserved nature of the Runx transcription factor family as well as the wealth of knowledge of the factors involved in transcriptional regulation, the Drosophila segmentation pathway provides a good starting point in investigating the mechanisms of transcriptional regulation and the role this family of transcription factors plays in this process.

During the segmentation process Runt functions to activate *fushi tarazu* (*ftz*) and repress *hairy* (*h*) and *even-skipped* (*eve*) (Aronson et al., 1997). Runt can also activate and repress the same target gene during the segmentation process such as *sloppy-paired* (*slp*), *wingless* (*wg*) and *engrailed* (*en*) (Fig. 1.2) (Aronson et al., 1997; Swantek and Gergen, 2004). DNA-binding by Runt is needed for regulation but in some cases, such as the initial repression of *en*, a DNA-binding defective form of Runt can function to establish repression but is not able to maintain this repression (Wheeler et al., 2002).

Runt works with other pair-rule transcription factors to establish the metameric pattern of the *Drosophila* embryo. The activities of segmentation and homeotic genes are responsible for positional identity along the anterior posterior axis of the *Drosophila* embryo (Harding et al., 1986). Segmentation genes are responsible for setting the polarity of each segment and the homeotic genes are responsible for selecting the identity of a certain segment that will eventually result in the adult phenotype. The homeotic gene family, first discovered by observing that their absence or misexpression resulted in

homeotic transformations of one embryonic segment to another, shares a 60 amino acid domain encoded by a 180 base-pair DNA sequence known as the homebox (Lewis, 1978). These proteins are transcription factors that bind the sequence TAAT and are important for patterning the anterior-posterior axis of both vertebrates and invertebrates.

The pair-rule genes *eve* and *ftz* play vital roles in the establishment of the segmented body plan of the *Drosophila* embryo and both contain a homebox. These proteins, are members of a subset of the super family of homeotic proteins that have a glutamine at position 50 of the homeodomain and have been shown to bind with similar affinity to the same wide range of DNA-binding sites, showing preferential binding to certain regions in the promoter of known target genes (Carr and Biggin, 1999). However, *in vivo* these proteins bind these sites throughout the length of their target genes along with binding at a somewhat lower affinity to genes not known to be regulated by these proteins (Walter et al., 1994). This is a unique property of the homeodomain transcription factors as a non-homeodomain transcription factor, Zeste, was detected only on short elements within a target promoter and not on other genes (Walter et al., 1994). The broad DNA recognition properties of these proteins *in vitro* are likely to be important determinants of their distribution on DNA *in vivo*, but it also shows that *in vitro* DNA binding specificity alone is not sufficient to explain the distribution of these proteins in embryos (Walter and Biggin, 1996). The similar DNA binding of Eve and Ftz *in vivo* supports the views that the functional specificities of the selector homeoproteins results from differences in the way they each activate or repress common target genes (Biggin and McGinnis, 1997).

A model that has been proposed by Biggin and McGinnis (Biggin and McGinnis, 1997) suggest that the Q50 homeoproteins bind many of their recognition sites without the aid of cofactors. In this model, cofactors merely aid in helping to distinguish the way in which homeoproteins regulate targets to which they are already bound. It is important to determine where these proteins are binding *in vivo* to resolve the mechanisms by which they function.

All three of the aforementioned transcription factors act to regulate the segmentation gene *slp1* (Fig. 1.3). The identification of *slp1* being involved in segmentation came from large-scale mutagenesis screens that revealed mutations in this locus resulted in embryonic lethality affecting pattern formation in the *Drosophila* embryo (Nusslein-Volhard, 1984). The *sloppy paired* locus consists of two genes, *slp1* and *slp2* that are structurally related containing a *fork head* domain that has been found in mammalian hepatocyte transcription factors (Grossniklaus et al., 1992). The Slp protein functions in maintaining segment polarity by acting to maintain *wg* expression and inhibit *en* expression (Cadigan et al., 1994). These two genes seem to be redundant in their function but *slp1* is expressed earlier then *slp2* and is required for early function during segmentation. In the blastoderm embryo the first metameric expression of *slp1* is apparent where it is expressed in seven stripes at double segment periodicity (Grossniklaus et al., 1992). At stage 6 when gastrulation begins, seven more stripes are expressed in between the first set resulting in expression in each parasegment (Grossniklaus et al., 1992). The final pattern of *slp1* during segmentation consists of 14 stripes, two cells in width, in the posterior most cells of each parasegment (Figure 1.3).

The majority of regulatory factors responsible for this early *slp1* regulation have been identified making this gene very amenable to experimental manipulation.

Runt in combination with Odd-paired (Opa) is responsible for *slp1* expression in odd parasegments and ectopic expression of Runt and Opa can activate *slp1* in all cells that do not express Ftz (Fig. 1.3) (Swantek and Gergen, 2004). Runt in combination with Ftz is able to repress *slp1* throughout the gastrula stage embryo (Swantek and Gergen, 2004). *Eve* and *ftz* are first expressed in 4 cell wide stripes in alternating segments at the blastoderm stage. As the pattern of expression of these two genes evolves the posterior two cells in each stripe lose expression resulting in alternating stripes in the anterior half of each parasegment with the highest levels of expression providing the anterior border (Fig. 1.3). Eve is thought to be responsible for repressing *slp1* in the anterior half of the odd parasegments while Ftz is needed to repress *slp1* in the anterior half of the even parasegments (Fig. 1.3) (Swantek and Gergen, 2004). One aspect of *slp1* expression that is not accounted for is activation of the even numbered *slp1* stripes. In the model shown in figure 1.3, this unknown component of *slp1* regulation is indicated as factor X. To understand *slp1* regulation fully we need to identify factor X. Although the majority of the factors responsible for *slp1* regulation have been identified, the mechanisms by which these factors act to regulate transcription is unknown. The most well understood target of Runt in the segmentation pathway is *slp1* and it is expressed at a stage that is accessible to experimental manipulation thereby making this an attractive model system for investigating Runt function and homeodomain response.

In an attempt to characterize the mechanism by which the above transcription factors function to regulate *slp1* it is important to identify the cis-regulatory elements that

mediate their action. Drosophila allows us to identify these elements by means of P-element mediated germline transformation to generate fly strains carrying chromosomally integrated reporter gene constructs. The work described here takes advantage of these tools to investigate the cis-regulatory sequences responsible for *slp1* expression. To do this the *slp* locus was dissected to identify regions able to mediate *slp1* regulation, concentrating our efforts upstream of *slp1,* as there is much evidence that the enhancers responsible for regulating both *slp1* and *slp2* are present 5' of *slp1* (Grossniklaus et al., 1992). Although P-element transformation is a valuable tool, the standard protocols for P-element mediated integration of reporter genes do not allow for control of where the transgene integrates resulting in position effects that can influence the strength if not also, the pattern of reporter gene expression. To overcome the problems associated with position effects on reporter gene expression it is common practice to examine expression from 3 to 5 independent lines before coming to a conclusion on the regulatory properties of a specific reporter gene. The differences in expression levels between lines make it very difficult, if not impossible to reliably detect subtle changes in expression. To analyze in detail the regulatory structure of a specific enhancer, the ability to have the reporters of interest insert into the same site in the genome is beneficial and allows for quantitative analysis of reporter gene expression. In an attempt to control the chromosomal location of insertion we have taken advantage of a site-specific integrase from *Streptomyces* phage, ϕC31.

Bacteriophages possess the ability to integrate into a specific site in the host bacterial genome when the environmental conditions do not permit their replication and release into the environment after infection. In *Streptomyces* phage ϕC31 this site-

specific recombination is carried out by a single subunit integrase enzyme encoded in the phage genome (Groth and Calos, 2004). Recent work has shown the ϕC31 integrase functions efficiently at site-specific, unidirectional integration in mammalian cells, mediating integration at introduced *att* sites or native sequence that have partial identity to *att* sites (Groth et al., 2000). Furthermore this system has also been shown to function efficiently in *Drosophila melanogaster* embryos (Groth et al., 2004). The Calos lab at Stanford University developed two independent fly lines that contain the *att*P sequence on the second and third chromosome, respectively. They have shown that co-injecting a reporter construct that contains an *att*B site along with ϕC31 RNA can allow for site-specific recombination at the desired location (Fig. 1.4). Using this system we have integrated the majority of the *slp1* cis-regulatory reporter gene constructs used in this work into the attP landing site on the third chromosome.

The work described here will first describe the identification of two distinct cis-regulatory modules (CRMs) that are able to recapitulate certain aspects of *slp1* regulation and when combined in a composite reporter respond properly to the pair-rule transcription factors responsible for *slp1* expression in a manner that can not be explained by the additive inputs of these two CRMs. This suggests that the proper integration of pair-rule regulatory cues involves interactions between these elements. Also, these CRMs respond differently to the homeodomain proteins Eve and Ftz, showing a functional difference in mediating transcriptional regulation by these factors. In the work presented in the following chapter I will focus on the distal element alone and investigate the importance of Runx binding sites within this element. These results demonstrate the importance of DNA-binding by Runt for proper regulation of *slp1* and *DESE-lacZ* and

show that Runx sites contribute to Runt dependent repression but are not critical for activation of this element. In the final chapter I will describe future experiments that address some of the unanswered questions on how these elements are able to elicit there function.

Figures

Figure 1.1

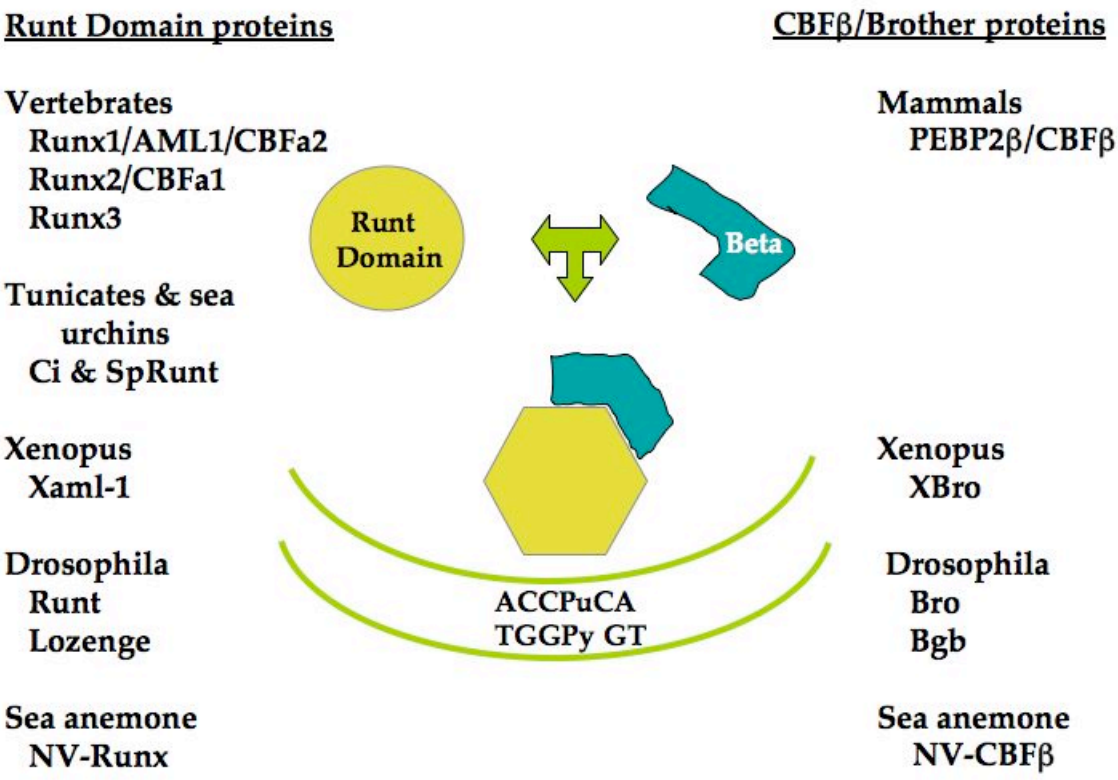


Figure 1.1. Runx Transcription Factor Family

Schematic diagram of conserved Runt domain transcription factors on the left and the corresponding Beta partner on the right. Runt domain transcription factors bind DNA as a heterodimer at the consensus ACCpuCA where the beta partner enhances DNA-binding.

Figure 1.2

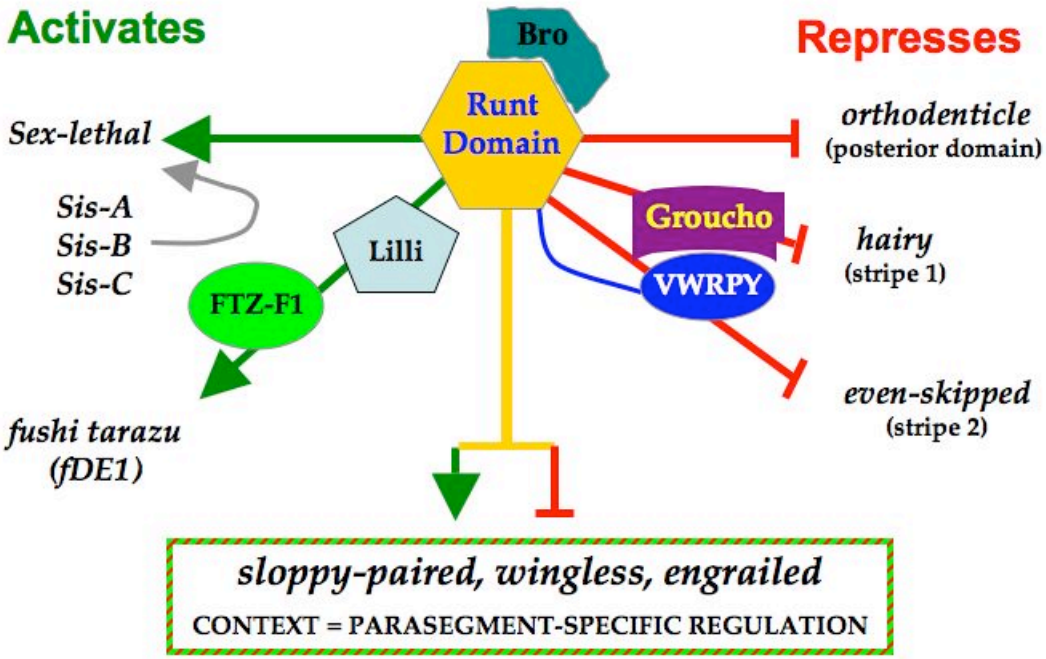

Activates
Bro
Represses
Runt Domain
Sex-lethal
orthodenticle
(posterior domain)
Sis-A
Sis-B
Sis-C
Lilli
Groucho
hairy
(stripe 1)
FTZ-F1
VWRPY
even-skipped
(stripe 2)
fushi tarazu
(fDE1)
sloppy-paired, wingless, engrailed
CONTEXT = PARASEGMENT-SPECIFIC REGULATION

Figure 1.2. Runt activates and Represses transcription

Schematic diagram of Runt target genes in Drosophila. Green arrows depict genes activated by Runt and red horizontal lines represent those genes that are repressed. Proteins on the lines of activation or repression, such as Lilli and Groucho are co-activators and repressors respectively. Groucho interacts with the VWRPY motif of Runt for repression. Genes in the green and red striped box at the bottom of the diagram are both activated and repressed by Runt.

Figure 1.3

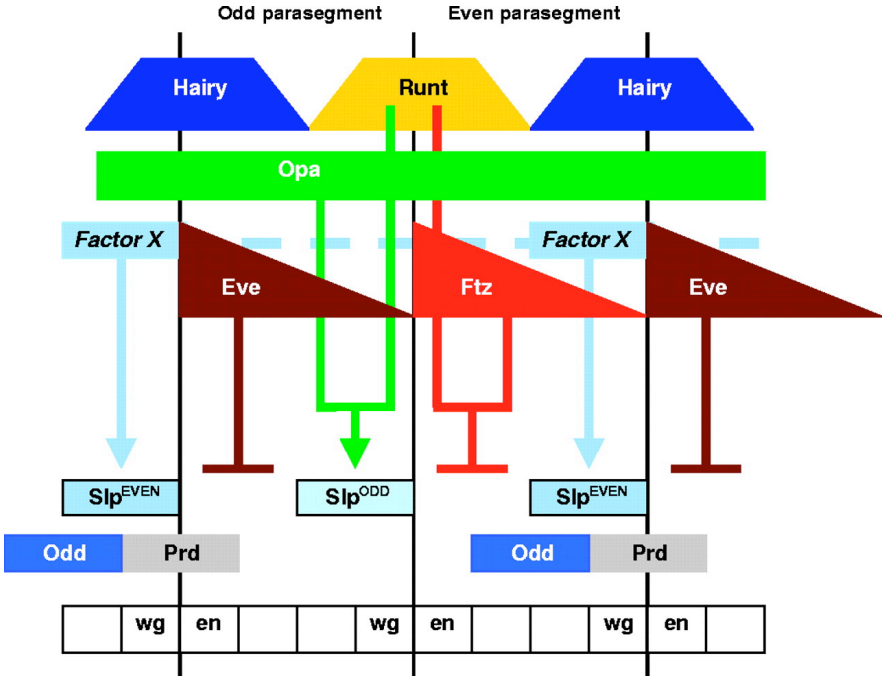

Odd parasegment
Even parasegment
Hairy
Runt
Hairy
Opa
Factor X
Eve
Ftz
Factor X
Eve
SlpEVEN
SlpODD
SlpEVEN
Odd
Prd
Odd
Prd
wg
en
wg
en
wg
en

Figure 1.3. Pair-rule to segment-polarity transition.

This figure diagrams the expression pattern of the pair-rule genes responsible for *slp1* regulation. The expression of *slp1* is in two cell wide stripes in the posterior half of each parasegment. Runt and Opa activate *slp1* in the odd-numbered parasegments, while Factor X activates *slp1* in the even parasegments. Runt and Ftz, present in the anterior half of the even parasegments repress *slp1* and Eve is responsible for repression of *slp1* in the anterior half of the odd parasegments.

Figure 1.4

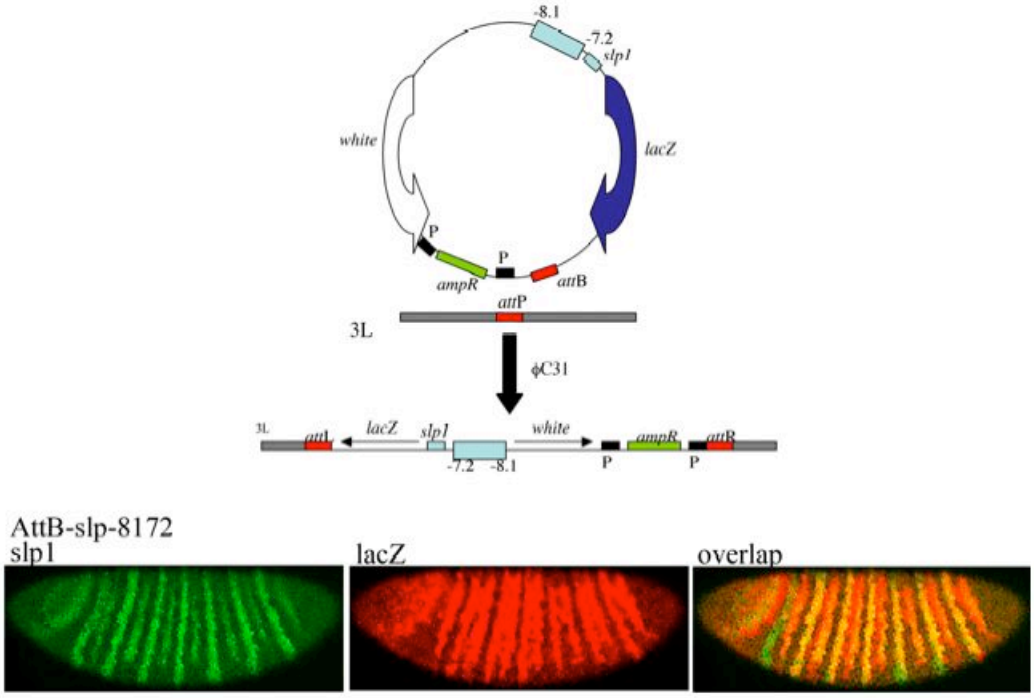

-8.1
-7.2
slp1
white
lacZ
P
P
ampR
attB
attP
3L
ϕC31
3L
attL
lacZ
slp1
white
ampR
attR
-7.2
-8.1
P
P
AttB-slp-8172
slp1
lacZ
overlap

Figure 1.4. Site-specific recombination using ΦC31.

The circle represents the CaSpeR based plasmid with an introduced *att*B site downstream of *lacZ* as well as the *slp1* basal promoter and distal regulatory sequence upstream of *lacZ*. In the presence of integrase the integration reaction proceeds, resulting in the reporter gene plasmid integrating into the left arm of the Drosophila chromosome, flanked by two hybrid sites, *att*L and *att*R with the resulting expression pattern of the reporter gene below showing expression of *slp1* (green), *lacZ* (red) and the merged image.

Chapter II

Materials and Methods

Chapter II

Materials and Methods

Construction of *slp1-lacZ* reporters

Reporter gene lines for the initial distal element reporter and composite reporter genes were generated using standard P-element germline transformation constructs (Figure 3.1 and 3.2). The full length DESE construct, pC:*slp1[8765]-lacZ*^{M}, contains sequence from 8,710 to 6,506 bp upstream of *slp1* PCR amplified with primers that add an XbaI site on the 5' end and a NotI site on the 3' end and cloned into the corresponding sites of a modified pCaSpeR vector (Fujioka et al., 1999) (Ellis et al., 1993; Fujioka et al., 1999) that also contains basal promoter sequences from 261 base pairs (bp) upstream to 121 bp downstream of the *slp1* transcription start site, including the first six codons of the protein. The composite DESE+PESE reporters were generated by insertion of a NotI fragment from pB:*slp1*[1839] into the NotI site of pC:*slp1[8765]-lacZ*^{M} to generate pC: *slp1[8765:3918]-lacZ* and pC:*slp1[8765:1839]-lacZ* . The pB:*slp1[1839]* subclone was generated by cloning a PCR product spanning from 3,926 to 1,774 bp upstream of *slp1*into pBluescript with flanking NotI restriction sites. Transgenic strains were generated by co-injecting these different reporter gene constructs into homozygous *y w[67c23]* embryos with the p:Δ2-3 helper plasmid (Rubin and Spradling, 1982). Multiple independent transgenic lines were examined for each construct.

Vectors that promote transgene integration into the same chromosomal site using ΦC31-mediated recombination were used for experiments with the *slp[3918]-lacZ*^{att} reporter and the different proximal element deletion constructs. Also, the majority of experiments with the distal element utilize site-specific integration with a truncated

slp1[8771]-lacZ^{att} distal element-containing reporter with the exception of ectopic Eve expression where a P-element vector containing the full length distal element described above was used. These constructs contain *slp1* basal promoter sequences spanning from -72 to +57, obtained by PCR amplification from genomic subclones with the addition of upstream *XhoI* and downstream *KpnI* sites and cloned into pBluescript to create pB:*slp1[BP]*. This basal promoter segment was moved into CaSpeR-AUG-βGal as an *Eco*RI + *Kpn*I fragment creating pC:*slp1*^{BP}*-lacZ*. To modify this vector for ΦC31-mediated transgenesis, the ΦC31 *attB* sequence was PCR amplified from pTA:attB (Groth et al., 2004) with primers that add flanking *PstI* sites and then cloned into the *PstI* site downstream of *lacZ* in pC:*slp1*^{BP}*-lacZ*^{att}. In the final pC:*slp1-link-lacZ*^{att} vector, a linker was introduced to replace Bluescript polylinker between the *Eco*RI and *Xho*I sites upstream of the basal promoter with unique *NotI*, *SphI, StuI*, and *SpeI* sites.

