



LIBRARY Michigan State University

This is to certify that the dissertation entitled

THE FUNCTION AND DESIGN OF CIS-ACTING ENHANCER ELEMENTS REGULATED BY SHORT-RANGE TRANSCRIPTIONAL REPRESSORS: GRAMMAR STUDIES FROM Drosophila melanogaster

presented by

Meghana Manohar Kulkarni

has been accepted towards fulfillment of the requirements for the

Doctoral degree in Genetics Major Professor's Signature 2.10.03

Date

MSU is an Affirmative Action/Equal Opportunity Institution

\_\_\_\_\_

# PLACE IN RETURN BOX to remove this checkout from your record. TO AVOID FINES return on or before date due. MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE
		6/01 c:/CIRC/DateDue.p65-p.15

-----

## THE FUNCTION AND DESIGN OF CIS-ACTING ENHANCER ELEMENTS REGULATED BY SHORT-RANGE TRANSCRIPTIONAL REPRESSORS: GRAMMAR STUDIES FROM Drosophila melanogaster

By

Meghana Manohar Kulkarni

# A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

# DOCTOR OF PHILOSOPHY

**Program in Genetics** 

2003

#### ABSTRACT

#### THE FUNCTION AND DESIGN OF CIS-ACTING ENHANCER ELEMENTS REGULATED BY SHORT-RANGE TRANSCRIPTIONAL REPRESSORS: GRAMMAR STUDIES FROM Drosophila melanogaster

By

#### Meghana Manohar Kulkarni

Given that the DNA of (most) all cells in an animal is identical, how do different cells acquire the unique morphologies and functional properties to create the diverse tissues and organs in multicellular organisms? With better understanding of the nature of genes and the process of gene regulation, the central role of transcriptional regulation in directing development is becoming evident. These exquisitely orchestrated gene regulatory programs are encoded in the DNA sequence of *cis*-regulatory control elements, whose features are only now becoming apparent. Morphological diversity has its origins in reshaped developmental processes, and these changes often reflect alterations in genetic regulatory programs and in particular the transcriptional cisregulatory regions. Thus, evolution of diversity is directly related to the evolution of these *cis*-regulatory regions or enhancers. To understand this aspect of evolution, we must first understand the structure-function relationships that apply to developmental cisregulatory enhancers. The studies described herein highlight general principles of the design and function of *cis*-acting enhancer elements that are involved in the control of complex patterns of gene expression and should in turn facilitate our understanding of the regulatory logic underlying morphological complexity and diversity.

*Cis*-regulatory enhancer elements are thought to function as information processing centers that integrate the negative and positive transcriptional inputs incident

upon them, in a computer-like fashion, resolving the multiple inputs into a single output that instructs the basal transcriptional machinery to either turn the linked gene on or off. In this type of computational model for gene regulation by an enhancer, the basal transcriptional machinery plays a passive role by simply responding to signals generated by the enhancer. In contrast, the 'Information Display' or 'Billboard' model for enhancer function proposed in this study, demonstrates that instead of integrating multiple inputs, the enhancer is capable of simultaneously displaying contrasting information that is interpreted by multiple successive or simultaneous interactions with the basal promoter. Thus, the basal transcriptional machinery plays an active role in processing regulatory information presented by the *cis*-element.

Key structural features of the internal organization of an enhancer control its ability to convert positional information specified by the transcription factors that bind it into differential gene activity. Although many functional analyses have indicated the presence of functional constraints, no work has been carried out to systematically define the spatial constraints between transcription factor binding sites within enhancers that are required for their proper function. The experiments described here take a first step towards systematically deducing the 'grammar' rules for an important class of transcriptional regulators, the short-range repressors. These rules define the internal organization of a functional module regulated by them.

These studies described here have important ramifications for the biochemistry and evolution of *cis*-regulatory elements, and should facilitate the development of more sophisticated computational algorithms for the identification of *cis*-regulatory elements and for the interpretation of their biological function. To my Dad (Mr. Manohar Kulkarni) and Mom (Mrs. Swati Kulkarni), without whose wisdom, love, and encouragement, this thesis would never have seen the light of day.

#### ACKNOWLEDGEMENTS

I want to take this opportunity to sincerely thank everyone who has at one time or another helped nurture this project from inception to its final birth. I am indebted to David Arnosti, my thesis advisor and mentor, for unreservedly believing in me and for helping me explore new scientific perspectives. In his unique way, he has enriched, enthused, and inspired me. I would like to extend a particularly deep thanks to my thesis guidance committee: Dr. Steven J. Triezenberg, Dr. Susan E. Conrad, and Dr. Bill Henry for their many substantive suggestions and for helping me remain focused through the Ph.D. program. I am especially grateful to the members of the Arnosti laboratory, past (Yifan Mao, Bethany Strunk, Anumeha Kumar, Jae-Ryeon Ryu, Jelani Thomas, and Montserrat Sutrias-Grau) and present (Scott Keller, Carlos Martinez, Paolo Struffi, Martin Buckley, Priya Mani, and Sandhya Payankaulam) for hours of scientific discussion, criticisms, and emotional support. I also extend my appreciation to other members of the Arnosti lab (Eduardo Fernandez-Villatoro, Chu Yin, and Hin) without whose technical support this work would not have been possible. I am especially grateful to my colleagues from various departments (Craig Hinkley, Heather Hirsch, and Kirsten Fertuck) for sharing their expertise and for providing an encouraging word. I am also grateful to the Genetics Program secretary, Jeannine Lee, and members of the Department of Biochemistry and Molecular Biology, Julie Oesterle, Carol McCutcheon and Pappan for all their help. My deepest thanks go to my family: my parents, brothers and sisters-in-law, without whom I would not have come this far; to my husband, Shailesh, who has so infinitely enriched my life.

# **TABLE OF CONTENTS**

List of Figures	vii	
Key to Abbreviations	ix	
CHAPTER I: Introduction	1	
CHAPTER II:		
Information Display by Transcriptional Enhancers	55	
Abstract	56	
Introduction	56	
Materials and Methods	60	
Results	62	
Discussion	70	
References	79	
<b>CHAPTER III:</b> Operating Principle of Short-range transcriptional repressors in melanogaster	Drosophila	
Introduction	83	
Materials and Methods	90	
Results	97	
Discussion	120	
References	131	
CHAPTER IV: Conclusions and Future Directions	139	
<b>APPENDIX A:</b> Composite element representing two different information states exhibits independent activity characteristic of enhancers	orientation- 177	
APPENDIX B: CtBP-dependent and CtBP-independent activities contribute quantitatively repressor function	y to Knirps 183	
<b>APPENDIX C:</b> Relative potencies of short-range transcriptional repressors	190	
<b>APPENDIX D:</b> Analysis of other <i>Drosophila</i> transcriptional repressors	195	

# LIST OF FIGURES

	Figure I-1: Sochastic versus rheostat function of enhancer elements	6
	Figure I-2: Anterior-Posterior patterning in Drosophila is initiated by a cas	scade of
	transcription factors	11
	Figure I-1: Cis-regulatory region of the pair-rule gene even-skipped	15
	Figure I-4: Model for the regulation of eve striped 2	18
	Figure I-5: Transcriptional repression in Drosophila	22
	Figure I-6: Expression pattern of the short-range repressors Giant and Knirps	25
	Figure I-7: Model for the regulation of the <i>eve</i> stripes $3 + 7$ and $4 + 6$	31
	Figure II-1: Simultaneous repression and activation from a compact regulatory	
	element	67
	Figure II-2: Compact regulatory element displays enhancer-like properties of	distance
	and orientation independence	71
	Figure II-3: Conversion of a multiple state element to a binary on or off switch	74
	Figure II-4: Enhanceosome versus Information Display enhancer models	78
	Figure III-1: Context dependence of short-range repression	102
	Figure III-2: Stoichometry between the number of activators and repressors in	fluences
	repression effectiveness	105
	Figure III-3: Effectiveness of repression correlates with the affinity of activator	binding
	sites	108
	Figure III-4: Repression not dependent on the nature of the activation domain	112
	Figure III-5: The arrangement or distribution of short-range repressor binary	sites is
(	critical in dictating repression effectiveness	115

-

Figure III-6: The ability to repress depends on enhancer configurations rather	than the
nature of the activation domain	118
Figure III-7: Repressor can be distinguished by their differential ability to a	epress in
different enhancer configurations	121
Figure IV-1: Evolutionary dynamics of transcriptional factor binding sites in a c	conserved
cis-regulatory element	155
Figure IV-2: Data and model for gap gene circuit	168
Figure A-1: Compact regulatory element displays enhancer-like property of o	rientation
independence	182
Figure B-1: Pattern of endogenous eve expression in embryos expressing for	ull length
Knirps 1-429 and CtBP-independent region of Knirps 1-330	189
Figure C-1: Relative potencies of the short-range repressors Giant, Knirps, and	
Krüppel	196
Figure D-1: The activity of the GAL4 activators from three high affinity is not	repressed
by Snail	201
Figure D-2: The activity of the GAL4 activators from three high affinity is not	repressed
by Engrailed (en), Runt (run), or Even-skipped (eve)	205
Figure D-3: The activity of transcription factors Brinker (brk) and Suppressor of	f Hairless
(Su(H)) on a reporter with three high affinity GAL4 sites	207
Figure D-4: The activity of the transcription factors Sloppy-paired (Slp1), 7	Tramtrack
(TTK), and Cubitus interruptus (Ci) on a reporter with three high affinity	
GAL4 sites	209

# **KEY TO ABBREVIATIONS**

A-P:	Anterior – Posterior
bp:	base pair
brk:	Brinker
ChIP:	Chromatin Immunoprecipitation
Ci:	Cubitus interruptus
CtBP:	C-terminal Binding Protein
dCtBP:	Drosophila C-terminal Binding Protein
dl:	Dorsal
DPE:	Downstream Promoter Element
GAL4AD:	GAL4 Activation Domain
GFP:	Green Fluorescent Protein
gt:	Giant
HDAC:	Histone Deacetylase Complex
Hh:	Hedgehog
hkb:	Huckebein
HMGI(Y):	High Mobility Group Protein I (Y)
hTBP:	Human TATA Binding Protein
IFN-β:	Interferon - Beta
IRF:	Interferon Regulatory Factor
JAK:	Janus Kinase
kb:	Kilobase
kr:	Krüppel
kni:	Knirps
NAD:	Nicotineamide Adenine Dinucleotide
NEE:	Neurectoderm Enhancer Element
NFkB:	Nuclear Factor Kappa B
PCR:	Polymerase Chain Reaction
PWM:	Position Weight Matrix
RTK:	Receptor Tyrosine Kinase
SAGE:	Serial Analysis of Gene Expression
Slp1:	Sloppy – paired 1
sna:	Snail
Su(H):	Suppressor of Hairless
Tor:	Torso
TTK:	Tramtrack
twi:	Twist
UAS:	Upstream Activating Sequence
VRR:	Ventral Repression Region
yp1, yp2:	Yolk Protein 1, 2

#### Chapter I

#### **INTRODUCTION**

Understanding the basis of the complexity and diverse morphologies of multicellular organisms remains a formidable challenge. According to the basic dogma of molecular biology, DNA is the ultimate repository of biological complexity. A clear revelation of the post-genome era is that organismal complexity does not simply correlate with gene number (Adams et al., 2000). Morphological complexity is clearly not the product of single genes. Rather, regulatory information encoded in the genome contains the key to the differences responsible for morphological complexity and diversity. An increasingly enormous bank of experimental data now confirms the *a priori* assumption that many hundreds, and often thousands, of genes must be differentially expressed as a function of time and space, in order to create any given tissue, body part, or multicellular structure. The carefully choreographed progression of temporal- and domain-specific gene expression is controlled by *cis*-regulatory elements, which constitute a fraction of that part of an organism's genome that does not encode proteins. Cis-regulatory elements are DNA sequences in the vicinity of each gene that contain sequence specific motifs bound by regulatory proteins or transcription factors that affect the expression of that gene. By binding multiple distinct regulatory proteins, the presence of which individually may depend on signaling events, cell cycle activity, temporal state, lineage, or spatial position, cis-regulatory elements integrate temporal and positional information to direct complex patterns of gene expression during development. Thus, the architecture of the cis-regulatory apparatus constitutes a discretely organized DNA map, which represents in

physical terms the different phases of gene expression that are to be installed throughout the life cycle, for every gene. Genomic changes that alter *cis*-regulatory element architecture have the power to create new developmental processes and thus different morphological outcomes (Belting et al., 1998; Shashikant et al., 1998). Thus, transcriptional regulatory regions themselves appear to be a driving force behind evolutionary changes that underlie morphological diversity. It follows that analysis of genomic *cis*-regulatory elements in terms of their structure and functional organization holds the key to understanding how genomes encode the properties of organisms. Despite the importance of the *cis*-regulatory apparatus in gene regulation, our ability to identify and uncover the meaning implicit in its DNA sequence is extremely limited.

In this work, I describe studies that shed light on how transcriptional regulatory information encoded within the *cis*-regulatory enhancer sequence is interpreted by the cellular machinery; what (if any) are the rules that govern the structural organization of *cis*-acting enhancer elements; how these architectural rules influence the biological output of the elements; and how *cis*-regulatory 'grammar' provides insights into the dynamics of developmental evolution.

#### THE REGULATORY APPARATUS ENCODED IN THE DNA

There are several classes of *cis*-regulatory DNAs that are involved in the control of gene expression. This includes (*a*) the basal promoter near the transcriptional initiation site. There are at least three different sequences (TATA box, initiator element [Inr], and the downstream promoter element [DPE]) that might be found in the basal promoter that

serve as binding sites for the basal transcriptional machinery and RNA polymerase, and are involved in the initiation of transcription (Arnosti, 2002; Arnosti, 2003); (*b*) Enhancer elements that contain binding sites for sequence specific transcriptional activators and repressors, and regulate levels of gene activity in a distance- and orientation-independent manner (Banerji et al., 1981; Blackwood and Kadonaga, 1998); (*c*) Silencer elements that suppress gene expression over long distances mainly through the creation of higher order chromatin structure (Ogbourne and Antalis, 1998); and (*d*) Boundary elements or insulators, which prevent enhancers/silencers associated with one gene from inappropriately regulating neighboring genes (Arnosti, 2002; Arnosti, 2003). These regulatory elements, enhancers, silencers, and insulators are scattered over distances of up to 100 kbp in metazoans. This elaborate organization of the regulatory DNAs permits the detailed control of gene expression. A defining feature of gene regulation in higher eukaryotes is the use of multiple enhancers, silencers, and promoters to control the activities of a protein-coding unit.

#### **ENHANCERS**

Enhancers represent the most thoroughly analyzed type of *cis*-regulatory DNA controlling gene expression. Enhancers were initially identified and characterized for promoters of mammalian viruses using cell culture assays (Banerji et al., 1981). They were shown to be DNA sequences that increased the expression of a linked gene in an orientation- and distance-independent manner (Banerji et al., 1981; Blackwood and Kadonaga, 1998). The characterization of enhancers in transgenic worms, flies, sea

urchins, ascidians, fish, frogs, chicks, and mice suggests that a typical enhancer is a discrete element of less than 1 kbp and is composed of multiple binding sites for different regulatory proteins, both sequence specific transcriptional activators and repressors (Davidson, 2001).

#### Stochastic versus Rheostat models for enhancer function

The analysis of transcriptional activation by enhancers on a single-cell basis indicates that the functional result of enhancer action can be to increase the probability that a gene will be activated in any particular cell, without influencing the rate of transcriptional initiation (Walters et al., 1996; Weintraub, 1988). This on or off (Stochastic) response is in contrast to the Rheostat model in which enhancers increase the rate of transcriptional initiation (Magis et al., 1996). Both of these effects are observed with reporter genes in the Drosophila embryo (Figure I-1). A weak activator drives lacZ reporter gene expression in a stochastic manner in which some nuclei fail to show any expression at all. In contrast, a strong activator drives expression in a more uniform Pattern, with overall higher levels of expression.

# The e Nature of Enhancer activity

Enhancers have been suggested to function mainly through two distinct pathways: through remodeling of chromatin (Blackwood and Kadonaga, 1998), thereby facilitating hibiting the binding of transcriptional machinery, and through direct interactions the general transcriptional machinery (Gaudreau et al., 1997). At least two factors

# Figure I-1: Stochastic versus rheostat function of enhancer elements.

(A) In the on or off, stochastic model, genes are either in the "on" state or the "off" state. Transcriptional enhancers act to increase the probability that their cognate genes will be transcribed, but do not affect the levels of transcription. The fraction of cells in which the gene is activated may reflect enhancer strength, which is a function of the type and number of its associated transcription factors. In the rheostat model, genes are uniformly activated by enhancers, and the amount of transcription is proportional to the strength of the enhancer (figure adapted from Blackwood and Kadonaga, 1998).

(B, C) A *lacZ* reporter gene containing five high affinity Gal4 binding sites is activated by Gal4-Sp1 (B) or by the activation domain of the Gal4 activator (C). In the case of the Gal4-Sp1 activator the expression of the reporter gene is switched on in some but not all nuclei containing the activator giving a punctate pattern indicating a stochastic on or off effect. In case of the Gal4-AD activator, the pattern of expression is not only more uniform but is also more intense indicating a rheostat type of effect. *LacZ* expression is visualized by *in situ* hybridization using digoxigenin labeled antisense *lacZ* mRNA probes. Embryos are oriented anterior to left and dorsal up; lateral (B) and ventrolateral view(C).

Images in this dissertation are presented in color.

Figure I-1: Stochastic versus rheostat function of enhancer elements.



contribute to the activity of enhancer elements: First, the sequence specific positively and negatively acting proteins that bind directly to sequences within the element and second, the cofactors (coactivators, corepressors, and chromatin modifying complexes) recruited by the enhancer bound factors.

Identification of enhancers was achieved originally through genetic approaches such as the classic mutations that altered the expression of the HOX genes in the *bithorax* complex (Lewis, 1998). The focus on the identification of regulatory elements for individual genes has included several other experimental approaches: the generation of deletion constructs to determine the minimal sequences necessary for transcription in cell-culture-based systems; DNase I hypersensitivity studies to identify sequences potentially available for transcription factor binding; and in vitro approaches, such as DNA footprinting and gel shifts, to determine sequences that bind various regulatory proteins. Screens to identify cis-regulatory elements have also been carried out in transgenic mice, albeit in an extremely laborious and low-throughput manner. In addition, a limited number of large-scale promoter and enhancer trapping studies have been done (A soh et al., 1994; Durick et al., 1999; Fukushige and Ikeda, 1996). Most of these gene regulatory studies have consisted of largely unguided searches of genomic sequence for those with gene regulatory properties. Much of our knowledge of *cis*-regulatory enhancer elements that function in the context of developmental specification at present derives from Drosophila, where sophisticated tools for characterizing the features of enhancer ele ments are available.

# THE FUNCTION OF CIS-REGULATORY ENHANCER ELEMENTS DERIVES FROM ITS STRUCTURAL ORGANIZATION

#### Modularity

A general feature of *cis*-regulatory architecture in higher eukaryotes is that the diverse phases of expression of developmental genes are frequently mediated by multiple cis-regulatory elements, referred to as regulatory 'modules' present in the DNA flanking the gene or in its introns. Cis-regulatory modules are discrete DNA sequences of a few hundred base pairs (100-1000 bp) that contain multiple binding sites for multiple distinct transcription factors. Each module drives a discrete portion of the overall expression profile of the linked gene. Modules can be moved from their native context and still recapitulate a portion of the native expression pattern, thus acting as autonomous units (Davidson, 2001). Thus, a gene may have many modules in its *cis*-regulatory region. Modularity allows each element to function independently of the other and therefore, each module signals the basal promoter independently. At a given moment, in a given nucleus, one module in a complex regulatory region may be inactive or silent, while at the same time another adjacent module may be actively firing the promoter. In this study (Chapter 2), I demonstrated that a similar kind of autonomy occurs within the tight confines of a single module. Modular regulatory organization provides the organism with the services of a given gene in multiple developmental contexts.

The modularity of *cis*-regulatory regions is puzzling, because modular structure be **argued** to be less optimal (precise) than nonmodular (elements that are interconnected and not autonomous) structures. Modules convey an advantage in

situations where the developmental specifications change from time to time, as in the evolution of diverse species. New modules can be easily constructed from existing, welltested modules and can be readily configured to adapt to new conditions. Changes in regulatory sequences within individual elements or modules may subtly affect the level, timing, or spatial pattern of gene expression, and may do so very selectively in terms of the tissues and stages of development involved. These quantitative, temporal, and spatial changes in the deployment of regulatory genes may affect the level, timing, and spatial expression of other developmental genes. A nonmodular element, in which every component is optimally linked to every other component, is effectively frozen and cannot evolve to meet new optimization conditions. Thus, modularity in *cis*-regulatory systems is **not** only essential to their developmental function, but also facilitates the evolution of developmental processes because individual elements can evolve independently (Carroll, S. B. 2001).

One of the most extensively studied modular *cis*-regulatory systems is that controlling the expression of the pair-rule gene even-skipped' during the segmentation **process** early during the embryogenesis of the fruitfly *Drosophila melanogaster*. Anterior-posterior patterning in *Drosophila* is initiated by a cascade of transcription factors. which culminates in the establishment of segment boundaries (Figure I-2). Positional information along the A-P axis of the syncitial blastoderm is encoded in a succession of different ways during development. A major determinant of anterior Positional information is the roughly exponential gradient of Bicoid protein imposed by

The the use of new Drosophila genetic nomenclature throughout this document: all gene names are

ital ste the use of new Drosophila genetic nomenciature infoughout this document. In our of the start with an upper case letters. Protein names are in regular font and start with an upper case

# Figure I-2: Anterior-Posterior patterning in *Drosophila* is initiated by a cascade of transcription factors.

Initially, maternally derived morphogenetic gradients of Bicoid, Hunchback and Caudal transcription factors provide gene regulatory inputs to activate the expression of the zygotic gap genes giant, knirps, krüppel, hunchback and tailless. Each of these zygotic genes is expressed in one or two broad domains and together with the maternally derived transcription factors regulate the expression of the pair-rule genes including *even-skipped*. The pair-rule genes together with the gap genes then further regulate the expression of the segment polarity genes that define the borders of the future segmental compartments of the adult body plan.

Figure I-2: Anterior-Posterior patterning in *Drosophila* is initiated by a cascade of transcription factors.



the mother fly, along with anterior maternal *hunchback* expression. Posterior positional information is provided by a similar gradient of maternal Caudal protein. These maternal gradients provide gene regulation inputs that directly or indirectly regulate the expression of the gap class of zygotic segmentation genes (*krüppel, knirps, giant, hunchback* and *tailless*), each of which is expressed in a simple pattern made up of one or two broad expression domains. Together with the maternal morphogens themselves, the gap genes act as more localized gradients to establish the A-P body plan. They directly or indirectly *control* the expression of the pair-rule class of segmentation genes, each of which is *expressed* in a series of seven transverse stripes that lie along the A-P axis and precisely *define* the segmental compartments of the embryo, which in turn correspond to the future compartments of the body of the adult organism. The primary pair-rule genes, including *even-skipped*, encode transcription factors. The location of each stripe of expression depends on the *cis*-regulatory system controlling these genes (Rivera-Pomar and Jackle, 1996).

eve encodes a homeodomain protein that is expressed in a series of seven transverse stripes (each of which is 4-5 nuclei in width) along the length of the embryo and plays a key role in the establishment of the metameric body plan (Frasch et al., 1987; Frasch and Levine, 1987; Harding et al., 1989; Harding et al., 1986; Macdonald et al., 1986; Macdonald and Struhl, 1986). The transcriptional regulation of *eve* in a sevenstripe pattern is complex, in that the *eve cis*-regulatory region is modular and contains arate stripe enhancer elements or modules that control the expression of individual pes or pairs of stripes (Goto et al., 1989; Harding et al., 1989; Small et al., 1992). Subsequent comprehensive analysis of the *eve* locus identified within a 16 kbp locus five separable elements 5' and 3' of the transcription unit, that together create the seven stripe pattern of *eve* (Fujioka et al., 1999) (Figure I-3). Stripe 2, stripe 3+7, stripe 4+6, stripe 1 and stripe 5 enhancer elements drive expression of *eve* in the corresponding stripes and corntain all necessary transcriptional information for their correct spatial distribution in the embryo. Each stripe enhancer element or module contains target sites for positively acting transcription factors which cause it to be expressed; and for negatively acting transcription factors which set the anterior and posterior boundaries within which expression is allowed. Thus, the general strategy for pair-rule gene expression is the initial widespread activation under the influence of generally distributed activators, followed shortly afterwards by repression in the interstripe regions (Frasch et al., 1987; Frasch and Levine, 1987).

#### Short-range transcriptional repression

A combination of genetic analyses (Frasch et al., 1987; Frasch and Levine, 1987), DNA binding experiments (Stanojevic et al., 1989), transient cotransfection assays (Small et al., 1991; Small and Levine, 1991), and expression assays in transgenic *Drosophila* embryos (Small et al., 1992; Stanojevic et al., 1991) have provided evidence for the following model of *eve* stripe 2 regulation. The broad Bicoid gradient emanating from anterior regions of the precellular embryo induces a steeper pattern of *hunchback* expression. Bicoid and Hunchback function synergistically to activate the *eve* stripe 2 embrancer in the anterior half of the embryo. The stripe borders are formed by two

## Figure I-3: Cis-regulatory region of the pair-rule gene even-skipped.

even-skipped is a pair-rule gene encoding a homeodomain transcription factor that is expressed in seven transverse stripes across the anterior-posterior axis of the blastoderm embryo. eve expression is visualized in the embryo above, by *in situ* hybridization using digoxigenin labeled antisense eve mRNA probe. Comprehensive analyses of the eve locus have identified within a 16 kb region, separable regulatory modules called stripe enhancer elements 5' and 3' of the transcription unit. Each of the stripe enhancer elements (stripe 3 + 7, stripe 2, stripe 4 + 6, stripe 1, and stripe 5) drives a discrete portion of the overall expression profile of the gene. Figure adapted from Fujioka et al., 1999; Sackerson et al., 1999. Figure I-3: Cis-regulatory region of the pair-rule gene even-skipped.



Fujioka et al. (1999) Development 12:2527 and Sackerson et al. (1999) Dev. Biol.211: 39

ı.

repressor gradients; an anterior Giant gradient (Capovilla et al., 1992; Kraut and Levine, 1991a; Kraut and Levine, 1991b) and a posterior Krüppel gradient (Gaul et al., 1987). The stripe 2 enhancer (Figure I-4) is an approximately 500 bp sequence of DNA located **bet**ween -1.6 kb and -1.1 kb upstream of the *eve* transcription start site and contains a **total** of eighteen binding sites, including eight activator sites (five for Bicoid and three for **Hunchback**) and ten repressor binding sites (three for Giant, six for Krüppel and one for **S10** ppy-paired) (Andrioli et al., 2002; Berman et al., 2002; Ludwig et al., 2000; Ludwig **and** Kreitman, 1995; Ludwig et al., 1998; Small et al., 1992). Mutations in these factor **binding** sites alter the normal expression pattern of a stripe 2-*lacZ* fusion gene in transgenic embryos; the abnormal patterns often mimic those seen when the wildtype stripe 2 enhancer is expressed in segmentation mutants (Small et al., 1992; Stanojevic et al., 1991).

The model for eve 2 regulation is incomplete in at least two respects. First, the combined activities of Bicoid and Hunchback do not seem to be sufficient for activation at the position of eve stripe 2. For example, reporter genes containing three high affinity Bicoid sites and three Hunchback sites cannot respond to the low levels of these proteins at the position of eve stripe 2 (Simpson-Brose et al., 1994). In addition, several attempts to construct artificial stripe 2 enhancers using up to ten Bicoid and/or Hunchback sites have been unsuccessful (S. Small and M. Levine, unpublished). These results suggest that cis-regulatory sequences other than the Bicoid and Hunchback sites are important for activation. Such sequences may contain low-affinity Bicoid- and Hunchback-binding

## Figure I-4: Model for the regulation of eve stripe 2.

The minimal *eve* stripe 2 enhancer element is ~480 bp long and contains binding sites for the maternally derived transcriptional activators Bicoid and Hunchback, and for transcriptional repressors encoded by the zygotic gap genes *giant* and *krüppel*. The maternal Bicoid and Hunchback proteins form roughly exponential anterior to posterior morphogenetic gradients, activating *eve* stripe 2 expression in the anterior regions of the embryo. The stripe 2 borders are established by the short-range transcriptional repressors Giant and Krüppel. Giant expressed in a broad band in the anterior regions of the embryo, sets the anterior border, while Krüppel expressed in the central region of the embryo sets up the posterior border. The expression of *eve* is visualized in the embryo above, by *in situ* hybridization using digoxigenin labeled antisense *eve* mRNA probe. Figure is adapted from D. Papatsenko (http://homepages.nyu.edu/~dap).





sites, or sites for other activator proteins. Alternatively, they may simply function to provide the correct spacing between known activator and repressor sites. Second, the mechanism(s) that control repression of eve 2 in anterior regions are not well understood. As the activators (Bicoid and Hunchback) are distributed throughout the anterior half of the embryo, repressive mechanisms other than that provided by Giant must exist that prevent activation in all nuclei anterior to the position of the stripe. Andrioli et al., show that the forkhead domain protein Sloppy-paired (Slp1), which is expressed in a broad anterior domain, binds to a site in the stripe 2 element and is sufficient for repression of stripe 2 of the endogenous eve gene. Further genetic experiments identify a separate repression activity near the anterior pole that is dependent on the terminal patterning gene torso. Thus, three position-specific activities are required for anterior repression of eve 2 (Andrioli et al., 2002).

Similarly a separate enhancer module contained within the complex eve cisregulatory region directs the expression of stripes 3 and 7. This enhancer is approximately 500 bp in length and maps ~ 3.3 kb upstream of the transcription start site (Frasch et al., 1987; Frasch and Levine, 1987; Goto et al., 1989; Harding et al., 1989; Small et al., 1993). The stripe 3 + 7 enhancer is regulated by one or more ubiquitously distributed activators, including components of a JAK-Stat pathway (Yan et al., 1996), which can switch on the gene along the entire length of the early embryo. The two-stripe **Pattern** is defined by multiple tiers of repression mediated by the gap proteins Knirps and Hunchback, which delimit the ubiquitous activation. The stripe 3 + 7 enhancer has at least five binding sites for Knirps and eleven binding sites for the Hunchback repressor. The zinc finger repressor Hunchback is responsible for establishing the anterior border of stripe 3 and the posterior border of stripe 7; Knirps, a member of the nuclear receptor family of transcription factors, establishes the posterior border of stripe 3 and the anterior border of stripe 7 (Small et al., 1996).

Thus, transcriptional repressors are critical in establishing localized patterns of gene expression during Drosophila embryogenesis that is essential for the development of this multicellular organism. A striking feature of the stripe enhancer elements is that their action is autonomous, such that the repression of one element does not lead to the gerneral repression of the entire locus. For example, repression of the eve stripe 2 enhancer in the central regions of the embryo by Krüppel does not prevent the more distal eve stripe 3 element from activating the promoter. Key to the functional autonomy of the stripe enhancer modules is the short-range of the transcriptional repressors that bind and regulate them. The short-range transcriptional repressors that are critical in the regulation of the even-skipped pair-rule gene are the products of the gap genes, giant, knirps and krüppel. The short-range of the repression activity plays an important role in the regulation of complex modular promoters (Figure I-5), by preventing inappropriate 'cross-talk' between enhancers and is essential in specifying the sharp borders of evenskipped expression. The stripe 2 and stripe 3 + 7 enhancers in the eve locus are separated by a 1.7 kb spacer sequence. The removal of this spacing caused repression signals from one enhancer to interfere with the activity of the adjacent enhancer to generate abnormal Patterns of expression in the early embryo (Small et al., 1993). Transgenic embryo assays carrying endogenous enhancers linked to the lacZ reporter gene have been used to study the range of this class of repressors. The zygotically active repressors Giant, Knirps and Krüppel were shown to repress either activators within enhancers or basal promoter

# Figure I-5: Transcriptional repression in Drosophila.

Transcriptional repressors in *Drosophila* are classified as Short-range repressors or Longrange repressors. The zygotic gap genes *giant*, *knirps*, *krüppel* and the mesoderm determinant *snail* encode the short-range repressors that are active during early embryogenesis. These proteins are capable of repressing the activity of enhancer elements when located within 100 bp of key activator sites, or of basal promoter elements when cognate sites are introduced close to the start site of transcription. The short-range of their activity allows multiple enhancers in a complex promoter to function autonomously such that the repression of one element does not lead to the general repression of the entire locus. In contrast long-range repressors, such as Hairy, can mediate repression over distances of > 1000 bp and thus function in a dominant manner to silence multiple enhancers in a complex promoter. Short-range repressors represent a flexible form of gene regulation as they show both enhancer-specific and promoterspecific effects depending on the location of their binding sites.





elements when bound within ~100 bp of their apparent targets (Arnosti et al., 1996a; Arnosti et al., 1996b; Gray and Levine, 1996; Gray et al., 1994; Hewitt et al., 1999; Keller et al., 2000). In contrast to the short-range repressors, long-range repressors such as Hairy, which are also active in the early embryo, can block enhancer elements over distances of more than 1 kb away (Figure I-5), leading to the dominant repression of multiple enhancer complexes (Barolo and Levine, 1997). Thus, short-range repressors clearly provide a more precise, tunable form of repression that can be used to direct complex patterns of gene expression during development.

The short-range transcriptional repressors, Giant, Knirps and Krüppel, are the products of the gap genes that are among the first zygotic genes to be expressed during the early development of Drosophila melanogaster (Figure I-6). Loss of gap gene function leads to missing segments, or gaps, in the embryo body plan. The gap gene product Giant has been characterized as a repressor of other gap genes, including krüppel and knirps, as well as pair-rule genes such as eve (Eldon and Pirrotta, 1991; Kraut and Levine, 1991a; Kraut and Levine, 1991b; Small et al., 1992). The Giant protein contains a C-terminal dimerization/DNA-binding domain of the basis/leucine zipper class (Capovilla et al., 1992; Vinson et al., 1989). giant mutants exhibit abdominal segment defects and loss of head structures, a pattern consistent with the regions in which it is expressed (Gergen and Wieschaus, 1986; Petschek et al., 1987). The Drosophila Knirps protein is a member of the nuclear receptor family of transcription factors and is expressed in the abdominal regions of the precellular embryos and anterior regions of the presumptive head (Rothe et al., 1989). The Knirps protein plays an essential role in the segmentation process, both by refining the expression patterns of other gap genes and by
### Figure I-6: Expression pattern of the short-range repressors Giant and Knirps.

(A) giant is expressed in two broad domains in the anterior and posterior regions of the early blastoderm embryo.

(B) knirps is also expressed in two broad domains in the head region and in the presumptive abdomen in the early blastoderm embryo.