The pC:*slp1[8771]-lacZ*^{att} and Runx binding site mutant constructs pC:*slp1[8771m2]-lacZ*^{att}, pC:*slp1[8771m4]-lacZ*^{att}, pC:*slp1[8771m5]-lacZ*^{att}, pC:*slp1[8771m3,4]-lacZ*^{att} and pC:*slp1[8771m1,2,3,4,5]-lacZ*^{att} were generated using the In-Fusion^{TM} Dry-Down PCR Cloning Kit (Clontech) utilizing primers that amplify from 8,710 to 7,136 bp upstream of *slp1* from pB:slp1[8765] such that these DNA segment could be cloned into the *Xho*I site of the above pC:*slp1*^{BP}*-lacZ*^{att}. Runx binding site mutants were generated by PCR of pB:slp1[8765] using complementary primers that change the two critical cytosines in the consensus sequence to adenine resulting in two fragments that extend to the flanking *M13* forward and reverse primers. These fragments were mixed and amplified with *M13* forward and reverse primers and cloned into the *XbaI* and *NotI* site of pBluescript creating pB:slp1[8765m2], pB:slp1[8765m4],

pB:slp1[8765m5], pB:slp1[8765m3,4] and pB:slp1[8765m1,2,3,4,5]. The Runx binding site mutant constructs were then moved into pC:*slp1*^{BP}-*lacZ*^{att} as described above.

PESE-containing pC:*slp1[3918]-lacZ*^{att} was generated by cloning the *NotI* fragment from pB:*slp1[1839]* into the *NotI* site of pC:*slp1-link-lacZ*^{att} and then sequenced to verify insert orientation. To make pC:*slp1[3118]-lacZ*^{att}, pC:*slp1[3925]-lacZ*^{att}, pC:*slp1[3125]-lacZ*^{att} and pC:*slp1[PESE:C1+]-lacZ*^{att}, PCR products with *NotI* sites at both ends that span from 3,140 to 1,800, 3,900 to 2,519, 3,140 to 2,519 and from 3,179 to 2,908 bp upstream of *slp1*, respectively were generated from pB:*slp1[1839]* and inserted in the *NotI* site of pC:*slp1-link-lacZ*^{att}. The pC:*slp1[PESE:ΔC1]-lacZ*^{att} construct has an internal deletion that removes sequences from 3,135 to 2,981 bp upstream of *slp1*. PCR on pB:*slp1[1839]*was used to create two fragments with a sequence overlap spanning the desired deletion breakpoint and that extend to the flanking *M13* forward and reverse primers. These fragments were mixed and amplified with the *M13* primers, the resulting PCR product cloned as a *NotI* fragment and then moved into pC:*slp1-link-lacZ*^{att}. To generate the central deletion of the proximal element pC:*slp1[3931/2518]-lacZ*^{att}, pB:*slp1[1839]* was digested with *SalI* and a blunt end was created using the klenow enzyme to fill in the 3' recessed end. This was then digested with *XhoI* to make pB:*slp1[3931]/SalI*-blunt/*XhoI*. The same full length PESE starting construct was also digested with *AatII* and the exonuclease activity of T4 DNA polymerase was used to generate a blunt end in parallel and then digested with *XhoI* to excise from 2,519 to 1,800 bp upstream of *slp1* to be cloned into the linearized pB:*slp1[3931]*/*SalI*-blunt/*XhoI* vector above generating pB:*slp1[3931/2518]*. The 3931/2518 fragment was then cloned into pC:*slp1-link-lacZ*^{att} as a *NotI* fragment. The orientation and sequence integrity of all of

the above constructs was verified by sequencing. Primer sequences are provided at the end of this chapter.

ΦC31-mediated transgenic lines were obtained using the *attP* integration site on the third chromosome (Groth et al., 2004). Constructs containing the *attB* sequence were co-injected with ΦC31 mRNA into *y w ; P{CaryP}attP2* embryos, the surviving adult progeny backcrossed to the parental line and the F1 generation was screened for $white^{+}$ transfromants. ΦC31 mRNA was generated from *BamHI* linearized pET-phiC31-polyA template with the mMessage mMachine high yield Capped RNA Transcription Kit (Ambion) and mRNA was recovered via LiCl precipitation without DNase treatment.

Drosophila mutants and genetics

The alleles used for mutant analysis were eve^{1} (= *ID19*), run^{29} (= *YP17*) and ftz^{11} (= *W20*). To generate embryos heterozygous for *runt[29]* and the reporter gene, females of the genotype *cv v runt[29]/ y w* were mated to males homozygous for the reporter gene of interest. To generate embryos homozygous for the *eve[1]* allele and containing at least one copy of the reporter gene, *eve[1]*/CyO males were out-crossed to females homozygous for different third chromosome-linked reporters creating flies doubly heterozygous for *eve[1]* and the reporter gene. Female and male progeny were then backcrossed generating embryos where 1/6 are homozygous for *eve[1]* and contain at least one copy of the reporter. Reporter gene expression in *ftz* mutants was determined by generating recombinant stocks containing the *ftz[11]* mutation and different third chromosome linked reporter genes balanced over *TM3*.

Ectopic expression of pair-rule transcription factors was achieved using the (*nanos-GAL4-tubulin)* maternal expression system. The second chromosome linked

P{GAL4-nos.NGT}40 (*NGT40*) driver and the *P{UAS-runt.T}15, P{UAS-runt.T}232,* and *P{UAS-opa.VZ}14* transgenes have been described previously (Li and Gergen, 1999; Swantek and Gergen, 2004; Tracey et al., 2000). The *P{UAS-ftz}263* and *P{UAS-eve}12* transgenes were provided to us by Leslie Pick (Lohr and Pick, 2005) and John Reinitz (Zallen and Wieschaus, 2004), respectively. Embryos were collected from crosses between females homozygous for *NGT40* and for the different third chromosome-linked reporter genes and males homozygous for these different *UAS* transgenes.

Whole-mount in situ hybridization

Embryos were collected as described (Tsai and Gergen, 1994). Embryos from experiments with temperature sensitive alleles were collected for 2 hours at 25^{o}C then allowed to develop at the permissive temperature of 18^{o}C for four hours then shifted to a non-permissive temperature of 30^{o}C for 20 minutes immediately prior to fixation. *In situ* hybridization was carried out as described (Swantek and Gergen, 2004) with the following modifications: embryos were digested with proteinase K (50ug/ml in PBS + 0.1% Tween-20) using the digoxigenin-labeled riboprobe for *slp1* described in (Wheeler et al., 2002). The digoxigenin-labeled riboprobe to detect *lacZ* was synthesized as described in (Tsai and Gergen, 1994). The probe for *slp2* was synthesized with T3 polymerase using BamH1 linearized pB:*slp2*, a plasmid containing a 586 bp Sau3A1 – DdeI fragment that spans the C-terminal 157 amino acids of the Slp2 protein and also includes 115 nucleotides from the 3' untranslated region. This DNA segment excludes regions with the highest homology to *slp1.* A digoxigenin-labeled probe for *CG3407* was synthesized using SP6 polymerase from EcoRI-digested LD31554 plasmid template (Rubin et al., 2000).

Fluorescent *in situ* hybridization was carried out as described (Janssens et al., 2005) with the following modification: After fixation embryos were cleared in xylenes:ethanol (9:1) then post-fixed in PBT + 5% formaldehyde. Embryos were permeabilized for 10 minutes in 80% acetone in H_2O at -20^{0}C. The fluorescein labeled *lacZ* riboprobe was synthesized with fluorescein-12-UTP (Roche) in place of digoxigenin-conjugated UTP. After hybridization, *lacZ* mRNA was visualized by sequential incubation with Rabbit Anti-fluorescein (1ug/ml final) and Alexa Fluor 647 Donkey Anti-rabbit (1ug/ml) antibodies (Molecular Probes). Digoxigenin labeled probes were deected using Mouse Anti-Digoxigenin antibody (Roche, 1.25ug/ml final) followed by Alexa Fluor 555 Goat Anti-mouse (1ug/ml) and Alexa Fluor 555 Donkey Anti-goat (1ug/ml) antibodies (Molecular Probes). Blocking was done in 2x Western Blocking Reagent (Roche) diluted in PBT. All antibodies were preabsorbed at a 10x concentration in PBT with 1/10 volume of 0 to 12 hour AED embryos and then diluted to 1X. PicoGreen (Molecular Probes) was used to stain nuclei at a 1:30,000 dilution. Prior to mounting, embryos were washed in PBS:glycerol (1:1) for 20 minutes then mounted in 45µl mounting medium (2.5% Dabco (Sigma), 50mM Tris (pH 8.0) and 90% glycerol) and covered with a Corning 22x40-mm cover glass (No. 1 ½). Images were obtained on a Leica TCS SP2 Spectral Confocal Microscope system as described (Janssens et al., 2005) each image was scanned 8x resulting in an averaged image to reduce noise.

Quantification of expression patterns

The images to be used for quantification were acquired as above on a Leica TCS SP2 Spectral Confocal Microscope system. Quantification of the expression of different reporters is essentially as described (Janssens et al., 2006) using PicoGreen (Molecular

probes) to mark nuclei for the first step of image processing as well as to find the correct focal plane (Janssens et al., 2005). In order to compare one embryo to the next the same gain and offset was used for each image by adjusting the scan settings to just below saturation on stage seven embryos, which have stronger *slp1* expression then those used for quantification. To reduce image noise each embryo was scanned 16 times and the average image was used for quantification. This was done with 2 sections and each channel is scanned sequentially producing an average image for each channel at two focal planes. The first step in the image processing method is segmentation, which results in the average fluorescence for each nuclei. Segmentation begins by creating a nuclear mask to identify individual nuclei and then a text file is obtained where each individual nucleus is characterized by a unique identification number, the *x* and *y* cooridinates of its centroid, and the average fluorescence levels of three proteins (Janssens et al., 2005). The *x* axis corresponds to the A-P axis of the embryo and the *y* axis to the D-V axis. In the segmented data files, *x* and *y* coordinates are expressed as percent of the maximum size of the embryo in the *x* and *y* directions (Myasnikova et al., 2005). Background staining of RNA signal was subtracted after a smoothing step by wavelets (Myasnikova et al., 2001; Myasnikova et al., 2005). Smoothed data at this step is only used for estimating background, which is obtained by finding individual non-expressing nuclei. For each gene the non-expressing areas of the embryo are detected and used to fit a quadratic paraboloid to the background signal, which is finally uses to remove background. Embryos were classified temporally belonging to stage 6 based on the *slp1* mRNA pattern and differential interference contrast (DIC) membrane images. Registration of expression patterns is as described (Kozlov et al., 2009). To register *lacZ*

expression patterns, *slp1* served as a template using 23 Ground Control Points (GCPs), which excludes stripes zero and one of *slp1*. The fast redundant dyadic wavelet transform was used for feature extraction (Kozlov et al., 2009; Myasnikova et al., 2001). All data were then averaged by collecting intensities from individual embryos according to the Anterior-Posterior (AP) position and then graphed.

Primers

Operon sequence 5' to 3'

PESEPD primer ATAAGAATGCGGCCGCGACGTCGTGGTAAAACGAATT

PESEDD primer ATAAGAATGCGGCCGCAGGACAATGACTGAAAGGCA

C1left primer ATAAGAATGCGGCCGCTTGTTGGGACCATAAAATCCA

C1right primer ATAGTTTAGCGGCCGCCTGTGGGTGAGCATTCAGTC

PESEID1L primer AATATCCTGTCGACTGTTGACAGCGAGGTTCCTCA

PESEID1R primer GTCGACAGGATATTCAGCATGTA

PESEID2L primer TCGCAGCGAGTCCCGGTGTCTACTTGGAATTC

PESEID2R primer GGACTCGCTGCGAGTCCTAT

PESEID3L primer ATTGTCTGTGGCCAGTCAAACAGCGTGGTCCTGAT

PESEID3R primer TGGCCACAGACAATGTGCATGT

PESEID4L primer TGAAACCATAATCTCCCGGCAAAAAAACACAT

PESEID4R primer ATTATGGTTTCATTGCCAAAAC

XhoI/12275 primer GGCCTCGAGCTCTTCGTGTAGACTTCGT

KpnI/12396 primer GTAGGTACCGACTTGGGATCGCTTGAGA

404T primer TCCGTTTAAATGATTTAAGCAACAGATTTGAGCTTT

404B primer AAAGCTCAAATCTGTTGCTTAAATCATTTAAACGGA

504T primer GCCTTTATGTTTTTGCTTTGTCCCATTAGTTTG

504B primer CAAACTAATGGGACAAAGCAAAAACATAAAGGC

962T primer GAAGGTCAATGTATGGAAACAGTTTTGCGTTAGATC

962B primer GATCTAACGCAAAACTGTTTCCATACAT

1048T primer GCCATTTACTTTGGAAACATATCGTTCTACGGGC

1048B primer GCCCGTAGAACGATATGTTTCCAAAGTAAATGGC

1370T primer GCCACCTCCGGTGCAAAGCTGGTGCAAATCGCCGG

1370B primer CCGGCGATTTGCACCAGCTTTGCACCGGAGGTGGC

PESERuntT primer GGAAACGCCTAAGTGTTTTTTTGGCAATGAAACCAT

PESERuntB primer ATGGTTTCATTGCCAAAAAAACACTTAGGCGTTTCC

IFforward primer GATACCGTCGACCTCGATAGAGCCTTCAAAGGTTTGCTA3

IFreverse primer CACGAAGAGAGCTCGAATCCTTTAGAATTCGTCGCAGT3

IFforward-8.1kb GATACCGTCGACCTCGATTTTTCAGGACTCGCAAAGGGA3

Chapter III

Interplay between distinct early stripe elements mediates combinatorial regulation by pair-rule transcription factors during Drosophila segmentation

The bulk of the work described in this chapter comprises a manuscript that has been submitted for publication to Developmental Biology. The authors on the submitted manuscript are Lisa Prazak, Miki Fujioka and J. Peter Gergen. Dr. Fujioka (Thomas Jefferson University) is responsible for the initial identification of the extended distal element as well as determining the expression pattern of the *slp1-[5534]-lacZ* and *slp1-[7150]-lacZ]* reporter genes in figure 3.1.

Chapter III

Summary

The relatively simple combinatorial rules responsible for establishing the initial metameric expression of *sloppy-paired-1 (slp1)* in the Drosophila blastoderm embryo make this system an attractive model for investigating the mechanism of regulation by pair rule transcription factors. This investigation of *slp1* cis-regulatory architecture identifies two distinct elements, a distal early stripe element (DESE) and a proximal early stripe element (PESE) located from -8.7 kb to -6.5 kb and from -3.9 kb to -1.8 kb upstream of the *slp1* promoter, respectively, that participate in generating this early pattern. The distal element expresses both odd and even-numbered stripes but also drives inappropriate expression in the anterior half of the odd-numbered parasegments due to an inability to respond to repression by Even-skipped (Eve). In contrast, the proximal element expresses only even-numbered stripes and mediates repression by Eve as well as by the combination of Runt and Fushi-tarazu (Ftz). A minimal PESE element retains Eve-dependent repression, but is expressed throughout the even-numbered parasegments due to the loss of repression by Runt and Ftz. Importantly, a composite reporter gene containing both cis-elements emulates *slp1* regulation in a manner beyond what is expected from combining their individual patterns. These results indicate that integration of pair-rule regulatory information involves interactions between these two distinct *cis*-elements. I propose a model whereby the metameric expression of *slp1* is achieved by pair-rule dependent regulation of enhancer-promoter interactions.

Introduction

The Drosophila segmentation pathway provides a valuable platform for investigating *in vivo* mechanisms of transcriptional regulation. Extensive molecular and genetic studies indicate three classes of genes, the gap, pair-rule and segment-polarity genes act in a hierarchical fashion to establish the segmented body pattern with cellular resolution in the three hour blastoderm embryo (Akam, 1987; Howard, 1990; Ingham, 1988; Nusslein-Volhard and Wieschaus, 1980). The broad expression domains of the transcription factors encoded by the gap genes provide positional information that generates the periodic expression patterns of the pair-rule genes. One principle that has emerged from studies on this gap to pair-rule transition is a modular cis-regulatory architecture with distinct cis-regulatory modules (CRMs) that independently respond to different combinations of DNA-binding transcription factors. A classic example of this mode of regulation is *eve* stripe number two which is activated in response to Bicoid and Hunchback and which has its anterior and posterior borders defined by repression from Giant and Kruppel, respectively (Small et al., 1992). These combinatorial cues are integrated by a 500 basepair (bp) CRM that contains multiple binding sites for each of these four different transcription factors (Small et al., 1992). There are distinct CRMs for other *eve* stripes, each of which mediates regulation in response to different combinations of gap gene transcription factors (Fujioka et al., 1999; Goto et al., 1989; Harding et al., 1989; Sackerson et al., 1999; Small et al., 1996; Stanojevic et al., 1991). Stripe-specific CRMs are also important for the early expression of the primary pair-rule genes *hairy* and *runt* (Butler et al., 1992; Hader et al., 1998; Klingler et al., 1996; La Rosee et al., 1997;

La Rosee-Borggreve et al., 1999; Langeland and Carroll, 1993), indicating that this mode of cis-regulatory architecture is central to the gap to pair-rule transition.

Pair-rule genes encode DNA-binding transcription factors that are critical for establishing the initial periodic expression of several different segment polarity genes. Although significant effort has gone into investigating segment-polarity gene regulaiton, the work has focused mostly on the intercellular signaling pathways responsible for maintaining expression at later developmental stages. In order to understand how pair-rule genes regulate segment-polarity gene expression it is necessary to focus on how the expression of this class is first established in the early embryo. One principle that has emerged from studies to date is a distinction in the regulation of the odd and even-numbered stripes, a direct consequence of the different periodicities of pair-rule and segment-polarity gene expression. Consistent with this, parasegment-specific CRMs have been identified for a couple of the segment-polarity genes. The *engrailed* (*en*) first intron mediates regulation that results in the timely expression of a *lacZ* reporter for the even-numbered stripes but does not express odd stripes until germ band extension (DiNardo et al., 1988; Kassis, 1990). Similarly, DNA sequences from 4.5 kilobasepairs (kb) upstream to the transcription start site of *wingless* (*wg*) drive early expression of the odd-numbered stripes but do not express even-numbered stripes until germ band extension (Lessing and Nusse, 1998). One hurdle in more fully dissecting the cis-regulatory logic of these two segment-polarity genes is the size of their prospective cis-regulatory regions. The function of *en* requires nearly 70 kb of flanking DNA (Kassis et al., 1985). Similar to *en*, the *wg* transcription unit is separated from the nearest upstream and downstream genes by more than 30 kb (Tweedie et al., 2009).

The *gooseberry* (*gsb*) locus provides an example of a segment-polarity gene with a more compact cis-regulatory region (Baumgartner et al., 1987; Bopp et al., 1986). In this case a reporter gene containing *gsb* upstream DNA that extends to the promoter of the divergently transcribed *gsb-neuro* transcription unit emulates the differential activation of the odd- and even-numbered stripes, although somewhat delayed with respect to the activation of the endogenous gene (Li et al., 1993). Further dissection of this 10 kb region identified a 514 bp CRM that drives expression in both odd and even parasegments in direct response to Paired, a pair-rule transcription factor that comes to be expressed in an every segment pattern during germ band extension (Bouchard et al., 2000). It remains an open question whether this well-defined CRM represents the full extent of *gsb* regulation as a direct response to pair-rule transcription factors.

The *slp1* gene provides several advantages for investigating regulation by pair-rule transcription factors. One prospective advantage is a relatively compact cis-regulatory region. The *slp* locus consists of two structurally related genes transcribed in the same direction, with *slp1* located 10 kb upstream of *slp2* (Grossniklaus et al., 1992). Although the genes have similar expression patterns, *slp1* is expressed more strongly in the early embryo and makes the major quantitative contribution in the early segmentation pathway (Grossniklaus et al., 1992). Indeed, characterization of rearrangements within the *slp* locus strongly suggests the cis-regulatory sequences that drive expression of both genes are located upstream of the *slp1* transcription unit (Grossniklaus et al., 1992). The 5' end of the divergently transcribed CG3407 gene, located 12 kb upstream from the *slp1* transcription start site defines a presumptive upstream boundary of this cis-regulatory region

A second key advantage offered by *slp1* is the relatively simple set of combinatorial rules responsible for its initial metameric expression (Swantek and Gergen, 2004). This pattern consists of 14 stripes, two cells in width, in the posterior half of each parasegment. In odd parasegments, the Eve homeodomain protein is important for repression in the two most anterior cells, whereas expression of the odd-numbered *slp1* stripes in adjoining posterior cells of these parasegments is driven by the combination of Odd-paired (Opa) and Runt. The two next-most posterior cells, comprising the anterior half of the even-numbered parasegments also express Runt and Opa, but the presence of the Fushi-tarazu (Ftz) homeodomain protein in these cells converts Runt from an activator to a repressor of *slp1*. One aspect of the early regulation of *slp1* that is currently not accounted for is the factor responsible for activating expression of the even-numbered stripes in the posterior half of the even-numbered parasegments, a factor or combination of factors referred to as Factor X.

This work investigates the cis-regulatory structure of *slp1* with a focus on understanding regulation by pair-rule transcription factors. The experiments discussed here identify two upstream CRMs, separated from each other by more than 2 kb that generate different subsets of the *slp1* pattern. The distal early stripe element (DESE) is capable of mediating both Runt-dependent activation and repression, but is insensitive to Eve-dependent repression. In contrast, the proximal early stripe element (PESE), which drives expression only in even parasegments, mediates repression by Eve, as well as by the combination of Runt and Ftz but is not activated by Runt and Opa. Further dissection of PESE identifies a minimal element for this Factor X-dependent activation, and

demonstrates a distinction between the cis-regulatory requirements for Eve- and Ftz-dependent repression. Importantly, a composite reporter containing both early stripe elements recapitulates pair-rule gene-dependent regulation in a manner beyond what is expected from the additive inputs of these two CRMs. This work reveals that integration of pair-rule regulatory cues involves functional interactions between distinct cis-regulatory elements, a phenomenon with important implications for understanding the actions of these transcription factors in segmentation and other developmental contexts.

Results

I utilized reporter gene constructs to identify cis-regulatory elements responsible for early *slp1* transcription. An initial panel of over-lapping constructs, each containing approximately 2 kb segments of *slp1* upstream DNA identified two distinct regions that drive early striped expression in the blastoderm embryo (Fig. 3.1). The proximal early stripe element (PESE), initially defined as extending from -3.9kb to -1.8kb upstream of *slp1*, expresses only even-numbered stripes at developmental stage 5 (Fig. 3.1), though the odd-numbered stripes do appear during germ band extension (Fig. 3.S3). The distal early stripe element (DESE), from -8.7kb to -6.5 kb, expresses both odd- and even-numbered stripes early but lacks expression of Stripe 0 (Fig. 3.2A). The *DESE-lacZ* reporter also shows early activation of the odd-numbered stripes with ectopic expression in inter-stripe regions anterior to the odd-numbered stripes. In the work that follows I first investigate the properties of each individual element and then examine the activity of composite reporters containing both regions.

DESE mediates Runt-dependent activation and repression.

Double fluorescent *in situ* hybridization comparing expression of a *DESE-lacZ* reporter with that of *slp1* confirms the loss of reporter gene repression in the anterior half of the odd-numbered parasegments (Fig. 3.2A, B). Repression of *slp1*in anterior odd-numbered parasegments is due to Eve (Swantek and Gergen, 2004), suggesting the reporter is unable to be repressed by Eve (Fig. 3.2C). I tested the ability of DESE to respond to Eve using the maternally expressed *NGT* GAL4 driver to express Eve throughout the embryo. *NGT*-driven Eve specifically represses the even-numbered *slp1*

stripes, but does not affect *DESE-lacZ* expression in this same embryo (Fig. 3.2D). The precocious activation of *DESE-lacZ* odd-stripes may also be due to the loss of Eve-dependent repression as activation of these stripes normally correlates with the loss of Eve in the posterior cells of odd parasegments.

Although the distal element does not respond to Eve, it mimics *slp1* expression in even-numbered parasegments of wild type embryos. Repression of *slp1* in the anterior half of the even parasegments is due to the combination of Runt and Ftz (Swantek and Gergen, 2004). In accord with this, *slp1* is expressed in broad six-cell wide stripes in *ftz* mutants (Fig. 3.3A). *DESE-lacZ* reporters show similar de-repression, which coupled with the ectopic expression in odd parasegments results in almost uniform expression throughout the pre-segmental region (Fig. 3.3A). The ability of DESE to mediate repression by Ftz is confirmed by elimination of the odd-numbered *DESE-lacZ* stripes in response to *NGT*-driven Ftz (Fig. 3.3B). As observed for *slp1*, the even-numbered stripes are not repressed in response to ectopic Ftz, resulting in *lacZ* expression that is comprised of these stripes in combination with the ectopic expression in the adjoining cells in the anterior portions of the odd parasegments. These results indicate a clear difference in the ability of DESE to mediate responses to Eve and Ftz.