Expression of *giant* and *knirps* in the embryos above are visualized by *in situ* hybridization using digoxigenin labeled antisense mRNA probes. Embryos are oriented anterior to left; dorsal up.

Figure I-6: Expression pattern of the short-range repressors Giant and Knirps.





establishing pair-rule stripes of gene expression (Nusslein-Volhard et al., 1987; Pankratz et al., 1989; Small et al., 1996). Mutations in the *knirps* gene are embryonic lethal, showing a characteristic gap phenotype of the larval cuticle. Knirps also plays important roles in tracheal and wing formation later in development. *krüppel* expression is regulated by the maternal Bicoid gradient (Hoch et al., 1992) and is restricted to the central regions of the embryo corresponding to the presumptive thorax and anterior abdomen (Gaul et al., 1987) and abutting the domains of Giant and Knirps repressor proteins. Krüppel is a zinc finger transcriptional repressor (Licht et al., 1993; Zuo et al., 1991) that regulates the expression of the pair-rule gene *even-skipped*.

#### Context-dependent multiple repression activities of short-range repressors

A common property of short-range transcriptional repressors is their interaction with the evolutionarily conserved corepressor CtBP (C-terminal Binding Protein). This transcription factor was originally identified in human cells through its interaction with the adenovirus E1A oncoprotein (Chinnadurai, 2002; Schaeper et al., 1995; Turner and Crossley, 2001). CtBP proteins are similar in sequence to NAD-dependent Dhydroxyacid dehydrogenases, and possess very similar overall structures to these enzymes (Kumar et al., 2002; Nardini et al., 2003). Recent studies have shown the CtBP has a weak dehydrogenase activity in vitro, although the physiological substrates of CtBP as well as the significance of this enzymatic activity in transcriptional repression remain unknown (Balasubramanian et al., 2003; Kumar et al., 2002). CtBP is recruited to promoters through interactions with a short PXDLS peptide motifs found in short-range transcriptional repressors (and in other interacting proteins) where it mediates repression through mechanisms not currently understood. CtBP has been shown to interact with chromatin modifying factors, including histone deacetylases and histone methyltransferases (Shi et al., 2003; Subramanian and Chinnadurai, 2003; Sundqvist et al., 1998). Thus, the major activity of CtBP might be to serve as a bridging molecule, recruiting chromatin or transcription factor modifying enzymes to the promoter. CtBP itself might possess an alternative enzymatic activity similar to the NAD-dependent Sir 2 repressor protein, which requires NAD to mediate deacetylation of histone proteins (Marmorstein, 2002).

Although CtBP-mediated repression is critical for full activity of short-range repressors, *Drosophila* short-range repressors also possess CtBP-independent repression activities (Keller et al., 2000; La Rosee et al., 1997; La Rosee-Borggreve et al., 1999; Nibu et al., 2003; Strunk et al., 2001). A part of the CtBP-independent activity of these repressors might be attributed to competition for activator binding sites but they also possess CtBP-independent repression activities that does not require direct competition (Keller et al., 2000; La Rosee-Borggreve et al., 1999; Nibu et al., 2003; Strunk et al., 2000; La Rosee-Borggreve et al., 1999; Nibu et al., 2003; Strunk et al., 2001). Studies on Giant repression activity in our lab demonstrated that the dCtBP cofactor is required for Giant repression of some, but not all target genes (Strunk et al., 2001). The results indicate that Giant can repress via both dCtBP-dependent and – independent pathways and dCtBP requirement can vary on a gene-to-gene as well as on an enhancer-to-enhancer basis.

Previous work in our lab has identified two repression regions of the Knirps protein (Keller et al., 2000). The C-terminal region from amino acids 202-358, appears to mediate repression through dCtBP and contains the dCtBP-binding motif PMDLSMK. The N-terminal region, from amino acids 139-330, does not bind dCtBP, and was shown to repress the *eve* stripe 2-*lacZ* in a dCtBP mutant background (Keller et al., 2000). Also, the loss of maternal dCtBP does not affect Knirps repression of *eve* stripe 3 + 7 enhancer, but abolishes repression of *eve* stripe 4+6 enhancer element (Keller et al., 2000). La Rosée et al. have shown that the other short-range repressor Krüppel can repress the *hairy* stripe 7 enhancer in the absence of dCtBP (La Rosee et al., 1997).

Thus, a growing body of evidence demonstrates that in addition to the short-range of activity, many, or perhaps all, short-range repressor proteins exhibit multiple repression activities. Multiple repression activities may allow for quantitative or qualitative effects on gene expression and may be context-dependent. Thus, qualitatively, a repressor may function selectively in a tissue-specific manner or in an activator-specific manner (Postigo and Dean, 1999) or in different promoter contexts (Lunyak et al., 2002). Quantitatively, dual activities may increase the overall level of repression. Consistent with a quantitative model, the CtBP-dependent and CtBP-independent repression activities of Knirps have been found to exhibit striking functional similarities in cell culture assays, indicating that they might utilize similar mechanisms of repression (Ryu and Arnosti, 2003). Additionally, we have shown that increasing the dose of the Knirps repressor may be sufficient to overcome the requirement for the dCtBP cofactor and that multiple repression activities within a single protein represent quantitative effects on gene expression (Paolo Struffi, Maria Corado, Meghana M. Kulkarni and David N. Arnosti. "Quantitative Contributions of CtBP-dependent and --independent repression activities of Knirps". Manuscript in revision).

Other studies have indicated that different concentrations of Giant, Knirps and Krüppel are important for the proper regulation of zygotic genes (Kosman and Small, 1997; Wu et al., 1998). An example is the regulation of *eve* stripes 3, 4, 6 and 7 (Figure I-7). As previously described, transcriptional repressors Knirps and Hunchback delimit expression borders of the *even-skipped* stripes 3,4, 6 and 7. Two corresponding *cis*-regulatory enhancer modules, the stripe 4+6 and stripe 3+7 enhancers, encode all sufficient transcriptional information for these four stripes. The *eve* stripes 4 and 6 are formed in the embryo zones with lower concentration of Hunchback and higher concentration of Knirps, conversely the *eve* stripes 3 and 7 are formed where Hunchback concentration is greater than that of Knirps. Sequence analysis of the *eve* gene indicates that there are more high affinity Knirps binding sites within the *eve* stripe 3 + 7 enhancer than in the 4 + 6 element (Berman et al., 2002; Papatsenko et al., 2002). This is also consistent with our observations that the *eve* 3 + 7 enhancer is more sensitive to Knirps repression than the *eve* 4 + 6 element (Struffi et al., manuscript submitted).

These observations suggest that in addition to spacing constraints, the relative affinity and the number of binding sites in the regulatory region are critical parameters in dictating short-range repression effectiveness. To analyze enhancer function in a setting in which activator-repressor stoichiometry and spacing can be exactly defined, I constructed chromosomally integrated, compact synthetic enhancer modules containing binding sites for endogenous short-range repressors Giant, Knirps, or Krüppel and chimeric Gal4 activators (Chapter 3). The construction and functional assessment of different synthetic elements allowed us to accurately quantify not only the exact distance requirements, but also parameters such as relative affinities, number and arrangements of

### Figure I-7: Model for the regulation of the *eve* stripes 3 + 7 and 4 + 6.

The expression of eve is visualized in the embryo above (A), by in situ hybridization using digoxigenin labeled antisense eve mRNA probe. (B) Two corresponding regulatory elements the 3 + 7 and 4 + 6 enhancers encode all the necessary transcriptional information for the precise expression of even-skipped in stripes 3, 7, 4 and 6. The transcriptional repressors Hunchback and Knirps, for which the enhancer elements contain binding sites, establish the stripe borders. The eve stripes 4 and 6 are formed in embryo zones with lower concentration of Hunchback and higher concentration of Knirps. Conversely, the eve stripes 3 and 7 are formed where Hunchback concentration is that of Knirps. Figure is greater than adapted from D. Papatsenko (http://homepages.nyu.edu/~dap).

Figure I-7: Model for the regulation of the eve stripes 3 + 7 and 4 + 6.



activator and repressor binding sites, features that could not be dissected in previous studies of short-range transcriptional repression.

### **CIS-REGULATORY INFORMATION PROCESSING**

The qualitative functional complexity of *cis*-regulatory enhancer elements is a manifestation of the complexity of the assemblage of target sites within the module. Thus, enhancers serve as binding platforms for multiple diverse regulatory proteins, both positively and negatively acting factors. In addition, each element may contain multiple target sites for the same or different transregulator thereby serving to increase the local concentration of regulatory proteins that affect the expression of the linked gene. Furthermore, the presence of each factor may additionally depend on signaling events, cell-cycle activity, temporal state, lineage or spatial position. Thus, enhancers can integrate environmental and developmental information to regulate the expression of genes in a biologically appropriate manner. Depending on the signals or cell type, the same enhancer element might activate or repress, and the magnitude of activation signals can be variable, in the manner of a rheostat (Arnosti, 2002; Arnosti, 2003; Barolo and Posakony, 2002; Biggar and Crabtree, 2001; Rossi et al., 2000). The combinations and arrangement of binding sites within the enhancer dictate the nature of the signal generated. This property of enhancer elements has led to the analogy of an enhancer as a molecular logic device or computer. Computational functions of an enhancer have been ascribed to its ability to receive multiple inputs in terms of the different transcription factors that bind it, to process these inputs resolving them into a single output, which is directed to the basal machinery, either turning the gene on or off.

The analysis of gene regulation in sea urchins by Eric Davidson and colleagues is the best known example of the "enhancer as a computer" analogy and showed the possibility of modeling gene regulation as logic circuits (Yuh et al., 1998). The same study also demonstrated that the endo 16 cis-regulatory region of the purple sea urchin Strongylocentrotus purpuratus works remarkably like a tiny analog computer. Like other cis-regulatory systems that mediate complex developmental patterns of expression, the endo 16 system is modular in organization, that is, it consists of sub-elements of DNA sequence, each of which can execute a certain regulatory function such as activation, repression, synergism, and integration. Separate portions of the regulatory region can be combined to recreate some or all, of the expression pattern of the gene, a characteristic of modular elements. To test their understanding of the endo 16 cis-regulatory region, Yuh and colleagues created a computer model with Boolean elements that simulates these regulatory interactions. The model made predictions about the consequences of specific promoter manipulations on transcription levels and when tested experimentally successfully simulate the output of the regulatory regions. The success of the model emphasizes the integrative, computer-like processing suggested to be a characteristic of enhancer elements.

Studies such as those described above suggest that the enhancer functions as the central information processing unit while the basal transcriptional machinery itself simply responds to the signals generated by these molecular logic circuits. Such integrative functions have been ascribed to the human interferon- $\beta$  (IFN- $\beta$ ) enhancer, which drives transcription of the IFN- $\beta$  gene in response to viral infection (Struhl, 2001). The IFN- $\beta$ 

enhancer is a small 65 bp region immediately upstream of the core promoter. It contains binding sites for activators of the NFkB, IRF, and ATF/Jun families as well as target sites for the architectural protein HMGI (Y) (Kim and Maniatis, 1997; Munshi et al., 2001; Thanos and Maniatis, 1995). These regulatory proteins assemble through cooperative interactions into a well-defined nucleoprotein complex called the 'enhanceosome'. Assembly of the enhanceosome is absolutely essential for the transcription of the IFN-B gene in response to viral infection in cells. Individual activators bound to their sites in the enhancer do not by themselves stimulate transcription. This is because the nucleoprotein complex provides a stereospecific interface for interaction with the basal transcriptional machinery, possibly engaging several components of the basal machinery simultaneously to effect synergistic activation (Carey et al., 1990; Chi et al., 1995). In this structured element, the presence of each transcription factor binding site and its precise arrangement within the regulatory element are critical for the various regulatory proteins (sequencespecific activators and architectural proteins) to assemble through cooperative interactions into a well-defined nucleoprotein complex called the "enhanceosome" and is essential in dictating the output of the element (Carey et al., 1990; Chi et al., 1995; Kim and Maniatis, 1997; Munshi et al., 2001; Thanos and Maniatis, 1995). The enhanceosome therefore imposes the restriction that the target gene would be activated only when all the regulatory proteins are present and are functionally active. The IFN-B enhanceosome is a paradigm for precision and functions as a precise on/off binary transcriptional switch in response to the appropriate stimulus (Struhl, 2001).

In contrast to the stringent organization of factor binding sites in the IFN-ß enhanceosome some flexibility in organization has been observed for the even-skipped

34

stripe 2 enhancer element (Arnosti et al., 1996a; Ludwig et al., 2000). A defective stripe 2 enhancer lacking the crucial Bicoid B1 site can be complemented with a high affinity Bicoid binding site inserted at a new location (Arnosti et al., 1996a). However, the same study also demonstrated that there are some constraints on enhancer organization since the insertion of the same high affinity Bicoid sequence at another location results in only a partial restoration of the stripe. In addition to flexibility in the organization of the *cis*element, the authors also observed that there was flexibility in the trans-regulation of the stripe. Activation does not seem to depend on a particular class of DNA binding protein (the Bicoid homeodomain can be replaced by the Gal4 zinc finger domain), nor does expression require a particular type of activation domain (Bicoid activation domain can be replaced either by the glutamine rich activation domain of Sp1 or the acidic activation domain of GCN4). The results from this study suggest that there may be no specific requirement for particular protein-protein interactions and that eve stripe 2 expression only requires the binding of a sufficient number of activation domains (Arnosti et al., 1996a). Mutations in individual binding sites further demonstrated that each bound transcription factor contributes to the overall output of the enhancer.

The functional analysis of the *eve* stripe 2 enhancer by Kreitman and colleagues has also suggested a more flexible arrangement of regulatory proteins to be the predominant pattern for *cis*-elements that provide diverse patterns in developing systems, *rather* than the more constrained architecture afforded by the enhanceosome. *Phylogenetic* comparisons among eve stripe 2 enhancer elements from disparate *Drosophilids* indicate that this enhancer has undergone some genetic drift in its internal organization. However, the redesign in internal architecture does not affect the overall output of the element (Ludwig et al., 2000; Ludwig and Kreitman, 1995; Ludwig et al., 1998). The plasticity of this enhancer suggests that much variation in spatial placement of individual transcription factors is possible, consistent with a model in which these factors contact the basal machinery in a flexible framework, not necessarily as a rigid complex.

We decided to test the functional significance of this flexibility in enhancer design on the way in which the enhancer communicates with the basal promoter. Our studies with composite enhancer elements (described in Chapter 2) containing binding sites for the short-range repressors Giant or Knirps and different activators, led us to propose a new model for enhancer function. We call this the'Information display' or 'Billboard' enhancer, where computation of regulatory inputs is not performed by the enhancer alone but results from multiple iterative/simultaneous interactions between enhancer subelements and the basal transcriptional machinery. Consistent with this model, functional analysis of many *cis*-regulatory elements has demonstrated the presence of redundant regulatory information which suggests that there are multiple configurations in which enhancer-bound factors can interact with the basal transcriptional machinery (Han et al., 1998; Hoch et al., 1992; Piano et al., 1999).

# CIS-REGULATORY SEQUENCE AND THE EVOLUTION OF MORPHOLOGICAL FEATURES

The morphological features of multicellular organisms are the products of developmental processes, which are controlled by genomic regulatory programs. The evolution of form has been suggested to largely reflect the evolution of these genomic regulatory programs (Carroll, S. B, 2001). Given that many regulatory genes encode highly conserved proteins, regulatory evolution is thought to be brought about primarily by substitutions in *cis*-regulatory sequences rather than in the proteins themselves. The creative potential of regulatory change and the comparatively greater constraints on protein evolution have always been recognized. However, little is known about the tempo and mode of *cis*-acting regulatory sequence evolution, in part a reflection of the technical difficulty in dissecting regulatory sequence structure and function.

Evolutionary analysis of non-coding sequences is more difficult than that of coding sequences given the paucity of structural information about *cis*-regulatory DNA (Dermitzakis and Clark, 2002; Leung et al., 2000; Wasserman et al., 2000). There are no straightforward properties in regulatory sequences analogous to the open reading frame and codons in coding sequences, making it difficult to define the position, amount, and strength of selective constraints on functional regulatory elements. In addition, the model of transcriptional regulation is not a simple one of activation or suppression by transcription factors, but rather, it includes competitive binding of proteins (Small et al., 1991), cooperative binding (Burz et al., 1998; Zhao et al., 2000), DNA bending, chromatin modifications and other molecular interactions that are not always reflected in the nucleotide sequence.

Understanding the evolutionary processes that regulatory sequences undergo will substantially improve our understanding of their functional constraints. Numerous comparative methods (termed "phylogenetic footprinting") have been developed that use sequence conservation to infer function (Hardison et al., 1997; Miller et al., 2001). It may be problematic, however, to assume that all functional sequences are conserved and all nonfunctional sequences have diverged. For instance, sequence comparisons of closely related species produce many false positive results, and comparison of distantly related species has low power to detect functional elements in the species compared (Dermitzakis and Clark, 2002). Moreover, it has been shown that regulatory sequences can maintain regulatory function despite structural reorganization as a result of species-specific loss and gain of transcription factor binding sites (Cuadrado et al., 2001; Ludwig et al., 2000; Piano et al., 1999). Given that transcription factor binding sites are the fundamental units of regulatory structure, they are also likely to be the fundamental units of regulatory evolution. Therefore, methods that can predict the binding site composition of a sequence should provide a powerful means for quantifying regulatory sequence divergence in comparative analyses. Application of predictive tools such as probability weight matrices (PWMs) has been useful in the annotation of regulatory regions (Berman et al., 2002; Stormo, 2000a; Stormo, 2000b; Stormo, 2000c), and such tools have been shown to be useful for predicting the evolution of regulatory regions as well (Chuzhanova et al., 2000; Liu et al., 2000; Ludwig et al., 2000). However, due to the low sequence specificity of transcription factors for their binding sites, only a fraction of the predicted binding sites are thought to be functionally significant.

Thus, meaningful evolutionary analysis of a regulatory sequence will require detailed information about the locations of transcription factor binding sites within a sequence, the functional specificity of the binding sequences, and the spatial requirements for their interaction. Therefore, if we knew in functional terms the components of the specific genomic *cis*-elements that result in different morphological outcomes in two animals of common ancestry, we could determine exactly what are the essential causal differences in the DNA of these animals and provide a mechanistic explanation of how the diverse forms actually arose during evolution.

## TOWARDS DEVELOPING A 'GRAMMAR' THAT DESCRIBES GENE REGULATORY INSTRUCTIONS CONTAINED IN GENOMIC DNA

*Cis*-regulatory function derives from *cis*-regulatory structure, therefore understanding why a given developmental process occurs as it does requires an understanding of the structure/function relations that exist within *cis*-regulatory enhancer elements, which can be thought of as a "genetic code" for development. Despite the importance of the genomic *cis*-regulatory enhancer sequences, our ability to identify and predict functions for this category of DNA is extremely limited. Deciphering how these elements interact with each other and the core promoter could in theory enable us to predict the spatial and temporal activity of genes and to simulate their expression pattern *in silico*. Thus, it is evident that from the structure and functional organization of *cis*regulatory DNA alone emerges an unprecedented explanatory and predictive power, in respect to understanding and even controlling developmental processes. In order to crack the '*cis*-regulatory code' that is implicit in the sequence of genomic regulatory DNA we need to 1) Identify *cis*-acting enhancer elements and characterize their internal architectural organization; 2) Determine the biological significance of their architectural design empirically and 3) Build and validate predictive models for differential gene expression as a function of time and space.

# Recent advances in computational methodologies for the identification and analysis of cis-regulatory enhancer elements

*Cis*-regulatory elements involved in gene regulation have been classically studied over the last forty years using traditional genetic and biochemical approaches on a promoter-by-promoter basis. Although the detailed inventory of single genes has offered spectacular successes, a more global understanding of transcriptional regulatory switch design is needed to comprehend the staggering complexity, versatility, and robustness of living systems. Accompanying the expansion of large data sets that have resulted from genomic sequencing and genome-wide expression profiling, new computational strategies have been developed to contribute to the creation of a vocabulary that describes gene regulatory instructions contained in genomic DNA on a global scale. Highthroughput sequencing technology, which has allowed the sequencing of complete genomes of a large variety of species (including human, mouse, and rat), makes genomewide comparisons or phylogenetic footprinting feasible endeavors. High-throughput expression profiling technology like microarrays (gene chip) and serial analysis of gene expression (SAGE) have allowed rapid parallel analysis of expression levels of hundreds of thousands of genes in a single assay. An underlying assumption of using highthroughput expression profiling technology in regulatory region analysis is that coregulated genes often share a similar set of regulatory motifs.

Computational sequence analysis provides three broadly different approaches for scanning genomic sequence to identify those regions predicted to participate in gene regulation. First, inter-species sequence comparisons have been used to identify non-coding sequences that have a reasonable likelihood of having gene regulatory properties (Duret and Bucher, 1997; Gottgens et al., 2000; Hardison et al., 1997; Loots et al., 2000). This is possible because sequences that mediate gene expression tend to be conserved between species. It may be problematic, however, to assume that all functional sequences are conserved and all nonfunctional sequences have diverged. For instance, sequence comparisons of closely related species produce many false positive results, and comparison of distantly related species has low power to detect functional elements in the species compared (Dermitzakis and Clark, 2002). Moreover, it has been shown that regulatory sequences can maintain function despite structural reorganization as a result of species-specific loss and gain of transcription factor binding sites (Cuadrado et al., 2001; Ludwig et al., 2000; Piano et al., 1999).

Sequence analysis of co-regulated genes within a species is a second approach for predicting regulatory elements. This strategy is based on the fact that few transcription factors exert their activity exclusively on single genes; rather, most bind to conserved sites in several genes to coordinate their expression. Accordingly, genes are thought to be co-regulated because they respond to similar regulatory pathways owing to shared noncoding sequence motifs that direct the binding of specific sets of shared transcription factors (Fickett and Wasserman, 2000; Wasserman et al., 2000). The correlation between gene cluster and regulatory motifs is imprecise for at least three reasons. First, not all coregulated gene promoters share a common motif, because some of the identified genes in a given cluster might in fact be secondary response genes. Second, because of the combinatorial nature of transcription factors, the same motif can be found in the promoter regions of genes that are not co-regulated. Third, some 'motifs' are likely to represent random noise and in fact do not bind the transcription factor at all.

The third approach for the identification of gene regulatory sequences involves generating and analyzing databases of known transcription-factor-binding sites and screening genomic sequences for the presence of clusters of the known transcription-factor-binding sites (Berman et al., 2002; Halfon et al., 2002; Halfon and Michelson, 2002; Markstein and Levine, 2002; Markstein et al., 2002).

One main difficulty with the output from computational searches for transcription factor binding sites is the large number of false-positive and false negative results. For example, Markstein *et al.* recently developed a computational method for identifying clusters of Dorsal-binding sites in the *Drosophila* genome. A Dorsal-responsive silencer from the target gene *zerknullt* was used to develop a specific model for a genome wide computational scan with the aim of identifying other novel *cis*-acting enhancers that are regulated by the Dorsal transcription factor. In this case clusters of at least three high affinity Dorsal binding sites were sought within a 400 bp window. The search yielded 15 matches of which three were associated with known Dorsal-responsive genes. However, many genes (for example, the *rhomboid* NEE) that are known to be regulated by the Dorsal transcription factor were not identified on the basis of optimal binding clusters (Markstein and Levine, 2002; Markstein et al., 2002). The short length and degenerate

nature of transcription-factor-binding sites account for most of these misleading predictions. Also genes are rarely controlled by a single transcription factor, and in fact, accumulating evidence suggests that specific combinations of transcription factors are required to achieve proper biological function of *cis*-elements involved in directing complex patterns of gene expression. One way to improve the computational identification of *cis*-regulatory DNAs is to search for clustering of two or more different classes of recognition sequences (Berman et al., 2002; Halfon et al., 2002; Halfon and Michelson, 2002; Rajewsky et al., 2002). Despite these various ways to minimize the number of false-positive binding sites and improve the hit rate, even those sequences that meet the most stringent criteria might still be non-functional in a genomic context. For instance, many cases are known where factors must interact with other factors or cofactors to be functional, and spatial correlations between their binding sites are observed.

# A Functional step towards understanding cis-regulatory architecture involved in development

Comparative sequence analysis, coupled with the development of algorithms to search genomic databases, have provided important tools for the identification of gene regulatory elements at a scale not previously possible, but they have only been partially successful at finding *cis*-regulatory motifs (as described above). In many cases, where *cis*-regulatory elements predicted by computational and statistical methods appear suitable, something in the arrangement of sites or a wider context renders the element non-functional. Many times the techniques do not take into consideration the combinatorial logic underlying enhancer architecture. Besides clustering of binding sites, it is necessary to factor in binding site type, number, affinities, spacing, orientations and order, in order to achieve better predictions and in the interpretation of their biological function. These parameters are difficult to predict accurately *a priori*. Thus, computational approaches would benefit from the availability of at least one well-defined representative of a particular regulatory network to serve as a starting paradigm.

Our study (described in Chapter 3 of this thesis) takes a first step towards providing just such a paradigm for early *Drosophila* developmental enhancers that are regulated by the short-range transcriptional repressors. The construction and functional assessment of synthetic enhancers to test particular combinatorial models of binding site type, number, affinities, spacing, and relative positioning that need to be incorporated to ensure the appropriate output aims to deduce the 'grammar' rules that define a functional module. Incorporating the architectural parameters defined in our study into formal clustering models might facilitate computational recognition of similar *cis*-regulatory elements and aid in the interpretation of their biological function.

#### **Regulatory Frontiers**

The prospect of obtaining a truly global picture of the regulatory control system of a complete eukaryotic organism with many thousands of genes seems daunting. Despite the challenges, an initial framework offering a rough roadmap appears to have been established. Current successes in large-scale sequencing and gene identification have provided the identity and physical location of individual genes. Increasingly more powerful genomic technologies will enable us to identify many of the changes in *cis*regulatory DNAs, the corresponding changes in gene expression patterns between organisms, and the role of these differences in speciation. The systematic integration of diverse data sets such as high throughput *in situ* hybridization screens that provide spatial and temporal information about transcription factor expression profiles, genome wide transcription factor binding analysis (Lee et al., 2002; Weinmann and Farnham, 2002; Weinmann et al., 2002), protein–protein interaction networks together with improved regulatory sequence identification strategies will provide an integrated platform for deciphering the transcriptional regulatory network. As we enter the post-genome era, it is possible to envision the elucidation of the transcriptional *cis*-regulatory code, whereby the information content of *cis*-regulatory DNAs can be predicted by simple sequence analysis.

### **REFERENCES**

Adams, M. D. Celniker, S. E. Holt, R. A. Evans, C. A. Gocayne, J. D. Amanatides, P. G. Scherer, S. E. Li, P. W. Hoskins, R. A. Galle, R. F. et al. (2000). The genome sequence of *Drosophila* melanogaster. *Science* 287, 2185-95.

Andrioli, L. P., Vasisht, V., Theodosopoulou, E., Oberstein, A. and Small, S. (2002). Anterior repression of a *Drosophila* stripe enhancer requires three position-specific mechanisms. *Development* 129, 4931-40.

Arnosti, D. N. (2002). Design and function of transcriptional switches in *Drosophila*. *Insect Biochem Mol Biol* **32**, 1257-73.

Arnosti, D. N. (2003). Analysis and function of transcriptional regulatory elements: Insights from *Drosophila*. Annu Rev Entomol 48, 579-602.

Arnosti, D. N., Barolo, S., Levine, M. and Small, S. (1996a). The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* 122, 205-14.

Arnosti, D. N., Gray, S., Barolo, S., Zhou, J. and Levine, M. (1996b). The gap protein knirps mediates both quenching and direct repression in the *Drosophila* embryo. *Embo J* 15, 3659-66.

Asoh, S., Lee-Kwon, W., Mouradian, M. M. and Nirenberg, M. (1994). Selection of DNA clones with enhancer sequences. *Proc Natl Acad Sci U S A* **91**, 6982-6.

Balasubramanian, P., Zhao, L. J. and Chinnadurai, G. (2003). Nicotinamide adenine dinucleotide stimulates oligomerization, interaction with adenovirus E1A and an intrinsic dehydrogenase activity of CtBP. *FEBS Lett* 537, 157-60.

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

**Barolo, S. and Levine, M.** (1997). hairy mediates dominant repression in the *Drosophila* embryo. *Embo J* 16, 2883-91.

**Barolo, S. and Posakony, J. W.** (2002). Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. *Genes Dev* 16, 1167-81.

Belting, H. G., Shashikant, C. S. and Ruddle, F. H. (1998). Modification of expression and *cis*-regulation of Hoxc8 in the evolution of diverged axial morphology. *Proc Natl Acad Sci U S A* 95, 2355-60.

Berman, B. P., Nibu, Y., Pfeiffer, B. D., Tomancak, P., Celniker, S. E., Levine, M., Rubin, G. M. and Eisen, M. B. (2002). Exploiting transcription factor binding site clustering to identify *cis*-regulatory modules involved in pattern formation in the *Drosophila* genome. *Proc Natl Acad Sci U S A* 99, 757-62.

**Biggar, S. R. and Crabtree, G. R.** (2001). Cell signaling can direct either binary or graded transcriptional responses. *Embo J* 20, 3167-76.

Blackwood, E. M. and Kadonaga, J. T. (1998). Going the distance: a current view of enhancer action. *Science* 281, 61-3.

Burz, D. S., Rivera-Pomar, R., Jackle, H. and Hanes, S. D. (1998). Cooperative DNAbinding by Bicoid provides a mechanism for threshold-dependent gene activation in the *Drosophila* embryo. *Embo J* 17, 5998-6009.

Capovilla, M., Eldon, E. D. and Pirrotta, V. (1992). The giant gene of *Drosophila* encodes a b-ZIP DNA-binding protein that regulates the expression of other segmentation gap genes. *Development* 114, 99-112.

Carey, M., Lin, Y. S., Green, M. R. and Ptashne, M. (1990). A mechanism for synergistic activation of a mammalian gene by GAL4 derivatives. *Nature* 345, 361-4.

Chi, T., Lieberman, P., Ellwood, K. and Carey, M. (1995). A general mechanism for transcriptional synergy by eukaryotic activators. *Nature* 377, 254-7.

Chinnadurai, G. (2002). CtBP, an unconventional transcriptional corepressor in development and oncogenesis. *Mol Cell* 9, 213-24.

Chuzhanova, N. A., Krawczak, M., Nemytikova, L. A., Gusev, V. D. and Cooper, D. N. (2000). Promoter shuffling has occurred during the evolution of the vertebrate growth hormone gene. *Gene* 254, 9-18.

Cuadrado, M., Sacristan, M. and Antequera, F. (2001). Species-specific organization of CpG island promoters at mammalian homologous genes. *EMBO Rep* 2, 586-92.

**Davidson, E. H.** (2001). Genomic Regulatory Systems: Development and Evolution: Academic Press: A Harcourt Science and Technology Company.

Dermitzakis, E. T. and Clark, A. G. (2002). Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover. *Mol Biol Evol* 19, 1114-21.

Duret, L. and Bucher, P. (1997). Searching for regulatory elements in human noncoding sequences. *Curr Opin Struct Biol* 7, 399-406.

Durick, K., Mendlein, J. and Xanthopoulos, K. G. (1999). Hunting with traps: genome-wide strategies for gene discovery and functional analysis. *Genome Res* 9, 1019-25.

Eldon, E. D. and Pirrotta, V. (1991). Interactions of the *Drosophila* gap gene giant with maternal and zygotic pattern-forming genes. *Development* 111, 367-78.

Fickett, J. W. and Wasserman, W. W. (2000). Discovery and modeling of transcriptional regulatory regions. *Curr Opin Biotechnol* 11, 19-24.

Frasch, M., Hoey, T., Rushlow, C., Doyle, H. and Levine, M. (1987). Characterization and localization of the even-skipped protein of *Drosophila*. *Embo J* 6, 749-59.

Frasch, M. and Levine, M. (1987). Complementary patterns of even-skipped and fushi tarazu expression involve their differential regulation by a common set of segmentation genes in *Drosophila*. Genes Dev 1, 981-95.

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.

Fukushige, S. and Ikeda, J. E. (1996). Trapping of mammalian promoters by Cre-lox site-specific recombination. DNA Res 3, 73-80.

Gaudreau, L., Schmid, A., Blaschke, D., Ptashne, M. and Horz, W. (1997). RNA polymerase II holoenzyme recruitment is sufficient to remodel chromatin at the yeast PHO5 promoter. *Cell* 89, 55-62.

Gaul, U., Seifert, E., Schuh, R. and Jackle, H. (1987). Analysis of Kruppel protein distribution during early *Drosophila* development reveals posttranscriptional regulation. *Cell* 50, 639-47.

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

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.

Gottgens, B., Barton, L. M., Gilbert, J. G., Bench, A. J., Sanchez, M. J., Bahn, S., Mistry, S., Grafham, D., McMurray, A., Vaudin, M. et al. (2000). Analysis of vertebrate SCL loci identifies conserved enhancers. *Nat Biotechnol* 18, 181-6.

Gray, S. and Levine, M. (1996). Transcriptional repression in development. Curr Opin Cell Biol 8, 358-64.

Gray, S., Szymanski, P. and Levine, M. (1994). Short-range repression permits multiple enhancers to function autonomously within a complex promoter. *Genes Dev* 8, 1829-38.

Halfon, M. S., Grad, Y., Church, G. M. and Michelson, A. M. (2002). Computationbased discovery of related transcriptional regulatory modules and motifs using an experimentally validated combinatorial model. *Genome Res* 12, 1019-28. Halfon, M. S. and Michelson, A. M. (2002). Exploring genetic regulatory networks in metazoan development: methods and models. *Physiol Genomics* 10, 131-43.

Han, W., Yu, Y., Su, K., Kohanski, R. A. and Pick, L. (1998). A binding site for multiple transcriptional activators in the fushi tarazu proximal enhancer is essential for gene expression *in vivo*. *Mol Cell Biol* 18, 3384-94.

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). Crossregulatory interactions among pair-rule genes in *Drosophila*. *Science* 233, 953-9.

Hardison, R., Slightom, J. L., Gumucio, D. L., Goodman, M., Stojanovic, N. and Miller, W. (1997). Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene* 205, 73-94.

Hewitt, G. F., Strunk, B. S., Margulies, C., Priputin, T., Wang, X. D., Amey, R., Pabst, B. A., Kosman, D., Reinitz, J. and Arnosti, D. N. (1999). Transcriptional repression by the *Drosophila* giant protein: cis element positioning provides an alternative means of interpreting an effector gradient. *Development* 126, 1201-10.

Hoch, M., Gerwin, N., Taubert, H. and Jackle, H. (1992). Competition for overlapping sites in the regulatory region of the *Drosophila* gene Kruppel. *Science* 256, 94-7.

Keller, S. A., Mao, Y., Struffi, P., Margulies, C., Yurk, C. E., Anderson, A. R., Amey, R. L., Moore, S., Ebels, J. M., Foley, K. et al. (2000). dCtBP-dependent and - independent repression activities of the *Drosophila* Knirps protein. *Mol Cell Biol* 20, 7247-58.S

Kim, T. K. and Maniatis, T. (1997). The mechanism of transcriptional synergy of an *in vitro* assembled interferon-beta enhanceosome. *Mol Cell* 1, 119-29.

Kosman, D. and Small, S. (1997). Concentration-dependent patterning by an ectopic expression domain of the *Drosophila* gap gene knirps. *Development* 124, 1343-54.

Kraut, R. and Levine, M. (1991a). Mutually repressive interactions between the gap genes giant and Kruppel define middle body regions of the *Drosophila* embryo. *Development* 111, 611-21.

Kraut, R. and Levine, M. (1991b). Spatial regulation of the gap gene giant during *Drosophila* development. *Development* 111, 601-9.

Kumar, V., Carlson, J. E., Ohgi, K. A., Edwards, T. A., Rose, D. W., Escalante, C. R., Rosenfeld, M. G. and Aggarwal, A. K. (2002). Transcription corepressor CtBP is an NAD(+)-regulated dehydrogenase. *Mol Cell* 10, 857-69.

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.

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I. et al. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* 298, 799-804.

Leung, J. Y., McKenzie, F. E., Uglialoro, A. M., Flores-Villanueva, P. O., Sorkin, B. C., Yunis, E. J., Hartl, D. L. and Goldfeld, A. E. (2000). Identification of phylogenetic footprints in primate tumor necrosis factor-alpha promoters. *Proc Natl Acad Sci U S A* 97, 6614-8.

Lewis, E. B. (1998). The bithorax complex: the first fifty years. Int J Dev Biol 42, 403-15.

Licht, J. D., Ro, M., English, M. A., Grossel, M. and Hansen, U. (1993). Selective repression of transcriptional activators at a distance by the *Drosophila* Kruppel protein. *Proc Natl Acad Sci U S A* 90, 11361-5.

Liu, T., Wu, J. and He, F. (2000). Evolution of *cis*-acting elements in 5' flanking regions of vertebrate actin genes. *J Mol Evol* 50, 22-30.

Loots, G. G., Locksley, R. M., Blankespoor, C. M., Wang, Z. E., Miller, W., Rubin, E. M. and Frazer, K. A. (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* 288, 136-40.

Ludwig, M. Z., Bergman, C., Patel, N. H. and Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403, 564-7.

Ludwig, M. Z. and Kreitman, M. (1995). Evolutionary dynamics of the enhancer region of even-skipped in *Drosophila*. Mol Biol Evol 12, 1002-11.

Ludwig, M. Z., Patel, N. H. and Kreitman, M. (1998). Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. *Development* 125, 949-58.

Lunyak, V. V., Burgess, R., Prefontaine, G. G., Nelson, C., Sze, S. H., Chenoweth, J., Schwartz, P., Pevzner, P. A., Glass, C., Mandel, G. et al. (2002). Corepressordependent silencing of chromosomal regions encoding neuronal genes. *Science* 298, 1747-52. Macdonald, P. M., Ingham, P. and Struhl, G. (1986). Isolation, structure, and expression of even-skipped: a second pair-rule gene of *Drosophila* containing a homeo box. *Cell* 47, 721-34.

Macdonald, P. M. and Struhl, G. (1986). A molecular gradient in early *Drosophila* embryos and its role in specifying the body pattern. *Nature* 324, 537-45.

Magis, W., Fiering, S., Groudine, M. and Martin, D. I. (1996). An upstream activator of transcription coordinately increases the level and epigenetic stability of gene expression. *Proc Natl Acad Sci U S A* 93, 13914-8.

Markstein, M. and Levine, M. (2002). Decoding *cis*-regulatory DNAs in the *Drosophila* genome. *Curr Opin Genet Dev* 12, 601-6.

Markstein, M., Markstein, P., Markstein, V. and Levine, M. S. (2002). Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila* embryo. *Proc Natl Acad Sci U S A* 99, 763-8.

Marmorstein, R. (2002). Dehydrogenases, NAD, and transcription--what's the connection? *Structure (Camb)* 10, 1465-6.

Miller, R. D., Taillon-Miller, P. and Kwok, P. Y. (2001). Regions of low singlenucleotide polymorphism incidence in human and orangutan xq: deserts and recent coalescences. *Genomics* 71, 78-88.

Munshi, N., Agalioti, T., Lomvardas, S., Merika, M., Chen, G. and Thanos, D. (2001). Coordination of a transcriptional switch by HMGI(Y) acetylation. *Science* 293, 1133-6.

Nardini, M., Spano, S., Cericola, C., Pesce, A., Massaro, A., Millo, E., Luini, A., Corda, D. and Bolognesi, M. (2003). CtBP/BARS: a dual-function protein involved in transcription co-repression and Golgi membrane fission. *Embo J* 22, 3122-30.

Nibu, Y., Senger, K. and Levine, M. (2003). CtBP-independent repression in the Drosophila embryo. Mol Cell Biol 23, 3990-9.

Nusslein-Volhard, C., Frohnhofer, H. G. and Lehmann, R. (1987). Determination of anteroposterior polarity in *Drosophila*. *Science* 238, 1675-81.

**Ogbourne, S. and Antalis, T. M.** (1998). Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem J* 331 (Pt 1), 1-14.

**Pankratz, M. J., Hoch, M., Seifert, E. and Jackle, H.** (1989). Kruppel requirement for knirps enhancement reflects overlapping gap gene activities in the *Drosophila* embryo. *Nature* 341, 337-40.

Papatsenko, D. A., Makeev, V. J., Lifanov, A. P., Regnier, M., Nazina, A. G. and Desplan, C. (2002). Extraction of functional binding sites from unique regulatory regions: the *Drosophila* early developmental enhancers. *Genome Res* 12, 470-81.

Petschek, J. P., Perrimon, N. and Mahowald, A. P. (1987). Region-specific defects in l(1)giant embryos of *Drosophila* melanogaster. *Dev Biol* 119, 175-89.

**Piano, F., Parisi, M. J., Karess, R. and Kambysellis, M. P.** (1999). Evidence for redundancy but not trans factor-cis element coevolution in the regulation of *Drosophila* Yp genes. *Genetics* 152, 605-16.

Postigo, A. A. and Dean, D. C. (1999). Independent repressor domains in ZEB regulate muscle and T-cell differentiation. *Mol Cell Biol* 19, 7961-71.

**Rajewsky, N., Vergassola, M., Gaul, U. and Siggia, E. D.** (2002). Computational detection of genomic *cis*-regulatory modules applied to body patterning in the early *Drosophila* embryo. *BMC Bioinformatics* **3**, 30.

Rivera-Pomar, R. and Jackle, H. (1996). From gradients to stripes in *Drosophila* embryogenesis: filling in the gaps. *Trends Genet* 12, 478-83.

Rossi, F. M., Kringstein, A. M., Spicher, A., Guicherit, O. M. and Blau, H. M. (2000). Transcriptional control: rheostat converted to on/off switch. *Mol Cell* 6, 723-8.

Rothe, M., Nauber, U. and Jackle, H. (1989). Three hormone receptor-like Drosophila genes encode an identical DNA-binding finger. Embo J 8, 3087-94.

Ryu, J. R. and Arnosti, D. N. (2003). Functional similarity of Knirps CtBP-dependent and CtBP-independent transcriptional repressor activities. *Nucleic Acids Res* 31, 4654-62.

Schaeper, U., Boyd, J. M., Verma, S., Uhlmann, E., Subramanian, T. and Chinnadurai, G. (1995). Molecular cloning and characterization of a cellular phosphoprotein that interacts with a conserved C-terminal domain of adenovirus E1A involved in negative modulation of oncogenic transformation. *Proc Natl Acad Sci U S A* 92, 10467-71.

**Carroll, S. B., Grenier, J. K., and Weatherbee, S. D.** (2001). From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design: Blackwell Science.

Shashikant, C. S., Kim, C. B., Borbely, M. A., Wang, W. C. and Ruddle, F. H. (1998). Comparative studies on mammalian Hoxc8 early enhancer sequence reveal a baleen whale-specific deletion of a *cis*-acting element. *Proc Natl Acad Sci U S A* 95, 15446-51.

Shi, Y., Sawada, J., Sui, G., Affar el, B., Whetstine, J. R., Lan, F., Ogawa, H., Luke, M. P. and Nakatani, Y. (2003). Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* 422, 735-8.

Simpson-Brose, M., Treisman, J. and Desplan, C. (1994). Synergy between the hunchback and bicoid morphogens is required for anterior patterning in *Drosophila*. Cell 78, 855-65.

Small, S., Arnosti, D. N. and Levine, M. (1993). Spacing ensures autonomous expression of different stripe enhancers in the even-skipped promoter. *Development* 119, 762-72.

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.

Small, S., Kraut, R., Hoey, T., Warrior, R. and Levine, M. (1991). Transcriptional regulation of a pair-rule stripe in *Drosophila*. *Genes Dev* 5, 827-39.

Small, S. and Levine, M. (1991). The initiation of pair-rule stripes in the Drosophila blastoderm. Curr Opin Genet Dev 1, 255-60.

Stanojevic, D., Hoey, T. and Levine, M. (1989). Sequence-specific DNA-binding activities of the gap proteins encoded by hunchback and Kruppel in *Drosophila*. *Nature* 341, 331-5.

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.

Stormo, G. D. (2000a). DNA binding sites: representation and discovery. *Bioinformatics* 16, 16-23.

Stormo, G. D. (2000b). Gene-finding approaches for eukaryotes. Genome Res 10, 394-7.

Stormo, G. D. (2000c). Identification of coordinated gene expression and regulatory sequences. *Pac Symp Biocomput*, 416-7.

Struhl, K. (2001). Gene regulation. A paradigm for precision. Science 293, 1054-5.

Strunk, B., Struffi, P., Wright, K., Pabst, B., Thomas, J., Qin, L. and Arnosti, D. N. (2001). Role of CtBP in transcriptional repression by the *Drosophila* giant protein. *Dev Biol* 239, 229-40.

Subramanian, T. and Chinnadurai, G. (2003). Association of class I histone deacetylases with transcriptional corepressor CtBP. *FEBS Lett* 540, 255-8.

Sundqvist, A., Sollerbrant, K. and Svensson, C. (1998). The carboxy-terminal region of adenovirus E1A activates transcription through targeting of a C-terminal binding protein-histone deacetylase complex. *FEBS Lett* **429**, 183-8.

**Thanos, D. and Maniatis, T.** (1995). Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell* **83**, 1091-100.

Turner, J. and Crossley, M. (2001). The CtBP family: enigmatic and enzymatic transcriptional co-repressors. *Bioessays* 23, 683-90.

Vinson, C. R., Conover, S. and Adler, P. N. (1989). A *Drosophila* tissue polarity locus encodes a protein containing seven potential transmembrane domains. *Nature* 338, 263-4.

Walters, M. C., Magis, W., Fiering, S., Eidemiller, J., Scalzo, D., Groudine, M. and Martin, D. I. (1996). Transcriptional enhancers act in cis to suppress position-effect variegation. *Genes Dev* 10, 185-95.

Wasserman, J. D., Urban, S. and Freeman, M. (2000). A family of rhomboid-like genes: *Drosophila* rhomboid-1 and roughoid/rhomboid-3 cooperate to activate EGF receptor signaling. *Genes Dev* 14, 1651-63.

Weinmann, A. S. and Farnham, P. J. (2002). Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods* 26, 37-47.

Weinmann, A. S., Yan, P. S., Oberley, M. J., Huang, T. H. and Farnham, P. J. (2002). Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev* 16, 235-44.

Weintraub, H. (1988). Formation of stable transcription complexes as assayed by analysis of individual templates. *Proc Natl Acad Sci U S A* **85**, 5819-23.

Wu, X., Vakani, R. and Small, S. (1998). Two distinct mechanisms for differential positioning of gene expression borders involving the *Drosophila* gap protein giant. *Development* 125, 3765-74.

Yan, R., Small, S., Desplan, C., Dearolf, C. R. and Darnell, J. E., Jr. (1996). Identification of a Stat gene that functions in *Drosophila* development. *Cell* 84, 421-30.

Yuh, C. H., Bolouri, H. and Davidson, E. H. (1998). Genomic *cis*-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science* 279, 1896-902.

Zhao, C., Dave, V., Yang, F., Scarborough, T. and Ma, J. (2000). Target selectivity of bicoid is dependent on nonconsensus site recognition and protein-protein interaction. *Mol Cell Biol* 20, 8112-23.

Zuo, P., Stanojevic, D., Colgan, J., Han, K., Levine, M. and Manley, J. L. (1991). Activation and repression of transcription by the gap proteins hunchback and Kruppel in cultured *Drosophila* cells. *Genes Dev* 5, 254-64.

## **Chapter II**

## Information Display by Transcriptional Enhancers

Kulkarni, M. M. and Arnosti, D. N. (2003). Information Display by Transcriptional

enhancer. Development 130, 6569-6575.

### ABSTRACT

Transcriptional enhancers integrate positional and temporal information to regulate the complex expression of developmentally controlled genes. Current models suggest that enhancers act as computational devices, receiving multiple inputs from activators and repressors and resolving them into a single positive or a negative signal that is transmitted to the basal transcriptional machinery. We show here that a simple, compact enhancer is capable of representing both repressed and activated states at the same time and in the same nucleus. This finding suggests that closely apposed factor binding sites, situated within compact *cis*-elements, can be independently interpreted by the transcriptional machinery, possibly through successive enhancer-promoter interactions. These results provide clear evidence that the computational functions usually ascribed to the enhancer itself are actually shared with the basal machinery. In contrast to the autonomous computer model of enhancer function, an information-display or "billboard" model of enhancer activity may better describe many developmentally regulated transcriptional enhancers.

### **INTRODUCTION**

Developmental programs of gene expression are controlled by "hard-wired" transcriptional circuits comprised of modular enhancers that communicate with basal promoter regions (Davidson, 2001). Prior studies in many systems have supported the general notion that an enhancer acts as an information-processing device, or computer,

receiving multiple inputs in the form of distinct transcription factors, both activators and repressors, that bind to it (Davidson, 2001; Ghazi and VijayRaghavan, 2000). The analogy of an enhancer as a computer is usually simply that of an element that sorts out inputs (processing) and resolves them into a single output that is instructive to the basal machinery, either turning the gene on or off. An important point is that computational functions - the decision to fire a promoter and at what level - have been ascribed to the enhancer. This is not to suggest that a given enhancer has only a single possible output; depending on signals or cell type, the same enhancer element might activate or repress, and the magnitude of activation signals can be variable, in the manner of a rheostat (Barolo and Posakony, 2002; Biggar and Crabtree, 2001; Rossi et al., 2000). However, it has been thought that enhancers do perform an integrative function and that in a particular nucleus, an enhancer represents a single information state at any given moment.

Such integrative functions have been ascribed to the human interferon- $\beta$  (IFN- $\beta$ ) enhancer, which drives transcription of the IFN- $\beta$  gene in response to viral infection (Struhl, 2001). The presence of each transcription factor binding site and its precise arrangement within the regulatory element are critical for the various regulatory proteins (sequence-specific activators and architectural proteins) to assemble through cooperative interactions into a well-defined nucleoprotein complex called the "enhanceosome". Assembly of the enhanceosome is essential for the transcription of the IFN- $\beta$  gene in response to viral infection in cells. In this structured element, the exact arrangement of factor binding sites is critical to dictating the output of the element, so the enhanceosome acts as a molecular computer, leading to a single output directed to the general machinery (Thanos and Maniatis, 1995; Kim and Maniatis, 1997; Munshi et al., 2001). Such a complex might provide a stereospecific interface for interaction with the basal transcriptional machinery, possibly engaging several components of the basal machinery simultaneously to effect synergistic activation (Carey et al., 1990; Chi et al., 1995). With such an enhancer, the target gene would be activated only upon the assembly of a "complete" complex, providing a precise on/off binary transcriptional switch in response to the appropriate stimulus.

Studies of developmentally regulated genes have also provided examples of enhancers as molecular computers. The developmentally regulated Drosophila evenskipped (eve) gene is regulated by developmental enhancers that are thought to act in a computational fashion. The reiterated stripe pattern of eve expression in the blastoderm embryo is generated by modular enhancers bound by broadly expressed transcriptional activators and regionally distributed repressors (Fujioka et al., 1999; Small et al., 1992; Small et al., 1996). These enhancers interpret gradients of regulatory factors and are active or inactive, depending on the particular set of regulatory proteins present in a given nucleus. The eve stripe 2 enhancer is active only in a narrow band of cells where activators Bicoid and Hunchback are present, but repressors Krüppel, Giant, and Sloppypaired are scarce or absent (Andrioli et al., 2002; Small et al., 1992). Key to the functional autonomy of the modular eve enhancers is the short-range of the repressors that regulate individual enhancers; for example, the short-range transcriptional repressor Krüppel bound to the stripe 2 enhancer in central regions of the embryo does not interfere with the activity of the adjacent eve stripe 3 enhancer (Small et al., 1993). An assumption is that each enhancer works as a single computational unit, not a redundant

set of independently acting elements. Consistent with this view is the finding that enhancer function is disrupted upon loss of a single activator or repressor site (Arnosti et al., 1996a; Small et al., 1992). However, these experiments have relied on minimal elements that may already represent a subset of the actual regulatory region (see Discussion).

A more detailed picture emerges from the functional dissection of the *endo 16 cis*regulatory region of *Strongylocentrotus purpuratus*. The *endo 16* gene is regulated during development by 2.3 kbp region containing binding sites for factors that contribute to distinct functions such as early widespread activation, late activation, repression of the early element, and potentiation of the repressor sites. Separate portions of the regulatory region can be combined to recreate some or all of the expression pattern, and models based on Boolean logical operators successfully simulate the output of these regulatory regions (Yuh et al., 1998). These studies emphasize the integrative, computer-like processing suggested to be a characteristic of developmental enhancers, and suggest that basal elements respond to signals generated by these molecular logic circuits.

In contrast to this view of the enhancer as an information-processing unit, we find that a single, compact enhancer can serve as an information display, representing on and off states, at the same time and in the same nucleus. This finding suggests that rather than acting as a computer that integrates various inputs, enhancers can simultaneously display both the active and repressed states, which maybe interpreted by successive or multiple, simultaneous interactions with the basal transcriptional machinery. In this case, the enhancer does not act in a concerted, computational fashion, and the basal transcriptional

59
machinery plays an active, rather than a passive, role in interpreting signals from the enhancer.

### **MATERIALS AND METHODS**

#### 1. Plasmid construction

### GAL4 (aa1-93) - GAL4 AD (aa753-881)

A KpnI-XbaI fragment from pSCTEV GAL4 (1-93)- GAL4 (Seipel et al., 1992) containing the reading frame for the yeast GAL4 activation domain (Gal4 AD) from amino acid residues 753-881, was cloned into KpnI-XbaI cut pTwiggy (Arnosti et al., 1996b) vector, which contains the *twist* enhancer (2xPEe-Et) element, *twist* basal promoter and the GAL4 DNA-binding domain from residues 1-93.

# Reporter genes.

The plasmid UAS-lacZ (Brand and Perrimon, 1993) was modified to contain two Giant sites (5' GGC CGC TAT GAC GCA AGA AGA CCC AGA TCT TTT TAT GAC GCA AGA GA 3') or two Knirps sites (5'GGC CGC ATC TGA TCT AGT TTG TAC TAG ACA TCT GAT CTA GTT TCA 3') twenty nucleotides upstream of the five GAL4 binding sites. The resulting vectors named M2g5u-lacZ or M2k5u-lacZ (Fig. 1C, D) respectively, consist of two Giant or Knirps binding sites, five tandemly arrayed GAL4 binding sites, followed by the *hsp 70* TATA box and transcriptional start driving *lacZ* expression. These reporters were further modified by introducing oligos containing two Twist (twi) and two Dorsal (dl) binding sites (Szymanski and Levine, 1995) at the *Not I* site upstream of the Giant or Knirps sites resulting in the **2twi.dl-M2g5u-lacZ** and **2twi.dl-M2k5u-lacZ** reporters (Fig. 1A, B, E and F).

The regulatory element from 2twi.dl-M2g5u-lacZ, containing two Twist sites, two Dorsal sites, two Giant sites and five GAL4 binding sites was introduced into the *EcoRI* site of the C4PLZ vector in both orientations. The C4PLZ vector lies between two divergently transcribed genes, the TATA-less *white* (*w*) gene and the *lacZ* gene. The *lacZ* gene is driven by the TATA containing *P* element *transposase* basal promoter (Fig. 2). Two additional Giant binding sites were introduced at the *SphI* site in the M2g5u-lacZ vector between the five GAL4 binding sites and the *hsp* 70 TATA box. The resulting vector was further modified by introducing oligos containing two Twist (twi) and two Dorsal (dl) binding sites (Szymanski and Levine, 1995) at the *Not I* site upstream of the Giant sites resulting in the **2twi.dl-M2g5u2g-lacZ** (Fig. 3).

# 2. P-element transformation, crosses to reporter genes, and whole-mount in situ hybridization of embryos.

P-element transformation vectors were introduced into the *Drosophila* germline by injection of  $yw^{67}$  embryos as described (Small et al., 1992). Embryos were collected either directly from each transgenic reporter line or from a cross between a reporter line and a line expressing the GAL4 activator in the ventral regions of the embryo. The embryos were fixed and stained using digoxigenin-UTP labeled antisense RNA probes to either *lacZ* or *w* as described (Small et al., 1992).

#### RESULTS

#### Limited ability of short-range repressors to block activators

The activity of developmental *cis*-regulatory elements has been studied mostly in the context of complex endogenous enhancers (Arnosti et al., 1996a; Grav et al., 1994; Kosman and Small, 1997; Small et al., 1993; Small et al., 1992). This approach is complicated by the functional complexity of many *cis*-regulatory elements where the identity and/or the stoichiometry of transacting factors is not always well defined. To analyze enhancer function in a setting in which activator-repressor stoichiometry and spacing can be exactly defined, we constructed chromosomally integrated, compact regulatory elements containing binding sites for endogenous short-range repressors Giant or Knirps, endogenous activators Twist and Dorsal, and chimeric Gal4 activators. The space between repressor and activator sites on these elements is less than 100 bp, a distance over which short-range repressors have been previously shown to be effective (Arnosti et al., 1996b; Gray et al., 1994; Hewitt et al., 1999; Keller et al., 2000). Twist and Dorsal drive gene expression in a ventral swathe approximately 22-24 cell in width, while the Gal4 activator protein, expressed under the control of the twist enhancer, drives reporter gene expression in a narrower 18-20 cell wide pattern. The protein product of the gap gene giant is present in broad anterior and posterior stripes, while the Knirps protein is present in a broad posterior stripe and more anterior regions in the early embryo. As anticipated, Giant and Knirps mediate repression of adjacent Dorsal and Twist activators, eliminating expression of the lacZ reporter gene in portions of the embryo where these repressor proteins are localized (Figure II-1A, B). Strikingly however, Giant and Knirps

are unable to repress an element containing five Gal4 activator sites, although these proteins also bind within 100 bp of the repressors, revealing a hitherto unknown limitation of short-range repressors (Figure II-1C, D). This lack of repression is not due to an inherent resistance of the Gal4 activation domain to repression, for Knirps and Giant can effectively repress an element containing only three Gal4 binding sites (M. Kulkarni, unpublished).

### Simultaneous repression and activation

When Gal4 activators are combined with Dorsal and Twist activators on a composite element, strongly enhanced staining is noted in the central regions of the embryo, indicative of additive or synergistic activation. In the regions of the embryo containing the repressors Giant or Knirps, the width of the nuclei stained (a 18-20 swathe of nuclei) is the same as the pattern of staining driven by the Gal4 protein alone. We conclude that in nuclei containing Giant and Knirps protein, the pattern of staining directed by Dorsal and Twist is being selectively repressed by the short-range repressors, while transcription driven by Gal4 (a narrower 18-20 nuclei swathe) is unimpeded (Figure II-1E, F). The pattern of gene expression indicates that, in nuclei where the activators and repressors are co-expressed, transcription is driven by one cluster of activators within the compact regulatory element, while at the same time other activators within the same element are being actively repressed by Giant or Knirps. This compact regulatory element therefore, has subelements that represent both "active" and "inactive"

# Figure II-1: Simultaneous repression and activation from a compact regulatory element.

(A) Knirps repression of adjacent Dorsal (dl) and Twist (twi) activators. Dorsal and Twist proteins, normally active in a broad (22-24 nuclei) ventral swathe of the blastoderm embryo, fail to activate a linked *hsp70 lacZ* transgene in regions containing Knirps (kni) protein (arrow).

(B) Giant repression of Dorsal and Twist. Repression is seen in anterior and posterior regions where the Giant (gt) repressor is expressed (arrows). (C, D) Gal4 activators, expressed in a narrower (18-20 nuclei) ventral swathe, are not inhibited by Knirps and Giant.

(E) Composite element containing Dorsal, Twist, and Gal 4 activators exhibits repression of Dorsal and Twist by Knirps, while narrower Gal4 driven expression pattern is unaffected.

(F) Composite element with Dorsal, Twist, and Gal4 activators, and Giant repressor, exhibits similar complex expression pattern.

(G) A similar pattern of selective repression of the Dorsal and Twist activators within the composite element used in 1F is seen when the activator Gal4 is driven throughout the embryo under the control of the *nanos* promoter (NGT40, Bloomington Stock #4442). In the central regions of the embryo more intense staining is visible, indicative of additive or synergistic gene activation by Dorsal, Twist and Gal4. In the regions of the embryo where the repressor Giant is expressed (arrows), the intensity of *lacZ* staining is the same as in the dorsal regions of the embryo where activation is driven by Gal4 alone.

The difference in lacZ staining intensity between cells containing or lacking Giant or Knirps is due to a difference in intensity in each cell, not the number of cells stained. Patterns of gene expression were visualized in 2-4 hour embryos by in situ hybridization with digU labeled antisense lacZ probes. Embryos are oriented anterior to left; ventrolateral views (A, B, C, D, E and, G) and ventral view (F) are shown.

Figure II-1: Simultaneous repression and activation from a compact regulatory element



states simultaneously, unlike the binary switch activity observed for many enhancers, where it appears that a single signal to activate or repress is present.

We make this conclusion based on the activity of the elements when only one set of activators is present (Figure II-1A-D), and on the characteristic narrower pattern driven by the Gal4 activators. Consistent with this conclusion, a similar pattern of exclusive repression of the Dorsal and Twist activators is seen when expression of Gal4 is driven in a ubiquitous pattern using the *nanos* promoter (Tracey et al., 2000). Here, we can compare promoter activity with Gal4 alone or in combination with Dorsal/Twist (Figure II-1G). In dorsal regions of the embryo, the only activator on the element is Gal4, and no repression by Giant is visible. In the ventral regions where Dorsal and Twist are present, but the repressor is absent, more intense staining is seen, consistent with synergistic or additive activation. Importantly, in the ventral regions also containing the Giant repressor (Figure II-1G, arrows), *lacZ* expression is similar to that observed in the dorsal regions of the embryo. This result indicates that Dorsal and Twist are not working together with Gal4, but are functionally independent and selectively repressed in the regulatory element.

#### Compact element functions in a distance, orientation, promoter-independent manner

To further evaluate the properties of this element, we tested whether it possessed classical characteristics of a transcriptional enhancer, namely, acting in a distance- and orientation-independent manner (Banerji et al., 1981). The element containing Giant binding sites was placed in either orientation between the divergently transcribed *white* (at -265 bp) and *lacZ* genes (at -130 bp). In both orientations tested, this element directed *white* expression from -265 bp in a manner closely resembling that seen for the *hsp 70 lacZ* reporter; Giant efficiently repressed Dorsal and Twist, while Gal4 activated transcription in a continuous ventral swathe (Figure II-1B, F; Figure II-2A, B, D, and E). A similar pattern of repression and activation is seen with the *transposase lacZ* gene (Figure II-2C, F). The identical results observed in Figure II-1F and Figure II-2B, C, E, and F indicate that the specific patterns of activation and repression are not dependent on the particular promoter context or orientation of activators and repressors.

#### See Appendix A

#### Conversion of enhancer output to a binary on/off state

The compact regulatory element assayed in Figure II-1 and Figure II-2 fits the classical definition of an enhancer, functioning in a distance- and orientation-independent manner. In addition, the size of this element resembles that of naturally occurring enhancers ~ 200-800 bp in length. However, the element does not function in the biphasic either "on or off" mode normally thought to be a characteristic feature of enhancers. We are unaware of documented cases where a single enhancer displays two different states at the same time and in the same nucleus, thus this dual activity appears to be unusual. It is possible that rather than being an inherent functional property of enhancers, the uniform output of enhancers might reflect evolutionary pressure to arrange repressor and activator binding sites to optimize a consistent output. To simulate this

# Figure II-2: Compact regulatory element displays enhancer-like properties of distance and orientation independence.

(A, D) The regulatory element shown in Figure II-1F was inserted in either orientation into a vector containing divergently transcribed *white* and *transposase lacZ* reporter genes. When situated at -265 bp, Dorsal/Twist activators within the element drive expression of the *white* reporter gene. Repression by Giant is evident in anterior and posterior regions (arrows).

(**B**, **E**) In the presence of Dorsal, Twist, and Gal4 activators, a composite pattern of gene regulation is seen as in Fig. 1**F** with inhibition of Dorsal/Twist and activation by Gal4.

(C, F) A similar expression pattern is observed with the divergently transcribed transposase lacZ promoter, with repression by Giant of Dorsal/Twist and activation by Gal4.

Embryos are oriented anterior to left; lateral views (A, D) and ventrolateral views (B, C, E, and F) are shown.

Figure II-2: Compact regulatory element displays enhancer-like properties of distance and orientation independence.



situation, two additional Giant repressor binding sites were introduced into this element 3' of the Gal4 binding sites. Now, complete loss of staining is evident in nuclei containing the Giant protein (arrows) yielding a classical biphasic "on or off" state (Figure II-3A, B, C, and D).

#### DISCUSSION

#### **Redundancy in enhancer function**

If an enhancer were an indivisible unit of transcriptional regulation, the functional independence of adjacent binding sites within the composite element (Figure II-1E, F) would suggest that this compact element is in fact two separate enhancers. However, this element is of similar size to natural enhancers and does conform to the classical definition of an enhancer, namely a compact element that functions to regulate transcription in a position and orientation independent manner (Banerji et al., 1981).

Functional analyses of *cis*-regulatory regions provide evidence for redundancy and hence divisibility, of natural enhancers, suggesting that they can also contain multiple, independently acting subelements. In the viral setting, the well-studied SV40 enhancer comprises two independently acting subelements that can be separately assessed (Herr and Clarke, 1986). In *Drosophila*, recent evidence suggests that *eve* enhancers possess redundant activities. Deletion of the entire 480 bp *eve* stripe 2 element within the *eve* locus fails to completely abrogate stripe 2 expression (M. Kreitman, personal

# Figure II-3: Conversion of a multiple state element to a binary on/off switch.

Two additional Giant binding sites were introduced at the 3' end of the Gal4 activator cluster.

(A, B) As observed previously, Dorsal/Twist activators are repressed in anterior and posterior regions of Giant expression (arrows).

(C, D) In the presence of Dorsal/Twist and Gal4 activators, complete repression of transcription is observed in areas of Giant expression (arrows).

Embryos are oriented anterior to the left. Lateral (A, C) and ventral (B, D) views are shown.





communication) indicating the presence of redundant regulatory sequences in the locus. Furthermore, tissue-specific expression of the yolk protein genes yp1 and yp2, is supported by flanking sequences after deletion of the 125 bp yolk protein enhancer (Piano et al., 1999). The resilience of natural enhancers to loss of single binding sites further supports the notion that these elements are built of redundantly acting sequences (Arnosti, 2003).

#### Selection for uniformity of enhancer output

A scenario of an enhancer with simultaneously displayed activation and repression states is reminiscent of the modular, autonomous pair rule stripe enhancers, such as even-skipped stripe elements, where separate enhancers represent different "states" of repression and activation in the same nucleus (Gray and Levine, 1996). An important distinction is that our findings suggest that a similar discrimination is taking place within the tight confines of a single enhancer, and that in order to establish a uniform signal output, enhancers require a proper stoichiometry or distribution of repressor and activator binding sites to ensure that all possible enhancer subelements provide the same information (Figure II-4). Indeed a distributed pattern of short-range transcriptional repressor binding sites is typical of many developmental enhancers that function in the early Drosophila embryo; this configuration would allow repressors to block multiple modes of enhancer-promoter interactions (La Rosee et al., 1997; Small et al., 1992; Small et al., 1996). In this study we actually measure the simultaneous independent activity of sub-elements (Figure II-1, Figure II-2) and show that they can be deployed to give a unitary response (Figure II-3) as is seen with natural enhancers. Thus,

the carefully designed internal organization of *cis*-regulatory modules can provide uniform information that closely simulates an integrative information processing capacity.

In contrast to the precision of the enhanceosome, a more flexible arrangement of regulatory proteins has been suggested to be the predominant pattern for elements that provide diverse patterns in developing systems (Struhl, 2001). Evolutionary and experimental studies of the *eve* stripe 2 enhancer suggest that this element can tolerate and has undergone considerable rearrangement, with great flexibility in the number and arrangement of individual sites (Arnosti et al., 1996a; Ludwig et al., 2000; Ludwig and Kreitman, 1995; Ludwig et al., 1998). For example, the recent acquisition of a strong Bicoid activator site appears to have been counterbalanced by the closer apposition of a nearby Giant binding site (Hewitt et al., 1999; Ludwig et al., 1998). The plasticity of this enhancer suggests that much variation in spatial placement of individual transcription factors is possible, consistent with a model in which these factors contact the basal machinery in a flexible framework, not necessarily as a rigid complex.

With such flexibility, the transcription factors of an enhancer might still engage the transcriptional machinery in simultaneous cooperative interactions, as is suggested with enhanceosomes. However, our studies suggest that an individual enhancer is capable of representing both the state of activation and repression, suggesting that the basal machinery may "sample" discrete regions, comprising a small number of transcription factor binding sites, within the enhancer (Figure II-4B, C). Successive interactions with the basal machinery, and the biochemical consequence of these multiple

74

#### Figure II-4: Enhanceosome versus Information Display enhancer models.

(A) In the enhanceosome model, the enhancer serves as an information processing center, receiving inputs from multiple transcription factors that bind it. A highly structured complex or enhanceosome, creates a stereospecific interface for docking with and recruiting the basal transcription machinery. Here the enhancer serves as a molecular computer, resolves multiple inputs and provides a single output to the basal transcription machinery. With such an enhancer, the target gene would be activated only upon the assembly of a complex, providing a precise on/off binary transcriptional switch in response to the appropriate stimulus. Graded responses from such an element could be achieved by varying the stability of the entire complex, possibly in response to activator concentrations.