The observation that *DESE-lacZ* is expressed in posterior odd parasegments and is repressed in anterior even parasegments of wild type embryos strongly suggests this element mediates both activation and repression by Runt. Elimination of *runt* results in an extremely abnormal *slp1* pattern due to changes in the expression of other genes such as *eve* and *ftz*. Therefore the role of Runt in *DESE-lacZ* regulation was investigated using a temperature sensitive allele that allows for elimination of *runt* activity in stages

subsequent to the initial establishment and refinement of the pair-rule expression patterns. As demonstrated previously (Swantek and Gergen, 2004), transient elimination of *runt* results in loss of *slp1* expression in odd-numbered parasegments and expanded expression in even-numbered parasegments (Fig. 3.3C). Although the *DESE-lacZ* reporter is de-repressed in even-numbered parasegments, expression of the reporter in the posterior regions of the odd parasegments is not completely lost in these embryos (Fig. 3.3C). Differences in mRNA turnover rates could in principle account for the perdurance of *lacZ* mRNA in these cells, although results presented below indicate that this expression instead reflects a difference in the response of DESE to pair-rule regulatory cues.

As a second approach for investigating the Runt-dependent regulation of DESE, I examined the response to ectopic expression. The combination of Runt and Opa is sufficient for *slp1* activation in all blastoderm cells that do not express Ftz, including regions anterior to the segmented region of the embryo (Swantek and Gergen, 2004). *DESE-lacZ* expression emulates that of *slp1* in embryos that have ectopic, *NGT*-driven expression of Runt and Opa, including clear anterior activation (Fig. 3.3D). The broad domains of *slp1* repression within the segmented region of these embryos form in Ftz-expressing cells (Swantek and Gergen, 2004), a response that is also observed for this *DESE-lacZ* reporter. Based on these results I conclude that DESE is capable of mediating both activation and repression in response to Runt.

PESE mediates repression by Eve and Ftz.

The *PESE-lacZ* reporter is initially expressed in only seven stripes. *In situ* hybridization confirms that these correspond to *slp1*-expressing cells in even-numbered

parasegments, including stripe zero (Fig. 3.4A). The factor responsible for the activation of the even-numbered stripes is not known and has been referred to as Factor X (Swantek and Gergen, 2004). These stripes are bordered anteriorly by cells expressing Runt and Ftz, and posteriorly by cells expressing Eve. I examined the response of *PESE-lacZ* to changes in pair-rule gene activity to investigate the roles of these factors in establishing these stripe borders. Transient elimination of *eve* results in six-cell wide *slp1* stripes due to de-repression in the anterior half of the odd-numbered parasegments (Fig. 3.4B). The *PESE-lacZ* reporter shows evidence of expanded expression in these same embryos, suggesting Eve plays a role in defining the posterior border of these stripes. I examined the response to ectopic Eve to further investigate whether PESE mediates Eve-dependent repression. *NGT*-driven Eve effectively represses *PESE-lacZ*, with the nearly complete elimination of *lacZ* expression in embryos that show incomplete repression of the even-numbered *slp1* stripes (Fig. 3.4C). This result indicates that PESE is sensitive, perhaps even super-sensitive to Eve-dependent repression. Super-sensitivity to Eve may explain the relatively modest expansion observed in the *eve[1]* embryos, especially if the residual effects of repression by Eve perdure for some time after the temperature shift. Similar experiments indicate PESE mediates repression by Runt and Ftz. There is evidence of some expansion in *ftz* mutants (Fig. 3.4D), and transient elimination of *runt* leads to expanded *PESE-lacZ* expression that resembles the altered *slp1* expression observed in the same embryo (Fig. 3.4E). Finally, *NGT*-driven co-expression of Runt and Ftz effectively represses *PESE-lacZ* (Fig. 3.4F), and as seen with Eve, the *PESE-lacZ* reporter is more sensitive to repression by Runt and Ftz than is *slp1* in the same embryo.

Differential sensitivities of a minimal Factor X-responsive element to Eve and Ftz.

The full-length PESE described above contains sequences extending from 3.9 to 1.8 kb upstream of the *slp1* promoter. This interval provides a starting point for identifying sequences necessary for Factor X-dependent activation, which in turn provides a route towards the molecular identification of this factor. In order to identify a minimal region that is able to express the even-numbered stripes I examined the expression of three truncated *PESE-lacZ* reporters, each deleted for approximately 700 bp from the left, central and right regions, respectively, of the full element (Fig. 3.5A, B). Deletion of the left and right regions had no effect on early stripe expression but the right region is necessary for expression of the odd-numbered stripes (Supplemental Fig. 3.S3). Deletion of the central region on the other hand, abolished expression of the early even-numbered stripes (Supplemental Fig. 3.S1). Importantly, the *slp1[3125]-lacZ*^{att} reporter expresses even-numbered stripes (Fig. 3.5C, D), indicating this 621 bp region is necessary and sufficient for even stripe activation.

A series of four deletions spanning the PESE central region within the context of a reporter containing the flanking left and right regions was generated to further define the sequences needed for Factor X-dependent activation. Three of the four internal deletions retained expression, the exception being *slp1[PESE:ΔC1]-lacZ,* which is not expressed in gastrula stage embryos (Fig. 3.5E, F). To determine if this region is sufficient, I generated a minimal reporter containing sequences from 3.1 to 2.9 kb upstream of the *slp1* promoter. This minimal *slp1[PESE:C1+]-lacZ* reporter not only drives expression of even-numbered stripes, but also shows expanded expression anterior to these stripes, through to the posterior edge of the neighboring odd-numbered

parasegment (Fig. 3.5G, H, I). One explanation for this expansion is that the minimal *slp1[PESE:C1+]* reporter is insensitive to repression by Runt and Ftz. Consistent with this explanation, *slp1[PESE:C1+]-lacZ* expression is retained in embryos in which *slp1* is nearly fully repressed by *NGT*-driven co-expression of these two factors (Fig. 3.5J). In contrast, this minimal reporter is effectively repressed by *NGT*-driven Eve (Fig. 3.5K). Moreover, transient elimination of *eve* results in *slp1[PESE:C1+]-lacZ* expression throughout the segmented region of the embryo (Fig. 3.5L). This observation confirms the sensitivity of this element to Eve-dependent repression, and also indicates that the factor responsible for activating this element should be expressed throughout the segmented region of the embryo at this stage of development. The finding that DNA sequences required for mediating repression by Runt and Ftz lie outside of this minimal Eve-responsive region provides another indication of the distinctive cis-regulatory requirements for repression by Eve and Ftz.

Interactions between DESE and PESE allow faithful integration of pair-rule cues.

The two early stripe elements described above each drive a subset of the early *slp1* pattern. The simple addition of these two partial patterns will not produce a normal pattern, primarily due to DESE-driven expression in the anterior regions of odd numbered parasegments. Expression in these cells is normally blocked by Eve. As shown above, PESE effectively mediates Eve-dependent repression. We generated a composite reporter containing both DESE and PESE in order to determine whether interactions between these two elements are capable of restoring repression in these cells. When these two elements are combined *lacZ* faithfully emulates *slp1* expression throughout the segmented region of the embryo, including restoration of repression in anterior odd-

numbered parasegments and a timely initial activation of the odd stripes (Fig. 3.6B). There are differences in expression in the anterior head region, an aspect of *slp1* expression that is not under control of the pair-rule transcription factors. Similar patterns are obtained for both orientations of the PESE in the composite [DESE+PESE] reporters in wild-type embryos (data not shown). Expression of *slp1[8765:3918]-lacZ* ^{att}not only emulates *slp1* in wild-type embryos but also faithfully recapitulates the response to manipulations in pair-rule activity. This composite reporter is expressed in the anterior half of odd parasegments in *eve* mutants (Fig. 3.6C), is repressed as effectively as *slp1* in response to *NGT*-driven Eve (Fig. 3.6D), and also emulates *slp1* de-repression in anterior even-numbered parasegments in *ftz* mutants (Fig. 3.6E). One aspect of reporter gene expression that did not match endogenous *slp1* for either the *DESE-* or *PESE-lacZ* reporters was the pattern in *runt* mutants. The *slp1[8765:3918]-lacZ*^{att} reporter also emulates *slp1* expression in these embryos (Fig. 3.6F). This result further indicates that the *DESE-lacZ* expression detected in the posterior half of the odd parasegments in *runt* mutants (Fig. 3.3E) is not due to difference in the turnover rates of the *lacZ* and *slp1* mRNAs, but instead reflects a functional difference between the DESE and [DESE+PESE] reporters. Based on these results I conclude that functional interactions involving distinct cis-regulatory elements that normally are separated by more than 2 kb are critical for generating the initial metameric expression of *slp1* in response to pair-rule transcription factors.

Discussion

This work identifies two distinct CRMs from the *slp1* gene that drive early expression in response to pair-rule gene regulation. The observation that a composite reporter gene containing both elements faithfully emulates the initial metameric expression of *slp1* in wild-type embryos as well as the response to manipulations in pair-rule activity strongly suggests these two CRMs together account for most of the early regulation of *slp1* in response to pair-rule transcription factors. The ability of the composite reporter to mimic the endogenous gene cannot be explained solely by the independent regulatory capabilities of the proximal and distal elements as a simple addition of the two patterns will include inappropriate DESE-driven expression in anterior even-numbered parasegments. This non-additive interaction potentially conflicts with the generally accepted paradigm for the modular and independent action of distinct CRMs, a point that will be discussed further below.

Although the early stripe elements need to be combined in order to fully recapitulate pair-rule regulation, studies on the independent elements provide new insights on the pair-rule to segment polarity gene transition. The homeodomain proteins Eve and Ftz both participate in *slp1* repression. Several lines of evidence indicate differences in the cis-regulatory requirements for repression by these two structurally related transcription factors. DESE is insensitive to repression by Eve, but is capable of mediating repression by Ftz. The exact opposite specificity is demonstrated by the PESE:C1+ element, which is repressed by Eve but not by Ftz. The DNA-binding specificity of Eve and Ftz is similar both *in vitro* and *in vivo* and their specificity of

action is thought to involve co-factor interactions that dictate the manner in which they regulate different targets (Biggin and McGinnis, 1997; Walter et al., 1994). An established co-factor for Ftz is the orphan nuclear receptor protein Ftz-F1 (Florence et al., 1997; Guichet et al., 1997; Yu et al., 1997). Indeed, elimination of maternally provided Ftz-F1 results in alterations in *slp1* expression that are identical to those seen in *ftz* mutants (data not shown). The Ftz-dependent repression of *slp1* also requires Runt, making this a second prospective co-factor for this activity of Ftz. Although this work does not investigate the histone acetylation state of the promoter region which correlates with an active or repressed state of transcription, previous work from the lab has shown that in wild type embryos versus those ectopically expressing Runt in combination with Ftz, there is no significant difference in the H3 acetylation pattern of *slp1* (Wang et al., 2007). This suggests that there is an alternative explanation to how these pair-rule transcription factors are able to function to regulate *slp1* transcription. Further studies on PESE should provide valuable information on the mechanisms that underlie repression by the Eve and Ftz proteins.

Our studies on the independent DESE and PESE reporters also provide information on the properties of the unidentified factor(s) that are responsible for *slp1* activation in posterior even-numbered parasegments. In the case of PESE, expression of the minimal *slp1[PESE:C1+]-lacZ* reporter throughout the entire pre-segmental region of *eve* mutant embryos provides an indication on the expression of the factor(s) responsible for this activation. The minimal PESE:C1+ element contains multiple putative binding sites for the transcription factors E2f, GATA-1, GATA-2 and Cdx, all of which have Drosophila homologues (*E2f*, *serpent*, *pannier* and *caudal*). Of these genes, only *E2f* is expressed

throughout the pre-segmental region at this stage of embryogenesis (Duronio et al., 1995; Seum et al., 1996). There are three prospective E2f binding sites within PESE:C1+, with three additional sites in the 600 bp PESE central region and a total of eight sites within the full length proximal element. E2F is a heterodimer of two proteins, E2F and DP1 (Girling et al., 1993) and in mammalian cells is crucial for entry into S phase. Drosophila mutant for *E2f* die as late larvae or early pupae with defects in cell cycle progression (Royzman et al., 1997). The lack of overt segmentation defects does not rule out a role in this developmental process as E2f is also maternally provided (Duronio and O'Farrell, 1994). It will be interesting to determine whether PESE is capable of mediating E2f-dependent activation, and if so, to then further investigate how this activity is modulated by pair-rule transcription factors.

DESE drives expression in the posterior regions of both odd- and even-numbered parasegments, but fails to generate stripe 0. This difference between DESE and PESE suggests there are differences in the factors responsible for activating these two elements in even-numbered parasegments. It is furthermore notable that the DESE contains no prospective binding sites for E2f, a dramatic contrast with the eight sites identified in the similarly sized PESE. The activation of *slp1* in odd parasegments is normally driven by the combination of Runt and Opa. Runt is normally expressed in the posterior half of only the odd parasegments and not in the posterior half of even-numbered parasegments. However, the observation that transient elimination of Runt does not abrogate *DESE*-driven expression in odd parasegments (Fig. 3.3E) suggests Opa may be capable of activation in the absence of Runt. This same proposal could account for the ectopic *DESE*-driven expression in the anterior half of the odd parasegments as Opa is expressed

uniformly in all cells within the pre-segmental region that are posterior to the cephalic furrow (Supplemental Fig. 3.S2) (Benedyk et al., 1994; Cimbora and Sakonju, 1995). Although the only Opa-expressing cells that do not activate the *DESE-lacZ* reporters are those that express the combination of Runt and Ftz, there are differences in the level of expression in different cells. The increased expression in posterior versus anterior odd-numbered parasegments may reflect a contribution from Runt in potentiating DESE-driven expression.

The central issue raised by these results is understanding the interactions involving two distinct CRMs that account for their ability to faithfully recapitulate the regulation of *slp1* in response to the pair-rule transcription factors. A major discrepancy between the expression of the composite [DESE+PESE] reporter and the pattern expected from the independent action of the separate CRMs is repression of the composite reporter in anterior odd parasegments. One potential explanation is that repression in these cells involves interactions that allow the Eve-sensitivity of PESE to be transmitted to DESE (Fig. 3.7A). Observations that the composite reporter lacks the 2.7 kb of intervening DNA that normally separates these two elements, and that the orientation of PESE within the composite reporter does not effect expression present challenging physical constraints for this explanation. A version of this model that would not require simultaneous communication between both upstream elements and the promoter is that Eve-interacting PESE sequesters DESE, thereby preventing DESE-dependent activation at the promoter (Fig. 3.7B). However, in this model ectopic Eve would be expected to repress both even and odd stripes, instead of just the even stripes.

There is an alternative model not involving interactions between the two CRMs, but

instead incorporating spatial differences in the expression of the transcription factors that interact with these elements to regulate CRM-promoter interactions. In this model I propose that Runt plays a role in switching the promoter from a default interaction with PESE to interacting with the further upstream DESE (Fig. 3.7C). This proposed role of Runt is bypassed in the *DESE-lacZ* reporter due to the lack of competition from PESE, thus accounting for the expression of this reporter in all cells within the segmented region of the embryo except for those that express both Runt and Ftz. The lack of expression of *slp1* and the composite [DESE+PESE] reporters in anterior odd parasegments is accounted for if interaction of DESE with the promoter requires Runt as these cells normally do not express Runt and thus will be restricted to PESE-dependent activation (Fig. 3.7D). The observation that *slp1* is expressed identically as *DESE-lacZ* in response to ectopic Runt (Fig. 3.3G) further supports the proposal that DESE is responsible for all aspects of *slp1* expression in Runt-expressing cells. Examples of enhancer-promoter interactions involving competition of different promoters for a single enhancer (Akbari et al., 2008; Zhou et al., 1996) as well as competition of different enhancers for a single promoter (Lin et al., 2007) have been described previously. This prior work focused on the roles of promoter tethering sequences and insulators in modulating enhancer-promoter interactions, and not on the role of enhancer-interacting transcription factors. The proposed regulation of *slp1* enhancer-promoter interactions by Runt during Drosophila segmentation opens the door for investigating transcription factor-dependent regulation of enhancer promoter interactions and should provide a valuable model for understanding the role of the Runt transcription factor family in other developmental contexts.

Figures

Figure 3.1

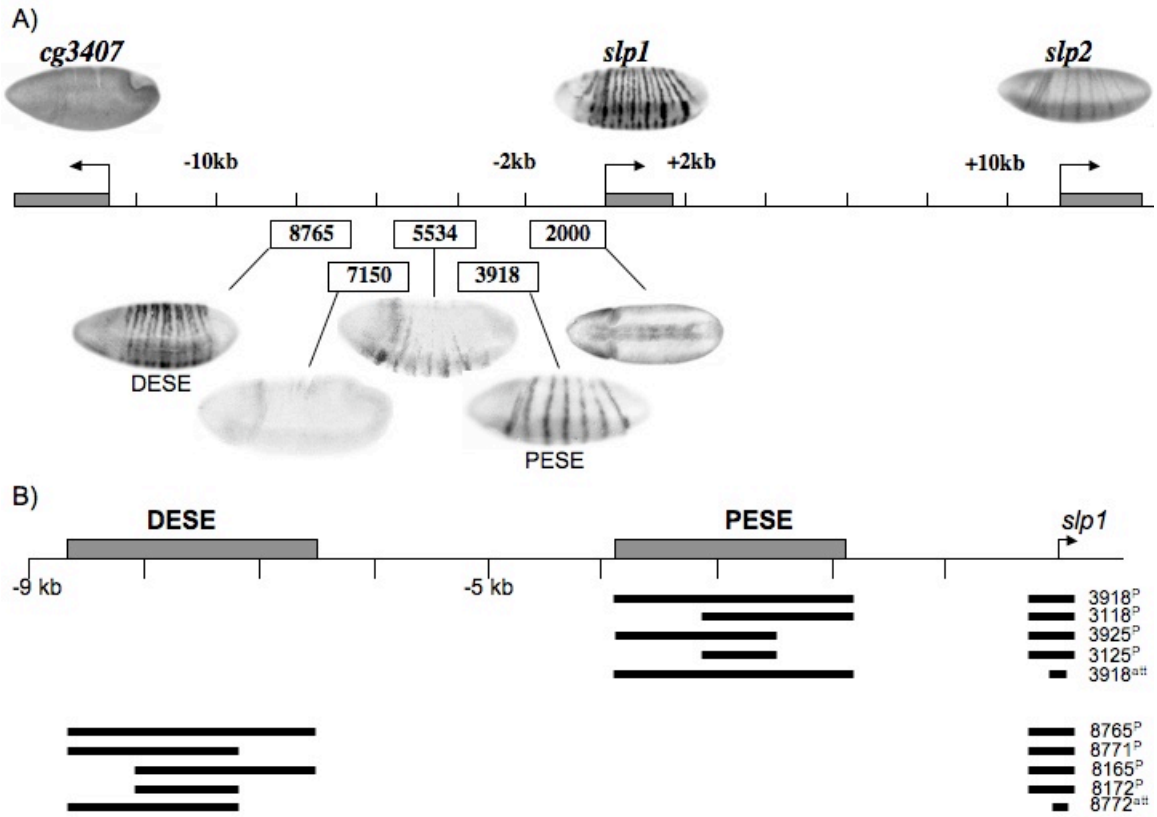


Figure 3.1. The *slp1* locus contains two distinct early stripe elements

Expression patterns of *slp1, slp2, cg3407* and a panel of *slp1-lacZ* reporters as visualized by *in situ* hybridization. The horizontal line depicts the chromosome region containing the *slp1* locus, extending to the flanking *cg3407* and *slp2* transcription units. The sequence coordinates are relative to *slp1* +1. The embryos above each of these three genes show their respective mRNA expression patterns in gastrula stage embryos. The upstream DNA segments contained in different reporter gene constructs are schematically represented by the over-lapping rectangles below the map, with the numbers in the boxes providing the coordinates for the respective reporter gene (e.g. 87665 extends from 8.7 to 6.5 kb upstream of *slp1*). The *lacZ* mRNA expression patterns generated by the different reporters are shown in embryos below the map. The *slp1[5534]-lacZ* reporter shows signs of partial stripes at this stage, but delayed relative to the early striped patterns of the *slp1[8765]-lacZ* and *slp1[3918]-lacZ* reporters. Embryos in this and the following figures are oriented anterior to the left, typically dorsal side up. A ventral view is provided of the *slp1[2000]-lacZ* to reveal the low level of mesodermal expression from this reporter. (B) Magnified representation of the *slp1* locus from -9 kb to +500 bp relative to the *slp1* transcription start site (+1). The location of each of the full-length early stripe elements is represented by grey rectangles above this map. The black rectangles below the map illustrate an initial panel of deletion constructs, with the numbers to the right indicating the cis-regulatory coordinates for each construct and the superscript P or att denoting whether the transgenic line was generated using standard P-element transformation (P) or ΦC31-mediated site-specific integration (att). The P-element constructs diagrammed in (B) contain a basal promoter extending from 261 bp upstream to 121 bp downstream of the *slp1* promoter. Similar striped expression patterns are produced for both PESE- and DESE-containing P-element based reporters with basal promoters extending further upstream (-941bp and -1.8 kb, data not shown). Reporter lines generated using site-specific integration contain *slp1* basal promoter sequences extending from -72 to +57 bp as indicated. All of the different PESE deletions, as well as the *slp1[8771]lacZ*^{att} and *slp1[8171]lacZ*^{att} reporters show expression comparable to that obtained with the larger basal promoters in the P-element transgenes. In contrast, inclusion of the DESE interval from -7.2 to -6.5 kb antagonizes expression from the

truncated basal promoter in the *slp1[8765]lacZ*^{att} reporter and can be rescued by incorporating the basal promoter used in the P-element constructs from – 261 bp to +121 bp (see supplemental figure 3.S4). The *slp1[8771]lacZ*^{att} and *slp1[3918]lacZ*^{att} reporters diagrammed in (B) represent the starting point for the use of site-specific transgenesis to investigate early *slp1* transcriptional regulation.

Figure 3.2

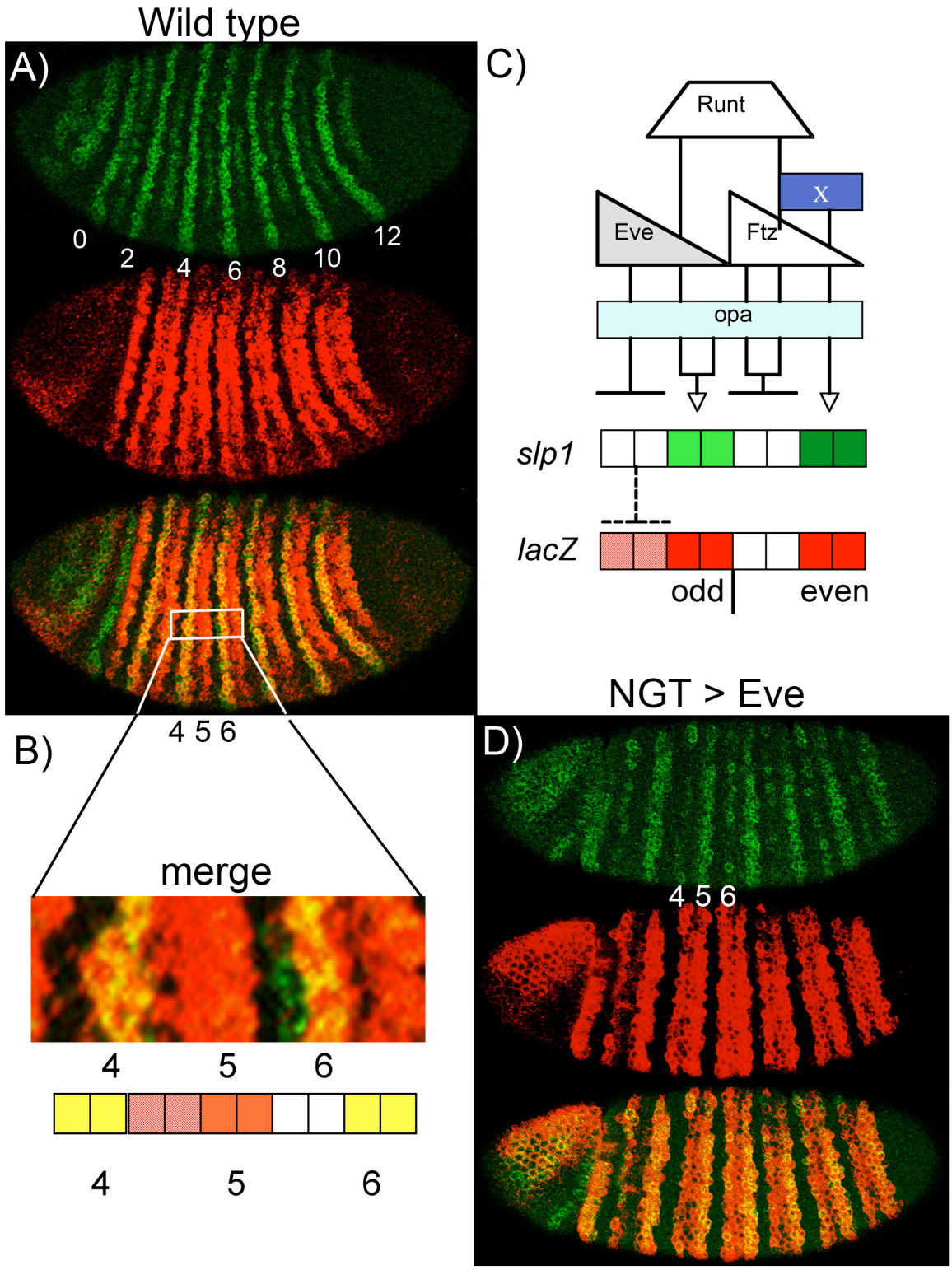

Wild type
A)
0
2
4
6
8
10
12
4 5 6
B)
merge
4
5
6
4
5
6
C)
Runt
X
Eve
Ftz
opa
slp1
lacZ
odd
even
NGT > Eve
D)
4 5 6

Figure 3.2. DESE is unable to mediate repression by Eve.

(A) Wild type mRNA expression of *slp1* (green), *lacZ* (red) and the merged image in a gastrula stage embryo homozygous for the *slp1[8765]-lacZ*^{M} reporter visualized using fluorescent *in situ* hybridization with the even-numbered *slp1* stripes labeled (B) Blow-up of parasegments 2, 3 and 4 showing ectopic *lacZ* expression in the anterior odd parasegments. The yellow-labeled cells in the schematic interpretation represent even numbered *slp1* stripes that co-express *lacZ*. Cells in odd-numbered stripes are labeled orange to reflect the relatively higher level of *lacZ* expression, whereas cells in anterior odd parasegments with ectopic reporter gene expression are stippled red. (C) Schematic representation of *DESE-lacZ* and *slp1* regulation by pair-rule transcription factors. The separate expression of *DESE-lacZ* and *slp1* is shown in two columns of cells spanning one double parasegment repeat across the bottom, with expressing cells shaded in red and green, respectively. The stippled red pattern identifies cells with ectopic *lacZ* expression. The relative domains of activity of different regulatory factors are provided above the cells, with arrows indicating activation of odd stripes by Runt and Opa and even stripes by Factor X. Repression in anterior even parasegments by Runt and Ftz is indicated by the horizontal bar, with the dashed lines indicating an insensitivity of *DESE-lacZ to* repression by Eve. The expression domains of Eve and Ftz are drawn as triangles with peaks at the anterior edges to reflect their loss of expression in posterior cells of each parasegment as the stripes narrow during cellularization. Runt expression is depicted as a trapezoid with peak expression levels in the two central cells. The expression pattern of Factor X is not known, with the rectangle indicating the requirement for Factor X in posterior even-numbered parasegments. (D) Expression of *slp1* (green) and the *slp1[8765]-lacZ*^{M} reporter (red) in a gastrula stage embryo in which the even-numbered *slp1* stripes are nearly fully repressed by *NGT*-driven Eve. The insensitivity of the *DESE-lacZ* reporter to this repression is apparent in the merged image in which cells in posterior even parasegments are now orange to red, instead of yellow.

Figure 3.3

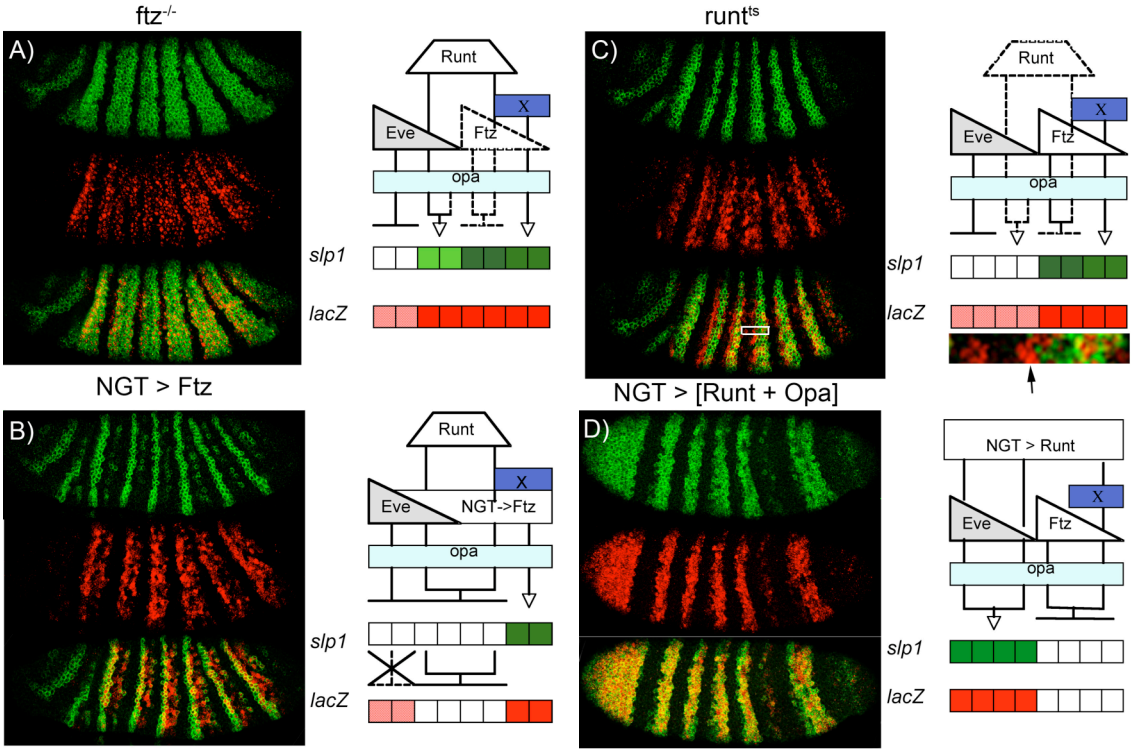

ftz-/-
A)
Runt
X
Eve
Ftz
opa
slp1
lacZ
runtts
C)
Runt
X
Eve
Ftz
opa
slp1
lacZ
NGT > Ftz
B)
Runt
X
Eve
NGT->Ftz
opa
slp1
lacZ
NGT > [Runt + Opa]
D)
NGT > Runt
X
Eve
Ftz
opa
slp1
lacZ

Fig. 3.3. DESE responds to regulation by Ftz, Runt and Opa. (A) Expression of *slp1* (green), *lacZ* (red) and the merged image in an embryo homozygous for the *ftz [11]* mutation and the *slp1[8771]-lacZ*^{att} reporter. This *DESE-lacZ* reporter, truncated by 600 bp relative to the extended DESE-containing reporter used in Figure 2, gives the same expression pattern in wild-type embryos. The reporter is expressed throughout the six cell-wide *slp1* stripes generated in *ftz* mutants. Cells that express *lacZ* but not *slp1* reflect the ectopic expression of the reporter gene in anterior odd parasegments. The accompanying Schematic representation of the *slp1* and *lacZ* indicates the response to the loss of repression by *ftz*, with ectopic *lacZ* expression in anterior odd parasegments indicated by stippled red cells. (B) Response of *slp1* and *slp1[8771]-lacZ*^{att} to *NGT*-driven Ftz. Cells expressing both genes (yellow) correspond to the even-numbered *slp1* stripes. The reporter retains ectopic expression in adjacent anterior odd parasegments but emulates the nearly complete repression of *slp1* odd stripes. (C) Expression of *slp1*, *lacZ* and the merged image of an embryo heterozygous for the *slp1[8771]-lacZ*^{att} reporter and hemizygous for the temperature sensitive *runt[29]* mutation. Transient elimination of *runt* results in *slp1* expression throughout even parasegments, with a concomitant loss of odd stripes. The reporter is expressed in *slp1*-expressing cells, but also in cells immediately anterior to these stripes where *slp1* is not expressed due to the loss of *runt*. (D) Response of *slp1[8771]-lacZ*^{att} to the *NGT*-driven co-expression of Runt and Opa. The endogenous gene and *lacZ* reporter show strong head activation and are expressed in the same pattern in the segmented region of the embryo as can be seen in the merged image. The schematic shows that this expression is in cells that do not express Ftz. The observation that the extended DESE reporter is insensitive to Eve (Fig. 2D), whereas this truncated DESE mediates regulation by Ftz, Runt and Opa underscores the differential sensitivity of DESE to these pair-rule transcription factors.

Figure 3.4

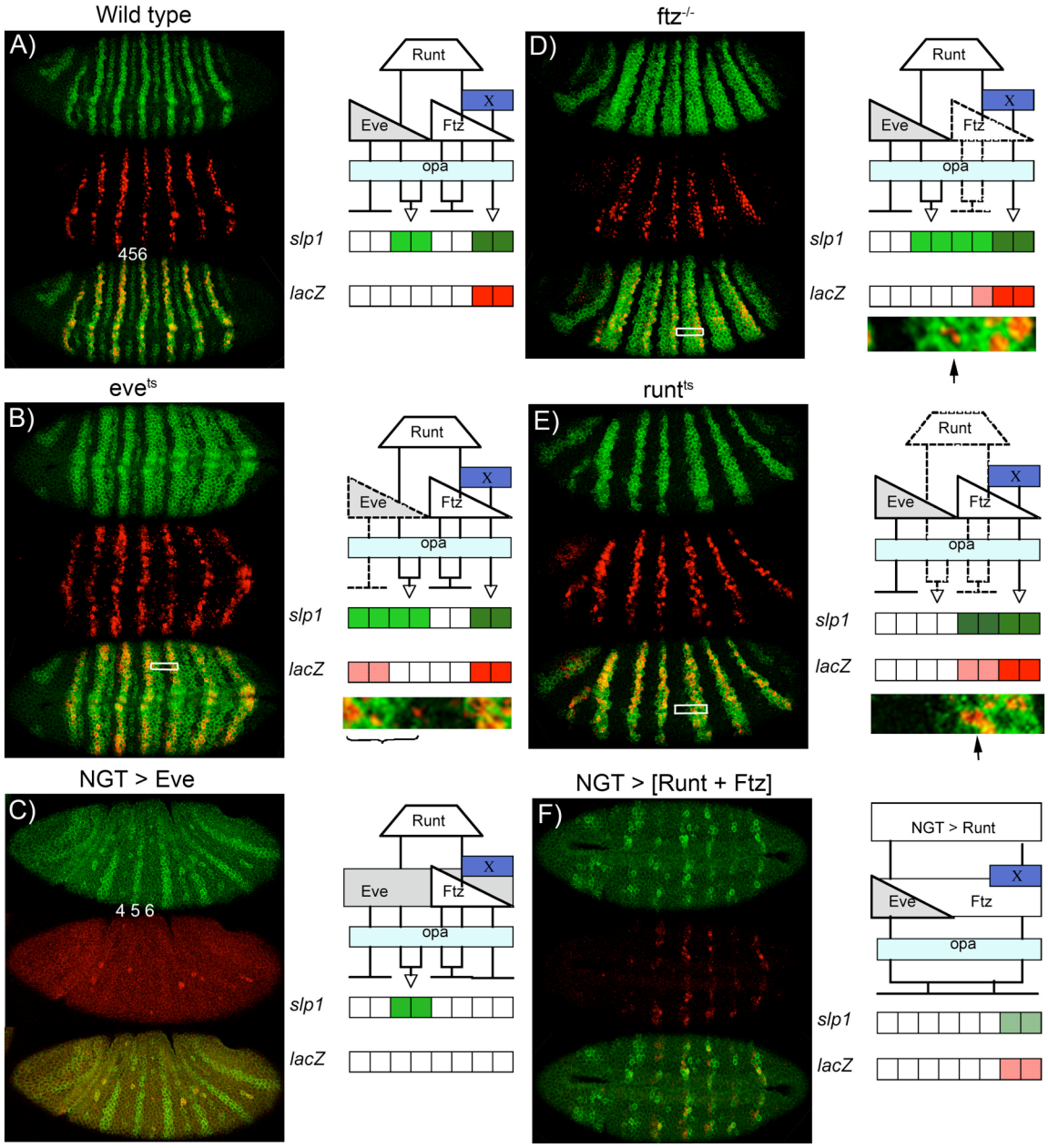

Wild type
A)
456
Runt
X
Eve
Ftz
opa
slp1
lacZ
ftz-/-
D)
eve^ts
B)
runt^ts
E)
NGT > Eve
C)
4 5 6
NGT > [Runt + Ftz]
F)
NGT > Runt

Fig. 3.4. PESE mediates repression by Eve and by Ftz in combination with Runt. (A) Expression of *slp1* (green), *lacZ* (red) and the merged image in an embryo homozygous for the *slp1[3918]-lacZ*^{att} reporter. Parasegments 4, 5 and 6 are labeled for the *lacZ* and merged image. The merged image demonstrates that *lacZ* expression overlaps with the more strongly expressed even-numbered *slp1* stripes. The accompanying schematic shows expression of *slp1* and *slp1[3918]-lacZ*^{att} in wild type embryos. Other panels show expression of *slp1* and this reporter transgene in response to manipulations in pair-rule gene activity with an accompanying schematic representation. (B) Ventral view of *lacZ* expression in an embryo homozygous for the temperature sensitive *eve[1]* allele shows some evidence of de-repression, but not as significant as endogenous *slp1* as can be seen in the magnified view of the merged image provided in the inset. (C) *PESE-lacZ* expression is nearly completely abolished in response to *NGT*-driven Eve. (D) An embryo homozygous for *ftz[11]* and the reporter shows modest expansion of *lacZ* stripes when compared to the endogenous gene. This can be better seen in the magnified view of the merged image provided in the inset with an arrow pointing to *lacZ* expressing cells in the anterior half of the even-numbered parasegments. (E) Transient elimination of *runt* gives expanded *PESE-lacZ* expression shown in the inset and stippled red cells of the schematic. (F) *PESE-lacZ* expression is nearly abolished in response to *NGT*-driven co-expression of Runt and Ftz.

Figure 3.5

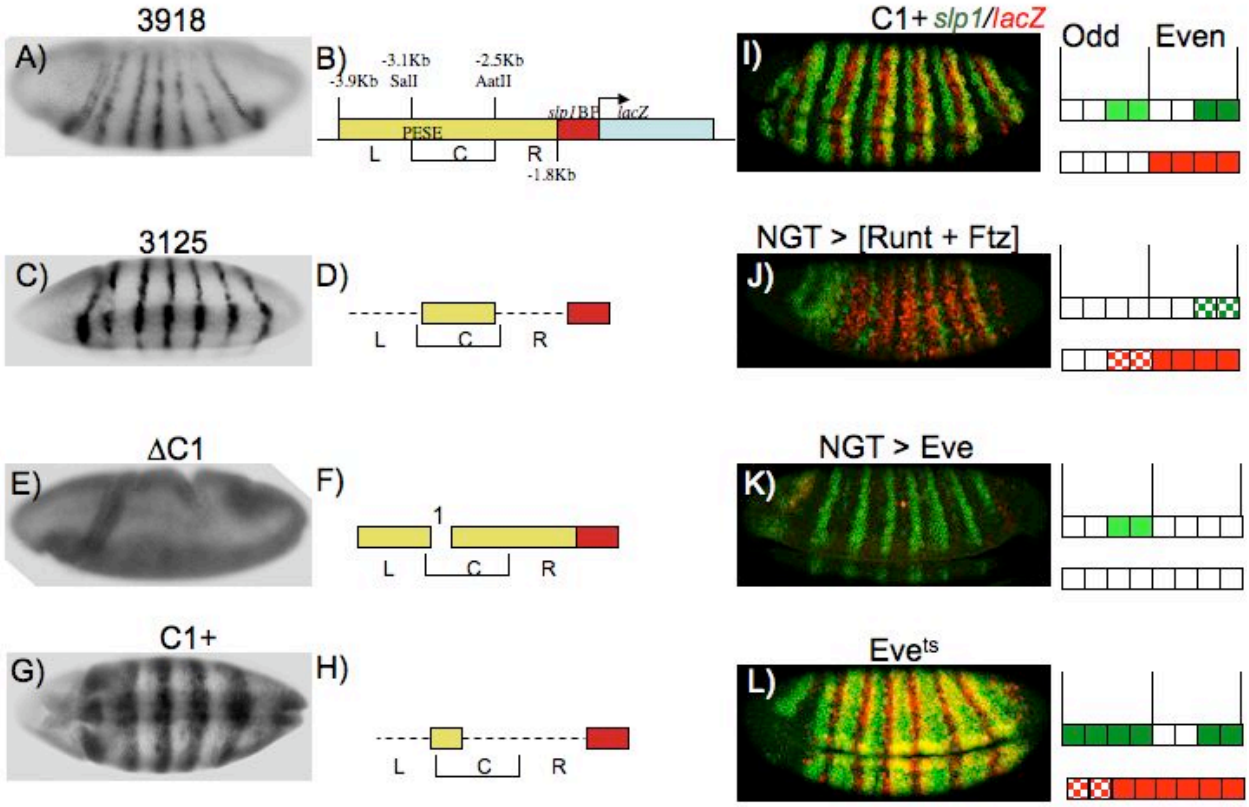
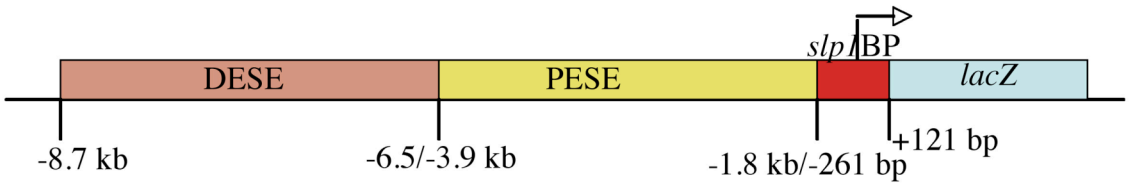

3918
A)
B)
-3.1Kb
-2.5Kb
-3.9Kb
SalI
AatII
slp1BE
lacZ
PESE
L
C
R
-1.8Kb
C1+ slp1/lacZ
I)
Odd
Even
3125
C)
D)
L
C
R
NGT > [Runt + Ftz]
J)
ΔC1
E)
F)
1
L
C
R
NGT > Eve
K)
C1+
G)
H)
L
C
R
Eve^ts
L)

Fig. 3.5. A minimal proximal element mediates even-stripe activation but proper repression by Ftz requires other sequence elements. (A), (C), (E), (G) Expression of *lacZ* mRNA from various PESE reporter gene constructs as visualized by in situ hybridization with the *slp1* upstream sequence present in each construct schematically diagrammed in (B), (D), (F) and (H), respectively. (I), (J), (K), (L) Merged images of *slp1* (green) and *lacZ* (red) mRNAs in embryos with the *slp1[PESE:C1+]-lacZ* reporter diagrammed in (H). (I) Merged image in a wild-type embryo demonstrates *lacZ* expression overlaps *slp1* even stripes and extends anteriorly to the posterior margin of the odd stripes. A schematic interpretation of this expression is given to the right. (J) This minimal PESE:C1+ reporter expresses four cell wide stripes in embryos where *slp1* expression is nearly completely repressed in response to *NGT*-driven co-expression of Runt and Ftz. The low level *lacZ* expression in odd parasegments of this embryo is likely an indirect effect due to partial repression of Eve by Runt and Ftz. (K) Merged image showing complete repression of *slp1[PESE:C1+]-lacZ* in an embryo in which *NGT*-driven Eve specifically represses even-numbered *slp1* stripes. (L) Transient elimination of *eve* results in *slp1[PESE:C1+]-lacZ* expression throughout the pre-segmental region of the embryo, overlapping with *slp1* expression throughout the odd parasegments and posterior even parasegments and filling the gap in *slp1* expression that is due to repression by Runt and Ftz in anterior even parasegments.

Figure 3.6

A)


slp1BP
DESE
PESE
lacZ
-8.7 kb
-6.5/-3.9 kb
-1.8 kb/-261 bp
+121 bp

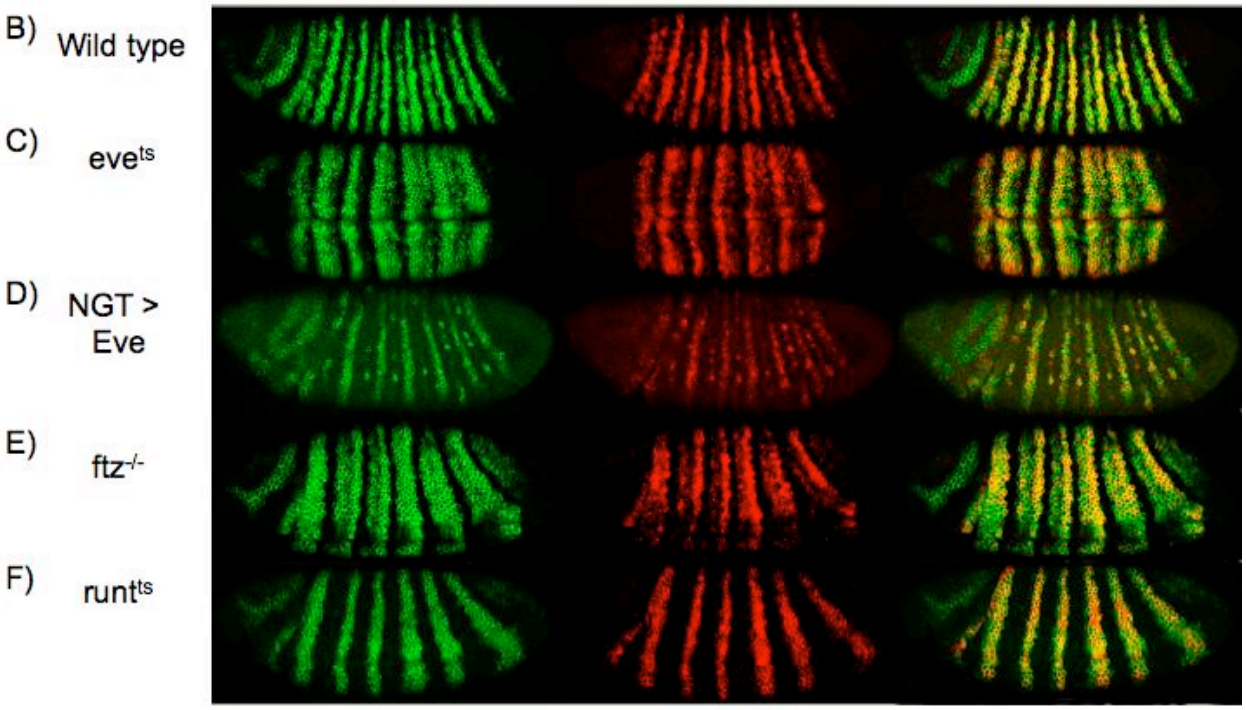

B) Wild type
C) eve^ts
D) NGT > Eve
E) ftz^-/-
F) runt^ts

Figure 3.6. A composite reporter faithfully recapitulates *slp1* expression in the segmented region of the embryo. (A) Diagram of the composite reporter gene construct with the coordinates of the elements relative to *slp1*. DESE and PESE are fused 5' of the *slp1* basal promoter with no spacer region between these elements. (B) Fluorescent *in situ* hybridization shows the expression of *slp1* (green) and the composite *slp1[8765:3918]-lacZ* reporter gene (red) in a wild-type gastrula stage embryo. The merged image on the right shows that the two patterns overlap throughout the segmented region of the embryo, with *slp1* showing an additional anterior band of expression in the un-segmented head region. (C) This composite reporter mimics the *slp1* response to the transient elimination of *eve* with expression in the anterior half of the odd-numbered parasegments (D) The composite reporter also responds faithfully to ectopic Eve expression with specific repression of the even stripes. (E) In a $ftz^{-/-}$ background, *slp1[8765:3918]-lacZ* expression fills the anterior half of the even parasegments. (F) The composite also fully recapitulates the response of *slp1* to the transient elimination of *runt*.