 $(\mathbf{B}, \mathbf{C})$  Information Display or "Billboard" enhancer. Rather than acting as a central processing unit, subelements can display contrasting information, which is then interpreted by basal transcription machinery. In this model, the basal machinery "samples" discrete regions of the enhancer each comprising of a small number of transcription factor binding sites, either iteratively (**B**) or simultaneously (**C**). Successive/multiple interactions with the basal machinery, and the biochemical consequence of these interactions, would dictate the overall output of the enhancer.



Figure II-4: Enhanceosome versus Information Display enhancer models.

interactions would dictate the overall output of the enhancer (Figure II-4B). Alternatively the enhancer may engage in multiple, simultaneous contacts with some or all of the enhancer bound proteins, with repressors such as Giant and Knirps preventing some of these interactions (Figure II-4C). In either case, multiple iterative sampling of the enhancer, or simultaneous readout, the enhancer would function as an information display element with computation at the level of enhancer-promoter interactions.

Our results suggest that a closer examination of enhancer classifications is warranted. The terms enhancer and enhanceosome are frequently used interchangeably to denote a complex of DNA bound regulatory proteins, yet there appear to be important functional distinctions between enhanceosomes, as typified by the IFN-B enhancer, and other regulatory elements. In the light of the functional differences outlined above, a distinction should be made between the terms enhanceosome, which requires the cooperative assembly of a higher order structure within an enhancer, and other cisregulatory elements that may or may not function in this manner. We propose a model, the information display or "billboard" model for enhancer action, in which an enhancer, rather than acting as a central processing unit, can display contrasting information, which is then interpreted by basal transcriptional machinery (Figure II-4B, C). The binary "on or off' decisions that appear to be transmitted by the enhancer to the basal machinery actually result from the basal machinery reading a series of redundant signals encoded within the enhancer. The model does not explicitly describe the molecular mechanisms of repression and activation, but direct contacts between the Drosophila activators used here and components of the basal machinery are supported by biochemical studies (Koh et al., 1998; Pham et al., 1999; Yuh et al., 1998; Zhou et al., 1998)

The billboard enhancer model appears to more accurately describe many developmentally regulated enhancers, whose internal architecture is subject to rapid evolutionary change, even as the overall output remains constant (Ludwig et al., 2000; Ludwig et al., 1998). Although studies such as those on the IFN- $\beta$  gene indicate that cells may commonly use enhanceosomes to achieve regulatory precision in gene expression, it is likely that eukaryotic organisms use the "billboard" type of enhancers to achieve diversity in gene expression patterns and evolutionary flexibility.

#### ACKNOWLEDGEMENTS

We thank R. W. Henry, S.J. Triezenberg, L. Kroos, and three anonymous reviewers for helpful suggestions on the manuscript, and E. Fernandez-Villatoro for technical assistance. This work was supported by grant GM56976 from the National Institutes of Health to D.N.A.

# REFERENCES

Andrioli, L. P., Vasisht, V., Theodosopoulou, E., Oberstein, A. and Small, S. (2002). Anterior repression of a Drosophila stripe enhancer requires three position-specific mechanisms. *Development* 129, 4931-40.

Arnosti, D. N., Barolo, S., Levine, M. and Small, S. (1996a). The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* 122, 205-14.

Arnosti, D. N., Gray, S., Barolo, S., Zhou, J. and Levine, M. (1996b). The gap protein knirps mediates both quenching and direct repression in the Drosophila embryo. *Embo J* 15, 3659-66.

Arnosti, D. N. (2003). Analysis and function of transcriptional regulatory elements: Insights from Drosophila. *Annu Rev Entomol* 48, 579-602.

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

**Barolo, S. and Posakony, J. W.** (2002). Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. *Genes Dev* 16, 1167-81.

**Biggar, S. R. and Crabtree, G. R.** (2001). Cell signaling can direct either binary or graded transcriptional responses. *Embo J* 20, 3167-76.

Brand, A. H. and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118, 401-15.

Carey, M., Lin, Y. S., Green, M. R. and Ptashne, M. (1990). A mechanism for synergistic activation of a mammalian gene by GAL4 derivatives. *Nature* 345, 361-4.

Chi, T., Lieberman, P., Ellwood, K. and Carey, M. (1995). A general mechanism for transcriptional synergy by eukaryotic activators. *Nature* 377, 254-7.

**Davidson, E. H.** (2001). Genomic Regulatory Systems: Development and Evolution: Academic Press: A Harcourt Science and Technology Company.

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. Ghazi, A. and VijayRaghavan, K. V. (2000). Developmental biology. Control by combinatorial codes. *Nature* 408, 419-20.

Gray, S. and Levine, M. (1996). Transcriptional repression in development. Curr Opin Cell Biol 8, 358-64.

Gray, S., Szymanski, P. and Levine, M. (1994). Short-range repression permits multiple enhancers to function autonomously within a complex promoter. *Genes Dev* 8, 1829-38.

Herr, W. and Clarke, J. (1986). The SV40 enhancer is composed of multiple functional elements that can compensate for one another. *Cell* 45, 461-70.

Hewitt, G. F., Strunk, B. S., Margulies, C., Priputin, T., Wang, X. D., Amey, R., Pabst, B. A., Kosman, D., Reinitz, J. and Arnosti, D. N. (1999). Transcriptional repression by the Drosophila giant protein: cis element positioning provides an alternative means of interpreting an effector gradient. *Development* 126, 1201-10.

Keller, S. A., Mao, Y., Struffi, P., Margulies, C., Yurk, C. E., Anderson, A. R., Amey, R. L., Moore, S., Ebels, J. M., Foley, K. et al. (2000). dCtBP-dependent and - independent repression activities of the Drosophila Knirps protein. *Mol Cell Biol* 20, 7247-58.

Kim, T. K. and Maniatis, T. (1997). The mechanism of transcriptional synergy of an in vitro assembled interferon-beta enhanceosome. *Mol Cell* 1, 119-29.

Koh, S. S., Ansari, A. Z., Ptashne, M. and Young, R. A. (1998). An activator target in the RNA polymerase II holoenzyme. *Mol Cell* 1, 895-904.

Kosman, D. and Small, S. (1997). Concentration-dependent patterning by an ectopic expression domain of the Drosophila gap gene knirps. *Development* 124, 1343-54.

La Rosee, A., Häder, T., Taubert, H., Rivera-Pomar, R. and Jäckle, 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.

Ludwig, M. Z. and Kreitman, M. (1995). Evolutionary dynamics of the enhancer region of even-skipped in Drosophila. *Mol Biol Evol* 12, 1002-11.

Ludwig, M. Z., Patel, N. H. and Kreitman, M. (1998). Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change. *Development* 125, 949-58.

Ludwig, M. Z., Bergman, C., Patel, N. H. and Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403, 564-7.

Munshi, N., Agalioti, T., Lomvardas, S., Merika, M., Chen, G. and Thanos, D. (2001). Coordination of a transcriptional switch by HMGI(Y) acetylation. *Science* 293, 1133-6.

**Pham, A. D., Muller, S. and Sauer, F.** (1999). Mesoderm-determining transcription in Drosophila is alleviated by mutations in TAF(II)60 and TAF(II)110. *Mech Dev* 84, 3-16.

**Piano, F., Parisi, M. J., Karess, R. and Kambysellis, M. P.** (1999). Evidence for redundancy but not trans factor-cis element coevolution in the regulation of Drosophila Yp genes. *Genetics* 152, 605-16.

Rossi, F. M., Kringstein, A. M., Spicher, A., Guicherit, O. M. and Blau, H. M. (2000). Transcriptional control: rheostat converted to on/off switch. *Mol Cell* **6**, 723-8.

Seipel, K., Georgiev, O. and Schaffner, W. (1992). Different activation domains stimulate transcription from remote ('enhancer') and proximal ('promoter') positions. *Embo J* 11, 4961-8.

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., Arnosti, D. N. and Levine, M. (1993). Spacing ensures autonomous expression of different stripe enhancers in the even-skipped promoter. *Development* 119, 762-72.

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.

Struhl, K. (2001). Gene regulation. A paradigm for precision. Science 293, 1054-5.

Szymanski, P. and Levine, M. (1995). Multiple modes of dorsal-bHLH transcriptional synergy in the Drosophila embryo. *Embo J* 14, 2229-38.

Thanos, D. and Maniatis, T. (1995). Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell* 83, 1091-100.

Tracey, W.D., Ning, X., Klingler, M., Kramer, S.G., and Gergen, J.P. (2000) Quantitative analysis of gene function in the *Drosophila* embryo. Genetics 2000 154, 273--284.

Yuh, C. H., Bolouri, H. and Davidson, E. H. (1998). Genomic *cis*-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science* 279, 1896-902.

Zhou, J., Zwicker, J., Szymanski, P., Levine, M. and Tjian, R. (1998). TAFII mutations disrupt Dorsal activation in the Drosophila embryo. *Proc Natl Acad Sci U S A* **95**, 13483-8.

# Chapter III<sup>1</sup>

# **Operating Principles of Short-range Transcriptional Repressors**

#### **INTRODUCTION**

The identification and characterization of functionally significant noncoding regions, especially those that control transcription, is one of the major challenges in understanding the regulatory language encoded in genome sequences. Although we now have genome sequences for many metazoans, our understanding of how this regulatory information is encoded is extremely limited.

The *cis*-regulatory information that orchestrates complex spatial and temporal patterns of gene expression in development of higher eukaryotes is typically organized into modular units termed enhancers or *cis*-regulatory modules, of a few hundred base pairs in size. A common feature of these modules is the presence of multiple binding sites for several distinct transcription factors including sequence-specific activators and repressors (Davidson, 2001; Guss et al., 2001; Carroll, S.B, 2001). The transcription factors in turn interact with each other, with cofactors, chromatin and with the basal transcriptional machinery to either activate or repress the corresponding target gene. Because a *cis* element can contain multiple copies of the same or different binding motifs, a limited number of transcription factors can be arranged in numerous possible combinations that can result in distinct transcriptional outputs. Such combinatorial action by transcription factors can confer temporal and spatial specificity (Struhl, 1991).

<sup>&</sup>lt;sup>1</sup> The data in Chapter 3 is presented in the form of a manuscript to be submitted soon.

Identification of a particular regulatory module within a promoter region often reveals little about its role in the expression of a given gene. Therefore empirical tests have been extensively employed to decipher how various regulatory elements within a promoter work together to modulate transcription. In such assays a normal or modified region of a promoter is fused to a reporter gene and introduced into a cell or whole organism, where it can be exposed to the shifting array of transcription factors that normally modulate the expression of the endogenous gene. The resulting pattern of reporter gene expression can reveal, for instance, whether a particular regulatory element can activate or repress transcription at a specific time and place. Because of the complexity of most regulatory regions, multiple experiments of this kind are needed to gain even a rough overview of how an expression pattern is generated (Arnosti et al., 1996a; Goto et al., 1989; Harding et al., 1989; Small et al., 1993; Small et al., 1996; Small et al., 1991; Yuh et al., 1998). As a result relatively few enhancers in all animal systems combined have been well characterized, and our understanding of general principles dictating interactions of *cis*-regulatory elements with each other and with the basal promoter remain elusive.

Recently, whole genome sequence assemblies have become available providing a powerful foundation to identify and analyze *cis*-regulatory modules function and organization on a global scale. Two current approaches to identify candidate regulatory regions from genomic data are computational methods that look for clusters of transcription factor binding sites (Berman et al., 2002; Markstein and Levine, 2002; Markstein et al., 2002; Rebeiz et al., 2002) and phylogenetic comparisons that identify

84

evolutionarily conserved sequences (Bergman and Kreitman, 2001; Maier et al., 1990). One main difficulty with the output from computational searches for clusters of transcription factor sites is the large number of false-positive results. The short length and degenerate nature of transcription-factor-binding sites account for most of these misleading predictions. Furthermore, genes are rarely controlled by a single transcription factor and accumulating evidence suggests that specific combinations of transcription factors are required to achieve the complex differential expression of genes in higher organisms (Davidson, 2001; Gray et al., 1994; Hewitt et al., 1999; Keller et al., 2000; Struhl, 2001; Yuh et al., 1998). The structural basis for combinatorial regulation exists in the specific organization of multiple transcription factor binding sites (Struhl, 2001). In cases where transcription factors bind cooperatively to form a higher order nucleoprotein complex that activates gene expression, spatial correlations or constraints between the binding sites have been observed (Kim and Maniatis, 1997; Merika and Thanos, 2001; Munshi et al., 2001; Thanos and Maniatis, 1995). Even in cases where a higher order structure is not required to regulate gene expression, spacing between binding sites with cis-elements can be of utmost important. For example, with short-range transcriptional repressors, spacing between the repressor and activator sites within the cis-regulatory element is critical in dictating repression effectiveness (Arnosti et al., 1996b; Gray et al., 1994; Hewitt et al., 1999; Keller et al., 2000). Thus, in many cases where cis-regulatory modules are predicted by computational methods to exist, something in the arrangement of sites (or a wider context) renders the *cis*-element non-functional. In order to achieve better predictions and eliminate false-positive and false-negative results, computational methods should include, in addition to binding site density and relative affinities, other

parameters for spacing and position of binding motifs within *cis*-elements in the search algorithms.

Alternatively, evolutionary conservation can often be used to indicate the existence of regulatory regions. Interspecific sequence comparisons (phylogenetic footprinting) of noncoding regions reveal conserved features, many of which are *cis*-regulatory elements. However, despite obvious indications of selective constraints (Hardison et al., 1997; Loots et al., 2000) the structure and sequences of *cis*-elements change over time, sometimes dramatically even in cases where expression patterns are conserved (Ludwig et al., 2000; Ludwig and Kreitman, 1995; Piano et al., 1999). Thus, phylogenetic comparisons that identify *cis*-regulatory elements would be greatly facilitated by empirical determination of spatial constraints between binding sites of *cis*-regulatory elements or between different elements themselves.

The early *Drosophila* embryo has provided a paradigmatic model for studying transcriptional control of development. Most of the important factors have been identified by exhaustive genetic analysis (Nusslein-Volhard et al., 1985; Nusslein-Volhard and Wieschaus, 1980), and there are sophisticated tools for characterizing the design and function of complex *cis*-regulatory DNA elements. Typical of complex regulatory systems in higher eukaryotes, the *cis*-regulatory elements of key patterning genes employ extensive sequences to translate broad patterns of maternal and early embryonic factors into precisely defined segmental distributions of transcription factors (Arnosti, 2002; Driever and Nusslein-Volhard, 1988; Driever et al., 1989; Nusslein-Volhard et al., 1987;

86

Pankratz and Jackle, 1990; Rivera-Pomar and Jackle, 1996). One of the best-studied complex loci in Drosophila is the pair-rule gene even-skipped, which encodes a homeodomain transcription factor that is expressed in a series of seven transverse stripes across the anterior-posterior length of the blastoderm embryo. This gene has several cisregulatory elements scattered within a 16 Kb region around the eve coding sequence. Five of these regions, the stripe enhancer elements, are responsible for early expression of eve in the form of seven regularly spaced stripes in the blastoderm embryo (Fujioka et al., 1999; Goto et al., 1989; Harding et al., 1989; Sackerson et al., 1999; Small et al., 1992; Small et al., 1996; Small et al., 1991; Small and Levine, 1991; Stanojevic et al., 1991). One key feature of these enhancers is that their action is autonomous; the repression of one element does not lead to the general repression of the entire locus (Arnosti et al., 1996b; Gray et al., 1994; Small et al., 1993). This autonomy is based on the properties of the short-range transcriptional repressors that regulate these enhancers and include the products of the gap genes, giant, knirps and krüppel. These transcriptional repressors block the activity of enhancer elements when bound within  $\sim 100$  bp of key activator sites or of basal promoter elements when cognate sites are introduced close to the start of transcription. The short-range of the repression activity provides a highly flexible, yet precisely tunable mechanism for specific gene regulation (Arnosti et al., 1996a; Arnosti et al., 1996b; Gray and Levine, 1996; Gray et al., 1994; Hewitt et al., 1999; Keller et al., 2000; Strunk et al., 2001). This flexibility contrasts with long-range repressors, typified by the Drosophila Hairy protein, which can block multiple enhancers over distances of several kilobases, regardless of location within a gene complex. The different activities of short- and long-range repressors probably reflect distinct mechanisms employed by

transcriptional repressors. The mechanisms by which short-range and long-range Drosophila repressors inhibit transcription are poorly understood, although one model of repression in the embryo suggests that the short-range/long-range distinction results from the recruitment of distinct cofactors (Nibu et al., 1998a; Nibu et al., 1998b; Zhang and Levine, 1999). Short-range repressors recruit the corepressor CtBP to mediate repression, whereas long-range repressors have been shown to interact with the Groucho corepressor (Barolo and Levine, 1997; Chen and Courey, 2000; Fisher and Caudy, 1998; Mannervik and Levine, 1999; Mannervik et al., 1999; Poortinga et al., 1998). One model for Groucho mediated long-range repression is through the recruitment of HDACs by Groucho, resulting in the production of a large transcriptionally silent chromosomal domain. Just as the Sir repressosome generates a transcriptionally silent chromatin structure that is able to spread along the chromatin fiber, it has been proposed that the Groucho protein nucleates a silenced chromosomal state that spreads to mediate longrange repression (Brantjes et al., 2001; Chen et al., 1999; Flores-Saaib and Courey, 2000).

Another striking feature of the *eve* stripe enhancers is that they contain multiple binding sites for both activators and transcriptional repressors. The best characterized *eve* enhancer controls the expression of *eve* stripe 2 and contains a total of eighteen known factor-binding sites, including eight activator (three for Hunchback, five for Bicoid) binding sites and ten repressor (three for Giant, six for Krüppel, and at least one for Sloppy-paired) binding sites (Andrioli et al., 2002; Ludwig et al., 1998; Small et al., 1991; Small and Levine, 1991; Stanojevic et al., 1991). It has been proposed that a high local density of transcription factor binding sites may be sufficient for the proper function of these cis-elements and can be used as a convenient signpost for computational identification novel *cis*-elements. However, several attempts to construct artificial stripe 2 enhancers by multimerizing Bicoid, Hunchback activator with Krüppel and Giant repressor sites have been unsuccessful (S. Small and M. Levine, unpublished). These results suggest that *cis*-regulatory sequences other than the known sites are important for activation. Such sequences may contain sites for other activator proteins. Alternatively, they may simply function to provide the correct spacing between known activator or repressor sites. Thus, the grammar of the *cis*-regulatory code is clearly more complex than simply the density of transcription factor binding sites. The relative affinity, spacing, and positioning of transcriptional activator and repressor binding sites within cisregulatory modules has been demonstrated to be significant in many cases (Arnosti et al., 1996a; Courey and Huang, 1995; Hanes et al., 1994; Hewitt et al., 1999; Lifanov et al., 2003; Szymanski and Levine, 1995). Previous analyses of short-range repressors on native enhancers had demonstrated that these proteins can block gene expression when within  $\sim 100$  bp of the target –either activators within an enhancer element or the core promoter (Arnosti et al., 1996b; Hewitt et al., 1999; Keller et al., 2000). However, since such studies have focused on the activity of short-range repressors in the context of complex, endogenous regulatory elements, the distance requirements have not been exactly defined as the positions of the nearest activators within the element is not known.

Using synthetic enhancer elements where the identity, stoichiometry and the exact arrangement of activator and repressor binding site are well-defined, we demonstrate that the previously held simple notion that short-range repressors block the activity of all protein complexes within a 100 bp is incorrect. The manipulation of these composite elements in terms of the number of activator and repressor binding sites, relative affinities, spacing and distribution of these binding sites further allowed us to define the contextual parameters that dictate repression effectiveness and include architectural features such as stoichiometry of activators and repressors, relative affinity, spacing and position of binding sites. These elements constitute the 'grammar' of short-range repressor function, and the empirical identification of such rules for different classes of transcriptional regulators will facilitate computational searches for *cis*-elements regulated by them.

#### **MATERIALS AND METHODS**

#### 1. Gal4- activator chimeric constructs

#### GAL4 (aa1-93) - GAL4 AD (aa753-881)

A KpnI-XbaI fragment from pSCTEV GAL4 (1-93)- GAL4 (Seipel et al., 1992) containing the reading frame for the yeast GAL4 activation domain (Gal4 AD) from amino acid residues 753-881, was cloned into KpnI-XbaI cut pTwiggy (Arnosti et al., 1996a) vector, which contains the *twist* enhancer (2xPEe-Et) element, *twist* basal promoter and the GAL4 DNA-binding domain from residues 1-93.

#### Insulated GAL4 (aa1-93) - GAL4 AD (aa753-881)

The following primers with *EcoRI* ends were used to PCR amplify a 420 bp fragment of DNA containing the gypsy insulator from the Green Pelican GFP transformation vector (Barolo et al., 2000).

DA639: 5' CGG AAT TCC GAA TTG TAA GCG TTA ATG ACT 3'

DA640: 5' CGG AAT TCC GAT ACA TAC TAG AAT TGA TCG 3'

The fragment containing the gypsy insulator was cloned into the pTwiggy transformation vector at the *EcoRI* site between the *twist* regulatory elements and the *white* gene. The vector was further modified to contain the reading frame for the yeast GALA activation domain (Gal4 AD) from amino acid residues 753-881 as a *KpnI-XbaI* fragment.

#### GAL4 (aa1-93) - VP16 (aa412-490)

The C-terminal transcriptional activation domain of the herpes simplex virus VP16 protein from residues 412-490 was amplified from pRevTet off (Ryu et al., 2001) which contains the bacterial Tet repressor DNA binding domain fused to the VP16 activation domain using the following primers-

DA410: 5' GGG TCG GTA CCG CAA CGG CCC CCC CGA CCG ATG TC 3'

DA411: 5' GGG GAA TCT AGA CTA ACT AAT TAC TAC CCA CCG TAC TCG TCA AT 3'

The PCR product was digested with KpnI and XbaI enzymes and cloned into KpnI- XbaI cut pTwiggy vector (Arnosti et al., 1996a).

#### GAL4 (aa1-93) - Sp1 (aa132-243)

A KpnI - XbaI fragment from pSCTEV Gal4 (1-93)-Sp1: Q1 (Seipel et al., 1992) containing the open reading frame for the activation domain (residues 132-243) of the transcription factor Sp1 was cloned into KpnI - XbaI cut pTwiggy vector (Arnosti et al., 1996a).

# GAL4 (aa1-93) - hTBP (aa1-339)

The following oligos with KpnI – XbaI ends were used to amplify full-length hTBP (aa1-339) a gift from J. Geiger.

DA162: 5' GGG TCG GTA CCG CAG CCG CAA TGG ATC AGA ACA ACA GCC TG 3'

DA164: 5'GGG GAA TCT AGA CTA ACT AAT TAC TAC GTC GTC TTC CTG AAT CCC TT 3'

The PCR product was digested with KpnI and XbaI enzymes and cloned into KpnI-XbaI cut pTwiggy vector (Arnosti et al., 1996a).

### 2. Fly stocks

Flies carrying a mutation in the *giant* gene on the X chromosome  $gt^{A8}/FM7c$ (Stock # 1004.1) and  $gt^{X11}/FM6$  (Stock # 1529) were obtained from the Bloomington Stock center. To analyze reporter gene expression in a *giant* mutant background, males carrying the reporter and the Gal4-activator transgenes were crossed to females carrying the *giant* mutation. Flies expressing the full-length yeast transcriptional activator Gal4 ubiquitously throughout the embryo under the control of the actin5C enhancer, act5cGAL4/CyO (Stock # 4414) were also obtained from Bloomington. In order to obtain ubiquitous activation of the *lacZ* reporter gene in the early (2-4 hour old) embryo, act5cGAL4/CyO females were crossed to males carrying the reporter transgene.

#### 3. Reporter genes

The stripe 2/2x UAS/eve-lacZ vector (Arnosti et al., 1996a) containing two GAL4 binding sites and the minimal eve basal promoter driving lacZ expression was modified to include two Giant (Capovilla et al., 1992) binding sites (DA127/128: 5'AAT TCG CAT GCT ATG ACG CAA GAA GAC CCA GAT CTT TTT ATG ACG CAA GAG CAT GCG 3') using *EcoRI-BssH*2 enzymes, upstream of the GAL4 sites. The vector was further modified to incorporate three additional GAL4 binding sites (DA139/140: 5' TCG GAT TAG AAG CCG CCG TCG CTA GAG GAA GAC TCT CCT CCG ACG TGA ACG CAG GAC ACT CCT GC GCT GCA 3') at the PstI site downstream of the existing GAL4 sites. Oligos containing a 50 bp spacer (DA125: 5'TCG CTA GAC GTG AAT CTC GTA GCT TCC GTA CCA AAT GCG TAT CAG CTG CA 3'; DA126: 5' GCT GAT ACG CAT TTG GTA CGG AAG CTA CGA GAT TCA CGT CTA GCG ATG CA 3') were introduced at the PstI site downstream resulting in the vector H2g5u-50 (Figure III-1E, F, G, and H) which contains two Giant binding sites, five tandemly arrayed GAL4 binding sites, a 50 bp spacer and minimal eve basal promoter driving lacZ expression.

The plasmid UAS-lacZ (Brand and Perrimon, 1993) was modified to contain two Giant sites (DA321/322: 5' GGC CGC TAT GAC GCA AGA AGA CCC AGA TCT TTT TAT GAC GCA AGA GA 3') or two Knirps sites (DA319/320: 5'GGC CGC ATC TGA TCT AGT TTG TAC TAG ACA TCT GAT CTA GTT TCA 3') or two Krüppel (DA694/695: 5'GGC CGC AAA ACG GGT TAA GCG ACC CAA AAC GGG TTA AGC A 3') sites or two Hairy (DA604/605: 5'GGC CGC GCG GCA CGC GAC ATG ACC CGC GGC ACG CGA CAT A 3') sites twenty nucleotides upstream of the five GAL4 binding sites (Arnosti et al., 1996c; Capovilla et al., 1992; Gray et al., 1994; Nibu et al., 2001). The resulting vectors named M2g5u-lacZ/M2k5ulacZ/M2kr5u-lacZ/M2h5u-lacZ respectively, consist of two Giant or Knirps or Krüppel or Hairy Binding sites, five tandemly arrayed GAL4 binding sites, followed by the *hsp 70* TATA box and transcriptional start driving *lacZ* expression (FigureIII-1A, B; Figure III-2A, B; 7B, D, F).

The vector M2g5u-lacZ was modified by introducing oligos containing a 55 bp neutral spacer (DA65/66: 5'TCC ATG ATA AAC GCG TGC TAG ACT ATT GCA GGT ACT GAT CGA ATG CCT CTG CAT G 3') at the *Sph*I site downstream of the Gal4 binding sites. The vector was further modified by introducing a 340 bp fragment of the Knirps open reading frame, which was amplified using oligos DA572/573 (DA 572: 5'ACA TGC ATG CAA CCG CTT TAG TCC CGC CAG 3'; DA 573: 5'ACA TGC ATG CTG TGC ACG GAG CTC CGC GAG 3') from the vector Gal4-kni F1 resulting in the spaced construct M2g5u-55-340 bp kni ORF-lacZ (Figure III-1C, D).

The vector M2g5u-lacZ was modified to replace the five tandemly arrayed GAL4 sites with *Hind*III-SphI oligos containing three high affinity GAL4 (Brand and

Perrimon, 1993) binding sites (DA469/470: 5'AGC TTG CCT GCA GGT CGG AGT ACT GTC CTC CGA GCG GAG TAC TGT CCT CCG AGC GGA GTA CTG TCC TCC GAG GCA TG 3') to give M2g3u-lacZ (Fig. 2C, D). This was further modified by introducing SphI spacer oligos (DA471: 5' TCA TAC AAC TGG TCA GTG AGC ATA CAA CTG GTC AGT GAG CAT G 3'; DA472: 5' CTC ACT GAC CAG TTG TAT GCT CAC TGA CCA GTT GTA TGA CAT G 3') equal to the length of two GAL4 sites downstream resulting in M2g3u2x-lacZ (Figure III-2E, F). The two Giant binding sites in M2g3u2x-lacZ were replaced by two Knirps sites (DA319/320: 5'GGC CGC ATC TGA TCT AGT TTG TAC TAG ACA TCT GAT CTA GTT TCA 3') or two Krüppel (DA694/695: 5'GGC CGC AAA ACG GGT TAA GCG ACC CAA AAC GGG TTA AGC A 3') sites or two Hairy (DA604/605: 5'GGC CGC GCG GCA CGC GAC ATG ACC CGC GGC ACG CGA CAT A 3') sites twenty nucleotides upstream of the three GAL4 binding sites. The resulting vectors named M2k3u2xlacZ/M2kr3u2x-lacZ/M2h3u2x-lacZ (Figure III-7A, C, and E) consist of two Knirps/Krüppel/Hairy binding sites respectively, three tandemly arrayed GAL4 binding sites, a spacer, followed by the hsp 70 TATA box and transcriptional start driving lacZ expression.

The vector M2g3u-lacZ was cut with *Hind*III to introduce spacer oligos (DA473: 5' AGC TTC ATA CAA CTG GTC AGT GAG CAT ACA ACT GGT CAG TG 3'; DA474: 5' AGC TCA CTG ACC AGT TGT ATG CTC ACT GAC CAG TTG TAT GA 3') equal to the length of two GAL4 sites in between the Giant and three GAL4 sites to result in M2g2x3u-lacZ (Figure III-6B, D, F). The five high affinity GAL4 binding sites in M2g5u-lacZ were replaced with five low affinity (LA) GAL4 sites (Burns and

95
Peterson, 1997; Johnston and Davis, 1984) (site no. 1 from the *Gal1,10 UAS* region) by sequentially cloning in *Hind*III-*Sph*I oligos containing three **low affinity sites** (DA600/601: 5'AGC TTG CCT GCA GGT CGG ATT AGA AGC CGC CGA GCG GAT TAG AAG CCG CCG AGC GGA TTA GAA GCC GCC GCA TG 3') followed by *Sph*I oligos containing two **low affinity Gal4 sites** (DA602/603: 5' TCG GAT TAG AAG CCG CCG AGC GGA TTA GAA GCC GCC ATG 3') resulting in the vector M2g5u (LA)-lacZ (Figure III-3).

Two additional Giant binding sites were introduced either at the *Sph*I site (DA50/51) in the M2g5u-lacZ vector between the five GAL4 binding sites and the *hsp* 70 TATA box resulting in M2g5u2g-lacZ (Figure III-5A, B), or at the *Not*I site (DA637/638) in the M2g5u-lacZ vector upstream of the two Giant sites resulting in M4g5u-lacZ (Figure III-5 E, F). Two additional binding sites for Giant (DA50/51) as well as two high affinity Gal4 sites (DA598/599) were introduced sequentially in the M2g3u-lacZ vector at the *Sph*I site resulting in M2g3u2g2u-lacZ (Figure III-5 C, D).

# 4. P-element transformation, crosses to reporter genes, and whole-mount in situ hybridization of embryos.

P-element transformation vectors were introduced into the *Drosophila* germline by injection of yw<sup>67</sup> embryos as described (Small et al., 1992). Embryos were collected either directly from each transgenic reporter line or from a cross between a reporter line and a line expressing the GAL4-activator chimeric proteins in the ventral regions or ubiquitously throughout the embryo. The embryos were fixed and stained using digoxigenin-UTP labeled antisense RNA probes to either *lacZ* or *w* as described (Small et al., 1992).

#### RESULTS

#### Context dependence of short-range repression

The activity of short-range transcriptional repressors has been studied mostly in the context of complex endogenous enhancers (Arnosti et al., 1996a; Arnosti et al., 1996b; Gray et al., 1994; Kosman and Small, 1997; Small et al., 1993; Small et al., 1992). This approach is complicated by the complexity of many *cis*-acting regulatory modules, where stoichiometry, spacing and even the identity of transacting factors are not always well defined. To analyze *cis*-acting element function in a setting in which activator-repressor composition, stoichiometry and spacing can be exactly defined, we constructed chromosomally integrated, compact regulatory modules containing binding sites for the endogenous short-range repressor Giant and chimeric Gal4 activators. The space between repressor and activator sites on these elements is less than 100 bp, a distance over which short-range repressors have been previously shown to be effective (Arnosti et al., 1996b; Gray et al., 1994; Hewitt et al., 1999; Keller et al., 2000). Strikingly, Giant was unable to repress the hsp70 lacZ reporter containing five high affinity Gal4 activator sites from an upstream position, although these proteins bind within 100 bp of the repressors, revealing a hitherto unknown limitation of short-range repressors (Figure III-1A, B). The close proximity of the Gal4 activators to the hsp 70 basal promoter may prevent Giant from mediating repression on this reporter. However, Giant is still unable to repress even after introduction of a 400 bp spacer that moves the

activators away from the transcriptional start site (Figure III-1C, D). The inability of Giant to repress is not due to an inherent resistance of the Gal4 activation domain, for Giant repressed a similar cluster of five Gal4 binding sites 5' of the *eve* basal promoter (Figure III-1E, F). The repression in anterior and posterior regions is relieved when this transgene is assayed in *giant* mutant embryos (Figure III-1G, H) confirming that repression is mediated by Giant.

These results (Figure III-1) overturn the simple notion that short-range repressors block the activity of all protein complexes within 100 bp. Clearly, mere proximity is not the only determinant affecting repression by Giant. We set out to systematically define other factors that dictate repression effectiveness to uncover a potential 'grammar' of repression.

Differences in activator site affinity or spacing, basal promoters, or repressor positioning with respect to the transcriptional start site may be the basis for the difference in repression effectiveness in the transgenes shown in Figure III-1. Activator site affinity or spacing appears to be the most likely cause of the difference in repression effectiveness, because both types of basal promoters have been shown to be directly repressed previously, and the relative spacing of the repressors to +1 should in fact favor repression in Figure III-1A and B, where the Giant repressor is closer to the start site of transcription. Weaker or suboptimal spacing of activator sites might decrease the average number of activators on the promoter, which may in turn favor repression (Figure III-1E, F).

## Figure III-1: Context dependence of short-range repression.

A schematic structure of the reporter transgene is shown below the corresponding embryos. Patterns of gene expression were visualized in 2-4 hour embryos by *in situ* hybridization with digU labeled antisense *lacZ* probes. Embryos are oriented anterior to left; lateral views (A, C, D, E, and G) and ventrolateral views (B, F, and H) are shown.

(A, B) LacZ expression driven by the Gal4- Gal4 AD activator fusion from a cluster of five high affinity Gal4 binding sites upstream of the *hsp70* basal promoter elements is not repressed by the short-range repressor Giant.

(C, D) LacZ expression driven by the Gal4- Gal4 AD activator fusion from a cluster of five high affinity Gal4 binding sites upstream ~ 450 bp upstream of the hsp70 basal promoter elements is also not repressed by the short-range repressor Giant.

(E, F) Lack of repression by Giant is not due to the inherent resistance of the Gal4 activation domain. The activity of the same Gal4 -activation domain from a similar cluster of five Gal4 binding sites 5' of the minimal *eve* basal promoter is repressed by the short-range repressor Giant in both the anterior and posterior domains of the embryo where the repressor is expressed (arrows).