Figure 3.7

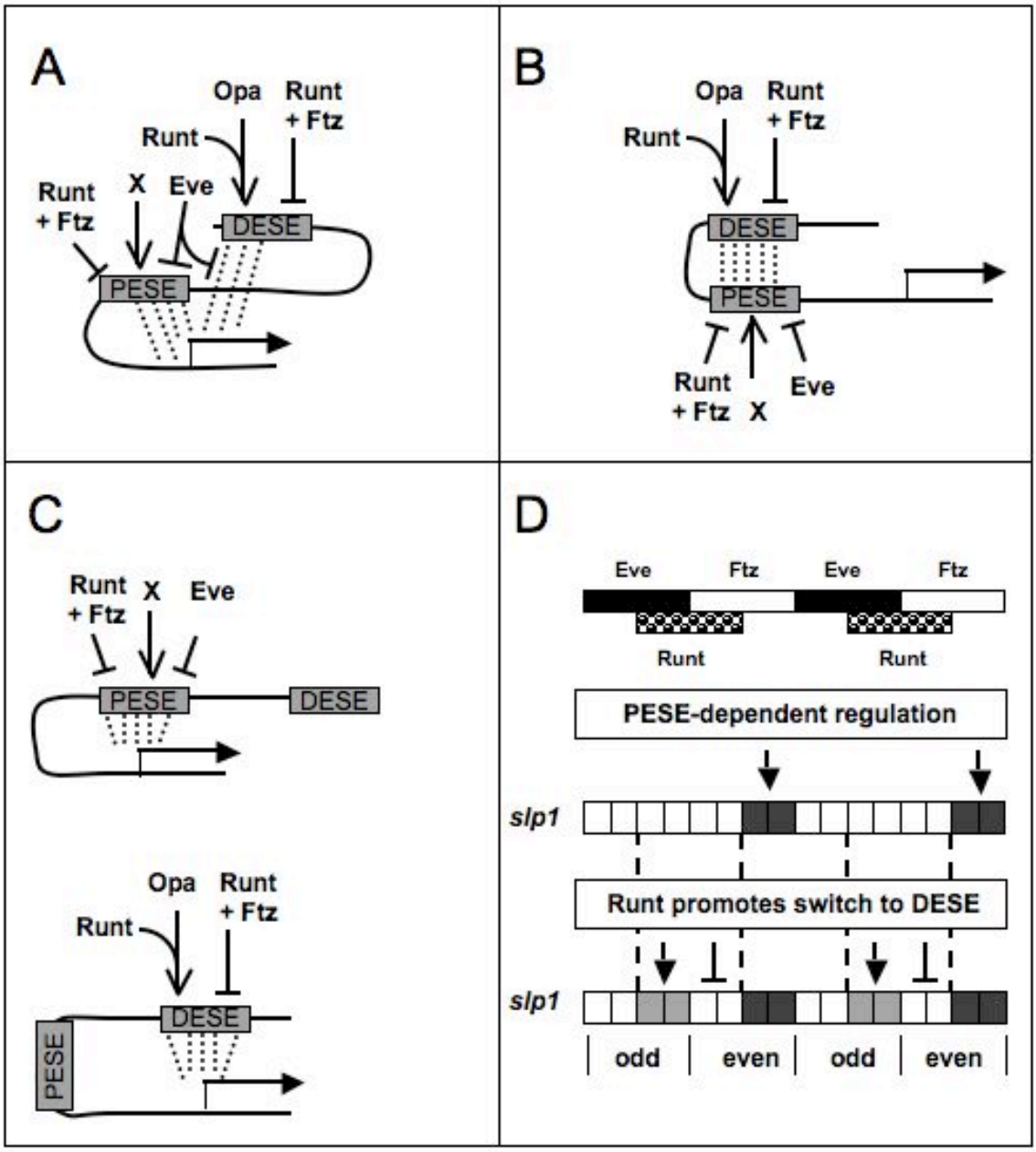

A
Opa
Runt + Ftz
Runt
Runt + Ftz
X
Eve
DESE
PESE
B
Opa
Runt + Ftz
Runt
DESE
PESE
Runt + Ftz
X
Eve
C
Runt + Ftz
X
Eve
PESE
DESE
Opa
Runt + Ftz
Runt
DESE
PESE
D
Eve
Ftz
Eve
Ftz
Runt
Runt
PESE-dependent regulation
slp1
Runt promotes switch to DESE
slp1
odd
even
odd
even

Figure 3.7. Integration of pair-rule cues by dynamic regulation of enhancer-promoter interactions. (A) Schematic representation of proposed interactions involving the DESE and PESE elements and the *slp1* promoter. The DESE and PESE elements are depicted as rectangles connected by a curved line representing upstream DNA that normally separates these elements from each other, as well as from the 5' end of the *slp1* transcription unit (indicated by the arrow). The regulation mediated in response to the pair-rule transcription factors is indicated above each element, with the dotted lines between the elements and the *slp1* promoter region indicating enhancer-promoter interactions. In this model PESE-dependent repression by Eve is capable of interfering with DESE-dependent activation via an unknown mechanism. (B) Schematic interpretation of an alternative model in which Eve-interacting PESE acts to sequester DESE, thereby preventing communication between DESE and the *slp1* promoter. (C) Schematic diagram of two alternative conformations of the *slp1* locus, each involving interaction of the promoter region with a single upstream CRM. The conformation on top has PESE interacting with the promoter, whereas the bottom diagram has DESE interacting with the promoter. The pair-rule dependent regulation pertinent to each conformation is indicated above the promoter-interacting CRM. (D) Relative expression patterns of Eve (black bar), Ftz (white bar) and Runt (stippled bar) are indicated above a column of 16 cells along the anterior-posterior axis. Cells that express *slp1* in response to cues mediated by the PESE-promoter conformation shown in panel C (i.e. all Factor X-expressing cells that do not have Eve, or the combination of Runt and Ftz) are shaded in black. A second column of cells provided at the bottom shows the aggregate expression pattern and parasegmental registration that is expected if the *slp1* locus switches to the DESE-promoter conformation specifically in Runt-expressing cells.

Supplemental Figures

Figure 3.S1

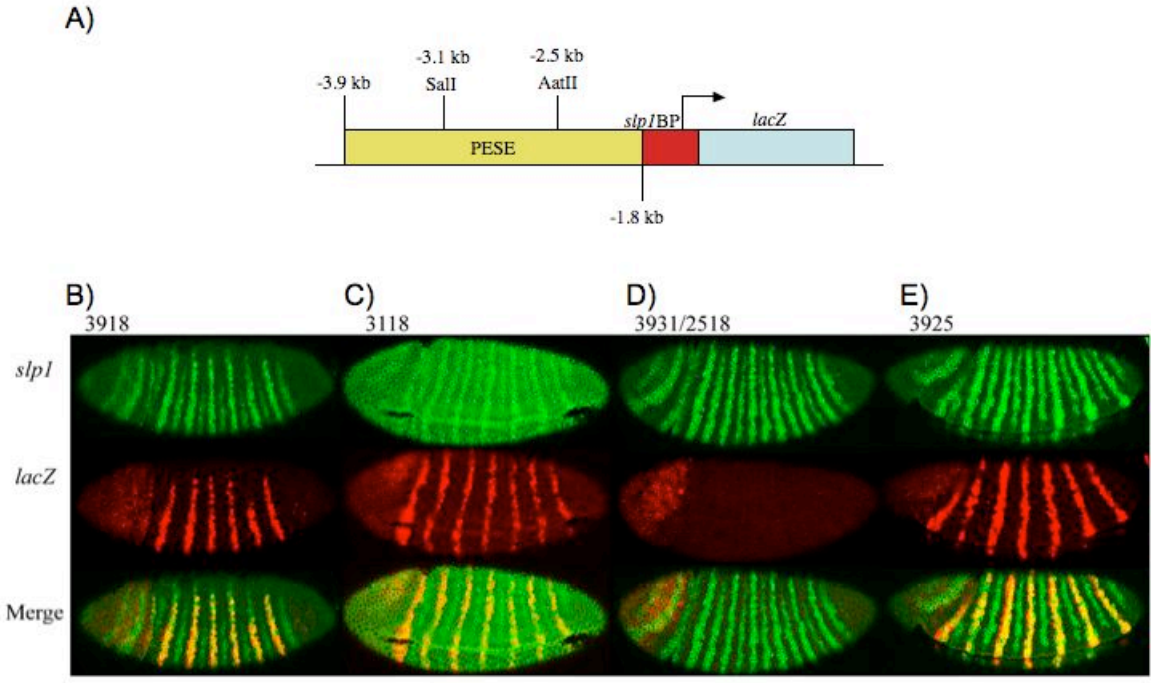


Figure 3.S1. The central region of PESE is needed for even stripe expression
(A) Schematic diagram of the full length proximal element from 3.9 kb to 1.8 kb upstream of *slp1*. The restriction site shown (*SalI* and *AatII*) were used to make the internal deletion of PESE by digestion followed by blunt end ligation. (B-E) Fluorescent in situ hybridization of reporter gene constructs showing the expression of *slp1* (green) at top, *lacZ* (red) middle row and the merge at the bottom. (D) A 600 bp deletion of the central region of PESE from 3.1 kb to 2.5 kb eliminates expression of the even-numbered stripes.

Figure 3.S2

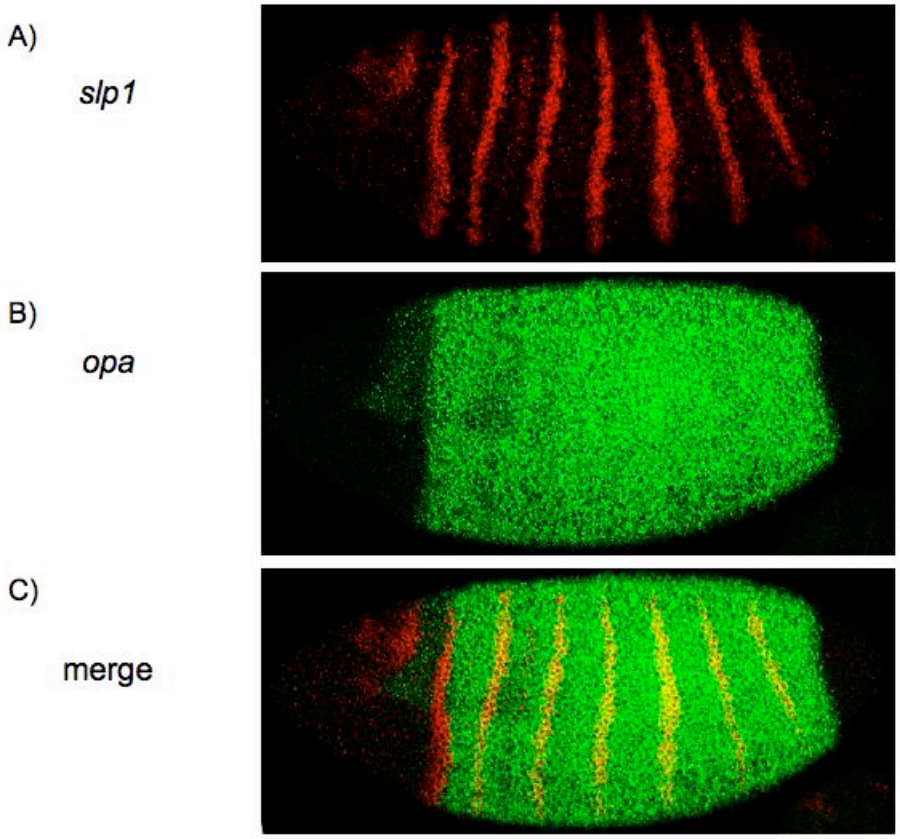


Figure 3.S2. Stripe zero of PESE does not overlap with *opa* expression

(A) Expression of *slp1* mRNA at stage 6, prior to the formation of the cephalic furrow where the even-numbered *slp1* stripes are expressed and the odd stripes are just beginning to be expressed. (B) Expression of *opa* (green) in the same embryo, throughout the presegmental region. (C) A merged image of *slp1* and *opa* showing an overlap of both expression patterns except stripe zero of *slp1* is just anterior to *opa* expression.

Figure 3.S3

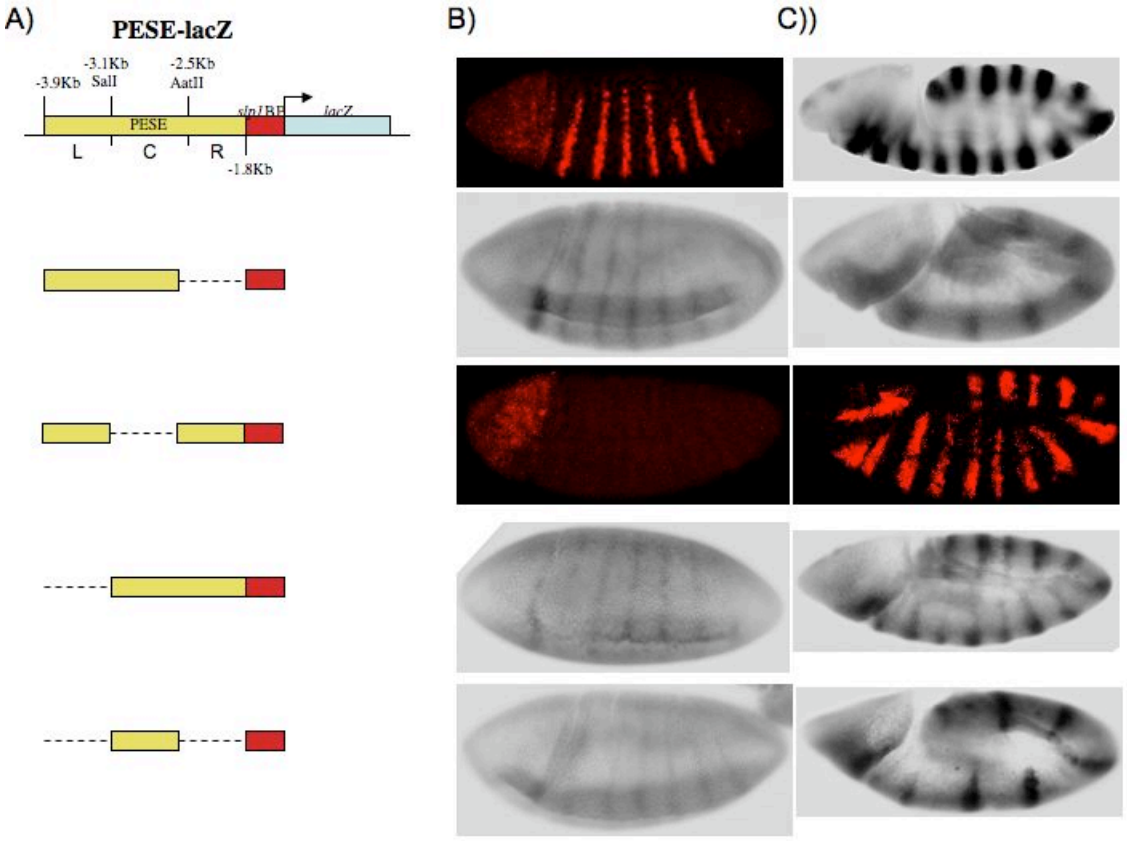

A)
PESE-lacZ
-3.1Kb
SalI
-2.5Kb
AatII
-3.9Kb
slp1BE
lacZ
PESE
L
C
R
-1.8Kb
B)
C))

Figure 3.S3 Right region of PESE needed for odd-stripes at germ band extension

(A) Schematic diagrams of a full length proximal element followed by the deletion constructs that were analyzed. (B) Expression of *lacZ* mRNA in stage 6 embros of the corresponding PESE constructs illustrated to the left of the images. A deletion of the central region of PESE lacks expression at this stage of development. (C) Expression of *lacZ* mRNA of the corresponding PESE constructs to the left at germ band extension showing that a deletion of the right region of PESE lacks odd-numbered stripes at this stage.

Figure 3.S4

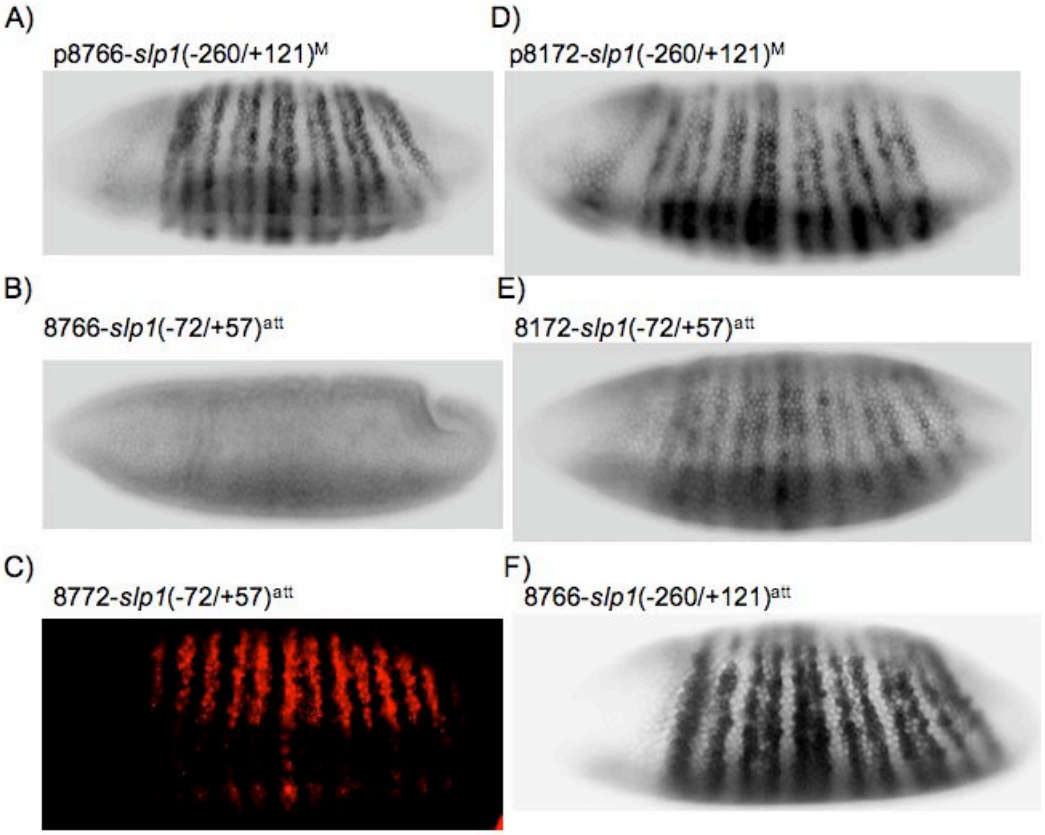

A)
p8766-slp1(-260/+121)M
D)
p8172-slp1(-260/+121)M
B)
8766-slp1(-72/+57)att
E)
8172-slp1(-72/+57)att
C)
8772-slp1(-72/+57)att
F)
8766-slp1(-260/+121)att

Figure 3.S4 A larger promoter is needed for proper expression of the extended distal element

Expression of *lacZ* mRNA of various distal element constructs in gastrula stage embrytos (A and D) Express of the original full length distal element construct as well as a minimal DESE construct from 8.1 kb to 7.1 kb upstream of *slp1* using P-element mediated transposition shows similar expression. (B and E) Expression from the same DESE elements used in A and B with a minimal basal promoter from -72 bp to +57 bp relative to the transcription start site of *slp1* using site-specfic integration. The full length element is not expressed using this construct. (C) Expression of a truncated form of DESE that is deleted for sequence from 7.1 kb to 6.5 kb 5' of *slp1* using the same promoter in B and E expresses *lacZ* in a manner similar to the full length construct in A. (F) Incorporation of a larger promoter as used in A and B allows expression of the full length distal element construct used for site specific integration.

Chapter IV

Importance of DNA-binding for transcriptional regulation by Runt

Chapter IV

Summary

The primary pair-rule gene *runt* is an important transcriptional regulator of segmentation in the Drosophila embryo. The work described here takes advantage of two experimental methods to examine the importance of DNA-binding by Runt in regulating transcription. This includes studies on the effects caused by mutating Runx binding sites in the *slp1* DESE enhancer as well as experiments utilizing a DNA-binding defective form of Runt. A distal early stripe element (DESE) from 8.7 kb to 7.1 kb upstream of the *slp1* promoter is able to mediate Runt dependent regulation and Runx binding sites within this element contribute to repression by Runt and Ftz but are not critical in mediating Runt dependent activation. Although these sites are not critical for activation of DESE, a DNA-binding defective form of Runt is unable to activate this element and proper regulation of *slp1* is dependent on DNA-binding by Runt. These results indicate a differential requirement for DNA-binding in the Runt-dependent regulation of the *DESE-lacZ* reporter gene and *slp1*.

Introduction

A key transcriptional regulator of segmentation in the early Drosophila embryo is the primary pair-rule gene *runt*. The Runt protein is the founding member of a family of heteromeric DNA-binding transcriptional regulatory proteins that share a 128 amino acid motif termed the Runt domain due to sequence homology to Runt. Insight into the molecular function of Runt domain proteins came from studies on the mammalian transcription factor polyomavirus enhancer-binding protein 2 (PEBP2) and core binding factor (CBF), two names for the same heteromeric DNA-binding complex. PEBP2 was shown to bind the polyoma virus enhancer at the consensus PuACCPuCA in NIH3T3 cells, a sequence motif which is important for transcription and replication and CBF was shown to interact with mammalian type C retroviruses (Kamachi et al., 1990; Satake, 1988; Wang and Speck, 1992). This same factor was also found to bind enhancers of genes expressed in lymphoid cells such as T cell receptors and PEBP2/CBF binding was found to be important as shown by decreased enhancer activity when these sites are mutated (Gottschalk and Leiden, 1990; Ogawa et al., 1993a; Redondo et al., 1991). Cloning of PEBP2/CBF revealed that it is a heterodimer of two subunits, α and β, the α subunit includes the Runt domain and mediates DNA binding, whereas the β subunit does not contact DNA but instead enhances the DNA binding of the α subunit (Ogawa et al., 1993b; Wang et al., 1993). Evidence that the Runt domain of this protein was responsible for DNA binding and protein-protein interactions was shown through a mutagenesis of the Runt domain that abolishes DNA binding by PEBP2, as well as by structural analysis (Ogawa et al., 1993b).

Runt domain proteins function in developmental pathways ranging from segmentation in Drosophila to hematopoeisis, bone and neural development in humans. Runt domain proteins can function as both transcriptional activators and repressors, depending on the specific target gene and developmental context. In Drosophila, Runt has been shown to both activate and repress its target genes during segmentation (Manoukian and Krause, 1993; Tsai and Gergen, 1994). Runt was first identified as a protein necessary for proper segmentation where zygotic elimination causes a series of deletions of pattern elements spaced at double segment intervals along the AP axis (Gergen and Wieschaus, 1986; Nusslein-Volhard and Wieschaus, 1980). Along with Runt being required for proper segmentation in the developing Drosophila embryo, it is also involved in sex determination and neurogenesis (Duffy and Gergen, 1991; Duffy et al., 1991). Runt expression is strongest at the blastoderm stage of development in a 7-stripe pattern characteristic of other pair-rule genes where it acts as a regulator of pair-rule and segment-polarity genes working with other pair-rule transcription factors to establish the metameric pattern of the Drosophila embryo (Gergen and Butler, 1988; Tsai and Gergen, 1994; Tsai and Gergen, 1995). The activities of these segmentation and homeotic genes are responsible for positional identity along the anterior posterior axis of the Drosophila embryo (Harding et al., 1986). A specific example of a target gene both activated and repressed by Runt during Drosophila development is the segment polarity gene *sloppy paired (slp1),* which is regulated in a relatively, well defined context dependent manner. The simple combinatorial rules responsible for the metameric striped pattern of *(slp1)* expression in the gastrula stage Drosophila embryo provide an attractive model for investigating Runt function.

Nuclear localization of the Runt protein is evidence that like the mammalian Runt domain homologs, Runt is acting as a sequence specific DNA-binding transcription factor (Kania et al., 1990). The homology of the Runt domain suggests that it has a conserved function. Runt has been shown to bind the same sequence as its mammalian homologs, with enhancement of binding in the presence of the mammalian CBF/PEBP2β protein (Kagoshima et al., 1993). Further evidence for the importance of DNA binding by Runt is the fact that DNA binding by Runt is needed for the initial activation of *Sex-lethal* and the maintenance of *engrailed* repression (Kramer et al., 1999; Wheeler et al., 2002). This work shows that DNA binding by Runt is needed for certain aspects of *runt*-dependent regulation. The identification of Runt binding sites in the regulatory region of target genes can aid in understanding how this protein elicits its function.

Work in the previous chapter characterized two discrete cis-regulatory elements that mediate the initial transcriptional response of *slp1* to Runt and the other pair-rule transcription factors. A proximal early stripe element (PESE) drives expression of the even-numbered *slp1* stripes but fails to generate the odd-numbered, Runt-dependent stripes. A distal early stripe element (DESE), although insensitive to Eve dependent repression, is able to mediate both activation by Runt and the Zn-finger transcription factor Opa, as well as repression by Runt and the homeodomain protein Ftz. When these two elements are combined the expression pattern of the reporter responds in a manner similar if not identical to *slp1* suggesting that the proper integration of pair-rule regulatory inputs involves functional interactions between these two elements.

Although the proximal element can mediate repression by Runt, the distal element alone can respond to both activation and repression by Runt in combination with Opa and

Ftz, respectively. The distal element provides a unique opportunity to study both aspects of Runt dependent regulation. The work described here investigates the role of DNA-binding in *DESE-lacZ* regulation. Results obtained with Runx binding site mutations in DESE, as well as with a DNA-binding defective form of Runt demonstrate that DNA-binding by Runt is critical for repression of DESE-dependent expression. Evidence is also presented that DNA-binding is important for activation of both *slp1* and *ftz*. The observation that mutagenesis of the Runx sites in DESE does not abrogate Runt-dependent activation further demonstrates differential requirements for DNA-binding in the activation of DESE versus the endogenous *slp1* gene.

Results

DESE contains Runx sites and associates with Runt *in vivo*

The distal early stripe element is able to mediate Runt dependent regulation and in Runt expressing cells is responsible for proper *slp1* activation. To further characterize this element and investigate the mechanisms by which Runt is able to elicit its function on DESE, I decided to investigate the role of DNA-binding by Runt in the regulation of DESE-dependent transcription. Experimental evidence for the association of Runt with CRMs within the *slp1* locus comes from whole genome ChIP on chip experiments done on early Drosophila embryos by Mark Biggin and co-workers at Lawerence Berkley National Laboratory (Fig. 4.1C) (Li et al., 2008). The consensus binding site for Runx proteins is ACCpuCA (Fig. 4.1A) (Kamachi et al., 1990; Satake et al., 1988). Two observations indicate the binding site for Runt is the same. The structure of Runx bound to DNA has been solved and all of the amino acid positions that make sequence specific contacts are identical in the Drosophila protein (Bravo, J. *et al*., 2001). Also, competition experiments with a series of oligonucleotides containing viral enhancer sequences of different affinities for mammalian Runx proteins (Lewis et al., 1999) indicate the Drosophila Runt protein has the same rank order of relative affinities for these sites (Kevin King, unpublished observations). A Runx position weight matrix was used to identify candidates for Runt binding sites in DESE (Fig. 4.1A). This search identified many sites in the distal element of varying quality based on Ri value (Fig. 4.1B). Of these sites, five were chosen that seem to be the best candidates, having a high Ri value as well as containing the core cytosines in the consensus that contribute most to Runx occupancy on DNA (Fig. 4.1A and B). In order to examine the importance of these sites,

I mutated the two critical cytosines in the consensus sequence to adenine, a change that has been shown to disrupt the ability of Runt to bind (Fig. 4.1A).

Runx sites three and four potentiate repression by Runt and Ftz

I used ΦC31 integrase for site-specific integration of reporter genes to be able detect subtle effects produced by mutation of these 5 sites alone as well as in combination. Compared to the extended distal element, a truncated reporter gene containing sequence from 8.7 kb to 7.1 kb 5' to *slp1* has no effect on the expression of the distal element where both the even and odd-numbered stripes are expressed at similar intensities and there is slight derepression in the anterior of the odd-numbered parasegments (Fig. 4.2A and B). This element removes Runx sites one and two, suggesting these sites are not critical to Runt mediated regulation of DESE. The observation that mutating site two alone has no effect on expression further confirms that this site is not critical to regulation (Fig. 4.2C). However single mutations of sites four and five, which show higher peaks for Runt association, also have no effect on expression (Fig. 4.1C and 4.2D, E). Since there was no change in DESE expression with single mutations when compared to the wild type reporter I combined Runx site mutants to see if this would affect DESE expression. A double mutant of sites 3 and 4 lacks sensitivity to repression by Runt and Ftz posterior to even stripe number 12 as there is ectopic expression in these cells (Fig. 4.2F and G). The significance of these sites is evident as they correspond to the highest peaks of Runt binding in DESE (Fig. 4.1C).

I quantified the expression patterns obtained from these various reporters to confirm this qualitative observation on the effect Runx sites three and four have on response to Runt and Ftz in the anterior half of parasegment number twelve. To quantify

expression patterns of these reporters an image processing method was used that takes a confocal image and converts it to a text file where the intensity of gene expression can be represented graphically in a format that allows comparison of embryos from one reporter gene to the next (Janssens et al., 2005). This is a step-wise process that consists of segmenting the image, background removal to account for differences in antibody specificity, registration and integration. The results of this method confirm the lack of repression in the anterior half of parasegment number 12 when sites 3 and 4 are mutated and further reveals evidence for de-repression anterior to stripe 10 when compared to wild type DESE (Fig 4.3A and B).

All five Runx sites are needed for repression throughout the segmental region

Given the partial loss of repression obtained in the site three and four double mutant I decided to generate a quintuple mutant to further investigate the role of DNA-binding by Runt. A distal element with mutations in these five Runx binding sites shows lack of sensitivity to repression by Runt and Ftz in all even numbered parasegments resulting in expression throughout the presegmental region of the embryo due to this loss of sensitivity combined with a lack of repression by Eve (Fig. 4.4C and D). It is interesting to note that the expression level of odd stripes in the quintuple mutant in wild type embyos does not seem to be affected, suggesting that these sites are not critical for Runt-dependent activation of the *DESE-lacZ* reporter (Figure 4.4C and D).

DNA-binding by Runt is necessary for repression but not activation of DESE

To further explore the contribution of Runx sites in DESE I investigated the response to pair-rule transcription factors. The combination of Runt and Ftz is able to repress *slp1* throughout the embryo (Swantek and Gergen, 2004). At slightly lower

levels of NGT driven Runt and Ftz we observe repression of the odd-numbered *slp1* stripes and partial repression of the even stripes (Fig 4.5B and E). The distal element responds similarly to these levels of Runt in combination with Ftz where we observe repression of the odd-numbered stripes and the even stripes are partially repressed (Fig. 4.5B and E). There is more *DESE-lacZ* expression when compared to the endogenous gene in these embryos (Fig. 4.5B). This can be attributed in part to the ectopic expression of *lacZ* in the anterior half of the odd parasegments (Fig. 4.5E). This expression occurs in cells that should be repressed by Eve, an aspect of *slp1* regulation this element is unable to mediate. In order to further investigate the role of DNA-binding in Runt-dependent regulation I utilized a DNA binding defective form of Runt, Runt[CK], which harbors mutations of two amino acids C127 and K199 that are important for DNA binding and is defective for Runt-mediated activation of *Sex-lethal* as well as the maintenance of *en* repression (Kramer et al., 1999; Wheeler et al., 2002). The Runt[CK] protein is unable to repress the even-numbered stripes of both the endogenous gene and the reporter when co-expressed with Ftz (Fig. 4.5C and F). Repression of the odd-numbered stripes is observed in these embryos due to the combination of ectopic Ftz in cells that express endogenous wild type Runt (Fig. 4.5C and F). This result, combined with the mutational analysis of Runx binding sites within DESE indicates that DNA-binding by Runt is need for proper repression of *slp1*.

In the posterior half of the odd parasegments the absence of Ftz allows Runt to function with Opa to activate *slp1* and these two factors are able to activate the corresponding stripes of *DESE-lacZ* (chapter III). In wild type embryos a reporter containing mutations within five Runx sites expresses the odd-numbered stripes,

indicating these sites are not critical for the Runt dependent activation of DESE (Fig. 4.4Cand D). This is further confirmed by the ability of this quintuple mutant to respond to NGT-driven Runt and Opa where both *DESE-lacZ* as well as *DESE-lacZm1,2,3,4,5* are both activated in the head (Fig. 4.6A and B). There is more expression of the quintuple mutant reporter compared to *DESE-lacZ* in the pre-segmental region due to expression in cells that should be repressed by Runt and Ftz further highlighting the importance of these sites to Runt dependent repression.

The combination of Runt and Opa is able to drive expression of *slp1* in the head and the quintuple Runx site mutant is activated in this region indicating that although theses sites are critical for Runt-dependent repression, they do not seem to be as important for activation. However, when we drive ectopic expression of Runt[CK] and Opa, the activation of *slp1* and *DESE-lacZ* in the head is not as strong (Fig 4.6C) implying that DNA-binding by Runt is important for activation of *slp1* and *DESE-lacZ* to a somewhat lesser extent. Taken together, these results suggest that in order for Runt to function in both activation and repression of *slp1* and *DESE-lacZ* DNA-binding by Runt is important.

Activation of *slp1* and *ftz* requires DNA-binding by Runt

Although DNA-binding by Runt is needed to mediate proper activation and repression of both *slp1* and the distal element reporter, there are significant differences in the responses within the pre-segmental region. During the segmentation process Runt also functions to activate *ftz* and repress *en* where the initial repression of *en* does not require Runts ability to bind DNA (Wheeler et al., 2002). Expression of Runt and Opa throughout the embryo at moderate levels results in expansion of *ftz* with stripes five and

six almost merging and expansion of the remaining stripes at this stage (Fig. 4.7A). This expansion is more obvious at high levels of NGT-driven Runt and Opa where *ftz* is now expressed throughout the majority of the presegmental region of the embryo in a pattern complementary to *slp1* due to the ability of Ftz to repress *slp1* in Runt expressing cells (Fig. 4.7B). The activation of *ftz* observed when these factors are ectopically expressed is dependent on DNA-binding by Runt as expansion of the Ftz stripes is only expanded at high levels of NGT-driven Runt [CK] and Opa (Fig 4.7C and D). To determine if the ectopic expression observed upon mutating Runx sites three and four is indeed in cells that express Ftz I looked at the expression of Ftz and *lacZ* in embryos that ectopically express Runt and Opa. The expression of Ftz and *DESE-lacZ* is mutually exclusive due to the regulatory effects of Ftz on DESE (Fig. 4.7E). Conversely when sites 3 and 4 are mutated there are cells that are expressing both Ftz and the reporter, unlike the wild type reporter where Ftz and lacZ are not co-expressed in the same cell (Fig. 4.7E and F). This result shows that in cells that have Ftz, Runt sites 3 and 4 in DESE are important for proper repression but not required for activation.

Discussion

Throughout Drosophila development Runt is involved in various aspects of transcriptional regulation. During sex determination Runt is needed for the initial activation of *Sxl* and then later on in development it is involved in the specification of neuroblasts during neurogenesis (Kania, 1990; Kramer, 1999). The work described here concentrates on the function of Runt during segmentation and in particular the role it plays in regulating *slp1*, with the work in this chapter focused on the importance of DNA-binding in Runt-dependent *slp1* regulation. Using the combination of a Runx position weight matrix to identify five Runx binding sites within DESE and the *in-vivo* occupancy of Runt on DESE as revealed by ChIP on chip data I have identified five Runx binding sites that contribute to repression by Runt and Ftz. Although Runt is found to associate with DESE there is also an association with PESE with the highest peaks of Runt binding present within DESE. These results suggest that DNA-binding by Runt is needed to regulate both of these elements. However PESE is not activated by Runt and a *PESE-lacZ* transgene with a mutation in a conserved Runx site has no effect on repression by Runt and Ftz (data not shown). The ChIP on chip data obtained from these experiments is on a mixed population of cells that are both repressing as well as activating *slp1* so one can not distinguish between the importance of DNA-binding by Runt for activation versus repression. Also, even though we see Runt association with this region this could be indirectly due to interactions with other DNA-binding factors associated with PESE and does not necessarily mean Runt needs to bind DNA to regulate PESE. Further experiments that investigate the importance of DNA-binding by Runt in the regulation of PESE will be interesting.

Although single mutations of Runx sites within DESE have no effect on Runt dependent regulation, mutagenesis of sites three and four, which are present in a region with the highest peaks of Runt binding leads to loss of repression anterior to stripes ten and twelve (Fig. 4.1 - 4.5). It is interesting that this region of the embryo is more susceptible to mutations in these sites and suggests interactions with other regionally expressed factors, such as the gap genes modulate transcription in these cells.

While Runx sites within DESE are important for proper repression they do not seem to affect activation. One explanation for this could be that *slp1* activation in Runt expressing cells requires different factors than those required for repression and that these factors allow Runt to utilize different sites for activation. Another explanation could be that although these sites do not contribute to activation of DESE they may be important for activation of the endogenous gene. Unlike *slp1*, *DESE-lacZm1,2,3,4,5* does not have to overcome 6.5 kb of intervening DNA in Runt expressing cells and therefore the requirements for DESE activation are different then *slp1*. In fact when I use a DNA-binding defective form of Runt I find that DNA-binding by Runt is important for activation of the endogenous *slp1* gene.

An interesting proposal that emerged from the work presented in Chapter III is that proper regulation of *slp1* DESE-driven expression involved Runt-dependent regulation of enhancer promoter interactions. In the context of this model the differential requirement for DNA-binding in the activation of *DESE-lacZ* versus endogenous *slp1* may provide an indication that DNA-binding by Runt is critical in the switch from PESE to DESE-*slp1* promoter interactions.

Figures

Figure 4.1

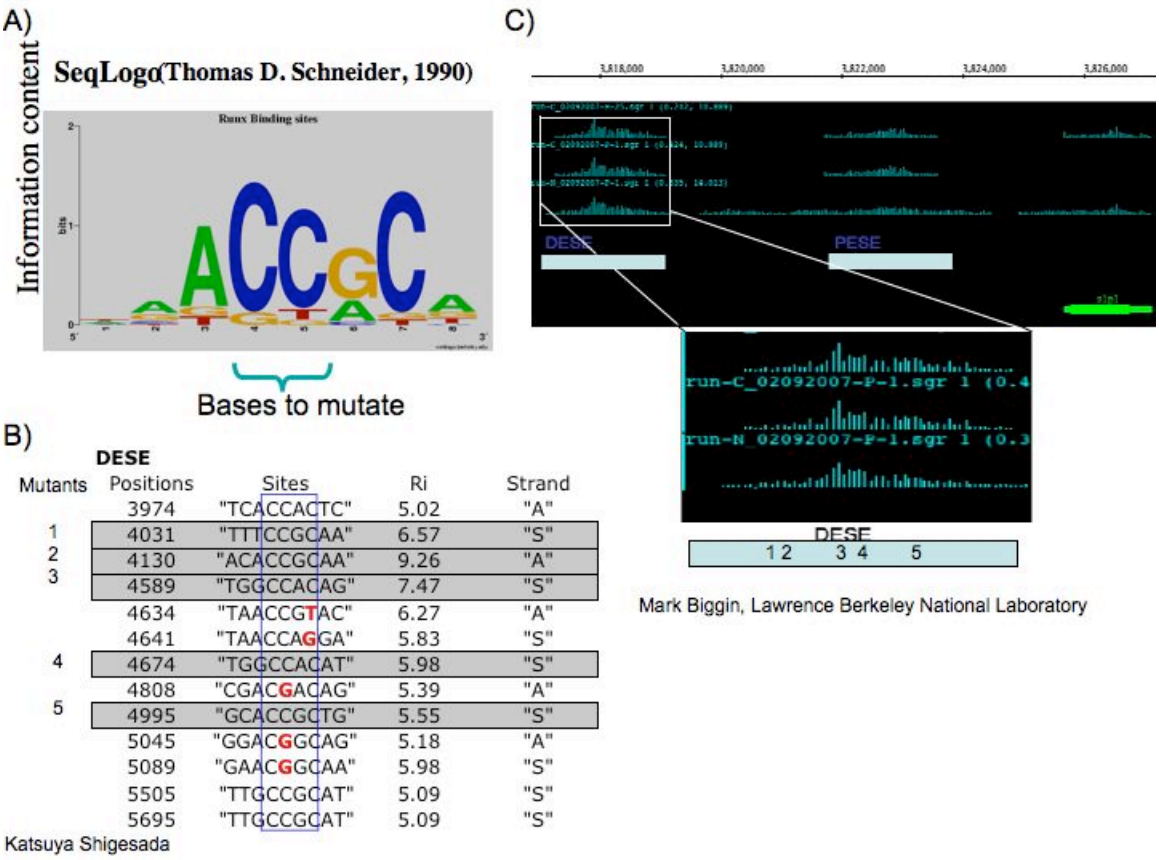


Figure 4.1. DESE contains Runx binding sites and is bound by Runt in vivo. (A) Sequence logo showing the Runx consensus ACCpuCA. (B) Runx binding sites within the distal element showing the location of these sites relative to the distal end of DESE, the sequence of each site with the Ri value and which strand. The sites highlighted are the best candidates based on Ri value and the cytosine bases present in the core. (C) ChIP on chip data of in vivo binding of Runt upstream of *slp1* with the position of DESE and PESE illustrated as rectangles in the peaks of Runt binding. Below is a blow up of this data relative to DESE and the position of the five sites chosen for mutation analysis.

Figure 4.2

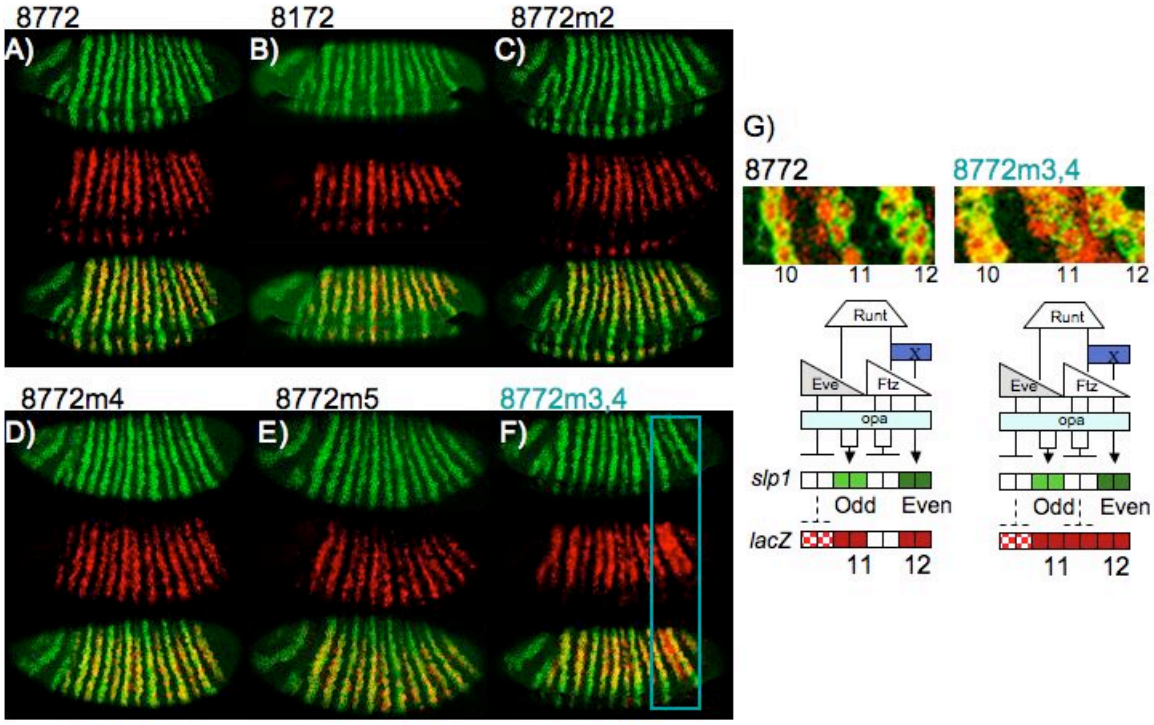

8772
A)
8172
B)
8772m2
C)
8772m4
D)
8772m5
E)
8772m3,4
F)
G)
8772
8772m3,4
10
11
12
Runt
X
Eve
Ftz
opa
slp1
Odd
Even
lacZ
11
12

Figure 4.2. DESE Runx sites three and four potentiate repression.

(A) Expression of *DESE-lacZ* (red) relative to *slp1* (green) in wild type embryos. (B) Runx sites one and two do not seem to mediate Runt dependent regulation through DESE as a minimal distal element from 8.1 kb to 7.2 kb upstream of *slp1* that deletes these sites is expressed similar to *DESE-lacZ*. (C-E) Mutating sites two, four and five alone has no affect on Runt dependent regulation of *DESE-lacZ* as the expression of these constructs is similar with activation of both the even and odd-numbered *slp1* stripes and proper repression by Runt and Ftz in the anterior half of the even parasegments. (F) Sites three and four contribute to repression as this reporter expresses *lacZ* in the anterior half of parasegment 12 which is highlighted by the blue rectangular box. (G) Blow up of stripes ten through twelve of *8772-lacZ* and *8772m3,4-lacZ* showing expression of *lacZ* in cells anterior to stripe twelve upon mutation of Runx sites three and four. Below the blown up images of these parasegments is a schematic diagram of expression of *slp1* and *lacZ* from these reporters where there is derepression in cells that should be repressed by Runt and Ftz upon mutating Runx sites three and four.

Figure 4.3

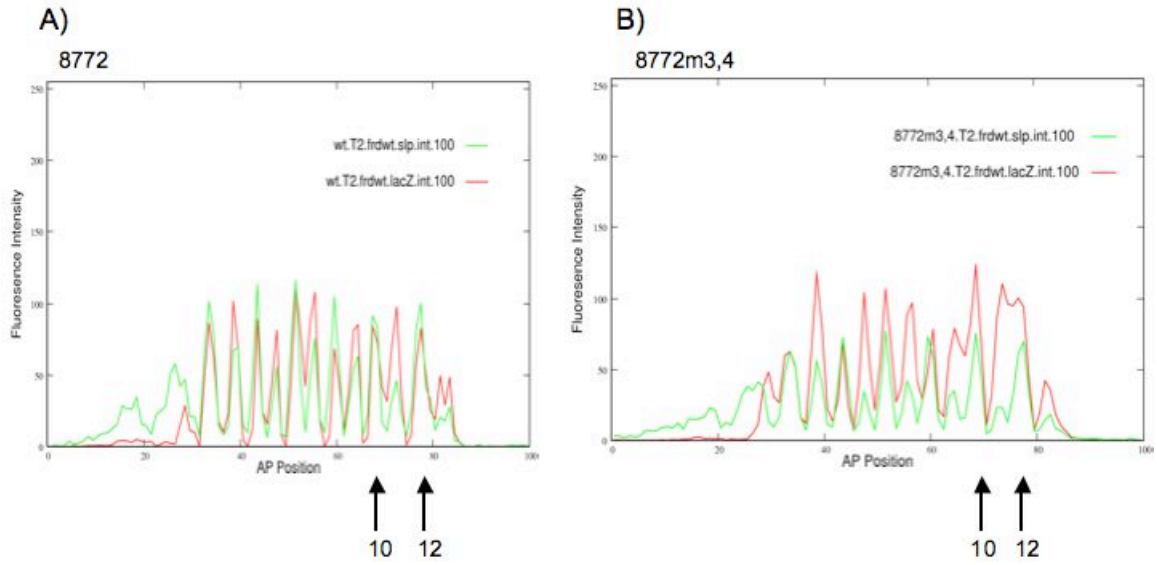

A)
8772
wt.T2.frdwt.slp.int.100
wt.T2.frdwt.lacZ.int.100
Fluorescence Intensity
AP Position
10
12
B)
8772m3,4
8772m3,4.T2.frdwt.slp.int.100
8772m3,4.T2.frdwt.lacZ.int.100
Fluorescence Intensity
AP Position
10
12

Figure 4.3. Runx sites three and four are needed for repression anterior to even stripes ten and twelve.