(G, H) The repression of lacZ in the anterior and posterior regions is relieved when the *eve lacZ* transgene is assayed in an embryo carrying the *gtA8* mutation.

Figure III-1: Context dependence of short-range repression.



#### Figure III-7: Repressors can be distinguished by their differential ability to repress in different enhancer configurations.

A schematic of the reporters used is shown below. The reporters contain two binding sites for either one of the short-range repressors Giant, Knirps or Kruppel; or the long-range repressor Hairy and five/three high affinity Gal4 binding sites.

All three of the short-range repressors Giant (A), Knirps (C), and Kruppel (E) and the long-range repressor Hairy (G) are able to repress the activity of the Gal4-activator in the context of three Gal4 binding sites.

While the short-range repressors Giant (B), Knirps (D), and Kruppel (F) are unable to repress in the context of five high affinity Gal4 sites the long-range repressor Hairy (H) is able to effectively repress.

Embryos are oriented with anterior to left; lateral views (A, B, C, D, and E) and ventrolateral views (F, G, and H) are shown.

### Specificity of regulatory 'grammar'

The contextual dependencies of repression described above were developed for the Giant repressor. To determine if similar rules applied to other types of repressors, we carried out parallel evaluations of the short-range repressors Giant, Knirps, and Krüppel. To test quantitative similarities or differences between these factors, we created reporters that would compare repressor activity on genes that represented permissive or nonpermissive contexts for the Giant protein. All three of these short-range repressors were unable to inhibit *lacZ* expression driven by the Gal4 activator from five high affinity Gal4 sites indicating a similar limitation of repression on even proximally bound activators (Figure III-7B, D, F). The Giant and Krüppel factors exhibited repression activity in the corresponding regions of the embryo when tested against three Gal4 sites (Figure III-7A, C). The Knirps repressor was also active in this context, although in general the levels of repression appeared to be lower (Figure III-7E). In contrast, the long-range repressor Hairy was able to mediate repression of transgenes containing three or five high affinity Gal4 sites (Figure III-7G, H). Interestingly as the embryo aged, repression of Gal4 by Hairy was attenuated and was completely abolished by the time germ band elongation is completed (data not shown), indicating that this type of repression, though potent, is also transient. The similarity in the activity of the short-range repressors Giant, Knirps and Krüppel, in contrast to that of Hairy, suggests that the contextual rules for repression are governed by the functional class of repressor, and likely reflect mechanistic differences.

Figure III-6: The ability to repress depends on enhancer configuration rather than the nature of the activation domain.



# Figure III-6: The ability to repress depends on enhancer configuration rather than the nature of the activation domain.

A schematic representation of the reporter genes are shown above and the chimeric Gal4activators used to drive expression from the reporter gene are indicated to the left of the embryos.

(A, C, E) Giant is able to repress all three activators when the number of Gal4 sites is reduced from five to three.

(**B**, **D**, **F**) However, when the three Gal4 sites are moved  $\sim$ 40 nucleotides away from the Giant sites, the ability to repress all three types of activators is abolished. In this context the three Gal4 sites are still within a 100 bp from the Giant binding sites.

Embryos are oriented with anterior to left; lateral views are shown.

Giant binding sites endowed a promoter with high or low sensitivity to repression (Hewitt et al., 1999). We tested whether Giant's ability to repress a smaller cluster of Gal4 sites could be attenuated by small changes in spacing between the activator and repressor binding sites. Moving the smaller cluster of three Gal4 sites 37 bp away from the Giant binding sites results in the loss of repression (Figure III-6A compare to Figure III-6B) suggesting that reducing the amount of activation potential does not guarantee repression by Giant in all cases even when the activators are located within 100bp of the repressor sites. In order to ascertain whether the spacing effects we see are specific to the activator or a general property of the repression activity of Giant, we tested the ability of Giant to block transcription mediated by the full length Gal4 protein expressed ubiquitously throughout the embryo (Figure III-6C, D), and another activation domain, namely Gal4-VP16 (Figure III-6E, F). As seen with the Gal4 activation domain, Giant is able to repress lacZ expression mediated by the full length Gal4 protein (Figure III-6C) and Gal4-VP16 (Figure III-6E) from three sites that are adjacent to the Giant binding sites. However, moving the three sites 37 bp away again results in the loss of repression by Giant of both Gal4 (Figure III-6D) and Gal4-VP16 (Figure III-6F) mediated activation. These results overturn the previously held simple notion that short-range repressors block all protein complexes within a 100 bp. The loss of repression caused by moving the activator binding sites 37 bp away from Giant sites is not specific to a particular activator, suggesting that the range over which short-range repressors function can be influenced by the strength of the DNA binding domain of the activator.

Har 12 165 63 **[**[-8 1323 165 ņε tr E (14 bie ï ŝă 105 <sup>h</sup> IN. 60 ų<sup>j</sup>ž 1110 a.C

Figure III-5: The arrangement/distribution of short-range repressor binding sites is critical in dictating repression effectiveness.



# Figure III-5: The arrangement/distribution of short-range repressor binding sites is critical in dictating repression effectiveness.

A schematic of the reporter gene is shown below the corresponding embryos.

The short-range repressor Giant is able to repress (arrows) both the proximal hsp70 lacZ gene (A) and the distal w (B) 4.5 kb away, when the repressor flanks the Gal4-activator.

Giant is also able to repress (arrows) in a reporter context where its binding sites are interspersed between the Gal4 sites and repression is seen both at the proximal hsp70 lacZ gene (C) and the distal w (D) 4.5 kb away.

However, Giant is unable to repress in a context in which all the repressor sites are clustered together on one side (E, F) of the five Gal4 sites although the number of repressor and activator sites is the same as in A, B, C and D.

Embryos are oriented anterior to left; lateral views are shown.

side resulted in repression of the proximal hsp 70 lacZ reporter gene (Figure III-5A). Interspersing the Giant repressor binding sites in between the Gal4 activator sites also resulted in inhibition of lacZ expression (Figure III-5C). However, placing all four Giant binding sites 5' of the five Gal4 sites prevented Giant from repressing the hsp 70 lacZ expression (Figure III-5E) suggesting that promoter response cannot be calculated simply from overall activator to repressor stoichiometries. We noted that the Giant binding sites in the reporter genes showing repression (Figure III-5A, C) are in close proximity to the basal promoter and therefore, it is possible that in these contexts Giant directly represses the basal promoter adding a confounding factor (Arnosti et al., 1996b; Gray and Levine, 1996; Hewitt et al., 1999). To distinguish between repressor-basal promoter and repressor-activator effects, we assayed for the activity of the w gene, which is  $\sim 4.5$  kb downstream of these *cis*-elements in each of the reporter gene constructs (Figure III-5B, D, and F). Here again we observed that Giant mediated repression only when flanking or interspersed with activators. The similar patterns of repression of proximal lacZ promoter and distal w promoter suggest that Giant is acting on the activator cluster rather than only on the basal promoter element (Figure III-5B, D). The preferential arrangement of transcription factor binding sites within regulatory regions might be considered as a specific type of functional information encoded in regulatory DNA that is critical in dictating the transcriptional output of the *cis*-element. Indeed a distributed pattern of short-range transcriptional repressor binding sites is typical of many developmental enhancers that function in the early Drosophila embryo.

Previous analysis of the short-range repressor Giant demonstrated that due to the extreme distance-dependent activity of this protein, subtle changes in the spacing of

Figure III-4: Repression not dependent on the nature of the activation domain.



#### Figure III-4: Repression not dependent on the nature of the activation domain.

A schematic representation of the reporter gene is shown below and the chimeric Gal4activators used to drive expression from the reporter gene are indicated to the left of the embryos. Giant is unable to repress the activity of strong activators Gal4-Gal4 AD (A), and Gal4-VP16 (B), or weaker activators Gal4-Sp1 (C), and Gal4-hTBP (D) on the *hsp70* reporter containing five high affinity Gal4 binding sites.

Embryos are oriented with anterior to left and lateral views are shown.

activator chimeric proteins were used to drive expression of the *hsp 70 lacZ* reporter from a cluster of five high affinity Gal4 sites (Figure III-4A, B, C, and D). Giant could inhibit neither the strong Gal4 (Figure III-4A) and VP16 (Figure III-4B) activators, nor the weak activation domains of Sp1 (Figure III-4C) and hTBP (Figure III-4D) (Seipel et al., 1992). These results indicate that the ability to repress does not depend on the strength of the activation domain or the activation pathway. Only those genes in which the number or affinity of Gal4 sites was reduced showed a response to Giant, suggesting that the Gal4 DNA binding domain provides a stable platform that can resist the activity of Giant and, and suggests a mechanism for short-range repression that involves blocking activator access to its cognate site.

# The arrangement/distribution of short-range repressor binding sites is critical in dictating repression effectiveness

Statistical models, based on motif clustering, are only partially successful at finding novel *cis*-regulatory elements in the genome, perhaps because they consider only site density and relative site affinity (Berman et al., 2002; Makeev et al., 2003; Markstein and Levine, 2002; Markstein et al., 2002). However, it is probable that specific arrangements of binding motifs are critical for proper biological function. We tested the effect of alternative arrangements of Giant repressor and Gal4 activator sites to determine if different arrangements or combinations resulted in distinct transcriptional outputs. In all reporter arrangements tested, we used four Giant binding sites and five high affinity Gal4 binding sites. Flanking the five Gal4 activator sites with two Giant sites on either

progressively refined into a two-stripe pattern, in regions where *giant* (Figure III-3A) is not expressed (arrows; Figure III-3D-E)). Analysis of the transgene in a *giant* mutant background, in the absence of Gal4, confirms that refinement of reporter gene expression is due to repression by Giant (Figure III-3G). These results suggest that five Bicoid binding sites are more susceptible to repression than are five high affinity Gal4 sites, indicating that stoichiometric relationships of repressors to activators may depend also on distinct DNA binding domains or activation domains.

#### Repression not dependent on the nature of activation domain

The differential effectiveness of Giant against five Gal4 or five Bicoid sites suggests that the nature of the activation domain itself or the DNA binding domain, may play a role in dictating the response to repressors. To distinguish between those two possibilities, we compared the activities of a variety of activation domains fused to the DNA binding domain of Gal4 (aa 1-93). In addition to the well characterized activation domain of the yeast transcriptional activator Gal4 (from aa 753-881), we tested the acidic transcriptional activation domain of the herpes simplex virus activator VP16 (from aa 412-490), the glutamine rich activation domain of the mammalian transcription factor Sp1 (from aa 132-243), and the full-length human TATA binding protein (hTBP) which has been shown to function as an activator when fused to the Gal4 DNA binding domain (Seipel et al., 1992). We also sought to test the activity of Gal4-Bicoid activators (Janody et al., 2001) but unfortunately these chimeras exhibit strong promoter specificity and are not active on the *hsp 70* promoter, precluding a side-by-side comparison. The Gal4-

## Figure III-3: Effectiveness of repression correlates with the affinity of activator binding sites.



## Figure III-3: Effectiveness of repression correlates with the affinity of activator binding sites.

Sequence of consensus Gal4 binding sites, high affinity Gal4 sites in the *hsp70* reporter (Figures III-1 and 2; not repressed), low affinity Gal4 binding sites, and high affinity bicoid (Eisen et al., 2001) binding sites (underlined with a red line) created while creating low affinity Gal4 sites are indicated on the left.

(A) giant expression in the early blastoderm embryo is dynamic and refines into two stripe anteriorly and one stripe posteriorly. giant expression is visualized by in situ hybridization using digoxigenin labeled antisense giant mRNA probe.

(**B**, **C**) Giant repression of the Gal4 activator. LacZ expression driven by the Gal4- Gal4 AD activator fusion from a cluster of five low affinity Gal4 binding sites upstream of the hsp70 basal promoter elements is now repressed by the short-range repressor Giant (arrows) in both the anterior and posterior regions of the embryo where the repressor is present.

(D, E) Giant represses bicoid mediated activation of the  $hsp70 \ lacZ$  reporter (arrows). Even in the absence of the Gal4 activator, LacZ expression is activated by the transcription factor bicoid in the anterior region of the embryo from five high affinity sites created in the reporter. Bicoid mediated activation gets progressively refined (D-E) into two stripes of expression as the embryo develops, in regions where *giant* is not expressed. A similar bicoid mediated activation of lacZ and subsequent refinement by the repressor Giant is also evident in the anterior regions of the embryo in the presence of the Gal4 activator (B, D).

(F, G) Analyses of lacZ expression from the hsp70 reporter in a giant mutant background.

(F) Giant repression of Gal4- Gal4 AD mediated activation of hsp70 lacZ is relieved in a *giant* mutant background and *lacZ* staining is evident in the posterior region where Giant is usually expressed (arrow).

(G) Analysis of the *hsp70 lacZ* reporter in the absence of the Gal4 activator. Bicoid mediated activation is no longer refined into a two stripe pattern in a *giant* mutant background.

Embryos are oriented anterior to left; ventral (F) and lateral (A, B, D, E, and G) views are shown.

Binding site affinity influences threshold responses to activator gradients in the embryo, (Jiang and Levine, 1993; Struhl, 1989; Szymanski and Levine, 1995) and indeed transcription factor binding sites of varying affinities are typically found in many developmental enhancers that function in the early *Drosophila* development. Such differences in activator site affinity might similarly influence responses to short-range repressors. We tested whether maintaining the number of activator sites but weakening their affinity would in fact change the reponse to repressors. We replaced the five high affinity Gal4 binding sites in the *hsp70 lacZ* reporter, with five low affinity sites from the endogenous yeast *S. cerevisiae* Gal1-Gal10 promoters that have been characterized to bind Gal4 poorly (Burns and Peterson, 1997; Johnston and Davis, 1984). The lower affinity Gal4 sites drive gene expression in a weaker striped pattern, but anterior and posterior repression by Giant is clearly evident (Figure III-3B, C; arrows). Repression is relieved in a *giant* mutant embryo (Figure III-3F), confirming that the gaps in expression seen in Figure III-3B, and C, are mediated by the Giant repressor.

In the process of weakening the Gal4 binding sites, we inadvertently created five high affinity binding sites for the Bicoid activator (Berman et al., 2002) providing an additional opportunity to assay Giant repression activity. Bicoid is maternally deposited in the anterior regions of the embryo, forming an anterior to posterior gradient (Driever and Nusslein-Volhard, 1988; Eldon and Pirrotta, 1991; Kraut and Levine, 1991). *LacZ* expression from the *hsp 70* reporter is activated even in the absence of the Gal4 activator by Bicoid transcription factor in anterior regions (Figure III-3D and E). As the embryo develops (Figure III-3D-E), Giant inhibits Bicoid mediated *lacZ* activation, which is Figure III- 2: Stoichiometry between the number of activators and repressors influences repression effectiveness.



## Figure III- 2: Stoichiometry between the number of activators and repressors influences repression effectiveness.

(A, B) LacZ expression driven by the Gal4- Gal4 AD activator fusion from a cluster of five high affinity Gal4 binding sites upstream of the *hsp70* basal promoter elements is not repressed by the short-range repressor Giant.

(C-F) Reducing the amount of activator by reducing the number of Gal4 sites from five to three restores the ability of Giant to block *lacZ* expression (arrows).

Embryos are oriented anterior to left; lateral views (A, C, E, and F) and ventrolateral views (B,andD) shown.

### Repression sensitivity correlated to the strength of the activating signal

Studies of the *cis*-regulatory elements of the *hairy* gene in *Drosophila* led to the suggestion that the overall stoichiometry, rather than the absolute number, of activators and repressors maybe critical in dictating enhancer output (La Rosee et al., 1997). To test whether the stoichiometry of enhancer- bound activators to repressors is a critical factor in determining short-range repression levels by Giant, we reduced from five to three the number of Gal4 activator binding sites on the hsp70 - lacZ reporter. As anticipated, the levels of transcriptional activation were lower in the transgene containing three Gal4 sites, leading to a less robust ventral staining pattern (Figure III-2A, B). In this context, Giant was able to block transcription of the *lacZ* gene (Figure III-2C, D). However, the removal of two Gal4 sites also positions the repressors closer to the start of transcription, which may facilitate direct basal promoter repression. Therefore, to maintain the distance between Giant binding sites and the start of transcription we tested a reporter gene where a neutral spacer was placed downstream of the three Gal4 sites (Figure III-2E, F). Again, Giant was also able to repress the Gal4 activators. These results demonstrate that repression is critically dependent on the number of activator binding sites but do not explicitly differentiate between overall level of transcriptional activation and binding site number, an issue explored below. These results are also consistent with previous analysis of the eve stripe 2 element, where the insertion of additional Bicoid binding sites in an otherwise normal stripe 2 enhancer causes a slight anterior expansion of its expression pattern, suggesting that an excess of Bicoid activators can 'overwhelm' the Giant repressor (Arnosti et al., 1996a).





#### DISCUSSION

Using defined synthetic enhancer elements to study short-range repressor action, we demonstrate that there is a rich contextual grammar that influences repressor activity. extending beyond the generalization that short-range repressors block the activity of all protein complexes within a 100 bp. Although the distance between short-range repressors and their targets is a critical factor in dictating repression effectiveness, it is not the only one, and in some cases, proximity is not sufficient. Activators can escape repression even when they are within the previously defined effective range of short-range repression. The manipulation of these composite elements in terms of the number of activator and repressor binding sites, relative affinities, spacing and distribution of binding sites further allowed us to define other contextual parameters that dictate repression effectiveness. First, we find that a balanced ratio of enhancer-bound activators and repressors is an important factor dictating repressibility. In the context of five high affinity Gal4 sites, two adjacent repressor sites are insufficient to permit the repression by Giant. Reducing the number of Gal4 sites from five to three allowed Giant to block Gal4 mediated activation of the *lacZ* reporter gene. Second, although the effectiveness of repression depends on stoichiometry between the number of activators and repressors, Giant repression of a smaller cluster of activators can be attenuated by subtle changes (< 40 bp) in the spacing between the repressor and activator binding sites even when binding sites are within the previously defined 100 bp "repression zone". A third parameter is that repression effectiveness correlates with the affinity of activator binding sites, and although binding affinity influences the strength of the activating signal, repression does

not depend on the chemical nature of the activation domain. Fourth, the short-range repressors need to be judiciously placed, either flanking activator sites or interspersed among them, possibly to block multiple modes of enhancer-promoter interactions. These contextual parameters constitute the 'grammar' of short-range repressor function. Empirical determination of such elements contributes to our understanding of enhancer design and should find application in bioinformatic analysis of novel gene regulatory sequences, as well as providing insights into the evolution and biochemical activity of short-range repressors as we describe below.

#### Computational analysis of cis-regulatory elements

Enhancers serve as binding platforms to localize multiple distinct transcription factors, including sequence-specific activators and repressors (Arnosti, 2002; Davidson, 2001). These factors act combinatorially to confer context-specific transcriptional activity. In effect, changing the combinations and arrangements of transcription factor binding sites within the enhancer can vary the nature of the transcriptional output generated by *cis*-regulatory elements. Thus, any computational algorithm designed to identify *cis*-regulatory modules must take into account the combinatorial logic underlying *cis*-regulatory module architecture.

Extensive analysis of the *Drosophila eve* stripe enhancers and their trans-acting factors has provided insights into the complexities inherent in decoding the combinatorial nature of *cis*-regulatory logic. The *eve* stripe 2 element is one of the best characterized

enhancer and contains a total of eighteen known factor-binding sites, including eight activator (three for Hunchback, five for Bicoid) binding sites and ten repressor (three for Giant, six for Krüppel and at least one for Sloppy-paired) binding sites (Andrioli et al., 2002; Ludwig et al., 1998; Stanojevic et al., 1991). Detailed site mutageneses studies have demonstrated that every (most) transcription factor binding site within the stripe 2 element contributes to the overall transcriptional output (Small et al., 1992) suggesting that stripe 2 expression, to a large extent, depends on a balanced ratio of repressors and transcriptional activators. However, studies have also indicated that this notion is an oversimplification and that simple multimerization of the known activators and repressor sites is not sufficient to recapitulate the expression of the native element. Other contextual features, such as spacing/orientation constraints between binding sites within the enhancer, may play an important role.

Different concentrations of the Giant, Knirps and Krüppel repressor proteins have also been shown to be important for proper regulation of many zygotic genes during early embryogenesis. For example, low concentrations of Giant are sufficient to repress the *Krüppel* promoter, while greater amounts appear to be required to repress the *eve* stripe 2 element (Wu et al., 1998). In a similar vein, higher levels of the Knirps protein are required to regulate the *eve* stripe 3 + 7 enhancer while lower levels are sufficient for the proper regulation of the *eve* stripe 4 + 6 element (Clyde et al., in press; Struffi et al., submitted). Sequence analysis of the *eve* gene indicates that there are more high affinity Knirps binding sites within the *eve* stripe 3 + 7 element than in the 4 + 6 enhancer, consistent with the relative sensitivities of these elements determined experimentally (Berman et al., 2002; Papatsenko et al., 2002). Removal of some of the Knirps binding sites from the *eve* 3 + 7 enhancer reduces the sensitivity of this element to the Knirps gradient (Clyde et al., in press). Thus, these studies indicate that in addition to the number of binding sites, binding site affinities within target genes define the molecular basis of differential responses to these transcriptional repressors and are crucial for the precise positioning of gene expression borders.

Transcription factors that function as morphogens generate unique threshold responses at different concentrations. This point is relevant to the clustering of cognate DNA binding sites, because the number and affinities of these sites provide a readout of the local transcription factor concentration. For example, a *cis*-regulatory module with a large number of sites or with high affinity sites would be activated in response to low levels of the corresponding factor. Additionally, it has been demonstrated that the same *cis*-regulatory module can generate two different transcriptional outputs depending on its position within the morphogenetic gradient of the transcription factor that regulates it (Hewitt et al., 1999).

Transcription factors bound to regulatory DNA are often involved in specific protein-protein interactions. Thus, the exact arrangement of binding motifs may control the formation of three dimensional protein complexes that are essential for the appropriate biological output of that element (Cai et al., 1996; Merika and Thanos, 2001). Therefore, the binding motifs are distributed in a non-random fashion within *cis*-elements leading to specific, functionally relevant arrangements. However, some plasticity in the

123

positioning of binding sites is tolerated in some situations and might reflect the way in which *enhancer-bound factors* interact with the basal transcriptional machinery (Kulkarni, 2003). For example, in the case of the information display enhancer, discrete portions of the element can be 'read' off by the basal transcriptional machinery. Individual subelements within the enhancer can signal independently of each other to the basal transcriptional machinery and therefore, the exact spatial arrangement between the autonomously acting subelements may not be critical. The independent information units might consist of one or a few factor binding sites, the enhancer substructure (Kulkarni, 2003) and thus, it is conceivable that within an independently acting information unit functional constraints between factor binding sites exist, which define the 'grammar' rules governing enhancer function.

Thus, the combinatorial logic underlying *cis*-regulatory module architecture requires not only consideration of binding site numbers and affinities, but also combinations of binding site type, spacing, and positioning within the *cis*-element. These parameters are difficult if not impossible to predict accurately *a priori*. So far structure-function analyses of *cis*-regulatory elements have focused on complex endogenous elements and therefore are complicated by the complexity of these elements where the exact identity, stoichiometry, relative spacing and arrangement of binding sites is not known. Although these studies have shed some light on the contextual features that might play a critical role in defining a function *cis*-regulatory module, the parameters have not been accurately determined. Computational approaches will benefit from the availability of at least one well-defined representative of a particular network as a starting paradigm.

Using synthetic well-defined genetic elements we have defined such a paradigm for the short-range transcriptional repressors. Incorporating the 'grammar' rules for short-range repressors into search algorithms will be better able to predict the possibility of Giant or Knirps regulating a given element given the relative the relative number, affinities, spacing and distribution of repressor and activator sites within the element. Empirical testing of identified candidates should enable further refinement of the model, which in turn can be utilized for another round of screening. This underscores the importance of combining bioinformatics with experimental biology.

#### **Evolution of transcriptional regulation**

It has been shown that regulatory sequences can maintain regulatory function despite structural reorganization as a result of species-specific loss and gain of transcription factor binding sites (Cuadrado et al., 2001; Ludwig et al., 2000; Piano et al., 1999). Thus, to decipher what kind of changes in *cis*-element structure cause a change in function, an underlying cause for morphological evolution, detailed information about the locations of transcription factor binding sites within the element, the functional specificity of the binding sites and the spatial requirements for their interaction is required. The empirical determination of the structural parameters that dictate enhancer function, such as those described here for elements regulated by short-range repressors, can facilitate the modeling of the evolutionary dynamics of *cis*-regulatory regions in a stepwise fashion, starting with sequence comparisons between closely related species, and then extending to more distantly related organisms. Thus, if we knew in functional terms the components of the specific genomic *cis*-elements that result in different **morphological** outcomes in two animals of common ancestry, we could see exactly what are the essential causal differences in the DNA of these animals and provide an explanation, in a mechanistically relevant way, of how the diverse forms of animal life actually arose during evolution.

#### Mechanism of short-range repression

A repressor that functions in all, or most, contexts is more likely to function via a 'general' mechanism that is inhibitory at all promoters, such as preventing activator binding by establishing a repressive chromatin domain. In contrast, a repressor that can inhibit transcription only in particular contexts is more likely to target a specific activator or proteins that mediate the action of that specific activator. Selective, context related repression affords an added layer of combinatorial control of gene expression by sequence specific transcription factors. Previous analyses of short-range repression have demonstrated that Giant, Knirps, Krüppel and Snail can block the activity of a number of distinct activators such as Bicoid, Hunchback, Dorsal, Twist, and D-Stat (Arnosti et al., 1996a; Gray et al., 1994; Small et al., 1996). Many biochemical and genetic analyses suggest that at least some of these activators activate transcription via distinct pathways (Koh et al., 1998; Pham et al., 1999; Yuh et al., 1998; Zhou et al., 1998). The apparent lack of activator specificity demonstrated by short-range repressors suggests that these proteins function via a general mechanism. However, the interpretation of the results from previous studies is complicated by the fact that the activators tested represent distinct classes of transcriptional activators not only in terms of the nature of their

**activation** domains and activation pathways, but also because they possess different DNA **binding** specificities. Using synthetic enhancer elements where the identity, stoichiometry and exact spacing of activator and repressor binding sites is well-defined, we have demonstrated that repression effectiveness does not depend on the nature of the activation domain but correlates instead with DNA binding site affinity. Thus, we propose that the short-range repressors inhibit transcription by blocking access to DNA by transcriptional activators through chromatin remodeling at a very local scale. Subtle (< 40 bp) changes in the spacing of Giant from the nearest activators sharply attenuates repression effectiveness suggesting that the action of short-range repressors is more localized than the previously defined range of a 100 bp.

A common property of short-range transcriptional repressors is their interaction with the evolutionarily conserved corepressor CtBP (C-terminal Binding Protein) through short PXDLS peptide motifs found in the transcriptional repressors. CtBP has been shown to interact with chromatin modifying factors, including histone deacetylases (HDAC1 and HDAC2) and histone methyltransferases (Chinnadurai, 2002; Chinnadurai, 2003; Shi et al., 2003; Subramanian and Chinnadurai, 2003; Sundqvist et al., 1998). Thus, the major activity of CtBP might be to serve as a bridging molecule, recruiting chromatin or transcription factor modifying enzymes to the promoter. In addition, studies in our laboratory have demonstrated a genetic and physical interaction between Rpd3, the *Drosophila* homolog of HDAC1, and the short-range repressor Knirps (Struffi, unpublished data). The Rpd3 protein in yeast is known to deacetylate histones at an extremely local level, consistent with its role in short-range repression in Drosophila (Deckert and Struhl, 2002).

Although the contextual elements defined above allow us to hypothesize about the molecular mechanisms of short-range repression, we do not yet understand the physical and biochemical changes in promoter complexes that accompany short-range repression. Two proposed mechanisms are consistent with their activity seen *in vivo*: Short-range repressors might "quench" activator proteins locally, within a short distance of their binding site by displacing them from the DNA through chromatin structure modification, or by preventing them from contacting their target in the basal transcriptional machinery; or the repressors might directly contact some component of the basal transcriptional machinery, but only when brought to the promoter by a closely linked activator (hitchhiking) or when bound near the start site of transcription. The simple transcriptional switch elements defined in this study will facilitate further molecular and biochemical characterization of short-range repressors. These elements can be used in whole *Drosophila* embryos chromatin immunoprecipitation assays to examine the nature of the promoter complexes and the chromatin state before and after repression.

Similar to the studies that have categorized activators according to transcriptional activation pathways, analyses of transcriptional repression in the fruit fly has revealed that repressors can be classified as short-range or long-range repressors on the basis of their range of action. The mechanism by which short-range and long-range *Drosophila* repressors inhibit transcription are poorly understood, although one model of repression in the embryo suggests that the short-range/long-range distinction results from the

recruitment of distinct cofactors (Nibu et al., 1998a; Nibu et al., 1998b; Zhang and Levine, 1999). Short-range repressors recruit the corepressor CtBP to mediate repression whereas long-range repressors have been shown to interact with the Groucho corepressor (Barolo and Levine, 1997; Chen and Courey, 2000; Fisher and Caudy, 1998; Mannervik and Levine, 1999; Mannervik et al., 1999; Poortinga et al., 1998).

We show that the long-range repressor Hairy can block transcription in a context where the short-range repressors cannot function and appears not to require a stoichiometric relationship of activators to repressors. Thus, the Hairy long-range repressor seems to be quantitatively stronger (as a repressor) than a short-range repressor, not only due to its longer range of action but also by the virtue of its ability to block the Gal4 activators, in the context of five high affinity sites, that were uninhibited by Giant, Knirps and Krüppel. However, we do not exclude the possibility that Hairy may be able to repress in a context where we do not see repression by the short-range repressors simply due to higher levels of Hairy protein.

Long-range repressors have been likened to silencers, as they function in a dominant fashion to shut down multiple enhancers in a complex modular promoter (Barolo and Levine, 1997; Cai et al., 1996). However, we demonstrate that Hairy mediated repression is transient and easily reversible as the embryo ages, much like the action of the short-range repressors Giant, Knirps and Krüppel. Another striking paradox is that like Groucho, a number of studies have suggested that CtBP may function, at least in part, by recruiting histone deacetylases (Criqui-Filipe et al., 1999; Sundqvist et al., 1998). Thus, if both long- and short-range corepressors function through histone deacetylation, why is Hairy is more potent with a longer range of action than Giant? The synthetic enhancer elements defined in this study can serve as useful tools to compare the biochemical activities of these two classes of transcriptional regulators in order to resolve this issue and distinguish between the molecular mechanisms of short-range and long-range repressors.
# REFERENCES

Andrioli, L. P., Vasisht, V., Theodosopoulou, E., Oberstein, A. and Small, S. (2002). Anterior repression of a *Drosophila* stripe enhancer requires three position-specific mechanisms. *Development* 129, 4931-40.

Arnosti, D. N. (2002). Design and function of transcriptional switches in *Drosophila*. *Insect Biochem Mol Biol* 32, 1257-73.

Arnosti, D. N., Barolo, S., Levine, M. and Small, S. (1996a). The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* 122, 205-14.

Arnosti, D. N., Gray, S., Barolo, S., Zhou, J. and Levine, M. (1996b). The gap protein knirps mediates both quenching and direct repression in the *Drosophila* embryo. *Embo J* 15, 3659-66.

**Barolo, S., Carver, L. A. and Posakony, J. W.** (2000). GFP and beta-galactosidase transformation vectors for promoter/enhancer analysis in *Drosophila*. *Biotechniques* 29, 726, 728, 730, 732.

**Barolo, S. and Levine, M.** (1997). hairy mediates dominant repression in the *Drosophila* embryo. *Embo J* 16, 2883-91.

Bergman, C. M. and Kreitman, M. (2001). Analysis of conserved noncoding DNA in *Drosophila* reveals similar constraints in intergenic and intronic sequences. *Genome Res* 11, 1335-45.

Berman, B. P., Nibu, Y., Pfeiffer, B. D., Tomancak, P., Celniker, S. E., Levine, M., Rubin, G. M. and Eisen, M. B. (2002). Exploiting transcription factor binding site clustering to identify *cis*-regulatory modules involved in pattern formation in the *Drosophila* genome. *Proc Natl Acad Sci U S A* 99, 757-62.

Brand, A. H. and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118, 401-15.

Brantjes, H., Roose, J., van De Wetering, M. and Clevers, H. (2001). All Tcf HMG box transcription factors interact with Groucho-related co-repressors. *Nucleic Acids Res* **29**, 1410-9.

**Burns, L. G. and Peterson, C. L.** (1997). The yeast SWI-SNF complex facilitates binding of a transcriptional activator to nucleosomal sites *in vivo*. *Mol Cell Biol* **17**, 4811-9.

Cai, H. N., Arnosti, D. N. and Levine, M. (1996). Long-range repression in the Drosophila embryo. Proc Natl Acad Sci US A 93, 9309-14.

Capovilla, M., Eldon, E. D. and Pirrotta, V. (1992). The giant gene of *Drosophila* encodes a b-ZIP DNA-binding protein that regulates the expression of other segmentation gap genes. *Development* 114, 99-112.

Chen, G. and Courey, A. J. (2000). Groucho/TLE family proteins and transcriptional repression. *Gene* 249, 1-16.

Chen, G., Fernandez, J., Mische, S. and Courey, A. J. (1999). A functional interaction between the histone deacetylase Rpd3 and the corepressor groucho in *Drosophila* development. *Genes Dev* 13, 2218-30.

Chinnadurai, G. (2002). CtBP, an unconventional transcriptional corepressor in development and oncogenesis. *Mol Cell* 9, 213-24.

Chinnadurai, G. (2003). CtBP family proteins: more than transcriptional corepressors. *Bioessays* 25, 9-12.

Courey, A. J. and Huang, J. D. (1995). The establishment and interpretation of transcription factor gradients in the *Drosophila* embryo. *Biochim Biophys Acta* 1261, 1-18.

Criqui-Filipe, P., Ducret, C., Maira, S. M. and Wasylyk, B. (1999). Net, a negative Ras-switchable TCF, contains a second inhibition domain, the CID, that mediates repression through interactions with CtBP and de-acetylation. *Embo J* 18, 3392-403.

Cuadrado, M., Sacristan, M. and Antequera, F. (2001). Species-specific organization of CpG island promoters at mammalian homologous genes. *EMBO Rep* 2, 586-92.

**Davidson, E. H.** (2001). Genomic Regulatory Systems: Development and Evolution: Academic Press: A Harcourt Science and Technology Company.

Driever, W. and Nusslein-Volhard, C. (1988). A gradient of bicoid protein in Drosophila embryos. Cell 54, 83-93.

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.

Eldon, E. D. and Pirrotta, V. (1991). Interactions of the *Drosophila* gap gene giant with maternal and zygotic pattern-forming genes. *Development* 111, 367-78.

Fisher, A. L. and Caudy, M. (1998). Groucho proteins: transcriptional corepressors for specific subsets of DNA-binding transcription factors in vertebrates and invertebrates. *Genes Dev* 12, 1931-40.

Flores-Saaib, R. D. and Courey, A. J. (2000). Analysis of Groucho-histone interactions suggests mechanistic similarities between Groucho- and Tup1-mediated repression. *Nucleic Acids Res* 28, 4189-96.

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.

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.

Gray, S. and Levine, M. (1996). Transcriptional repression in development. Curr Opin Cell Biol 8, 358-64.

Gray, S., Szymanski, P. and Levine, M. (1994). Short-range repression permits multiple enhancers to function autonomously within a complex promoter. *Genes Dev* 8, 1829-38.

Guss, K. A., Nelson, C. E., Hudson, A., Kraus, M. E. and Carroll, S. B. (2001). Control of a genetic regulatory network by a selector gene. *Science* 292, 1164-7.

Hanes, S. D., Riddihough, G., Ish-Horowicz, D. and Brent, R. (1994). Specific DNA recognition and intersite spacing are critical for action of the bicoid morphogen. *Mol Cell Biol* 14, 3364-75.

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.

Hardison, R., Slightom, J. L., Gumucio, D. L., Goodman, M., Stojanovic, N. and Miller, W. (1997). Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene* 205, 73-94.

Hewitt, G. F., Strunk, B. S., Margulies, C., Priputin, T., Wang, X. D., Amey, R., Pabst, B. A., Kosman, D., Reinitz, J. and Arnosti, D. N. (1999). Transcriptional repression by the *Drosophila* giant protein: cis element positioning provides an alternative means of interpreting an effector gradient. *Development* 126, 1201-10.

Janody, F., Sturny, R., Schaeffer, V., Azou, Y. and Dostatni, N. (2001). Two distinct domains of Bicoid mediate its transcriptional downregulation by the Torso pathway. *Development* 128, 2281-90.

Jiang, J. and Levine, M. (1993). Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* 72, 741-52.

Johnston, M. and Davis, R. W. (1984). Sequences that regulate the divergent GAL1-GAL10 promoter in Saccharomyces cerevisiae. *Mol Cell Biol* **4**, 1440-8.

Keller, S. A., Mao, Y., Struffi, P., Margulies, C., Yurk, C. E., Anderson, A. R., Amey, R. L., Moore, S., Ebels, J. M., Foley, K. et al. (2000). dCtBP-dependent and - independent repression activities of the *Drosophila* Knirps protein. *Mol Cell Biol* 20, 7247-58.

Kim, T. K. and Maniatis, T. (1997). The mechanism of transcriptional synergy of an *in vitro* assembled interferon-beta enhanceosome. *Mol Cell* 1, 119-29.

Koh, S. S., Ansari, A. Z., Ptashne, M. and Young, R. A. (1998). An activator target in the RNA polymerase II holoenzyme. *Mol Cell* 1, 895-904.

Kosman, D. and Small, S. (1997a). Concentration-dependent patterning by an ectopic expression domain of the *Drosophila* gap gene knirps. *Development* 124, 1343-54..

Kraut, R. and Levine, M. (1991a). Mutually repressive interactions between the gap genes giant and Kruppel define middle body regions of the *Drosophila* embryo. *Development* 111, 611-21.

Kraut, R. and Levine, M. (1991b). Spatial regulation of the gap gene giant during *Drosophila* development. *Development* 111, 601-9.

Kulkarni, M. M. and Arnosti, D. N. (2003). Information Display by Transcriptional enhancer. *In press*.

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.

Lifanov, A. P., Makeev, V. J., Nazina, A. G. and Papatsenko, D. A. (2003). Homotypic regulatory clusters in *Drosophila*. Genome Res 13, 579-88.

Loots, G. G., Locksley, R. M., Blankespoor, C. M., Wang, Z. E., Miller, W., Rubin, E. M. and Frazer, K. A. (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* 288, 136-40.

Ludwig, M. Z., Bergman, C., Patel, N. H. and Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403, 564-7.

Ludwig, M. Z. and Kreitman, M. (1995). Evolutionary dynamics of the enhancer region of even-skipped in *Drosophila*. *Mol Biol Evol* 12, 1002-11.

Ludwig, M. Z., Patel, N. H. and Kreitman, M. (1998). Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. *Development* 125, 949-58.

Maier, D., Preiss, A. and Powell, J. R. (1990). Regulation of the segmentation gene fushi tarazu has been functionally conserved in *Drosophila*. *Embo J* 9, 3957-66.

Makeev, V. J., Lifanov, A. P., Nazina, A. G. and Papatsenko, D. A. (2003). Distance preferences in the arrangement of binding motifs and hierarchical levels in organization of transcription regulatory information. *Nucleic Acids Res* 31, 6016-26.

Mannervik, M. and Levine, M. (1999). The Rpd3 histone deacetylase is required for segmentation of the *Drosophila* embryo. *Proc Natl Acad Sci U S A* **96**, 6797-801.

Mannervik, M., Nibu, Y., Zhang, H. and Levine, M. (1999). Transcriptional coregulators in development. *Science* 284, 606-9.

Markstein, M. and Levine, M. (2002). Decoding *cis*-regulatory DNAs in the *Drosophila* genome. *Curr Opin Genet Dev* 12, 601-6.

Markstein, M., Markstein, P., Markstein, V. and Levine, M. S. (2002). Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila* embryo. *Proc Natl Acad Sci U S A* 99, 763-8.

Merika, M. and Thanos, D. (2001). Enhanceosomes. Curr Opin Genet Dev 11, 205-8.

Munshi, N., Agalioti, T., Lomvardas, S., Merika, M., Chen, G. and Thanos, D. (2001). Coordination of a transcriptional switch by HMGI(Y) acetylation. *Science* 293, 1133-6.

Nibu, Y., Zhang, H., Bajor, E., Barolo, S., Small, S. and Levine, M. (1998a). dCtBP mediates transcriptional repression by Knirps, Kruppel and Snail in the *Drosophila* embryo. *Embo J* 17, 7009-20.

Nibu, Y., Zhang, H. and Levine, M. (1998b). Interaction of short-range repressors with *Drosophila* CtBP in the embryo. *Science* 280, 101-4.

Nibu, Y., Zhang, H. and Levine, M. (2001). Local action of long-range repressors in the *Drosophila* embryo. *Embo J* 20, 2246-53.

Nusslein-Volhard, C., Frohnhofer, H. G. and Lehmann, R. (1987). Determination of anteroposterior polarity in *Drosophila*. *Science* 238, 1675-81.

Nusslein-Volhard, C., Kluding, H. and Jurgens, G. (1985). Genes affecting the segmental subdivision of the *Drosophila* embryo. *Cold Spring Harb Symp Quant Biol* 50, 145-54.

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

Pankratz, M. J. and Jackle, H. (1990). Making stripes in the Drosophila embryo. Trends Genet 6, 287-92.

Papatsenko, D. A., Makeev, V. J., Lifanov, A. P., Regnier, M., Nazina, A. G. and Desplan, C. (2002). Extraction of functional binding sites from unique regulatory regions: the *Drosophila* early developmental enhancers. *Genome Res* 12, 470-81.

**Pham, A. D., Muller, S. and Sauer, F.** (1999). Mesoderm-determining transcription in *Drosophila* is alleviated by mutations in TAF(II)60 and TAF(II)110. *Mech Dev* 84, 3-16.

**Piano, F., Parisi, M. J., Karess, R. and Kambysellis, M. P.** (1999). Evidence for redundancy but not trans factor-cis element coevolution in the regulation of *Drosophila* Yp genes. *Genetics* 152, 605-16.

**Poortinga, G., Watanabe, M. and Parkhurst, S. M.** (1998). *Drosophila* CtBP: a Hairyinteracting protein required for embryonic segmentation and hairy-mediated transcriptional repression. *Embo J* 17, 2067-78.

**Rebeiz, M., Reeves, N. L. and Posakony, J. W.** (2002). SCORE: a computational approach to the identification of *cis*-regulatory modules and target genes in whole-genome sequence data. Site clustering over random expectation. *Proc Natl Acad Sci U S A* **99**, 9888-93.

Rivera-Pomar, R. and Jackle, H. (1996). From gradients to stripes in *Drosophila* embryogenesis: filling in the gaps. *Trends Genet* 12, 478-83.

Ryu, J. R., Olson, L. K. and Arnosti, D. N. (2001). Cell-type specificity of short-range transcriptional repressors. *Proc Natl Acad Sci U S A* 98, 12960-5.

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.

Sean B. Carroll, J. K. G., Scott D. Weatherbee. (2001). From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design: Blackwell Science.

Seipel, K., Georgiev, O. and Schaffner, W. (1992). Different activation domains stimulate transcription from remote ('enhancer') and proximal ('promoter') positions. *Embo J* 11, 4961-8.

Shi, Y., Sawada, J., Sui, G., Affar el, B., Whetstine, J. R., Lan, F., Ogawa, H., Luke, M. P. and Nakatani, Y. (2003). Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* 422, 735-8.

Small, S., Arnosti, D. N. and Levine, M. (1993). Spacing ensures autonomous expression of different stripe enhancers in the even-skipped promoter. *Development* 119, 762-72.

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.

Small, S., Kraut, R., Hoey, T., Warrior, R. and Levine, M. (1991). Transcriptional regulation of a pair-rule stripe in *Drosophila*. Genes Dev 5, 827-39.

Small, S. and Levine, M. (1991). The initiation of pair-rule stripes in the Drosophila blastoderm. Curr Opin Genet Dev 1, 255-60.

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.

Struhl, G. (1989). Morphogen gradients and the control of body pattern in insect embryos. *Ciba Found Symp* 144, 65-86; discussion 86-91, 92-8.

Struhl, K. (1991). Mechanisms for diversity in gene expression patterns. *Neuron* 7, 177-81.

Struhl, K. (2001). Gene regulation. A paradigm for precision. Science 293, 1054-5.

Strunk, B., Struffi, P., Wright, K., Pabst, B., Thomas, J., Qin, L. and Arnosti, D. N. (2001). Role of CtBP in transcriptional repression by the *Drosophila* giant protein. *Dev Biol* 239, 229-40.

Subramanian, T. and Chinnadurai, G. (2003). Association of class I histone deacetylases with transcriptional corepressor CtBP. *FEBS Lett* 540, 255-8.

Sundqvist, A., Sollerbrant, K. and Svensson, C. (1998). The carboxy-terminal region of adenovirus E1A activates transcription through targeting of a C-terminal binding protein-histone deacetylase complex. *FEBS Lett* **429**, 183-8.

Szymanski, P. and Levine, M. (1995). Multiple modes of dorsal-bHLH transcriptional synergy in the *Drosophila* embryo. *Embo J* 14, 2229-38.

Thanos, D. and Maniatis, T. (1995). Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell* 83, 1091-100.

Wu, X., Vakani, R. and Small, S. (1998). Two distinct mechanisms for differential positioning of gene expression borders involving the *Drosophila* gap protein giant. *Development* 125, 3765-74.

Yuh, C. H., Bolouri, H. and Davidson, E. H. (1998). Genomic *cis*-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science* 279, 1896-902.

**Zhang, H. and Levine, M.** (1999). Groucho and dCtBP mediate separate pathways of transcriptional repression in the *Drosophila* embryo. *Proc Natl Acad Sci U S A* **96**, 535-40.

Zhou, J., Zwicker, J., Szymanski, P., Levine, M. and Tjian, R. (1998). TAFII mutations disrupt Dorsal activation in the *Drosophila* embryo. *Proc Natl Acad Sci U S A* 95, 13483-8.

#### **Chapter IV**

## **Conclusions and Future Directions**

#### Enhancers: Molecular Computers or Information displays- Biochemical implications

Our analysis of structure-function relationships within enhancers has demonstrated that there are at least two models for enhancer function- the enhanceosome model and the 'Information Display' or 'Billboard' model. The differences between the enhanceosome and billboard enhancers reflect the differences in the mechanism by which regulatory information incident on the enhancer element is converted into transcriptional output.

In the enhanceosome model, the enhancer functions like a 'molecular computer' that is solely responsible for directly sensing and computing both the quality and quantity of regulatory signals (transcription factor inputs), while the core promoter element simply responds to the instructions generated by these molecular logic circuits. Although demonstrated for only a handful of elements, this conceptual formulation and terminology is rife in most contemporary literature (Flores-Saaib and Courey, 2000; Halfon et al., 2002; Halfon and Michelson, 2002; Xu et al., 2000; Yuh et al., 1998). The best studied example of computer-like integration of signaling inputs to provide highly specific gene activation is the human interferon-ß (IFN-ß) enhancer, which drives transcription of the IFN-ß gene in response to viral infection (Struhl, 2001). The regulatory proteins (activators and architectural proteins) assemble through cooperative interactions into a well-defined nucleoprotein complex called the 'enhanceosome'.

Assembly of the enhanceosome is absolutely essential for the transcription of the IFN-B gene in response to viral infection in cells. Individual activators bound to their sites in the enhancer do not by themselves stimulate transcription; enhancers that contain multiple copies of any one activator site are less inducible by the virus and respond nonspecifically to other signals. These observations indicate that the higher order nucleoprotein complex or enhanceosome provides a stereospecific interface for interaction with the basal transcriptional machinery, possibly engaging several components of the basal machinery simultaneously to effect synergistic and highly specific gene activation (Cai et al., 1996; Carey et al., 1990; Chi et al., 1995). In this structured element, the presence of each transcription factor binding site and its precise arrangement within the regulatory element are critical in dictating the output of the element (Kim and Maniatis, 1997; Munshi et al., 2001; Thanos and Maniatis, 1995). The enhanceosome therefore imposes the restriction that the target gene would be activated only under very precise conditions, when all the regulatory proteins are present and are functionally active. Thus, the IFN-B enhanceosome functions as a precise on/off binary transcriptional switch in response to the appropriate stimulus (Struhl, 2001).

In my study (described in Chapter 2) I find clear evidence for the first time that the computational functions usually ascribed to the enhancer itself are actually shared with the basal transcriptional machinery and suggest instead, that the basal machinery plays an active, rather than passive role in interpreting transcriptional signals from enhancers. Using synthetic compact enhancer elements I demonstrated that closely spaced factor binding sites, situated within compact cis-elements, are actually able to be independently "read off" by the transcriptional machinery. In such an enhancer, exact binding site locations are not critical and each bound factor contributes to the overall transcriptional output. Thus, the removal or addition of individual binding sites might attenuate the overall level of the transcriptional output, but will not completely abolish enhancer function, as is the case with the enhanceosome element. The 'on or off' binary switch decisions that appear to be transmitted from the enhancer to the basal machinery may actually result from the basal transcriptional machinery reading a series of redundant signals encoded within the enhancer. Thus, instead of acting in a concerted, all- ornothing computational fashion, the enhancer regulates target gene expression through successive dialogues/interactions with the basal transcriptional machinery. This feature provides a novel view of biochemical mechanisms of enhancer-promoter interactions, suggesting that activators are not all making simultaneous contacts with the basal machinery but that discrete portions of the enhancer can be successively "sampled" by the basal transcriptional machinery from different directions.

What is the nature of the interaction(s) between the enhancer bound regulatory proteins and the cognate promoter? One idea is that the 'effect' produced by the transcription factors bound at the enhancer spreads along the DNA to the target promoter. Just as the Sir repressosome generates a transcriptionally silent chromatin structure that is able to spread along the chromatin fiber, it is possible that the enhancer bound factors nucleate an open or active chromosomal state that spreads to the core promoter allowing the binding of the basal transcriptional machinery. Alternatively, the enhancer-bound proteins and their associated cofactors establish a productive interaction(s) with the cognate promoter by looping out the DNA between the enhancer and promoter such that the regulatory proteins at the enhancer make multiple direct contacts with the basal promoter elements. The information display model clearly favors the latter, where multiple successive or simultaneous interactions between promoter and enhancer subelements lead to a transcriptional readout.

If the enhancer subelements are indeed making direct contacts with basal promoter elements, then what are the timescales for these interactions? Again, the billboard model would favor transient interactions versus stable enhancer-promoter complexes. The dynamic nature of the interactions would allow different enhancer subelements to communicate the regulatory information contained within them to the basal transcriptional machinery. Such a situation is of critical importance in the case of modular cis-regulatory regions controlling the expression of developmental genes where multiple independently acting enhancer elements either turn the gene off or on at the same time.

Multiple, temporal 'snapshots' of the precise physical architecture of enhancerpromoter complexes would provide definitive information about the molecular aspects of the interactions between cis-regulatory element and the core promoter. However, current structural techniques have not yet succeeded in capturing the complexity of multiple activators and repressors interacting with the basal machinery on a chromatinized gene. An important consideration towards attempting to solve a 'crystal structure' for the entire transcriptional complex, is whether regulatory proteins are stably bound to their cognate

142

sites with the enhancer element or are in constant flux between the bound state and unbound state. It is conceivable that such a 'snapshot' of an enhancer-promoter complex is possible in the case of the enhanceosome, which activates expression of the linked gene only when all the regulatory proteins are present and bound in the correct conformation to their sites within the element.

## Modularity within the 'Billboard' enhancer substructure

I have demonstrated that the 'billboard' enhancer is capable of simultaneously representing both the states of activation and repression at the same time and in the same nucleus. In this case, the independent information units might consist of one or a few factor binding sites, the enhancer substructure. On a larger scale, this ability to display contrasting information at the same time is similar to the information content of the multiple modular stripe enhancers that regulate even-skipped (Goto et al., 1989; Harding et al., 1989; Small et al., 1992). In this case a separate enhancer module within a complex promoter can be repressed while an adjacent enhancer module that regulates the same promoter can be activated, at the same time and in the same nucleus. Modular organization of cis-regulatory regions is typical of many developmentally regulated genes in higher eukaryotes and allows the gene to be expressed in temporally and spatially complex expression patterns. The modularity of cis-regulatory regions of genes also facilitates evolution because individual elements can evolve to either acquire additional function(s) or lose function(s) independently without destroying the primary function of the gene. In a similar vein, on a local level in the 'billboard' model a single enhancer can be thought to contain a similar modular organization where separate modules signal independently to the basal promoter.

## Enhanceosome and 'Billboard' enhancer – two extremes of a continuum

The 'enhanceosome' and the 'billboard' enhancer are likely to represent two extremes of a continuum that describes the range of function of cis-regulatory elements. Most enhancers are likely to lie within this spectrum, exhibiting properties that are more like one or the other. A number of experimental studies have revealed enhancers with properties of both enhanceosomes and billboards. The Drosophila zen ventral repression region (VRR) is approximately 600 bp in length and is responsible for both the repression of zen expression in the ventral region of the embryo and for its expression in the dorsal region. The VRR contains important binding sites for the Dorsal protein (Doyle et al., 1989; Jiang et al., 1992), which normally activates transcription. On this element however, adjacent AT-rich binding sites bind the Cut and Dead Ringer factors, which in turn allow the recruitment of the Groucho corepressor that converts Dorsal from an activator to a repressor. Exact spacing between the AT-rich sites and the Dorsal sites is required for Dorsal to function as a repressor. A 180-bp region from the zen VRR containing three AT-rich sites, AT1 to AT3, and three Dorsal binding sites, dll to dl3, was altered by the insertion of a 5-bp spacer between the AT2 and dl2 sites. This change resulted in the elimination of the repression activity of the element, while the insertion of a 10-bp spacer restored repression (Cai et al., 1996; Valentine et al., 1998). This result implies that the correct stereospecific positioning of Dorsal relative to AT2 bound

proteins is critical for repression, a feature suggesting a high degree of cooperativity characteristic of the enhanceosome model. At the same time, a separate portion of the 600 bp *zen* VRR binds transcriptional activators responsible for transcription of *zen* in the dorsal regions of the embryo. This regulation is independent of the repression module and can function autonomously, reminiscent of the 'Billboard' model. Thus, the *zen* VRR exhibits properties that are enhanceosome-like and at the same time functions like an information display.

Another example of an enhancer with properties of both the enhanceosome and billboard models is the element that is involved in the transcriptional control of the Drosophila terminal gap gene huckebein (hkb) that depends on Torso (Tor) receptor tyrosine kinase (RTK) signaling and the Rel/NFkappaB homolog Dorsal. Analysis of the interplay between Dorsal, Groucho and Dead ringer on the hkb enhancer demonstrates that when the Dorsal and Dead Ringer binding sites are separated by a distance of  $\sim 90$ bp. Groucho recruited by Dead ringer blocks rather than converts Dorsal activator function, which is in contrast to what is observed in case of the repressosome described above. In this case, Groucho quenches or blocks the Dorsal activator from signaling to the basal machinery. Removing the binding site for Dead ringer allows Dorsal to activate transcription, indicating that the element is essentially bipartite and constitutes two independently acting subelements. Reducing the distance between Dorsal and Dri binding sites, however, switches Dorsal into a Gro-dependent repressor resulting in a repressosome type of activity that overrides activation of transcription (Hader et al., 2000).

In the enhanceosome model transcription factor binding sites are highly constrained, while the billboard enhancer affords more plasticity in the arrangement of transcription factor binding sites. A defective eve stripe 2 enhancer lacking a crucial Bicoid binding site can be complemented with a high affinity Bicoid binding site inserted at a new location, therefore exhibiting structural flexibility that is a characteristic of the billboard enhancer. However, there are some constraints on enhancer organization; insertion of the same high affinity Bicoid sequence at a different location within the element results in only partial restoration of the defective stripe (Arnosti et al., 1996a). In addition, several attempts to construct an artificial stripe 2 enhancer element by simply clustering binding sites for Bicoid and Hunchback activators that regulate the native element have been unsuccessful, indicating that there may be some spacing or orientation constraints between the factor binding sites. Thus, in this case although there is some flexibility in the arrangement of transcription factor binding sites, it is clearly not without any functional constraints. As described before, independent information units within the information display type of enhancer might consist of one or a few factor binding sites. It is conceivable that within an independently acting information unit functional constraints between factor binding sites exist, which are the 'grammar' rules for enhancer substructure as described in Chapter 3.

At the other end of the spectrum is the transcriptional regulation of *cut* expression in the wing imaginal disc of *Drosophila melanogaster*. The transcription factor Scalloped potentiates the transcriptional activation of this gene throughout the disc. Specific Notch signaling along the dorsal/ventral boundary of the disc activates the Suppressor of Hairless [Su(H)] transcription factor, and Su(H) together with Scalloped bind to the *cut* enhancer to activate Cut expression specifically along the dorsal/ventral boundary of the disc. A simple cluster of Scalloped and Su(H) binding sites without regard to exact distance and arrangement, is sufficient to recapitulate this pattern of expression (Guss et al., 2001). In a similar manner, a simple cluster of Dorsal and Twist binding sites can recapitulate the expression pattern driven by endogenous regulatory elements that bind these transcription factors (Szymanski and Levine, 1995). These observations suggest that some cis-regulatory elements do not require the formation of a higher order nucleoprotein complex to regulate gene expression, and are free of detectable constraints in their internal design. The lack of 'grammar' rules might reflect the nature and the mechanism(s) of function of the transcription factors that bind to such elements. Clearly, elements that bind short-range repressors will exhibit some constraints in internal design that are dictated by the way in which these proteins function.

A number of studies (Bouallaga et al., 2000; Ellwood et al., 2000; Read et al., 1994) suggest that enhancer elements designed for sudden inducibility, such as those involved in immune response, might tend to function as enhanceosomes. Thus, it is likely that eukaryotic organisms use 'billboard' enhancers to achieve diversity in gene expression patterns and evolutionary flexibility, and enhanceosomes to achieve high levels of specific gene activation (Struhl, 2001).

# 'Grammar' rules for repressor function: Insights into cis-regulatory evolutionary dynamics

Cis-regulatory DNA function may evolve either by de novo evolution of cisregulatory elements via changes in nonfunctional DNA or through the evolution of cisregulatory elements from preexisting functional elements. These modifications include duplications and DNA rearrangements of existing functional elements, and the loss or gain of binding sites through changes in nucleotide sequence. The ability of regulatory DNA to evolve is greatly facilitated by the modularity of cis-regulatory elements. Individual elements can act, and therefore evolve, independently of others (Fink and Scandalios, 2002). The typical organization of the cis-regulatory regions of developmental regulatory genes, composed of many independent elements, is tacit evidence for the expansion and diversification of cis-regulatory systems in evolution. In a similar manner, the independently acting subelements of the billboard enhancer have a greater capacity for evolutionary flexibility than the nonmodular element such as the enhanceosome, where every component (transcription factor binding site) is optimally linked to every other component and is thus effectively frozen.

To understand the role played by cis-regulatory DNA in evolution it is important to appreciate two common features of cis-acting elements. First, cis-regulatory elements are regulated by multiple distinct transcription factors that interact with each other, with cofactors, with chromatin and with components of the basal transcriptional machinery. Second, the spatial relationship of binding sites for transcription factors within ciselements can be critically important, as with enhanceosomes, but also for billboard

enhancers that still are subject to contextual grammar. Thus, a meaningful evolutionary analysis of a cis-regulatory sequence requires detailed information about the locations of transcription factor binding sites within a sequence, the functional specificity of the binding sequences, and the spatial requirements for their interactions. Without a deep understanding of the structure-function correlations that exist with cis-regulatory elements it is impossible to discern functional consequences of evolutionary changes in regulatory sequence. Using defined synthetic enhancer elements I have systematically deduced the contextual parameters in terms of the number, affinity, relative spacing and arrangements of binding sites that define the enhancer 'grammar' required for the function of an important class of regulators, the short-range transcriptional repressors (described in Chapter 3). Although previous studies had provided evidence that some parameters might be important, my work with defined elements has provided a quantitative look at these factors. This study has important ramifications on the field of regulatory evolution as it provides quantitative measures to further assess the evolutionary dynamics of enhancers regulated by short-range repressors.

From our limited perspective enhancer elements appeared to have haphazard assemblages of regulatory factor binding sites that were adequate to get the job done with little logic to their internal organization. However, I have demonstrated (Chapter2 and 3), that there are internal design principles that direct how different features of transcription factor binding sites provide varied transcriptional outputs. Discrete regions of the billboard enhancer are 'read' by the basal transcriptional machinery through multiple successive or simultaneous interactions and the sum of these multiple interactions dictates the transcriptional output of the element. Therefore, the 'billboard' enhancer can display contrasting regulatory information, such that both the activated and repressed states are represented within the same enhancer element at the same time and in the same nucleus. This would appear to be in contrast to the action of natural enhancers, which are seen to function as binary on or off transcriptional switches. Resolving this apparent paradox, I demonstrated that the billboard enhancer can provide uniform information in the manner of a binary transcriptional switch if the stoichiometry of repressor and activator binding sites is adjusted to ensure that all possible enhancer subelements provide the same information. Thus, using simple well defined synthetic enhancer elements we were able to highlight that stoichiometry between activators and repressors is one of design principles of cis-acting elements that allow them to function as binary transcriptional switches.

Further manipulation of these composite elements to alter the number of activator and repressor binding sites, relative affinities, spacing and distribution of these binding sites allowed us to define other contextual parameters that dictate repression effectiveness. First, we find that although a close proximity between the short-range repressors and their targets is a critical factor in dictating repression effectiveness, it is not sufficient. I found that the general notion that short-range repressors block the activity of all protein complexes within a 100 bp is an oversimplification and that in some contexts (greater number of activators or high affinity of activator binding sites) activators can escape repression even when bound close to short-range repressors. Second, the effectiveness of repression correlates with the number and affinity of activator binding sites. Although stronger binding of activators correlates with greater transcriptional output, I was able to demonstrate that repression is not inversely proportional to activator strength, as even weak activators with good binding sites are not repressed. A third general parameter is the importance of proper distribution of activator and repressor binding sites for effective repression. The 'Billboard' model of enhancer action, which suggests that subelements of a cis-regulatory enhancer can be independently sampled by the basal machinery, can explain this finding. Effective repression of an entire enhancer would thus require distributed pattern of repressors that would interfere with the activity of each subelement. Fourth, short-range repressors do not exhibit activator specificity, indicating that the repression mechanism may be targeted to the binding potential of the activator rather than its activation domain. These contextual parameters constitute a 'grammar' of repressor function. The quantitative description of these elements will facilitate the identification of functional changes in cis-regulatory sequences that are the raw material for morphological evolution.

Interspecific sequence comparisons of noncoding cis-regulatory regions have shown that inspite of selective constraints, the structure and sequences of cis-elements change over time, sometimes dramatically so, even in cases where expression patterns are conserved (Hardison et al., 1997; Loots et al., 2000; Ludwig et al., 2000; Ludwig and Kreitman, 1995; Ludwig et al., 1998). Homologous sequences to the *D. melanogaster eve* stripe 2 element have been isolated from a host of other *Drosophila* species. These elements all drive an accurately positioned stripe of reporter gene expression in the *eve* stripe 2 domain in *D. melanogaster*, although considerable sequence divergence occurs among the elements (Figure IV-1). Interestingly, some of the sequence changes in the other Drosophila species abolish sites that are known to be essential in the D. melanogaster element. For instance, the third Bicoid binding site (bcd-3) and the first Hunchback binding site found in D. melanogaster lack counterparts in D. pseudobscura, D. erecta, and D. yakuba. This analysis suggests that the bcd-3 site may be a new site in D. melanogaster (Kreitman, 1996; Ludwig et al., 2000; Ludwig and Kreitman, 1995; Ludwig et al., 1998). In light of the contextual rules defined for the short-range repressors, we predict that the D. melanogaster stripe 2 element requires higher levels of repression to compensate for the additional activation potential. Indeed, sequence analysis of the homologous elements demonstrates that a 30 bp deletion has moved a nearby Giant site closer to the site of this Bicoid activator, presumably compensation for the increase in activation potential. A theoretical framework for interpreting the evolutionary changes in any cis-regulatory elements, although still in its infancy, will be greatly facilitated by the empirical determination of such spatial rules for other classes of transcriptional regulators.

# 'Grammar' rules for repressor function: Improved computational identification of functional cis-regulatory modules

Recently, whole genome sequence assemblies have become available providing a powerful foundation to identify and analyze cis-regulatory module function and organization on a global scale. Two current approaches to identify candidate regulatory regions from genomic data are computational methods that look for clusters of transcription factor binding sites (Berman et al., 2002; Markstein and Levine, 2002;

Kr6	Kr5	G	Gt2	Kr4		Gt	Kr3	Kr2	Krl	
	Bcd5	Bcd4	B	cd3	Bcd2	Hb <b>3</b> ₽	Bedl	Hb2		Hbl

# B

A

**Conservation of Transcription Factor Binding Sites in Eve Stripe 2 Enhancer** 

	Sites								
	Kr	Bcd	Hb	Gt					
Species	123456	12345	123	123					
D. simulans	PPPPP	ΡΡSSP	SPP	<b>ΡΡ</b> S					
D. yakuba	ΡΡSPPΡ	SPWSP	SSP	PSS					
D. erecta	ΡΡSPPP	SPSSP	SSP	PSS					
D. pseudobscura	SSSSPP	SSASP	APS	SWS					
D. picticornis	WSWSPP	WSASP	A S W	w w w					

# Conservation

P = perfect

S = strong (1-2 changes)

W =weak (3 or more changes)

A = absent

# Figure IV-1: Evolutionary dynamics of transcription factor binding sites in a conserved cis-regulatory element.

(A) Binding sites for the Krüppel (Kr), Giant (Gt), Bicoid (Bcd), and Hunchback (Hb) proteins in the 670 bp *D. melanogaster eve* stripe 2 cis-regulatory element are shown. (B) The conserved binding sites in five different *Drosophila* species are tabulated. The degree of sequence conservation within each site is indicated (P, S, W, A). Note that certain sites such as Kr5, Kr6, and Bcd5 are perfectly conserved, whereas other sites such as Hb1 and Bcd3 are missing from certain species. (Figure adapted from Carroll, SB, 2001).

Markstein et al., 2002; Rebeiz et al., 2002) and phylogenetic comparisons that identify evolutionarily conserved sequences (Bergman and Kreitman, 2001; Maier et al., 1990). One main difficulty with the output from computational searches for clusters of transcription factor sites is the large number of false-positive and false-negative results. The short length and degenerate nature of transcription-factor-binding sites account for some of these misleading predictions. Furthermore, genes are rarely controlled by a single transcription factor and accumulating evidence suggests that specific combinations of transcription factors are required to achieve the complex differential expression of genes in higher organisms. In addition, as I have demonstrated in Chapters 2 and 3, the proper spatial organization of transcription factor binding sites within the element may be critical for their proper biological function. The structural basis for combinatorial regulation exists in the specific organization of multiple transcription factor binding sites. Transcription factors interact with each other or with cofactors to mediate their function, therefore spatial correlations/constraints between their binding sites need to be taken into account. Thus, in many cases where cis-regulatory modules predicted by computational methods appear suitable, something in the arrangement of sites (or a wider context) renders the cis-element non-functional. In order to achieve better predictions and eliminate false-positive and false-negative results, computational methods should include, in addition to binding site density and relative affinities, parameters for spacing and position of binding motifs within cis-elements in the search algorithms. These parameters are difficult to predict accurately a priori. Thus, computational approaches will benefit from the availability of at least one well-defined representative of a particular

regulatory network to serve as a starting paradigm. An example of how this will actually work is described on p. 170.

At the other end of the spectrum, researchers are studying cis-regulatory element function using extensive deletion and mutational analyses and experimentally determining the consequence of these changes on expression pattern *in vivo*. While such an approach is useful for discovering the functional interactions between the different components of a cis-regulatory element, it is constrained in its focus to a specific *cis* element. Furthermore, the analyses of complex endogenous regulatory elements is often confounded by the lack of information about the internal organization of the element with regard to the number, relative affinities, order, spacing and orientation of binding sites, and even the identity of the transacting regulatory proteins. Because of this complexity of cis-regulatory regions, the contributions of individual parameters to the overall transcriptional output cannot be accurately quantified and multiple experiments are required to gain even a rough overview of how an expression pattern is generated.

My study seeks to find a bridge between the two approaches by using simple well-defined synthetic enhancer elements containing binding sites for transcriptional activators and repressors. This approach allowed us to systematically identify the general operating principles in terms of cis-element architecture for the short-range transcriptional repressors. We defined a set of 'grammar' rules in terms of the number, relative affinities, spacing and arrangement of binding sites that are required for proper regulation by short-range repressors. Since this approach allowed us to systematically

155

alter one parameter at a time, we were able to dissect with greater accuracy the contribution of each parameter to the transcriptional output of the element. The empirical determination of these elements and the relative contribution of each element to repressor function will facilitate bioinformatic analysis of novel gene regulatory sequences. For example, suppose that a computational search for novel cis-elements regulated by Giant based on clustering of binding sites for transcription factors that are known to act together with, it identifies a putative regulatory module. Given the relative number, affinities, spacing and distribution of repressor and activator sites within the module we would be better able to predict the possibility of Giant regulating that element.