Graphical representation of averaged expression patterns utilizing an image processing method with *slp1* expression graphed in green and *DESE-lacZ* in red. (A) Overlap of *slp1* and *DESE-lacZ* shows the stronger intensity of the odd stripes of DESE compared to the endogenous gene with evidence of ectopic expression in the anterior of parasegments eleven, seven and five. (B) Mutating sites three and four of DESE confirms the loss of repression anterior to stripe twelve and also reveals a loss of repression anterior to stripe number ten.

Figure 4.4

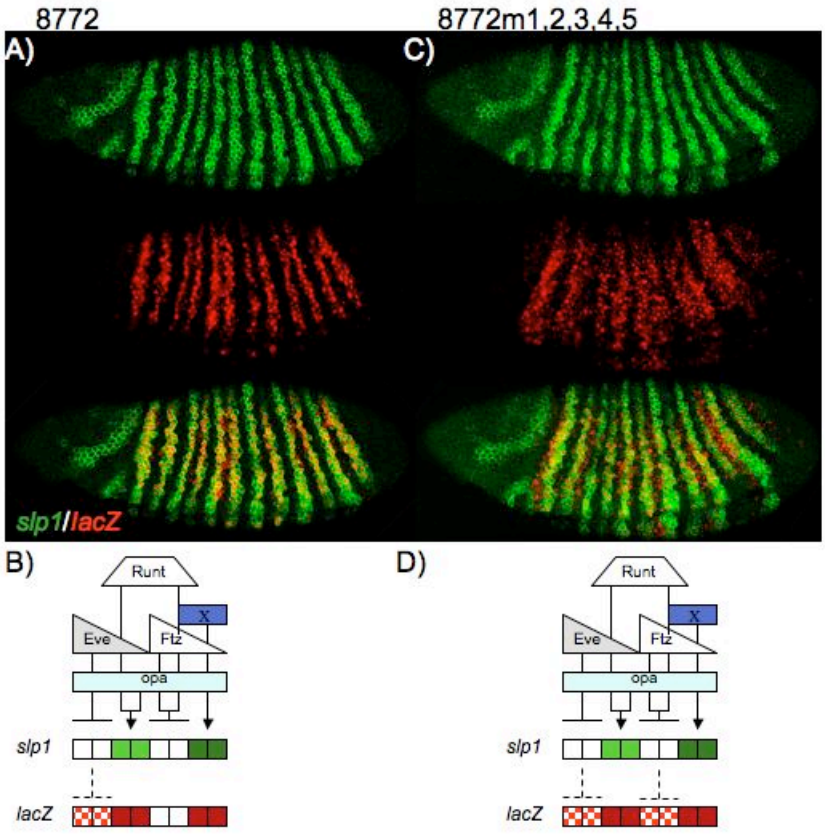


Figure 4.4. All five Runx sites contribute to repression by Runt and Ftz throughout the segmented region of the embryo

(A) Expression of *slp1*, *DESE-lacZ* and a merged image of these patterns at the bottom. (B) Schematic representation of the regulation of both *slp1* and *DESE-lacZ* showing ectopic expression of *lacZ* anterior to the odd numbered stripes. (C) Runx sites one though five are important for repression by Runt and Ftz as *DESE-lacZm1,2,3,4,5* is expressed in the anterior half of all even numbered parasegments. (D) This is illustrated in the schematic diagram showing ectopic expression in the anterior half of both the odd and the even numbered parasegment due to the inability of *DESE-lacZm1,2,3,4,5* to respond to repression by Eve in the anterior half of the odd-numbered paraegents and Runt in combination with Ftz in the anterior of the even parasegments.

Figure 4.5

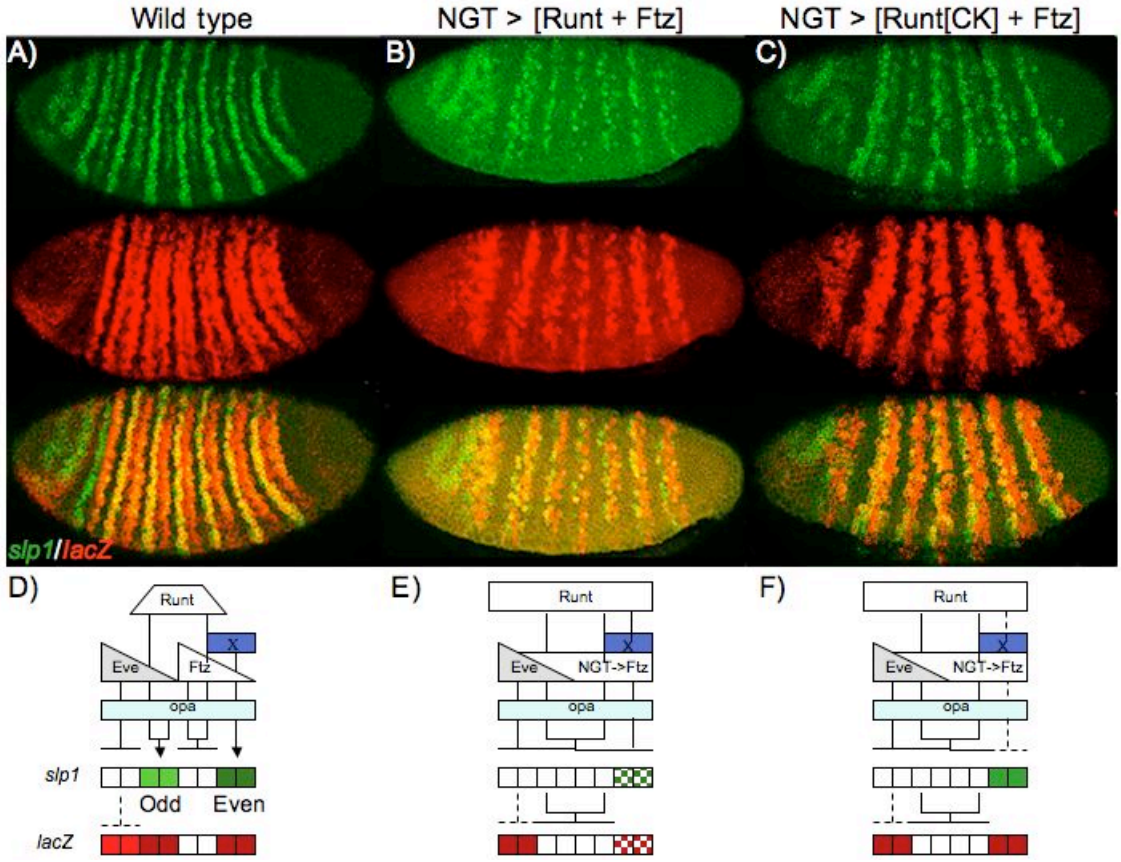

Wild type
NGT > [Runt + Ftz]
NGT > [Runt[CK] + Ftz]
A)
B)
C)
slp1/lacZ
D)
E)
F)
Runt
X
Eve
Ftz
NGT->Ftz
opa
slp1
Odd
Even
lacZ

Figure 4.5. DNA binding by Runt is needed for proper repression of *slp1* and *DESE-lacZ*.

All embryos in this figure and the ones to follow are oriented anterior to the left dorsal side up. (A) Fluorescent *in-situ* hybridization of gastrula staged embryos showing expression of *slp1* (green) and *lacZ* (red) at the blastoderm stage in wild type embryos. (B) Upon NGT-driven Runt and Ftz both the odd and to a lesser extent the even-numbered stripes of *slp1* and *DESE-lacZ* are repressed. There is more expression of *lacZ* compared to the endogenous gene, but this is in cells that express Eve, which fails to fully repress *DESE-lacZ*. (C) Expression of Runt[CK], a DNA-binding defective form of Runt, in combination with is unable to mediate proper repression of DESE and *slp1* as a the odd-numbered stripes are not fully repressed and the even-numbered stripes do not respond. (D-F) Schematic representation of the response of *slp1* and *DESE-lacZ* in (D) wild type, (E) embryos ectopically expressing Runt and FTZ and (F) expression of Runt[CK] and Ftz. The *UAS-runt*^{232} and *UAS-runt[CK]*^{77} transgenes used in these experiments are expressed at similar levels and have nearly equivalent activities on DNA-binding independent targets of Runt.

Figure 4.6

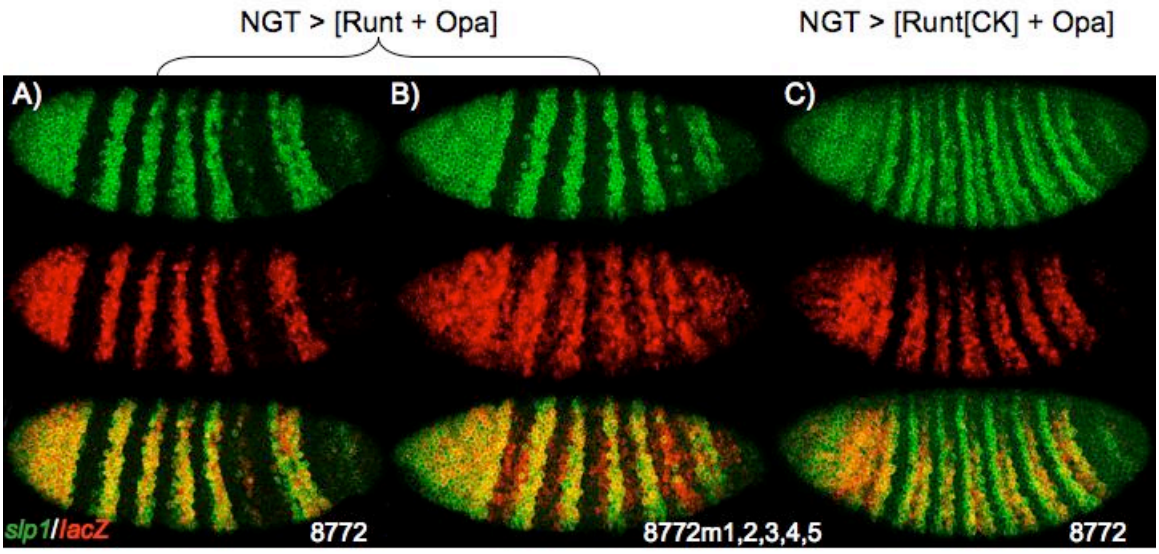

NGT > [Runt + Opa]
NGT > [Runt[CK] + Opa]
A)
B)
C)
slp1/lacZ
8772
8772m1,2,3,4,5
8772

Figure 4.6. Activation of both *slp1* and *DESE-lacZ* requires DNA-binding by Runt. Expression of *slp1* (green) and *DESE-lacZ* (red) in embryos driving expression of two forms of Runt in combination with Opa. (A) The distal element can mediate Runt dependent activation in the head in embryos ectopically expressing Runt in combination with Opa in a manner similar to the endogenous gene. (B) Although these sites are important for proper repression *DESE-lacZm1,2,3,4,5* can be activated by Runt and Opa as shown by expression of this reporter in the head in embryos with NGT-driven Runt and Opa. The expanded expression of this reporter is due to a decreased sensitivity to repression by Runt and Ftz. (C) DNA-binding by Runt is needed for activation of *slp1* and *DESE-lacZ* as embryos expressing Opa and a DNA-binding defective form of Runt are not activated in the head to the extent as those in A and B. Runt[CK] and Ftz. These experiments use the UAS-runt[232] and UAS-runt[CK][77] transgenes with similar activity.

Figure 4.7

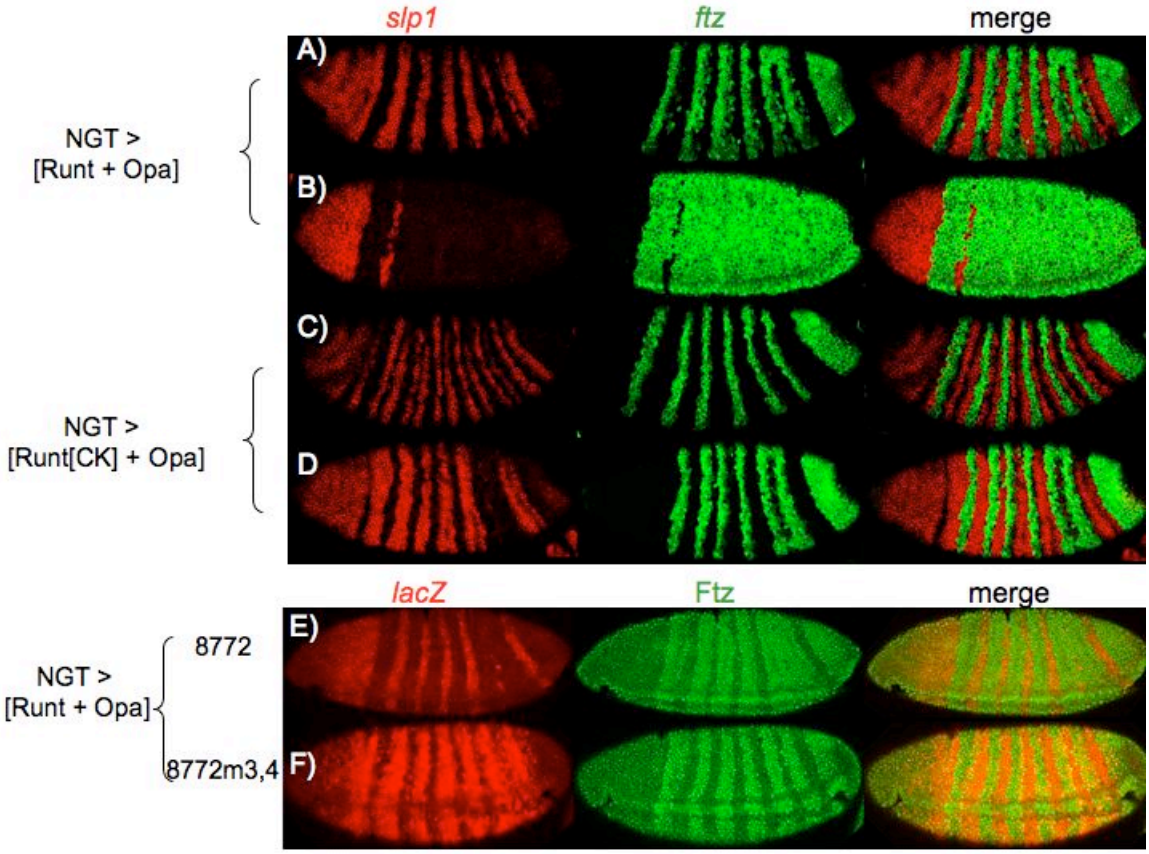

slp1
ftz
merge
A)
B)
C)
D)
NGT >
[Runt + Opa]
NGT >
[Runt[CK] + Opa]
lacZ
Ftz
merge
E)
F)
NGT >
[Runt + Opa]
8772
8772m3,4

Figure 4.7. Activation of *ftz* also requires DNA-binding by Runt and Runx sites three and four contribute to repression in Ftz expressing cells.

(A-D) Expression of *slp1* (red) and *ftz* (green) mRNA in gastrula stage embryos diving expression of Runt and Opa at (A and C) moderate and (B and D) high levels. (A) NGT-driven Runt and Opa at moderate levels is sufficient to activate *slp1* in the head, with expansion of Ftz stripes in the segmented region of the embryo. (B) At high levels of NGT-driven Runt and Opa *slp1* is repressed throughout the segmented region, with the exception of some cells due to the presence of Ftz throughout this region of the embryo. (C) Expression of Runt[CK] and Opa at levels similar to a show an inability of *slp1* to be activated in the head and Ftz expression does not seem to expand. (D) In order to activate *slp1* in the head and start to observe expansion of Ftz we need to use high levels of NGT-Driven Runt[CK] and Opa. (E) Expression of *DESE-lacZ* mRNA (red) and Ftz protein (green) in embryos expressing Runt in combination with Opa at the same levels used in A showing the ability of this element to be activated by Runt and Opa in all cells where Ftz is not present. (F) Runx binding sites three and four contribute to repression by Runt and Ftz as there is expression of *DESE-lacZm3,4* in cells that also contain Ftz confirming the importance of these sites for proper repression in these cells. Runt[CK] and Ftz. The UAS-runt^{232} and UAS-runt[CK]^{77} transgenes were used for these experiments.

Chapter V

Summary

and

Future directions

Summary

This work identifies two distinct elements able to mediate different aspects of *slp1* regulation at the blastoderm stage. The combination of these CRM's mediates most, if not all aspects of *slp1* regulation during this time of development. I propose that initial expression of *slp1* is dependent on a proximal early stripe element and then slightly later, during early gastrulation, expression of the odd-numbered stripes relies on a distal early stripe element in Runt expressing cells. The proximal element can mediate repression by Eve and is able to respond to Runt dependent repression but not activation. The distal element on the other hand can mediate Runt dependent repression and activation but is insensitive to Eve dependent repression. The combination of these two elements in a composite reporter results in appropriate activation and repression in both even and odd-numbered parasegments, a result which can not be explained by the simple addition of the response of these elements alone. One explanation for this result that does not contradict the modularity of CRMs and can account for the structural constraints of these two elements in a composite reporter is a differential interaction of these two enhancer elements with the *slp1* promoter depending on the context of the cell.

The switch model that I propose is dependent on Runt. I have shown that Runx binding sites within the distal element contribute to repression by Runt and Ftz but are not critical in mediating Runt dependent activation as the odd-numbered stripes are expressed in a quintuple Runx site mutant of DESE and this reporter is activated by Runt in combination with Opa. This activation is slightly less then the endogenous gene but there is observable *DESE-lacZm1,2,3,4,5* expression in the head upon ectopic expression of these factors. There are two proposed roles for Runt in activation through DESE. One

involves augmenting Opa-dependent *slp1* activation and the other relies on Runt interacting with DESE to switch enhancer promoter/interactions. In both cases the ability of *DESE-lacZm1,2,3,4,5* to be activated could be a result of the difference between the reporter gene construct and the endogenous locus. In all distal element constructs the proximity of DESE to the promoter varies when compared to this same element relative to the *slp1*transcription start site of the endogenous gene. The lack of 6.5 kb of intervening DNA in this construct does not require an enhancer switch as this reporter brings Opa closer to the promoter and possibly reduces the need for proper DNA binding by Runt suggesting flexibility with this reporter gene for Runx binding sites. Although these sites are not critical for activation of DESE, a DNA-binding defective form of Runt is unable to activate *slp1* as well as *DESE-lacZm1,2,3,4,5* to the extent of wild type Runt indicating that although there may be some flexibility regarding Runx sites within this element, Runt in fact does need to bind DNA for proper activation.

Another observation from this work is that a proximal early stripe element expresses the even-numbered factor X dependent *slp1* stripes and in particular a 260 bp fragment of the central region of PESE is necessary and sufficient for this activation. This not only reveals a minimal even stripe element but also uncovered differential requirements for repression by the homeodomain proteins Eve and Ftz. Along with these results, I have found that large deletions to the left or the right of the central region of PESE do not affect expression of *lacZ* in even-numbered parasegments at stage six but a proximal deletion that removes from 2.5 kb to 1.8 kb of DNA upstream of *slp1* results in the loss of the odd-numbered *slp1* stripes at germband extension (Fig. 3.S3). Although this work concentrates on the establishment of *slp1* expression, regulation at later stages

of development is important to understand the maintenance of segment polarity gene expression and suggests that these two processes are separable. Both DESE and PESE express both sets of *slp1* stripes at germ band extension, which suggests that there is more than one element that is responsible for this expression and suggests different or redundant regulatory mechanism for expression at these later stages. The *gsb* gene provides a precendent for different elements being responsible for early versus late expression where regulation of *gsb* relies on the seperate *gsb*-early and *gsb*-late elements for the establishment and maintenance of regulation, respectively (Li et al., 1993).

Future directions – some unanswered questions

Can Eve block DESE dependent activation?

One explanation for a proper response of *slp1* and the composite reporters to Eve dependent repression put forward in chapter III was that PESE blocks DESE activation in Eve expressing cells (Fig. 3.7). The ability of ectopic Eve to repress only the even-numbered *slp1* stripes suggests that Eve is not blocking activation. However it may be the case that the levels of ectopic Eve obtained in these experiments were insufficient to prevent DESE dependent activation. This alternative explanation can be further investigated using lines in which *UAS-{Eve}* transgenes have been combined that will allow for expression of Eve at higher levels.

Investigate the role of DNA-binding by Runt in activation of a composite reporter

One proposed role for Runt that has emerged from this work is a switch from PESE-driven to DESE-driven expression in Runt expressing cells. This model explains the interaction of these two elements and the differential response to pair-rule transcription factors in a manner that does not upset the modularity of enhancer promoter interactions. An explanation for the observation that a distal element containing mutations in five Runx binding sites being able to mediate Runt dependent activation is that these sites are not critical for activation of DESE alone because it does not have to switch from PESE. Importantly, experiments with the DNA-binding defective Runt[CK] protein indicate that DNA-binding by Runt is important for *slp1* activation. To investigate this further it will be interesting to combine the quintuple Runx site mutant with the proximal element to determine how DNA-binding contributes to expression of a composite *slp1* reporter.

In chapter IV the distal element constructs investigating Runx binding sites used a distal element from 8.7 kb to 7.1 kb 5' of *slp1* as the extended DESE element which adds sequence from -7.1 kb to -6.5 kb did not express well in the context of the pC:*slp1*^{BP}-*lacZ*^{att} vector used for ΦC31 mediated transgenesis (Fig. 3.S4B). There are a couple of differences between this construct and the original distal element construct. One of these differences is that the new constructs used site-specific integration. Also the newly made constructs utilize a linker sequence for easier cloning. Neither of these factors seem to be responsible for the lack of expression as removal of the linker did not restore expression and a minimal element from 8.1 kb to 7.1 kb as well the element used in chapter IV from 8.7 kb to 7.1 kb expressed using the site-specific integration technique (Fig. 3.S4E and C). Another difference between these two constructs is the size of the basal promoter used. The original distal element construct used a promoter from -260 to +121 bp relative to the *slp1* transcription start site (Fig. 3.S4A). When this larger promoter is incorporated into pC:*slp1*^{BP}-*lacZ*^{att} the extended distal element expresses *lacZ* comparable if not better then the original distal element construct (Fig. 3.S4F). Therefore, further experiments with composite reporters should use a *slp1* promoter from -260 to +121 relative to the transcription start site.

It is interesting to consider possible explanations for the effects of the basal promoter on expression. One explanation is that the extra DNA in the extended DESE element contains binding sites for a short range repressor that antagonizes expression from the short basal promoter but not capable of inhibiting transcription when the larger basal promoter is included in the transgene. However, this model does not account for the inactivity of a composite reporter with 2 kb of proximal element sequence separating

this region from the promoter. These results make it clear that much is yet to be learned about the structure and function of cis-regulatory DNA.

Can we detect differential enhancer/promoter interactions?

The Chromosome Conformation Capture technique (3C) allows for investigation of enrichment of interactions between different DNA regions in a particular locus (Dekker et al., 2002). This techinique has been successfully used in uncovering interactions at the β-globin locus control region (LCR) (Bender et al., 2000; Epner et al., 1998; Tolhuis et al., 2002). Considering the relatively small size of the *slp* locus in comparison to β-globin, an enzyme that is able to digest more frequently could be beneficial in further characterizing interactions within the *slp1* locus. I will utilize *ApoI* to investigate the interactions between the *slp1* promoter and flanking cis-regulatory sequences. Along with using *ApoI* for digestion one can utilize the NGT-system to manipulate transcription factors thought to regulate enhancer/promoter interactions. In particular, we can investigate if Runt is responsible for a switch from a PESE/promoter interaction to DESE by ectopically expressing Runt. If the presence of Runt leads to a switch then we would expect to see a decrease in the frequency of PESE association with the promoter and possibly an increase in DESE.