#### **FUTURE DIRECTIONS**

## Molecular mechanism(s) of short-range transcriptional repression

Gene-specific repression of transcription plays a central role in gene regulation. This is true not only for the spatial control of gene activity in development, during which boundaries of gene expression are determined by the spatially restricted localization or activity of transcriptional repressors (Mannervik and Levine, 1999; Mannervik et al., 1999) but also true for the control of gene expression by extracellular signals, in which genes are often maintained in an off state by repressor proteins until signal transduction alleviates the repression (Barolo and Posakony, 2002; Roose and Clevers, 1999). As introduced in Chapter 1, studies on transcriptional repression in the *Drosophila* embryo suggest that there may be two basic forms of repression, long-range and short-range repression (Arnosti et al., 1996b; Barolo and Levine, 1997; Gray and Levine, 1996; Gray et al., 1994; Hewitt et al., 1999; Keller et al., 2000). Short-range repression represents a flexible form of gene regulation exhibiting either enhancer- or promoter- specific effects depending on the position of the repressor binding sites. This flexibility contrasts with long-range repressors, which can block multiple enhancers over distances of several kilobases regardless of location within a gene complex. The different activities of each class probably reflect distinct mechanisms employed, however, the molecular events that differentiate one from the other are not well understood.

Two modes of short-range repression have been proposed that are consistent with the activity of these proteins seen in vivo: Short-range repressors might "quench" activator proteins locally, within a short distance of their binding site by displacing them from the DNA through chromatin structure modification, or by preventing them from contacting their target in the basal transcriptional machinery; or the repressors might directly contact some component of the basal transcriptional machinery, but only when brought to the promoter by a closely linked activator (hitchhiking) or when bound near the start site of transcription. Studies on the *Drosophila* short-range repressor Snail prompted the proposal that quenching might involve direct protein-protein interactions between repressors and upstream activators (Gray et al., 1994). This type of mechanism has been proposed for the *Drosophila* Krüppel protein where transient transfection assays have suggested that Krüppel can selectively repress transcription activated by a glutamine-rich activator but not by an acidic activator (Licht et al., 1993). Further studies identified two evolutionarily conserved repression domains in the Krüppel protein that differ in activator specificity (Hanna-Rose et al., 1997). Knirps and Giant have been shown to repress heterologous activators (Arnosti et al., 1996a; Arnosti et al., 1996b; Hewitt et al., 1999; Kulkarni, 2003), and therefore do not appear to be "dedicated" repressors.

As discussed in Chapter 1, a common property of short-range transcriptional repressors is their interaction with the evolutionarily conserved corepressor CtBP (C-terminal Binding Protein) through short PXDLS peptide motifs found in the transcriptional repressors (and in other interacting proteins). CtBP has been shown to interact with chromatin modifying factors, including Histone deacetylases (HDAC1 and HDAC2) and Histone methyltransferases (Chinnadurai, 2002; Chinnadurai, 2003; Subramanian and Chinnadurai, 2003; Sundqvist et al., 1998). Thus, the major activity of CtBP might be to serve as a bridging molecule, recruiting chromatin or transcription factor modifying enzymes to the promoter. CtBP has also been shown to have a weak dehydrogenase activity in vitro, but the role of this activity if any in repression is yet unknown (Balasubramanian et al., 2003; Kumar et al., 2002). Alternatively, CtBP might possess an enzymatic activity similar to the NAD-dependent Sir 2 repressor protein, which requires NAD to mediate deacetylation of histone proteins (Marmorstein, 2002).

Drosophila short-range repressors also possess CtBP-independent repression activities and the dCtBP requirement can vary on a gene-to-gene as well as on an enhancer-to-enhancer basis (Keller et al., 2000; Nibu et al., 2003; Strunk et al., 2001). Multiple repression activities may allow for quantitative or qualitative effects on gene expression and may be context-dependent. Thus, qualitatively, a repressor may function selectively in a tissue-specific manner or in a activator-specific manner (Postigo and Dean, 1999) or in different promoter contexts (Lunyak et al., 2002). Quantitatively, dual activities may increase the overall level of repression.

My analysis of transgenes with varying number and affinities of activator binding sites indicate a mechanism of short-range repression that is highly dependent on a critical ratio of enhancer bound activators and repressors. However, I also find that stoichiometry by itself is insufficient for proper regulation by the Giant short-range repressor and subtle (< 40 bp) changes in the spacing from the nearest activators sharply attenuates repression effectiveness. This study clearly demonstrated for the first time that three parameters, the number, affinity and relative distance of activator binding sites, contribute equally and are critical in order for Giant to mediate repression. These contextual parameters together with the observation that Giant's ability to repress does not depend on the nature of activation domains suggest that short-range repressors might function by preventing access to DNA by activators perhaps through chromatin remodeling on a very local scale. As indicated previously, biochemical purification of CtBP complexes from HeLA cells has shown that such complexes contain chromatin modifying factors, including Histone deacetylases (HDAC1 and HDAC2) (Shi et al., 2003). Preliminary biochemical studies in

our laboratory (Struffi, unpublished) also indicate that the Knirps protein interacts with the *Drosophila* homolog of HDAC1, Rpd3, in embryos. Recruitment of Rpd3 has been shown to create a localized domain of deacetylated histones that extends one to two nucleosomes to either side of the recruitment site (Kadosh and Struhl, 1998a; Kadosh and Struhl, 1998b; Rundlett et al., 1998). In yeast, it has been demonstrated that Rpd3dependent repression occurs only when the recruitment site is located at a distance of <200 bp relative to the region containing the activator binding site and core promoter elements, which is also in accord with the size of the domain of histone deacetylation. Mapping experiments with target promoters indicated that histone deacetylation peaks at the Rpd3 recruitment site and extends 200 to 300 bp in both directions (Deckert and Struhl, 2002). Our hypothesis is that Giant and other short-range repressors may prevent activator binding by remodeling chromatin on a local scale to establish a repressive chromatin domain. High affinity or greater number of activator binding sites may allow the activator to bind in spite of such repressive chromatin remodeling.

Although we have defined the contextual elements of cis-regulatory enhancers that allow short-range repressors to mediate repression effectively, we do not yet understand the physical and biochemical changes in promoter complexes that accompany short-range repression. Chromatin Immunoprecipitation (ChIP) assays in *Drosophila* embryos using the simple genetic switch elements defined in this study will facilitate our understanding of the molecular mechanisms of repression employed at developmentally regulated genes. These experiments can be designed to directly test mechanistic questions about the nature of short-range transcriptional repression, specifically whether repression

blocks activator access to the DNA, alters chromatin structure, or whether changes in the composition or presence of the basal machinery is observed. The use of the synthetic genetic elements defined in this study for ChIP analysis offers two advantages. First, the endogenous targets of short-range repressors such as the pair-rule gene even-skipped, are expressed in discrete domains in the early blastoderm embryo and therefore, are not useful targets for ChIP experiments as such assays would result in high background noise. Using transgenes as those described in Chapter 2 and 3 of this study, would allow one to turn the expression of the reporter transgene on or off throughout the embryo using inducible and ubiquitous expression of the activators and repressors. Second, the use of enhancer elements where the exact identity of all the components and their spatial positions within the element are known, will allow the determination of the molecular and biochemical changes that accompany small changes in the internal design (in terms of activator-repressor stoichiometry, relative affinities, spacing and positioning of binding sites) of the enhancer that either result in repression or the lack thereof. Such studies would provide quantitative data that could facilitate the predictive modeling studies described in the last section.

The biochemical identification and analysis of protein complexes associated with short-range transcriptional repressors or their cofactor, CtBP, from *Drosophila* extracts will also provide insights on the molecular aspects of short-range repression.

The availability of both short- and long-range repressors adds another layer of flexibility to gene regulation and may serve specific gene regulatory needs. Long-range

repression provides the possibility of shutting down an entire locus regardless of the number of separate regulatory modules that control the activity of that locus. This kind of repression has often been referred to as silencing because an entire chromosomal locus is inactivated. On the other hand, short-range repression provides a way to control the activity of one enhancer without interfering with the activity of others and is useful for creating precise, tunable repression required to establish intricate patterns of expression. In addition to having a longer range of action, I find that the long-range repressor Hairy is more potent than Giant even on a local scale and can block the activity of a greater number of activators as compared to Giant.

The mechanisms by which short-range and long-range *Drosophila* repressors inhibit transcription are poorly understood, although one model of repression in the embryo suggests that the short-range/long-range distinction results from the recruitment of distinct cofactors (Nibu et al., 1998a; Nibu et al., 1998b; Zhang and Levine, 1999). Short-range repressors recruit the corepressor CtBP to mediate repression whereas longrange repressors have been shown to interact with the Groucho corepressor (Barolo and Levine, 1997; Chen and Courey, 2000; Fisher and Caudy, 1998; Mannervik and Levine, 1999; Mannervik et al., 1999; Poortinga et al., 1998). One model for Groucho mediated long-range repression is through the recruitment of HDACs by Groucho resulting in the production of a large transcriptionally silent chromosomal domain. Just as the Sir repressosome generates a transcriptionally silent chromatin structure that is able to spread along the chromatin fiber, it has been proposed that the Groucho repression (Brantjes et al., silent chromosomal state and spreads to mediate long-range repression (Brantjes et al., 2001; Chen et al., 1999; Flores-Saaib and Courey, 2000). However, I find that the repression mediated by Hairy is transient, similar to the activity of the short-range repressors and is easily reversed as the embryo develops. Thus, if Hairy mediates transcriptional repression by creating a repressive heterochromatin domain that can spread over large distances, then the nature of this heterochromatin must be such that it can be easily modified back to a more open conformation.

Another similarity between short-range and long-range repressors is that like Groucho, a number of studies have suggested that CtBP may function, at least in part, by recruiting histone deacetylases (Criqui-Filipe et al., 1999; Sundqvist et al., 1998). Thus, if both long- and short-range corepressors function through histone deacetylation, why is Hairy is more potent with a longer range of action than Giant? There are a number of possibilities. First, as mentioned previously, long-range corepressors may have the ability to spread along the template recruiting histone deacetylases and/or other chromatin modifying activities to a large domain, but short-range repressors may lack the capacity to spread. Alternatively, the differences between long- and short-range corepressors could relate to the different properties of different histone deacetylases. Groucho has thus far only been found to bind class I histone deacetylases, whereas CtBP appears to bind both class I and class II histone deacetylases (Bertos et al., 2001). Perhaps the different repertoires of histone deacetylases recruited by different corepressors result in different histone acetylation patterns in the surrounding chromatin. The synthetic enhancer elements that I have defined in my study can thus serve as useful tools to compare the

biochemical activities of these two classes of transcriptional regulators in order to distinguish between the molecular mechanisms of short-range and long-range repressors.

# Next generation Biology: Predictive modeling of differential gene expression as a function of time and space.

Although computational methodologies have so far focused on the simple identification of cis-regulatory elements involved in gene regulation, a more challenging goal has been set to predict from the sequence of a cis-element gene expression patterns over time and space. While some contextual parameters such as number, relative affinities, spacing and arrangements of transcription factor binding sites can be tested by manipulating these parameters individually or in combination, in designing the experimental cis-element, in vivo gradients of transcription factors, such as the gap repressor gradient will provide the opportunity to test different concentrations of transacting factors on the output of a given cis-element. Crucial to the formulation of a quantitative predictive model for differential gene expression is the ability to quantitate regulation by the 'response characteristics', that is the level of gene expression as a function of the concentrations of transcription factors (Bower, 2001). We will require the development of more quantitative approaches to interpret very subtle differences in gene expression seen with small changes in the sequence of the cis-element. For example, in order to be able to build a robust predictive tool, quantitative methods to detect subtle changes in transcriptional readout caused by 10 bp changes in the spacing of transcription factors sites within a cis-element will have to be established.

The first computer experiments fitting analog gene regulation models to real expression data, were used to understand the network of gap genes expressed in bands (or domains) along the anterior-posterior axis of the very early (syncitial blastoderm) of the Drosophila melanogaster embryo (Reinitz and Sharp, 1995). Positional information along the A-P axis of the syncitial blastoderm is encoded in a succession of different ways during development. At first the main encoding is a roughly exponential gradient of Bicoid protein imposed by the mother fly, along with the maternal Hunchback expression. These provide gene regulation inputs to the gap gene: giant, knirps, krüppel, tailless, and hunchback. These each establish one or two broad domains of expression along the A-P axis. The gap genes then serve as network inputs to the pair-rule genes, including even-skipped, which establish narrow, precise stripes of expression along the A-P length of the embryo. The predictive model of Reinitz and colleagues was concerned with the establishment of the broad gap gene domains (excluding the extreme ends of the Anterior-Posterior axis) from maternally supplied initial conditions, by a gene regulation network in which all gap genes interact with all others and Bicoid provides input to, but does receive any input from the gap genes. Figure IV-2 shows the experimentally observed (Figure IV-2A) and model (Figure IV-2B) fitted curves for gap gene expression (Reinitz and Sharp, 1995). They are in qualitative agreement, which is the most that could be expected from the data (expression data and cis-regulatory structure-function) that was available at the time. The extra dip in Giant expression could not be predicted by the model, which can be interpreted as an indication of the role of circuit components not included in the model. Further experimental analyses of this regulatory cascade point to at least one additional level of complexity that needs to be included in such predictive



Figure IV-2: Data and Model for gap gene circuit.

Horizontal axes are nuclei along lateral from anterior to posterior. Vertical axes are relative concentrations. (A) Data estimated from immunofluorescence images of gap gene expression. (B) Output of a circuit model fit to expression data. (Figure adapted from Reinitz et al., 1995).
models of gene regulatory networks, which is a detailed understanding of the cisregulatory substructure and its functional significance as we have defined for the shortrange transcriptional repressors (Giant, Knirps and Krüppel) in Chapter 3.

High resolution in vivo and in vitro imaging technologies, combined with largescale gene expression assays, are providing vast amounts of quantitative data on gene regulatory events during development. With genome wide in situ hybridization analysis, together with global transcription factor binding analysis (Lee et al., 2002; Weinmann and Farnham, 2002; Weinmann et al., 2002) we now have the technology to track the expression of thousands of genes during the lifetime of a cell, and to trace the interactions of many of the products of these genes. Thus, by combining quantitative methods to measure the concentrations of transcription factors over space and time and the expression of a downstream target gene with extensive input (variations in the internal design of the cis-element) - output (target gene off or on) patterns we can build a predictive modeling tool that will allow us not only to identify novel cis- elements that are regulated by the same suite of transcription factors, in this case the short-range repressors, but also to predict its expression pattern. As predictions of these models can be subjected to experimental testing, erroneous predictions maybe overcome by updating the modeling tool with new empirically obtained corrective structure-function data. Achieving robust and quantitatively correct predictions is an integrated, iterative process, with models feeding experimental design, while experimental data in turn feed models. Thus, the predictive model can be 'trained' with input-output data of iteratively increasing resolution to achieve higher predictive reliability.

A simplified example of how exactly such a predictive tool can be used is as follows: We would like to identify putative targets of short-range transcriptional repressors in the Drosophila melanogaster genome and predict how these candidate genes will respond over space and time. Computational algorithms based on binding site sequence data from transcription factor binding site databases (for example, TRANSFAC), can be used to scan the genome for clusters of short-range repressor binding sites together with binding sites for other proteins that are known to work with them. Combining datasets from a number of large-scale analyses will then help us prefilter the set of putative regulatory elements obtained. Using gene expression data from microarray analysis we could ask whether the putative regulatory element is in the vicinity of a target gene that is known to be regulated by our factors or if they are near genes that are expressed at approximately the same time as the regulatory factors. Using data from genome-wide in situ hybridization analysis we could ask if the putative target genes are expressed in patterns consistent with the pattern of expression of short-range repressors. We could also probe whether the expression pattern of other factors for which binding sites occur in the identified regulatory elements correlates with that of our factor of interest both spatially and temporally and with that of the candidate target gene. Once the dataset has been scaled down to the most likely candidates, we can compare the internal organization of the cis-elements with the 'grammar' rules defined in this study to further eliminate those that do not seem likely to be regulated by our factor of interest on the basis of enhancer design.

A fully quantitative modeling of animal transcriptional control may appear unrealistic, but the progress made over the last 30 years has been nothing short of remarkable. The enormous and growing amount of detailed quantitative information available from large-scale analysis of the dynamics and localization of transcription factors, genome-wide analysis of transcription factor binding, improved computational and phylogenetic methods for identifying cis-regulatory element structure and function, and genome wide gene expression data may bring us much closer to our goal than currently seems possible. With a complete understanding of animal transcription systems, it should be possible to predict biological output directly from regulatory sequence and the physiological effects on an organism of mutating cis-regulatory sequence or misexpressing a transcription factor. Ideally, such an understanding of the cis-regulatory code and the ability to predict the consequences of changes in its sequence, will enable us to decipher the evolutionary dynamics of transcriptional regulation underlying morphological complexity and diversity of living organisms. Also, it should allow the design of novel heterologous promoters that express selected gene products in a specified group of cells at a controlled level. Such progress would be important both for human gene therapy and for the development of improved transgenic animals. Quantitative genome-wide analyses are already changing the way we think of transcriptional control and will continue to do so in the future.

169

#### REFERENCES

Arnosti, D. N., Barolo, S., Levine, M. and Small, S. (1996a). The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* 122, 205-14.

Arnosti, D. N., Gray, S., Barolo, S., Zhou, J. and Levine, M. (1996b). The gap protein knirps mediates both quenching and direct repression in the Drosophila embryo. *Embo J* 15, 3659-66.

Balasubramanian, P., Zhao, L. J. and Chinnadurai, G. (2003). Nicotinamide adenine dinucleotide stimulates oligomerization, interaction with adenovirus E1A and an intrinsic dehydrogenase activity of CtBP. *FEBS Lett* 537, 157-60.

Barolo, S. and Levine, M. (1997). hairy mediates dominant repression in the Drosophila embryo. *Embo J* 16, 2883-91.

**Barolo, S. and Posakony, J. W.** (2002). Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. *Genes Dev* 16, 1167-81.

Bergman, C. M. and Kreitman, M. (2001). Analysis of conserved noncoding DNA in Drosophila reveals similar constraints in intergenic and intronic sequences. *Genome Res* 11, 1335-45.

Berman, B. P., Nibu, Y., Pfeiffer, B. D., Tomancak, P., Celniker, S. E., Levine, M., Rubin, G. M. and Eisen, M. B. (2002). Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. *Proc Natl Acad Sci U S A* 99, 757-62.

Bertos, N. R., Wang, A. H. and Yang, X. J. (2001). Class II histone deacetylases: structure, function, and regulation. *Biochem Cell Biol* 79, 243-52.

Bouallaga, I., Massicard, S., Yaniv, M. and Thierry, F. (2000). An enhanceosome containing the Jun B/Fra-2 heterodimer and the HMG-I(Y) architectural protein controls HPV 18 transcription. *EMBO Rep* 1, 422-7.

**Bower, J. a. B., Hamid.** (2001). Computational Modeling of Genetic and Biochemical Networks: A Bradford Book, The MIT press.

Brantjes, H., Roose, J., van De Wetering, M. and Clevers, H. (2001). All Tcf HMG box transcription factors interact with Groucho-related co-repressors. *Nucleic Acids Res* 29, 1410-9.

Cai, H. N., Arnosti, D. N. and Levine, M. (1996). Long-range repression in the Drosophila embryo. *Proc Natl Acad Sci U S A* 93, 9309-14.

Carey, M., Lin, Y. S., Green, M. R. and Ptashne, M. (1990). A mechanism for synergistic activation of a mammalian gene by GAL4 derivatives. *Nature* 345, 361-4.

Chen, G. and Courey, A. J. (2000). Groucho/TLE family proteins and transcriptional repression. *Gene* 249, 1-16.

Chen, G., Fernandez, J., Mische, S. and Courey, A. J. (1999). A functional interaction between the histone deacetylase Rpd3 and the corepressor groucho in Drosophila development. *Genes Dev* 13, 2218-30.

Chi, T., Lieberman, P., Ellwood, K. and Carey, M. (1995). A general mechanism for transcriptional synergy by eukaryotic activators. *Nature* 377, 254-7.

Chinnadurai, G. (2002). CtBP, an unconventional transcriptional corepressor in development and oncogenesis. *Mol Cell* 9, 213-24.

Chinnadurai, G. (2003). CtBP family proteins: more than transcriptional corepressors. *Bioessays* 25, 9-12.

Criqui-Filipe, P., Ducret, C., Maira, S. M. and Wasylyk, B. (1999). Net, a negative Ras-switchable TCF, contains a second inhibition domain, the CID, that mediates repression through interactions with CtBP and de-acetylation. *Embo J* 18, 3392-403.

Deckert, J. and Struhl, K. (2002). Targeted recruitment of Rpd3 histone deacetylase represses transcription by inhibiting recruitment of Swi/Snf, SAGA, and TATA binding protein. *Mol Cell Biol* 22, 6458-70.

Doyle, H. J., Kraut, R. and Levine, M. (1989). Spatial regulation of zerknullt: a dorsalventral patterning gene in Drosophila. *Genes Dev* 3, 1518-33.

Ellwood, K. B., Yen, Y. M., Johnson, R. C. and Carey, M. (2000). Mechanism for specificity by HMG-1 in enhanceosome assembly. *Mol Cell Biol* 20, 4359-70.

Fink, R. C. and Scandalios, J. G. (2002). Molecular evolution and structure--function relationships of the superoxide dismutase gene families in angiosperms and their relationship to other eukaryotic and prokaryotic superoxide dismutases. *Arch Biochem Biophys* 399, 19-36.

Fisher, A. L. and Caudy, M. (1998). Groucho proteins: transcriptional corepressors for specific subsets of DNA-binding transcription factors in vertebrates and invertebrates. *Genes Dev* 12, 1931-40.

Flores-Saaib, R. D. and Courey, A. J. (2000). Analysis of Groucho-histone interactions suggests mechanistic similarities between Groucho- and Tup1-mediated repression. *Nucleic Acids Res* 28, 4189-96.

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.

Gray, S. and Levine, M. (1996). Transcriptional repression in development. Curr Opin Cell Biol 8, 358-64.

Gray, S., Szymanski, P. and Levine, M. (1994). Short-range repression permits multiple enhancers to function autonomously within a complex promoter. *Genes Dev* 8, 1829-38.

Guss, K. A., Nelson, C. E., Hudson, A., Kraus, M. E. and Carroll, S. B. (2001). Control of a genetic regulatory network by a selector gene. *Science* 292, 1164-7.

Hader, T., Wainwright, D., Shandala, T., Saint, R., Taubert, H., Bronner, G. and Jackle, H. (2000). Receptor tyrosine kinase signaling regulates different modes of Groucho-dependent control of Dorsal. *Curr Biol* 10, 51-4.

Halfon, M. S., Grad, Y., Church, G. M. and Michelson, A. M. (2002). Computationbased discovery of related transcriptional regulatory modules and motifs using an experimentally validated combinatorial model. *Genome Res* 12, 1019-28.

Halfon, M. S. and Michelson, A. M. (2002). Exploring genetic regulatory networks in metazoan development: methods and models. *Physiol Genomics* 10, 131-43.

Hanna-Rose, W., Licht, J. D. and Hansen, U. (1997). Two evolutionarily conserved repression domains in the Drosophila Kruppel protein differ in activator specificity. *Mol Cell Biol* 17, 4820-9.

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.

Hardison, R., Slightom, J. L., Gumucio, D. L., Goodman, M., Stojanovic, N. and Miller, W. (1997). Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene* 205, 73-94.

Hewitt, G. F., Strunk, B. S., Margulies, C., Priputin, T., Wang, X. D., Amey, R., Pabst, B. A., Kosman, D., Reinitz, J. and Arnosti, D. N. (1999). Transcriptional repression by the Drosophila giant protein: cis element positioning provides an alternative means of interpreting an effector gradient. *Development* 126, 1201-10.

Jiang, J., Rushlow, C. A., Zhou, Q., Small, S. and Levine, M. (1992). Individual dorsal morphogen binding sites mediate activation and repression in the Drosophila embryo. *Embo J* 11, 3147-54.

Kadosh, D. and Struhl, K. (1998a). Histone deacetylase activity of Rpd3 is important for transcriptional repression in vivo. *Genes Dev* 12, 797-805.

Kadosh, D. and Struhl, K. (1998b). Targeted recruitment of the Sin3-Rpd3 histone deacetylase complex generates a highly localized domain of repressed chromatin in vivo. *Mol Cell Biol* 18, 5121-7.

Keller, S. A., Mao, Y., Struffi, P., Margulies, C., Yurk, C. E., Anderson, A. R., Amey, R. L., Moore, S., Ebels, J. M., Foley, K. et al. (2000). dCtBP-dependent and - independent repression activities of the Drosophila Knirps protein. *Mol Cell Biol* 20, 7247-58.

Kim, T. K. and Maniatis, T. (1997). The mechanism of transcriptional synergy of an in vitro assembled interferon-beta enhanceosome. *Mol Cell* 1, 119-29.

Kreitman, M. (1996). The neutral theory is dead. Long live the neutral theory. *Bioessays* 18, 678-83; discussion 683.

Kulkarni, M. M. a. A., D. N. (2003). Information Display by Transcriptional enhancer, (ed.

Kumar, V., Carlson, J. E., Ohgi, K. A., Edwards, T. A., Rose, D. W., Escalante, C. R., Rosenfeld, M. G. and Aggarwal, A. K. (2002). Transcription corepressor CtBP is an NAD(+)-regulated dehydrogenase. *Mol Cell* 10, 857-69.

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I. et al. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* 298, 799-804.

Licht, J. D., Ro, M., English, M. A., Grossel, M. and Hansen, U. (1993). Selective repression of transcriptional activators at a distance by the Drosophila Kruppel protein. *Proc Natl Acad Sci U S A* 90, 11361-5.

Loots, G. G., Locksley, R. M., Blankespoor, C. M., Wang, Z. E., Miller, W., Rubin, E. M. and Frazer, K. A. (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* 288, 136-40.

Ludwig, M. Z., Bergman, C., Patel, N. H. and Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403, 564-7.

Ludwig, M. Z. and Kreitman, M. (1995). Evolutionary dynamics of the enhancer region of even-skipped in Drosophila. *Mol Biol Evol* 12, 1002-11.

Ludwig, M. Z., Patel, N. H. and Kreitman, M. (1998). Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change. *Development* 125, 949-58.

Lunyak, V. V., Burgess, R., Prefontaine, G. G., Nelson, C., Sze, S. H., Chenoweth, J., Schwartz, P., Pevzner, P. A., Glass, C., Mandel, G. et al. (2002). Corepressordependent silencing of chromosomal regions encoding neuronal genes. *Science* 298, 1747-52.

Maier, D., Preiss, A. and Powell, J. R. (1990). Regulation of the segmentation gene fushi tarazu has been functionally conserved in Drosophila. *Embo J* 9, 3957-66.

Mannervik, M. and Levine, M. (1999). The Rpd3 histone deacetylase is required for segmentation of the Drosophila embryo. *Proc Natl Acad Sci U S A* **96**, 6797-801.

Mannervik, M., Nibu, Y., Zhang, H. and Levine, M. (1999). Transcriptional coregulators in development. *Science* 284, 606-9.

Markstein, M. and Levine, M. (2002). Decoding cis-regulatory DNAs in the Drosophila genome. Curr Opin Genet Dev 12, 601-6.

Markstein, M., Markstein, P., Markstein, V. and Levine, M. S. (2002). Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the Drosophila embryo. *Proc Natl Acad Sci U S A* 99, 763-8.

Marmorstein, R. (2002). Dehydrogenases, NAD, and transcription--what's the connection? *Structure (Camb)* 10, 1465-6.

Munshi, N., Agalioti, T., Lomvardas, S., Merika, M., Chen, G. and Thanos, D. (2001). Coordination of a transcriptional switch by HMGI(Y) acetylation. *Science* 293, 1133-6.

Nibu, Y., Senger, K. and Levine, M. (2003). CtBP-independent repression in the Drosophila embryo. *Mol Cell Biol* 23, 3990-9.

Nibu, Y., Zhang, H., Bajor, E., Barolo, S., Small, S. and Levine, M. (1998a). dCtBP mediates transcriptional repression by Knirps, Kruppel and Snail in the Drosophila embryo. *Embo J* 17, 7009-20.

Nibu, Y., Zhang, H. and Levine, M. (1998b). Interaction of short-range repressors with Drosophila CtBP in the embryo. *Science* 280, 101-4.

**Poortinga, G., Watanabe, M. and Parkhurst, S. M.** (1998). Drosophila CtBP: a Hairyinteracting protein required for embryonic segmentation and hairy-mediated transcriptional repression. *Embo J* 17, 2067-78.

Postigo, A. A. and Dean, D. C. (1999). Independent repressor domains in ZEB regulate muscle and T-cell differentiation. *Mol Cell Biol* 19, 7961-71.

**Read, M. A., Whitley, M. Z., Williams, A. J. and Collins, T.** (1994). NF-kappa B and I kappa B alpha: an inducible regulatory system in endothelial activation. *J Exp Med* **179**, 503-12.

**Rebeiz, M., Reeves, N. L. and Posakony, J. W.** (2002). SCORE: a computational approach to the identification of cis-regulatory modules and target genes in whole-genome sequence data. Site clustering over random expectation. *Proc Natl Acad Sci U S A* **99**, 9888-93.

Reinitz, J. and Sharp, D. H. (1995). Mechanism of eve stripe formation. *Mech Dev* 49, 133-58.

Roose, J. and Clevers, H. (1999). TCF transcription factors: molecular switches in carcinogenesis. *Biochim Biophys Acta* 1424, M23-37.

Rundlett, S. E., Carmen, A. A., Suka, N., Turner, B. M. and Grunstein, M. (1998). Transcriptional repression by UME6 involves deacetylation of lysine 5 of histone H4 by RPD3. *Nature* **392**, 831-5.

Shi, Y., Sawada, J., Sui, G., Affar el, B., Whetstine, J. R., Lan, F., Ogawa, H., Luke, M. P. and Nakatani, Y. (2003). Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* 422, 735-8.

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

Struhl, K. (2001). Gene regulation. A paradigm for precision. Science 293, 1054-5.

Strunk, B., Struffi, P., Wright, K., Pabst, B., Thomas, J., Qin, L. and Arnosti, D. N. (2001). Role of CtBP in transcriptional repression by the Drosophila giant protein. *Dev Biol* 239, 229-40.

Subramanian, T. and Chinnadurai, G. (2003). Association of class I histone deacetylases with transcriptional corepressor CtBP. *FEBS Lett* 540, 255-8.

Sundqvist, A., Sollerbrant, K. and Svensson, C. (1998). The carboxy-terminal region of adenovirus E1A activates transcription through targeting of a C-terminal binding protein-histone deacetylase complex. *FEBS Lett* **429**, 183-8.

Szymanski, P. and Levine, M. (1995). Multiple modes of dorsal-bHLH transcriptional synergy in the Drosophila embryo. *Embo J* 14, 2229-38.

Thanos, D. and Maniatis, T. (1995). Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell* 83, 1091-100.

Valentine, S. A., Chen, G., Shandala, T., Fernandez, J., Mische, S., Saint, R. and Courey, A. J. (1998). Dorsal-mediated repression requires the formation of a multiprotein repression complex at the ventral silencer. *Mol Cell Biol* 18, 6584-94.

Weinmann, A. S. and Farnham, P. J. (2002). Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods* 26, 37-47.

Weinmann, A. S., Yan, P. S., Oberley, M. J., Huang, T. H. and Farnham, P. J. (2002). Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev* 16, 235-44.

Xu, C., Kauffmann, R. C., Zhang, J., Kladny, S. and Carthew, R. W. (2000). Overlapping activators and repressors delimit transcriptional response to receptor tyrosine kinase signals in the Drosophila eye. *Cell* 103, 87-97. Yuh, C. H., Bolouri, H. and Davidson, E. H. (1998). Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science* 279, 1896-902.

Zhang, H. and Levine, M. (1999). Groucho and dCtBP mediate separate pathways of transcriptional repression in the Drosophila embryo. *Proc Natl Acad Sci U S A* 96, 535-40.

#### APPENDIX A

# Composite element representing two different information states exhibits orientation-independent activity characteristic of enhancers.

As described before in Chapter 2 (Figure II-1F) a composite enhancer element, containing binding sites for the endogenous short-range repressor Giant, endogenous activators Twist and Dorsal, and chimeric Gal4 activators, behaves like an 'information display', simultaneously representing contrasting information of activation and repression at the same time and in the same nucleus. This enhancer element drives lacZ reporter gene expression from the basal hsp 70 promoter in a complex pattern such that in the nuclei where the activators and the repressor Giant are co-expressed, transcription is driven by the cluster of Gal4 activators within the compact regulatory element, while at the same time the Dorsal and Twist activators within the same element are being actively repressed by Giant. Thus, this compact regulatory element, has subelements that simultaneously represent two types of information states, activation and repression, unlike the binary switch activity observed for many transcriptional enhancers, where usually a single signal to activate or repress is present. Since this dual activity is unlike that of 'real' enhancers, we tested to see if the element we had created possessed any of the qualities that are attributed to natural enhancers, namely the ability to function in a distance- and orientation- independent manner (Banerji et al., 1981), by placing it in either orientation between the divergently transcribed white (at -265 bp) and lacZ genes (at -130 bp). We showed that although this element did not function in the binary on or off mode, its activity was independent of both orientation and distance with respect to the start of transcription (Figure II-2). The distance- and orientation- independent activity of the element was tested on the *transposase lacZ* gene, which is different from the *hsp* 70 basal promoter used in Figure II-1. In order to confirm that these activities were an inherent property of the element and not specific for the promoter being tested, I also tested the orientation-independent activity of the element on the *hsp* 70 *lacZ* reporter gene as shown in Figure A-1A, B. A similar pattern of repression and activation of the *hsp* 70-*lacZ* gene is seen when the element is tested in the opposite orientation (Figure A-1A, B).



# Figure A-1: Compact regulatory element displays enhancer-like property of orientation independence.

(A, B) The regulatory element shown in Figure II-1F was inserted in the opposite orientation into the UAS-lacZ vector containing the hsp 70 basal promoter. In the presence of Dorsal, Twist, and Gal4 activators, a composite pattern of gene regulation is seen as in Figure II-1F with inhibition of Dorsal/Twist and activation by Gal4.

Patterns of gene expression were visualized in 2-4 hour embryos by *in situ* hybridization with digU labeled antisense lacZ probes. Embryos are oriented anterior to left; ventral view (A) and lateral view (B) are shown.

#### **MATERIALS AND METHODS**

#### 1. Plasmid construction

#### GAL4 (aa1-93) - GAL4 AD (aa753-881)

A KpnI-XbaI fragment from pSCTEV GAL4 (1-93)- GAL4 (Seipel et al., 1992) containing the reading frame for the yeast GAL4 activation domain (Gal4 AD) from amino acid residues 753-881, was cloned into KpnI-XbaI cut pTwiggy (Arnosti et al., 1996b) vector, which contains the *twist* enhancer (2xPEe-Et) element, *twist* basal promoter and the GAL4 DNA-binding domain from residues 1-93.