What is Factor(s) X?

The results presented in chapter III suggest that there is possibly more then one factor capable of activating the even-numbered stripes. DESE is able to express all even-numbered stripes except stripe zero and PESE expresses even-numbered stripes zero to twelve. Two candidates for this activation are E2f for PESE and Opa for DESE. To determine if *de2f* and Opa play a role in the segmentation process we have obtained

UAS{*de2f*}; *UAS{dp}* and *UAS{Opa}* transgenes and will investigate the response to ectopic expression of these factors on *slp1* and the proximal and distal element reporter genes. It will also be interesting to investigate the effect of mutating these factors on expression of *slp1* and the different reporter genes.

Cis-regulatory factors controlling enhancer promoter interactions

Composite DESE/PESE constructs provide a powerful tool for investigating the cis-regulatory sequences important for preventing inappropriate DESE-dependent activation in the anterior odd parasegments. I proposed that PESE is the default enhancer potentially due to its proximity to its promoter. It will be interesting to see what happens if DESE is inserted between PESE and the promoter. In addition to investigating the Runx sites in DESE-dependent composite reporter gene expression it will be interesting to examine the expression of a composite deleted for central region 1 of the proximal element as well as a distal element combined with a minimal C1+ region. These experiments will indicate if the sequences responsible for PESE expression have a role in modulating the ability of DESE to interact with the promoter.

References

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

Akbari, O. S., Bae, E., Johnsen, H., Villaluz, A., Wong, D. and Drewell, R. A. (2008). A novel promoter-tethering element regulates enhancer-driven gene expression at the bithorax complex in the Drosophila embryo. *Development* **135**, 123-31.

Aronson, B. D., Fisher, A. L., Blechman, K., Caudy, M. and Gergen, J. P. (1997). Groucho-dependent and -independent repression activities of Runt domain proteins. *Mol Cell Biol* **17**, 5581-7.

Barges, S., Mihaly, J., Galloni, M., Hagstrom, K., Muller, M., Shanower, G., Schedl, P., Gyurkovics, H. and Karch, F. (2000). The Fab-8 boundary defines the distal limit of the bithorax complex iab-7 domain and insulates iab-7 from initiation elements and a PRE in the adjacent iab-8 domain. *Development* **127**, 779-90.

Baumgartner, S., Bopp, D., Burri, M. and Noll, M. (1987). Structure of two genes at the gooseberry locus related to the paired gene and their spatial expression during Drosophila embryogenesis. *Genes Dev* **1**, 1247-67.

Bender, M. A., Bulger, M., Close, J. and Groudine, M. (2000). Beta-globin gene switching and DNase I sensitivity of the endogenous beta-globin locus in mice do not require the locus control region. *Mol Cell* **5**, 387-93.

Benedyk, M. J., Mullen, J. R. and DiNardo, S. (1994). odd-paired: a zinc finger pair-rule protein required for the timely activation of engrailed and wingless in Drosophila embryos. *Genes Dev* **8**, 105-17.

Biggin, M. D. and McGinnis, W. (1997). Regulation of segmentation and segmental identity by Drosophila homeoproteins: the role of DNA binding in functional activity and specificity. *Development* **124**, 4425-33.

Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. and Noll, M. (1986). Conservation of a large protein domain in the segmentation gene paired and in functionally related genes of Drosophila. *Cell* **47**, 1033-40.

Bouchard, M., St-Amand, J. and Cote, S. (2000). Combinatorial activity of pair-rule proteins on the Drosophila gooseberry early enhancer. *Dev Biol* **222**, 135-46.

Butler, B. A., Soong, J. and Gergen, J. P. (1992). The Drosophila segmentation gene runt has an extended cis-regulatory region that is required for vital expression at other stages of development. *Mech Dev* **39**, 17-28.

Cadigan, K. M., Grossniklaus, U. and Gehring, W. J. (1994). Localized expression of sloppy paired protein maintains the polarity of Drosophila parasegments. *Genes Dev* **8**, 899-913.

Canon, J. and Banerjee, U. (2000). Runt and Lozenge function in Drosophila development. *Semin Cell Dev Biol* **11**, 327-36.

Canon, J. and Banerjee, U. (2003). In vivo analysis of a developmental circuit for direct transcriptional activation and repression in the same cell by a Runx protein. *Genes Dev* **17**, 838-43.

Carr, A. and Biggin, M. D. (1999). A comparison of in vivo and in vitro DNA-binding specificities suggests a new model for homeoprotein DNA binding in Drosophila embryos. *EMBO J* **18**, 1598-608.

Chung, J. H., Whiteley, M. and Felsenfeld, G. (1993). A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell* **74**, 505-14.

Cimbora, D. M. and Sakonju, S. (1995). Drosophila midgut morphogenesis requires the function of the segmentation gene odd-paired. *Dev Biol* **169**, 580-95.

Cohen, M. M., Jr. (2001). RUNX genes, neoplasia, and cleidocranial dysplasia. *Am J Med Genet* **104**, 185-8.

Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002). Capturing chromosome conformation. *Science* **295**, 1306-11.

DiNardo, S., Sher, E., Heemskerk-Jongens, J., Kassis, J. A. and O'Farrell, P. H. (1988). Two-tiered regulation of spatially patterned engrailed gene expression during Drosophila embryogenesis. *Nature* **332**, 604-9.

Driever, W. and Nusslein-Volhard, C. (1989). The bicoid protein is a positive regulator of hunchback transcription in the early Drosophila embryo. *Nature* **337**, 138-43.

Driever, W., Thoma, G. and Nusslein-Volhard, C. (1989). Determination of spatial domains of zygotic gene expression in the Drosophila embryo by the affinity of binding sites for the bicoid morphogen. *Nature* **340**, 363-7.

Drissen, R., Palstra, R. J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S. and de Laat, W. (2004). The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev* **18**, 2485-90.

Duffy, J. B. and Gergen, J. P. (1991). The Drosophila segmentation gene runt acts as a position-specific numerator element necessary for the uniform expression of the sex-determining gene Sex-lethal. *Genes Dev* **5**, 2176-87.

Duffy, J. B., Kania, M. A. and Gergen, J. P. (1991). Expression and function of the Drosophila gene runt in early stages of neural development. *Development* **113**, 1223-30.

Duronio, R. J. and O'Farrell, P. H. (1994). Developmental control of a G1-S transcriptional program in Drosophila. *Development* **120**, 1503-15.

Duronio, R. J., O'Farrell, P. H., Xie, J. E., Brook, A. and Dyson, N. (1995). The transcription factor E2F is required for S phase during Drosophila embryogenesis. *Genes Dev* **9**, 1445-55.

Ellis, M. C., O'Neill, E. M. and Rubin, G. M. (1993). Expression of Drosophila glass protein and evidence for negative regulation of its activity in non-neuronal cells by another DNA-binding protein. *Development* **119**, 855-65.

Epner, E., Reik, A., Cimbora, D., Telling, A., Bender, M. A., Fiering, S., Enver, T., Martin, D. I., Kennedy, M., Keller, G. et al. (1998). The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. *Mol Cell* **2**, 447-55.

Flanagan, P. M., Kelleher, R. J., 3rd, Sayre, M. H., Tschochner, H. and Kornberg, R. D. (1991). A mediator required for activation of RNA polymerase II transcription in vitro. *Nature* **350**, 436-8.

Florence, B., Guichet, A., Ephrussi, A. and Laughon, A. (1997). Ftz-F1 is a cofactor in Ftz activation of the Drosophila engrailed gene. *Development* **124**, 839-47.

Fu, W. and Noll, M. (1997). The Pax2 homolog sparkling is required for development of cone and pigment cells in the Drosophila eye. *Genes Dev* **11**, 2066-78.

Fujioka, M., Emi-Sarker, Y., Yusibova, G. L., Goto, T. and Jaynes, J. B. (1999). Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. *Development* **126**, 2527-38.

Gergen, J. P. and Butler, B. A. (1988). Isolation of the Drosophila segmentation gene runt and analysis of its expression during embryogenesis. *Genes Dev* **2**, 1179-93.

Gergen, J. P. and Wieschaus, E. (1986). Dosage requirements for runt in the segmentation of Drosophila embryos. *Cell* **45**, 289-99.

Girling, R., Partridge, J. F., Bandara, L. R., Burden, N., Totty, N. F., Hsuan, J. J. and La Thangue, N. B. (1993). A new component of the transcription factor DRTF1/E2F. *Nature* **365**, 468.

Goto, T., Macdonald, P. and Maniatis, T. (1989). Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. *Cell* **57**, 413-22.

Gottschalk, L. R. and Leiden, J. M. (1990). Identification and functional characterization of the human T-cell receptor beta gene transcriptional enhancer: common nuclear proteins interact with the transcriptional regulatory elements of the T-cell receptor alpha and beta genes. *Mol Cell Biol* **10**, 5486-95.

Grossniklaus, U., Pearson, R. K. and Gehring, W. J. (1992). The Drosophila sloppy paired locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. *Genes Dev* **6**, 1030-51.

Groth, A. C. and Calos, M. P. (2004). Phage integrases: biology and applications. *J Mol Biol* **335**, 667-78.

Groth, A. C., Fish, M., Nusse, R. and Calos, M. P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics* **166**, 1775-82.

Groth, A. C., Olivares, E. C., Thyagarajan, B. and Calos, M. P. (2000). A phage integrase directs efficient site-specific integration in human cells. *Proc Natl Acad Sci U S A* **97**, 5995-6000.

Grummt, I. (1999). Regulation of mammalian ribosomal gene transcription by RNA polymerase I. *Prog Nucleic Acid Res Mol Biol* **62**, 109-54.

Guichet, A., Copeland, J. W., Erdelyi, M., Hlousek, D., Zavorszky, P., Ho, J., Brown, S., Percival-Smith, A., Krause, H. M. and Ephrussi, A. (1997). The nuclear receptor homologue Ftz-F1 and the homeodomain protein Ftz are mutually dependent cofactors. *Nature* **385**, 548-52.

Hader, T., La Rosee, A., Ziebold, U., Busch, M., Taubert, H., Jackle, H. and Rivera-Pomar, R. (1998). Activation of posterior pair-rule stripe expression in response to maternal caudal and zygotic knirps activities. *Mech Dev* **71**, 177-86.

Hagstrom, K., Muller, M. and Schedl, P. (1996). Fab-7 functions as a chromatin domain boundary to ensure proper segment specification by the Drosophila bithorax complex. *Genes Dev* **10**, 3202-15.

Harding, K., Hoey, T., Warrior, R. and Levine, M. (1989). Autoregulatory and gap gene response elements of the even-skipped promoter of Drosophila. *EMBO J* **8**, 1205-12.

Harding, K., Rushlow, C., Doyle, H. J., Hoey, T. and Levine, M. (1986). Cross-regulatory interactions among pair-rule genes in Drosophila. *Science* **233**, 953-9.

Herr, A. J., Jensen, M. B., Dalmay, T. and Baulcombe, D. C. (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**, 118-20.

Howard, K. (1990). The blastoderm prepattern. *Semin Cell Biol* **1**, 161-72.

Ingham, P. W. (1988). The molecular genetics of embryonic pattern formation in Drosophila. *Nature* **335**, 25-34.

Ingham, P. W., Baker, N. E. and Martinez-Arias, A. (1988). Regulation of segment polarity genes in the Drosophila blastoderm by fushi tarazu and even skipped. *Nature* **331**, 73-5.

Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3**, 318-56.

Janssens, H., Hou, S., Jaeger, J., Kim, A. R., Myasnikova, E., Sharp, D. and Reinitz, J. (2006). Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even skipped gene. *Nat Genet* **38**, 1159-65.

Janssens, H., Kosman, D., Vanario-Alonso, C. E., Jaeger, J., Samsonova, M. and Reinitz, J. (2005). A high-throughput method for quantifying gene expression data from early Drosophila embryos. *Dev Genes Evol* **215**, 374-81.

Kagoshima, H., Shigesada, K., Satake, M., Ito, Y., Miyoshi, H., Ohki, M., Pepling, M. and Gergen, P. (1993). The Runt domain identifies a new family of heteromeric transcriptional regulators. *Trends Genet* **9**, 338-41.

Kamachi, Y., Ogawa, E., Asano, M., Ishida, S., Murakami, Y., Satake, M., Ito, Y. and Shigesada, K. (1990). Purification of a mouse nuclear factor that binds to both the A and B cores of the polyomavirus enhancer. *J Virol* **64**, 4808-19.

Kania, M. A., Bonner, A. S., Duffy, J. B. and Gergen, J. P. (1990). The Drosophila segmentation gene runt encodes a novel nuclear regulatory protein that is also expressed in the developing nervous system. *Genes Dev* **4**, 1701-13.

Kassis, J. A. (1990). Spatial and temporal control elements of the Drosophila engrailed gene. *Genes Dev* **4**, 433-43.

Kassis, J. A., Wong, M. L. and O'Farrell, P. H. (1985). Electron microscopic heteroduplex mapping identifies regions of the engrailed locus that are conserved between Drosophila melanogaster and Drosophila virilis. *Mol Cell Biol* **5**, 3600-9.

Kelleher, R. J., 3rd, Flanagan, P. M. and Kornberg, R. D. (1990). A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* **61**, 1209-15.

Kellum, R. and Schedl, P. (1992). A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol Cell Biol* **12**, 2424-31.

Kim, Y. J., Bjorklund, S., Li, Y., Sayre, M. H. and Kornberg, R. D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599-608.

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

Kozlov, K. N., Myasnikova, E., Samsonova, A. A., Surkova, S., Reinitz, J. and Samsonova, M. (2009). GCPReg package for registration of the segmentation gene expression data in Drosophila. *Fly (Austin)* **3**, 151-6.

Kramer, S. G., Jinks, T. M., Schedl, P. and Gergen, J. P. (1999). Direct activation of Sex-lethal transcription by the Drosophila runt protein. *Development* **126**, 191-200.

La Rosee, A., Hader, T., Taubert, H., Rivera-Pomar, R. and Jackle, H. (1997). Mechanism and Bicoid-dependent control of hairy stripe 7 expression in the posterior region of the Drosophila embryo. *EMBO J* **16**, 4403-11.

La Rosee-Borggreve, A., Hader, T., Wainwright, D., Sauer, F. and Jackle, H. (1999). hairy stripe 7 element mediates activation and repression in response to different domains and levels of Kruppel in the Drosophila embryo. *Mech Dev* **89**, 133-40.

Langeland, J. A. and Carroll, S. B. (1993). Conservation of regulatory elements controlling hairy pair-rule stripe formation. *Development* **117**, 585-96.

Lessing, D. and Nusse, R. (1998). Expression of wingless in the Drosophila embryo: a conserved cis-acting element lacking conserved Ci-binding sites is required for patched-mediated repression. *Development* **125**, 1469-76.

Levings, P. P. and Bungert, J. (2002). The human beta-globin locus control region. *Eur J Biochem* **269**, 1589-99.

Lewis, A. F., Stacy, T., Green, W. R., Taddesse-Heath, L., Hartley, J. W. and Speck, N. A. (1999). Core-binding factor influences the disease specificity of Moloney murine leukemia virus. *J Virol* **73**, 5535-47.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565-70.

Li, L. H. and Gergen, J. P. (1999). Differential interactions between Brother proteins and Runt domain proteins in the Drosophila embryo and eye. *Development* **126**, 3313-22.

Li, X., Gutjahr, T. and Noll, M. (1993). Separable regulatory elements mediate the establishment and maintenance of cell states by the Drosophila segment-polarity gene gooseberry. *EMBO J* **12**, 1427-36.

Li, X. Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D. A., Iyer, V. N., Hechmer, A., Simirenko, L., Stapleton, M., Luengo Hendriks, C. L. et al. (2008). Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. *PLoS Biol* **6**, e27.

Lin, Q., Chen, Q., Lin, L., Smith, S. and Zhou, J. (2007). Promoter targeting sequence mediates enhancer interference in the Drosophila embryo. *Proc Natl Acad Sci U S A* **104**, 3237-42.

Lohr, U. and Pick, L. (2005). Cofactor-interaction motifs and the cooption of a homeotic Hox protein into the segmentation pathway of Drosophila melanogaster. *Curr Biol* **15**, 643-9.

Manoukian, A. S. and Krause, H. M. (1993). Control of segmental asymmetry in Drosophila embryos. *Development* **118**, 785-96.

Matsui, T., Segall, J., Weil, P. A. and Roeder, R. G. (1980). Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J Biol Chem* **255**, 11992-6.

Myasnikova, E., Samsonova, A., Kozlov, K., Samsonova, M. and Reinitz, J. (2001). Registration of the expression patterns of Drosophila segmentation genes by two independent methods. *Bioinformatics* **17**, 3-12.

Myasnikova, E., Samsonova, M., Kosman, D. and Reinitz, J. (2005). Removal of background signal from in situ data on the expression of segmentation genes in Drosophila. *Dev Genes Evol* **215**, 320-6.

Nimmo, R. and Woollard, A. (2008). Worming out the biology of Runx. *Dev Biol* **313**, 492-500.

Noordermeer, D. and de Laat, W. (2008). Joining the loops: beta-globin gene regulation. *IUBMB Life* **60**, 824-33.

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

Nusslein-Volhard, C., Wieschaus, E. and Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster.I. Zygotic loci on the second chromosome. *Roux's Archive of Developmental Biology* **193**, 267-282.

Ogawa, E., Inuzuka, M., Maruyama, M., Satake, M., Naito-Fujimoto, M., Ito, Y. and Shigesada, K. (1993a). Molecular cloning and characterization of PEBP2 beta, the heterodimeric partner of a novel Drosophila runt-related DNA binding protein PEBP2 alpha. *Virology* **194**, 314-31.

Ogawa, E., Maruyama, M., Kagoshima, H., Inuzuka, M., Lu, J., Satake, M., Shigesada, K. and Ito, Y. (1993b). PEBP2/PEA2 represents a family of transcription factors homologous to the products of the Drosophila runt gene and the human AML1 gene. *Proc Natl Acad Sci U S A* **90**, 6859-63.

Redondo, J. M., Pfohl, J. L. and Krangel, M. S. (1991). Identification of an essential site for transcriptional activation within the human T-cell receptor delta enhancer. *Mol Cell Biol* **11**, 5671-80.

Roeder, R. G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* **21**, 327-35.

Royzman, I., Whittaker, A. J. and Orr-Weaver, T. L. (1997). Mutations in Drosophila DP and E2F distinguish G1-S progression from an associated transcriptional program. *Genes Dev* **11**, 1999-2011.

Rubin, G. M., Hong, L., Brokstein, P., Evans-Holm, M., Frise, E., Stapleton, M. and Harvey, D. A. (2000). A Drosophila complementary DNA resource. *Science* **287**, 2222-4.

Rubin, G. M. and Spradling, A. C. (1982). Genetic transformation of Drosophila with transposable element vectors. *Science* **218**, 348-53.

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

Satake, M., Ibaraki, T., and Ito, Y. (1988). Modulation of polyomavirus enhancer binding protein by Ha-ras oncogene. *Oncogene*, 69-78.

Seum, C., Spierer, A., Pauli, D., Szidonya, J., Reuter, G. and Spierer, P. (1996). Position-effect variegation in Drosophila depends on dose of the gene encoding the E2F transcriptional activator and cell cycle regulator. *Development* **122**, 1949-56.

Small, S., Blair, A. and Levine, M. (1992). Regulation of even-skipped stripe 2 in the Drosophila embryo. *EMBO J* **11**, 4047-57.

Small, S., Blair, A. and Levine, M. (1996). Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo. *Dev Biol* **175**, 314-24.

Stanojevic, D., Small, S. and Levine, M. (1991). Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. *Science* **254**, 1385-7.

Stein, G. S., Lian, J. B., van Wijnen, A. J., Stein, J. L., Montecino, M., Javed, A., Zaidi, S. K., Young, D. W., Choi, J. Y. and Pockwinse, S. M. (2004). Runx2 control of organization, assembly and activity of the regulatory machinery for skeletal gene expression. *Oncogene* **23**, 4315-29.

Swantek, D. and Gergen, J. P. (2004). Ftz modulates Runt-dependent activation and repression of segment-polarity gene transcription. *Development* **131**, 2281-90.

Sweetser, D., Nonet, M. and Young, R. A. (1987). Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proc Natl Acad Sci U S A* **84**, 1192-6.

Tautz, D. (1988). Regulation of the Drosophila segmentation gene hunchback by two maternal morphogenetic centres. *Nature* **332**, 281-4.

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* **10**, 1453-65.

Tracey, W. D., Jr., Ning, X., Klingler, M., Kramer, S. G. and Gergen, J. P. (2000). Quantitative analysis of gene function in the Drosophila embryo. *Genetics* **154**, 273-84.

Tsai, C. and Gergen, J. P. (1994). Gap gene properties of the pair-rule gene runt during Drosophila segmentation. *Development* **120**, 1671-83.

Tsai, C. and Gergen, P. (1995). Pair-rule expression of the Drosophila fushi tarazu gene: a nuclear receptor response element mediates the opposing regulatory effects of runt and hairy. *Development* **121**, 453-62.

Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S., Millburn, G., Osumi-Sutherland, D., Schroeder, A., Seal, R. et al. (2009). FlyBase: enhancing Drosophila Gene Ontology annotations. *Nucleic Acids Res* **37**, D555-9.

Vakoc, C. R., Letting, D. L., Gheldof, N., Sawado, T., Bender, M. A., Groudine, M., Weiss, M. J., Dekker, J. and Blobel, G. A. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell* **17**, 453-62.

Walter, J. and Biggin, M. D. (1996). DNA binding specificity of two homeodomain proteins in vitro and in Drosophila embryos. *Proc Natl Acad Sci U S A* **93**, 2680-5.

Walter, J., Dever, C. A. and Biggin, M. D. (1994). Two homeo domain proteins bind with similar specificity to a wide range of DNA sites in Drosophila embryos. *Genes Dev* **8**, 1678-92.

Wang, S., Wang, Q., Crute, B. E., Melnikova, I. N., Keller, S. R. and Speck, N. A. (1993). Cloning and characterization of subunits of the T-cell receptor and murine leukemia virus enhancer core-binding factor. *Mol Cell Biol* **13**, 3324-39.

Wang, S. W. and Speck, N. A. (1992). Purification of core-binding factor, a protein that binds the conserved core site in murine leukemia virus enhancers. *Mol Cell Biol* **12**, 89-102.

Wang, X., Lee, C., Gilmour, D. S. and Gergen, J. P. (2007). Transcription elongation controls cell fate specification in the Drosophila embryo. *Genes Dev* **21**, 1031-6.

Wheeler, J. C., Shigesada, K., Gergen, J. P. and Ito, Y. (2000). Mechanisms of transcriptional regulation by Runt domain proteins. *Semin Cell Dev Biol* **11**, 369-75.

Wheeler, J. C., VanderZwan, C., Xu, X., Swantek, D., Tracey, W. D. and Gergen, J. P. (2002). Distinct in vivo requirements for establishment versus maintenance of transcriptional repression. *Nat Genet* **32**, 206-10.

Wierzbicki, A. T., Ream, T. S., Haag, J. R. and Pikaard, C. S. (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet* **41**, 630-4.
Willis, I. M. (1993). RNA polymerase III. Genes, factors and transcriptional specificity. *Eur J Biochem* **212**, 1-11.

Yu, Y., Li, W., Su, K., Yussa, M., Han, W., Perrimon, N. and Pick, L. (1997). The nuclear hormone receptor Ftz-F1 is a cofactor for the Drosophila homeodomain protein Ftz. *Nature* **385**, 552-5.

Zallen, J. A. and Wieschaus, E. (2004). Patterned gene expression directs bipolar planar polarity in Drosophila. *Dev Cell* **6**, 343-55.

Zhou, J., Barolo, S., Szymanski, P. and Levine, M. (1996). The Fab-7 element of the bithorax complex attenuates enhancer-promoter interactions in the Drosophila embryo. *Genes Dev* **10**, 3195-201.