#### **Reporter gene**

The vector M2g5u-lacZ (Chapter 2) was cut with *Hind*III-*Sph*I enzymes to remove the five Gal4 sites. These reporters were further modified by introducing oligos containing two **Dorsal** (dl) and two **Twist** (twi) binding sites (DA481/482: 5'AGC TTG AGG GAT TTT CCC AAA TCG AGG GAA AAC CCA ACT CGC ATA TGT TGA GCA TAT GGC ATG 3') (Szymanski and Levine, 1995). Annealed oligos containing a neutral 55 bp spacer DNA (DA65/66: 5'TCC ATG ATA AAC GCG TGC TAG ACT ATT GCA GGT ACT GAT CGA ATG CCT CTG CAT G 3') was placed downstream of the Dorsal and Twist sites and upstream of the *hsp 70* basal promoter (Hewitt et al., 1999). Five Gal4 binding sites were PCR amplified from the *UAS-lacZ* vector (Brand and Perrimon, 1993) using oligos DA487/488 (5' AAG GAA AAA AGC GGC CGC GCG CTC GCT AGA GTC 3') and were introduced upstream of the two Giant binding sites at the *Not*I site resulting in the vector **M5u2g2dl.twi-55-lacZ**.

# 2. P-element transformation, crosses to reporter genes, and whole-mount in situ hybridization of embryos.

P-element transformation vectors were introduced into the *Drosophila* germline by injection of  $yw^{67}$  embryos as described (Small et al., 1992). Embryos were collected from a cross between a reporter line and a line expressing the GAL4 activator in the ventral regions of the embryo. The embryos were fixed and stained using digoxigenin-UTP labeled antisense RNA probes to either *lacZ* or *w* as described (Small et al., 1992).

#### REFERENCES

Arnosti, D. N., Gray, S., Barolo, S., Zhou, J. and Levine, M. (1996b). The gap protein knirps mediates both quenching and direct repression in the Drosophila embryo. *Embo J* 15, 3659-66.

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

Brand, A. H. and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118, 401-15.

Hewitt, G. F., Strunk, B. S., Margulies, C., Priputin, T., Wang, X. D., Amey, R., Pabst, B. A., Kosman, D., Reinitz, J. and Arnosti, D. N. (1999). Transcriptional repression by the Drosophila giant protein: cis element positioning provides an alternative means of interpreting an effector gradient. *Development* 126, 1201-10

Seipel, K., Georgiev, O. and Schaffner, W. (1992). Different activation domains stimulate transcription from remote ('enhancer') and proximal ('promoter') positions. *Embo J* 11, 4961-8.

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

Szymanski, P. and Levine, M. (1995). Multiple modes of dorsal-bHLH transcriptional synergy in the Drosophila embryo. *Embo J* 14, 2229-38.

#### **APPENDIX B**

#### CtBP-dependent and CtBP-independent activities contribute quantitatively to Knirps repressor function<sup>1</sup>.

In this study I have defined the contextual elements of *cis*-regulatory enhancers that allow short-range repressors to mediate repression effectively. However, we do not yet understand the physical and biochemical changes in promoter complexes that accompany short-range repression. Chromatin Immunoprecipitation (ChIP) assays in *Drosophila* embryos using the simple genetic switch elements defined in this study will facilitate our understanding of the molecular mechanisms of repression employed at developmentally regulated genes. These experiments can be designed to directly test mechanistic questions about the nature of short-range transcriptional repression, specifically, whether repression blocks activator access to the DNA alters chromatin structure, or whether changes in the composition or presence of the basal machinery are observed.

Chromatin Immunoprecitation (ChIP) has been extensively used to study transcriptional regulation in cell culture, where a homogeneous population of cells is available for analysis. In the *Drosophila* embryo however, the endogenous targets of the short-range repressors such as the pair-rule gene *even-skipped*, are expressed in discrete domains in the early blastoderm embryo (Frasch et al., 1987; Harding et al., 1986; Macdonald and Struhl, 1986) and therefore, are not useful targets for ChIP experiments as such assays would result in high background noise. Using transgenes, such as those

<sup>&</sup>lt;sup>1</sup>This data is included in the following manuscript submitted to *Development*: Paolo Struffi, Maria Corado, Meghana M. Kulkarni and David N. Arnosti. Quantitative Contributions of CtBP-dependent and – independent repression activities of Knirps.

described in Chapter 2 and 3 of this study would allow one to regulate the expression of the reporter transgene throughout the embryo using inducible and ubiquitous expression of activators and the repressors.

The first step towards performing Drosophila embryo ChIP assays is to obtain a homogeneous population of nuclei/cells. This involves engineering embryos to carry a transgene that can be activated in all (or almost all) nuclei and then repressed in all (or almost all) nuclei at a given time. In the embryo however, the repressors are expressed in spatially localized domains (Gergen and Wieschaus, 1986; Hoch et al., 1992; Petschek et al., 1987; Rothe et al., 1989) and are therefore, typically active on a given gene only in a small fraction of the nuclei. Thus a ChIP analysis of the whole embryo will represent a mixture of promoters that are not repressed and those that are actively repressed. To overcome this difficulty and to increase the signal to noise ratio, I decided to utilize transgenic flies that express different versions (full length and N-terminal dCtBPindependent repression domain) of recombinant Knirps proteins, ubiquitously and under a heat-shock inducible promoter (constructed and generated by Paolo Struffi in the lab). As a first step, it was important to determine if the recombinant proteins were able to function as repressors in vivo on an endogenous Knirps target. I looked at the expression pattern of one of Knirps targets, the pair-rule gene even-skipped (eve) in transgenic embryos before and after heat shock. Knirps is required for correct regulation of the eve stripe 3/7 and 4/6 enhancers, as demonstrated by the expression patterns of lacZ reporter genes in kni mutant embryos (Fujioka et al., 1999; Small et al., 1996). It has been previously demonstrated that (Keller et al., 2000) the posterior portion of the eve stripe 3 pattern is not derepressed in a CtBP mutant, consistent with the CtBP-independent

activity of Knirps on this enhancer. In contrast, Knirps repression of *eve* stripe 4/6 is compromised in a *CtBP* mutant background, indicating that the CtBP-independent repression activity of Knirps is insufficient to regulate this enhancer (Struffi et al., submitted). Therefore, depending on which portion of the *eve* gene is bound by the Knirps protein, its repression activity is either dependent or independent of the CtBP cofactor.

2-4 hour old transgenic embryos carrying transgenes for either the full-length recombinant Knirps protein or the dCtBP-independent repression domain were heatshocked at 38 ° C for 30 minutes, recovered for 30 minutes in a water bath at room temperature before fixing with formaldehyde. *In situ* hybridization was performed on both heat-shocked (Figure B-1B, D) and non heat-shocked transgenic (Figure B-1A, C) embryos using antisense mRNA probe to *eve*. The N-terminal region of Knirps (kni 1-330) is a weak repressor compared to the full-length protein (kni 1-429), but is able to repress the previously known dCtBP-dependent target, the *eve* stripe 4+6 enhancer, when over-expressed (Figure B-1D). This suggests that increasing the dose of the repressor may be sufficient to overcome the requirement for the dCtBP cofactor and that multiple repression activities within a single protein represent quantitative effects on gene expression. Upon heat-shock both the full-length (Figure B-1B) and the dCtBP-independent (Figure B-1D) versions of the recombinant Knirps protein were able to abolish the expression of *eve* stripes 3/7 and the *eve* stripes 4/6.



### Figure B-1: Pattern of endogenous *eve* expression in embryos expressing full length Knirps 1-429 (B) and CtBP-independent region of Knirps 1-330 (D).

Structure of proteins expressed from *hsp70* promoter: 1-429, full-length Knirps protein; 1-330, CtBP-independent Knirps repression domain are depicted above the corresponding embryos. Endogenous *eve* pattern is visualized by *in situ* hybridization using digoxigenin labeled eve mRNA probe before and after heat-shock of embryos carrying transgenes that express either the full-length recombinant Knirps protein or the CtBP-independent Knirps repression domain. Embryos are oriented anterior to right and lateral views are shown.

(A, C) Endogenous *eve* expression in non heat-shocked embryos carrying the transgene for full-length recombinant Knirps protein (A) or the CtBP-independent Knirps repression domain (C).

(B) Endogenous *eve* expression in heat-shocked embryos expressing the full-length recombinant Knirps protein. Overexpression of full-length Knirps abolishes the expression of all *eve* stripes except for stripe 5.

(D) Endogenous *eve* expression in heat-shocked embryos expressing the CtBPindependent Knirps repression domain. Overexpression of CtBP-independent Knirps repression domain abolishes the expression of *eve* stripes 3, 4, 6, and 7. The experiments described above also demonstrated that recombinant Knirps protein expressed ubiquitously throughout the embryo under heat shock conditions was functional as a repressor and could be used in future ChIP assays.

#### **MATERIALS AND METHODS**

#### Transgenic flies expressing recombinant Knirps proteins

Transgenic flies carrying heat-shock inducible versions of the Knirps proteins, hsKni 1-429 (full-length protein) and hsKni 1-330 (dCtBP-independent domain) with a hexahistidine tag at the N-terminus and a double FLAG tag at the C-terminus were obtained from Paolo Struffi in the lab. These proteins can be expressed throughout the blastoderm embryo under conditions of heat-shock. Construction of these constructs is described in Struffi et al., (submitted).

#### Heat-shock treatment

To induce expression of recombinant Knirps proteins, 2-4 hour old embryos collected on apple-juice plates at room temperature (22-23°C) were incubated for 30 minutes at 38°C in a 10-liter water bath to ensure rapid and even heating. After induction, embryos were allowed to recover in a water bath at room temperature for 30 minutes prior to fixation.

#### In situ hybridization

In situ hybridizations were performed using digoxigenin-UTP-labeled antisense RNA probes to eve on both heat-shocked and non-heat-shocked transgenic embryos (Small et al., 1992).

#### REFERENCES

Frasch, M., Hoey, T., Rushlow, C., Doyle, H. and Levine, M. (1987). Characterization and localization of the even-skipped protein of Drosophila. *Embo J* 6, 749-59.

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

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

Hoch, M., Gerwin, N., Taubert, H. and Jackle, H. (1992). Competition for overlapping sites in the regulatory region of the Drosophila gene Kruppel. *Science* 256, 94-7.

Keller, S. A., Mao, Y., Struffi, P., Margulies, C., Yurk, C. E., Anderson, A. R., Amey, R. L., Moore, S., Ebels, J. M., Foley, K. et al. (2000). dCtBP-dependent and - independent repression activities of the Drosophila Knirps protein. *Mol Cell Biol* 20, 7247-58.

Macdonald, P. M. and Struhl, G. (1986). A molecular gradient in early Drosophila embryos and its role in specifying the body pattern. *Nature* 324, 537-45.

Petschek, J. P., Perrimon, N. and Mahowald, A. P. (1987). Region-specific defects in l(1)giant embryos of Drosophila melanogaster. *Dev Biol* 119, 175-89.

Rothe, M., Nauber, U. and Jackle, H. (1989). Three hormone receptor-like Drosophila genes encode an identical DNA-binding finger. *Embo J* 8, 3087-94.

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.

#### **APPENDIX C**

#### **Relative potencies of short-range repressor function**

The analyses of short-range transcriptional repressors, Giant, Knirps and Krüppel, have so far indicated that they function in very similar ways. These proteins are capable of repressing the activities of enhancer elements when bound within approximately 100 bp of key activator sites, or of a basal promoter element when cognate sites are introduced close to the start of transcription (Arnosti et al., 1996; Gray and Levine, 1996; Gray et al., 1994; Hewitt et al., 1999; Keller et al., 2000). Another common property of short-range transcriptional repressors is their interaction with the evolutionarily conserved corepressor CtBP (C-terminal Binding Protein) (Nibu et al., 1998a; Nibu et al., 1998b) Drosophila short-range repressors also possess CtBP-independent repression activities and the requirement for dCtBP can vary on a gene-to-gene as well as on an enhancer-to-enhancer basis. We have successfully shown that increasing the dose of the Knirps repressor may be sufficient to overcome the requirement for the dCtBP cofactor and that multiple repression activities within a single protein represent quantitative effects on gene expression (Paolo Struffi, Maria Corado, Meghana M. Kulkarni and David N. Arnosti. Quantitative Contributions of CtBP-dependent and -independent repression activities of Knirps. Manuscript submitted).

Using synthetic enhancer elements where the identity, stoichiometry, and the exact arrangement of activator and repressor binding site are well-defined, we demonstrated that the previously held simple notion that short-range repressors block the activity of all protein complexes within a 100 bp is incorrect. The manipulation of these

190

composite elements in terms of the number of activator and repressor binding sites, relative affinities, spacing and distribution of these binding sites further allowed us to define the contextual parameters that dictate repression effectiveness.

The contextual dependencies of repression described in Chapter 3 were developed for the Giant repressor. To determine if similar rules applied to other types of repressors, we carried out parallel evaluations of the short-range repressors Giant, Knirps, and Krüppel. To test quantitative similarities or differences between these factors we created reporters that would compare repressor activity on genes that represented permissive or non-permissive contexts for the Giant protein. The short-range repressor Knirps was unable to inhibit *lacZ* expression driven by the Gal4 activator from five high affinity Gal4 sites indicating a similar limitation of repression on even proximally bound activators as Giant (Figure C-1E and F). The Krüppel protein was also unable to repress the activity of five high affinity Gal4 sites (Figure C-1G), although many precellular embryos showed a narrowing in the lacZ expression pattern (Figure C-1H; arrow) in the central regions where Krüppel is expressed. Thus, in general, the short-range repressor Krüppel appears to be a more potent transcriptional repressor than Giant or Knirps. The Giant (Figure C-1A; arrows) and Krüppel (Figure C-1D; arrow) factors exhibited repression activity in the corresponding regions of the embryo when tested against three Gal4 sites. The Knirps repressor was also active in this context although in general the levels of repression appeared to be lower (Figure C-1B and C; arrows).

The results described above indicate that although Giant, Knirps, and Krüppel function similarly in most respects, they may have different repression potencies. This

difference may reflect subtle differences in the mechanism of repression employed or may likely represent differences in the levels of the protein or DNA binding specificity. Whatever the case, these observations are important with ramifications not only in the efficient design of reporters for ChIP assays but also for computational modeling of gene regulatory networks involving the short-range repressors.

#### **MATERIALS AND METHODS (SEE CHAPTER 3)**



Figure C-1: Relative potencies of the short-range repressors Giant, Knirps and Krüppel

A schematic of the reporters used is shown below. The reporters contain two binding sites for either one of the short-range repressors Giant, Knirps or Kruppel and five/three high affinity Gal4 binding sites.

All three of the short-range repressors Giant (A), Knirps (B, C), and Kruppel (D) are able to repress the activity of the Gal4-activator in the context of three Gal4 binding sites (arrows). However, the level of repression mediated by the Knirps protein was in general lower.

All the short-range repressors Giant (E), Knirps (F), and Kruppel (G, H) are unable to repress in the context of five high affinity Gal4 sites. However, in the case of the Krüppel protein, we observe a narrowing in the lacZ expression pattern in the central regions of the embryo where the repressor is present (H).

Embryos are oriented with anterior to left; lateral views are shown. LacZ expression was visualized in 2-4 hr old embryos using *in situ* hybridization with digU labeled antisense mRNA probes to LacZ.

# 

#### REFERENCES

Arnosti, D. N., Gray, S., Barolo, S., Zhou, J. and Levine, M. (1996). The gap protein knirps mediates both quenching and direct repression in the Drosophila embryo. *Embo J* 15, 3659-66.

Gray, S. and Levine, M. (1996). Short-range transcriptional repressors mediate both quenching and direct repression within complex loci in Drosophila. *Genes Dev* 10, 700-10.

Gray, S., Szymanski, P. and Levine, M. (1994). Short-range repression permits multiple enhancers to function autonomously within a complex promoter. *Genes Dev* 8, 1829-38.

Hewitt, G. F., Strunk, B. S., Margulies, C., Priputin, T., Wang, X. D., Amey, R., Pabst, B. A., Kosman, D., Reinitz, J. and Arnosti, D. N. (1999). Transcriptional repression by the Drosophila giant protein: cis element positioning provides an alternative means of interpreting an effector gradient. *Development* 126, 1201-10.

Keller, S. A., Mao, Y., Struffi, P., Margulies, C., Yurk, C. E., Anderson, A. R., Amey, R. L., Moore, S., Ebels, J. M., Foley, K. et al. (2000). dCtBP-dependent and - independent repression activities of the Drosophila Knirps protein. *Mol Cell Biol* 20, 7247-58.

Nibu, Y., Zhang, H., Bajor, E., Barolo, S., Small, S. and Levine, M. (1998a). dCtBP mediates transcriptional repression by Knirps, Kruppel and Snail in the Drosophila embryo. *Embo J* 17, 7009-20.

Nibu, Y., Zhang, H. and Levine, M. (1998b). Interaction of short-range repressors with Drosophila CtBP in the embryo. *Science* 280, 101-4.

#### **APPENDIX D**

#### Analysis of other Drosophila transcriptional repressors

Studies on transcriptional repression in the Drosophila embryo indicate that there may be two basic forms of repression: Long-range and Short-range repression. Longrange repressors, for example typified by the Drosophila Hairy protein, function over distances of at least 500 bp to silence the transcriptional complex or to inhibit upstream activators that are bound to promoter-proximal regions. In principle, long-range repressors can function in a dominant fashion to block multiple enhancers in a complex modular promoter (Barolo and Levine, 1997; Gray and Levine, 1996). Short-range repressors present in the early Drosophila embryo include the products of the gap genes snail (sna), krüppel (Kr), giant (gt), and knirps (kni). These proteins are capable of repressing the activities of enhancer elements when bound within approximately 100 bp of key activator sites, or of a basal promoter element when cognate sites are introduced close to the start of transcription (Arnosti et al., 1996b; Gray and Levine, 1996; Gray et al., 1994; Hewitt et al., 1999; Keller et al., 2000). The mechanisms by which short-range and long-range Drosophila repressors inhibit transcription are poorly understood, although one model of repression in the embryo suggests that the short-range/long-range distinction results from the recruitment of distinct cofactors (Nibu et al., 1998a; Nibu et al., 1998b; Zhang and Levine, 1999). Short-range repressors recruit the corepressor CtBP to mediate repression whereas long-range repressors have been shown to interact with the Groucho corepressor (Barolo and Levine, 1997; Fisher and Caudy, 1998; Mannervik et al., 1999; Poortinga et al., 1998). An interesting observation is that both the short-range corepressor, CtBP, and the long-range corepressor, Groucho, have been shown to interact

with and recruit Histone deacetylase complexes (Flores-Saaib and Courey, 2000; Nibu et al., 2001; Sundqvist et al., 1998).

In addition to having a longer range of action, we find that the long-range repressor Hairy is more potent than Giant even on a local scale and can block the activity of a greater number of activators as compared to Giant. However, contrary to previous studies that suggest that long-range repressors function by spreading along the chromatin to silence a large chromosomal locus, we find that the repression mediated by Hairy is transient similar to the activity of the short-range repressors and is easily reversed as the embryo develops.

We decided to compare the activities of other *Drosophila* transcriptional repressors to determine if we could detect similar such differences and similarities in their repression function. Since several of these transcriptional repressors are not well characterized, we first decided to test the activity of different repressors on genes that represented permissive contexts for the Giant protein. We constructed different *hsp* 70-*lacZ* reporters containing two binding sites for the repressor, three Gal4 sites and a spacer similar to the M2g3u2x-lacZ construct used in Chapter 3.

**Snail (sna):** Snail is a zinc finger protein and is a key mesoderm determinant in the *Drosophila* embryo. It is activated in the ventral regions of the embryo by the transcription factor Dorsal. Snail has been characterized to be a short-range transcriptional repressor like Giant, Knirps and Krüppel and has been shown to inhibit activators within enhancers or the basal promoter itself over distances of a 100 bp (Gray and Levine, 1996; Gray et al., 1994). The Snail repression domain contains both a

196

conserved copy of the P-DLS-k motif, as well as the slightly divergent sequence P-DLS-R. Like other short-range repressors Snail, interacts with and requires the CtBP corepressor to mediate repression.

However, we find that in a context where the other short-range repressors are able to mediate repression, Snail is unable to block the activity of three high affinity Gal4 sites (Figure D-1A, B). We consider several possibilities for this lack of Snail repressor function: First, the binding sites do not bind the protein efficiently. The binding sites used here were previously used in the context of the *rhomboid NEE* enhancer (Gray and Levine, 1996; Gray et al., 1994; Nibu et al., 1998a; Nibu et al., 1998b), which is an endogenous target of Snail. In this context Snail was shown to mediate repression of the linked *lacZ* gene. It is possible that the *rhomboid NEE* enhancer has cryptic Snail binding sites that cooperate with the synthetic Snail sites used in this study to mediate repression by Snail. Second, Snail protein levels may be low compared to that of Giant, Knirps and Krüppel and therefore two Snail binding sites may be insufficient. Third, Snail is a weak transcriptional repressor and the contextual 'rules' defined for the short-range repressors may not apply to the Snail repressor.

Engrailed, Runt and Even-skipped: Runt is a potent repressor of the pair-rule gene *even-skipped* (Jimenez et al., 1996; Manoukian and Krause, 1993) and the segmentation gene *engrailed* and is expressed in seven transverse stripes in the cellular blastoderm stage embryo (Wheeler et al., 2002). Runt has also been shown to activate the expression of the *sex lethal* gene in Drosophila (Kramer et al., 1999). Runt has been

197



# Figure D-1: The activity of the Gal4 activators from three high affinity sites is not repressed by Snail.

Schematic of the reporter gene is shown below. The reporter contains two Snail binding sites (sna), three high affinity Gal sites (colored circles), a neutral spacer (open circles), *hsp 70* basal promoter driving *lacZ* expression.

(A) Snail is unable to repress the Gal4 activators in early blastoderm embryo.

(B) Snail is unable to repress the Gal4 activators even at later stages of embryogenesis.

LacZ expression is activated throughout the embryo by ubiquitous expression of the full length Gal4 activator protein under the control of the actin5C enhancer. Patterns of gene expression were visualized in 2-4 hour embryos by *in situ* hybridization with digU labeled antisense *lacZ* probes. Embryos are oriented anterior to left; lateral view (A) and field shot of embryos at various stages in development (B) are shown. shown to interact physically and genetically with the corepressor Groucho (Aronson et al., 1997). Interestingly, only a subset of genes, for example *even-skipped*, repressed by Runt require the recruitment of Groucho for their repression (Aronson et al., 1997; Jimenez et al., 1996).

*engrailed* encodes for a homeodomain protein and is expressed in fourteen stripe (each one cell wide) along the anterior-posterior axis and plays a key role during segmentation in the Drosophila embryo. Engrailed has been characterized as a long-range transcriptional repressor and has been shown to interact with and recruit the corepressor Groucho to mediate repression of its target genes such as *even-skipped* (Jimenez et al., 1997). Engrailed like Runt has also been shown to have Groucho-independent repression activities (Tolkunova et al., 1998).

even-skipped encodes a homeodomain transcription factor that is required for the expression of both odd- and even-numbered Engrailed stripes, which are activated by distinct mechanisms (DiNardo and O'Farrell, 1987; Howard and Ingham, 1986). Previous data suggests that the role of Eve in the activation of *engrailed* might be, at least in part, indirect. Early Eve stripes repress *paired* at high concentrations, and *sloppy-paired*, a repressor of *engrailed* (Cadigan et al., 1994a; Cadigan et al., 1994b; Grossniklaus et al., 1992) at low concentrations (Fujioka et al., 1995). Eve interacts with and recruits the Groucho corepressor to mediate repression of both *paired* and *sloppy-paired*. Eve also represses another repressor of engrailed, *odd-skipped* in a Groucho-independent manner (Fujioka et al., 1995; Fujioka et al., 2002; Manoukian and Krause, 1992).

Thus, all three proteins Runt, Engrailed and Even-skipped mediate repression via Grouch-dependent and -independent pathways like the CtBP-dependent and - independent activities of the short-range repressors. However, when tested against three high affinity Gal4 sites, reporters carrying binding sites for these repressors did not show corresponding repression of *lacZ* expression (Figure D-2A, B, C, D, E, F, G, and H). Again, it is possible that the failure to mediate repression because the binding sites used do not bind the protein efficiently or because protein levels are insufficient to mediate repression in this context.

The experiments described above tested the activity of repressors that recruit the CtBP corepressor or repressors that recruit the Groucho corepressor to mediate repression. In *Drosophila* there are at least two known transcription factors (Brinker and Suppressor of Hairless) that interact with and require both CtBP and Groucho to mediate repression.

**Brinker (Brk):** Brinker is a transcription repressor and is expressed in ventrolateral regions of the embryo abutting the dorsal *decapentaplegic* (dpp) expression domain. Brinker is itself one of the downstream targets of *dpp* signaling and in turn represses some Dpp-responsive genes (Campbell and Tomlinson, 1999; Jazwinska et al., 1999a; Jazwinska et al., 1999b; Minami et al., 1999). It has been demonstrated that Brinker harbors a functional and transferable repression domain, through which it recruits the corepressors Groucho and CtBP. The mechanism of Brinker repression is dependent on promoter context and requires either or both Groucho and CtBP for switching of some target genes, whereas for silencing of others, it requires neither of these cofactors in which case it might be involved in direct competition with activators for DNA binding

sites (Hasson et al., 2001; Saller and Bienz, 2001). When tested against three high affinity Gal4 sites, no repression was observed in ventolateral stripes, leading to the conclusion that Brinker was unable to block *lacZ* expression (Figure D-3A, B).



## Figure D-2: The activity of the Gal4 activators from three high affinity sites is not repressed by Engrailed (en), Runt (run) or Even-skipped (eve).

Schematic of the reporter gene is shown below the corresponding embryos. The reporter contains two Engrailed (en) or two Runt (run) or two Even-skipped (eve) binding sites, three high affinity Gal sites (colored circles), a neutral spacer (open circles), hsp 70 basal promoter driving *lacZ* expression.

(A, B, C) Engrailed is unable to repress the Gal4 activators at the different stages of embryogenesis examined.

(D, E, F) Runt is unable to repress the Gal4 activators at various stages of embryogenesis.

(G, H) Even-skipped is unable to repress the Gal4 activators at various stages of embryogenesis.

LacZ expression is activated by the Gal4-Gal4 AD chimeric protein expressed in the ventral regions of the embryo. Patterns of gene expression were visualized in 2-4 hour embryos by *in situ* hybridization with digU labeled antisense *lacZ* probes. Embryos are oriented anterior to left; lateral view (A, B, C, D, E, F, and G) and field shot of embryos at various stages in development (H) are shown.
Suppressor of Hairless Su(H): Notch is the receptor for a conserved signaling pathway that regulates numerous cell fate decisions during development (Qi et al., 1999). Signal transduction involves the presenilin-dependent intracellular processing of Notch and the nuclear translocation of the intracellular domain of Notch, NICD (Lecourtois and Schweisguth, 1998; Struhl and Adachi, 1998; Struhl and Greenwald, 1999). NICD associates with Suppressor of Hairless [Su(H)], a DNA binding protein, and Mastermind (Mam), a transcriptional coactivator (Petcherski and Kimble, 2000). In the absence of Notch signaling, Su(H) acts as a transcriptional repressor (Barolo et al., 2000; Dou et al., 1994; Morel and Schweisguth, 2000). It has been shown that Hairless, an antagonist of Notch signaling (Bang et al., 1995; Bang et al., 1991; Bang and Posakony, 1992; Schweisguth and Posakony, 1994), and is required to repress the transcription of the singleminded gene in the Drosophila embryo. Hairless forms a DNA-bound complex with Su(H). Furthermore, it directly binds the Drosophila C-terminal Binding Protein (dCtBP) and Groucho, which act as transcriptional corepressors (Barolo et al., 2002; Morel et al., 2001). In the case of Su(H), we see some interesting patterns of lacZ expression when the protein is tested against three high affinity Gal4 sites (Figure D-3C, D, and E). In some embryos we see a posterior patch of *lacZ* expression in the dorsal region of the embryo (Figure D-3C; arrow). This may reflect activation by Su(H) in the absence of the Gal4 activator in nuclei where Notch signaling is active. In other embryos (Figure D-3D and E; arrow), we observe more intense *lacZ* staining, with a gap in expression in the central regions of the embryo that is wider in the more ventral regions. In this case we believe that lacZ



# Figure D-3: The activity of the transcription factors Brinker (brk) and Suppressor of Hairless [Su(H)] on a reporter with three high affinity Gal4 sites.

Schematic of the reporter gene is shown below the corresponding embryos. The reporter contains two Brinker (brk) or two Suppressor of Hairless [Su(H)] binding sites, three high affinity Gal sites (colored circles), a neutral spacer (open circles), hsp 70 basal promoter driving *lacZ* expression.

(A, B) Brinker expressed in ventrolateral stripes in the blastoderm embryo is unable to repress the Gal4 activators at the different stages of embryogenesis examined.

(C, D, E) Activity of Suppressor of Hairless [Su(H)]on hsp 70-lacZ reporter. (C) Posterior patch of lacZ expression (arrow) may result from activation by Su(H) alone in the absence of Gal4 activators. (D, E) LacZ staining is more intense, suggesting that Su(H) is synergizing with the Gal4 activator to drive lacZ expression. In the nuclei that show lacZ expression presumably Notch signaling is active. LacZ staining shows a gap in the central regions of the embryo (arrow). The gap is much wider in the ventrolateral and ventral regions of the embryo. This gap in lacZ expression may be mediated by Su(H) acting as a repressor presumably because Notch signaling is absent in the nuclei in this region.

LacZ expression is driven by full length Gal4, which is expressed ubiquitously throughout the embryo under the control of the actin5C enhancer. Patterns of gene expression were visualized in 2-4 hour embryos by *in situ* hybridization with digU labeled antisense *lacZ* probes. Embryos are oriented anterior to left; lateral view (A, C, D, and E)) and field shot of embryos at various stages in development (B) are shown.

staining is synergistically activated by both Gal4 and Su(H) in nuclei where Notch signaling is active, and in regions where *lacZ* expression is absent Su(H) is actively repressing the Gal4 activator in the absence of a Notch signal. Notch signaling in early embryogenesis is not well characterized. Therefore further investigation of Su(H) activity on this reporter construct will have to be examined later in development, for example in the wing imaginal disc where Notch signaling is well studied.

In addition to the repressors described above, I also tested the activity of the transcription factors Sloppy-paired, Tramtrack and Cubitus interruptus on the reporter with three high affinity Gal4 sites. The activity of these proteins as repressors is not well characterized and not much is known about their mode of action in gene regulation.

**Sloppy-paired (Slp 1):** Sloppy-paired (Slp 1) is a forkhead domain protein and has been recently shown to be involved in setting up the anterior border of *eve* stripe 2 expression in the *Drosophila* embryo (Andrioli et al., 2002). Two kinds of embryos were frequently seen in experiments testing the activity of this protein (Figure D-4A and B). We observed embryos that showed *lacZ* expression in a broad domain in the anterior region of the embryo, (Figure D-4A) which is consistent with the pattern in which Sloppy-paired is expressed. We also observed embryos showing *lacZ* expression in a dition to the anterior broad domain (Figure D-4B). These results indicate that Slp 1 activates rather than represses in the context tested.

Figure D-4: The activity of the transcription factors Sloppy-paired (Slp1), Tramtrack (ttk) and Cubitus interruptus (Ci) on a reporter with three high affinity Gal4 sites.

Schematic of the reporter gene is shown below the corresponding embryos. The reporter contains two Sloppy-paired (Slp1) or two Tramtrack (ttk) or two Cubitus interruptus (Ci) binding sites, three high affinity Gal sites (colored circles), a neutral spacer (open circles), hsp 70 basal promoter driving lacZ expression.

(A, B) Activity of the forkhead domain protein Sloppy-paired. (A) LacZ expression is activated in a broad anterior domain consistent with the pattern of Slp1 expression in the blastoderm embryo (arrow). (B) LacZ expression is also activated in a ventral swath by the Gal4 activator. Thus, Slp1 appears to activate rather than repress in this context.

(C, D) Activity of Tramtrack (ttk) on *hsp 70-lacZ* reporter. Tramtrack is unable to repress the Gal4 activators at various stages of embryogenesis.

(D, E) Activity of Cubitus interruptus (Ci) on *hsp 70-lacZ* reporter. Ci is unable to repress the Gal4 activators at various stages of embryogenesis.

LacZ expression in A, B, C, and D is driven by the Gal4-Gal4 AD chimeric activator in ventral regions where it is expressed and in E and F by full length Gal4, which is expressed ubiquitously throughout the embryo under the control of the actin5C enhancer. Patterns of gene expression were visualized in 2-4 hour embryos by *in situ* hybridization with digU labeled antisense *lacZ* probes. Embryos are oriented anterior to left; lateral view (A, C, D, and E)) and field shot of embryos at various stages in development (B) are shown.



Figure D-4: The activity of the transcription factors Sloppy-paired (Slp1), Tramtrack (ttk) and Cubitus interruptus (Ci) on a reporter with three high affinity Gal4sites.

CCI.

100

**Tramtrack (TTK):** TTK has been proposed to act as a maternally provided repressor of several pair-rule genes, such as even-skipped (eve). *eve* contains in its promoter region binding sites for the trithorax-like transcription factor GAGA and TTK. In transient expression experiments, it was demonstrated that GAGA activates transcription from the *eve* stripe 2 promoter element and that TTK inhibits this GAGA-dependent activation. Repression by TTK of the *eve* promoter requires its activation by GAGA and depends on the presence of the POZ/BTB domains of TTK and GAGA (Pagans et al., 2002). In the context of three high affinity Gal4 sites, TTK failed to block lacZ expression (Figure D-4C, D). It is possible that GAGA-TTK interaction is required for TTK function *in vivo*.

**Cubitus interruptus (Ci):** The Drosophila Gli homolog Cubitus interruptus (Ci) controls the transcription of Hedgehog (Hh) target genes. A repressor form of Ci arises in the absence of Hh signalling by proteolytic cleavage of intact Ci, whereas an activator form of Ci is generated in response to the Hh signal. These different activities of Ci regulate overlapping but distinct subsets of Hh target genes (Muller and Basler, 2000). We did not observe any activity mediated by Ci at the stages of embryogenesis examined (Figure D-4E, F). It is possible that in order to observe Ci activity, would require investigation at later stages in development for example in the imaginal discs where Hh signaling has been well characterized.

### MATERIALS AND METHODS

## 1. GAL4 (aa1-93) - GAL4 AD (aa753-881)

A *KpnI-XbaI* fragment from pSCTEV GAL4 (1-93)- GAL4 (Seipel et al., 1992) containing the reading frame for the yeast GAL4 activation domain (Gal4 AD) from amino acid residues 753-881, was cloned into *KpnI-XbaI* cut pTwiggy (Arnosti et al., 1996a; Arnosti et al., 1996c) vector, which contains the *twist* enhancer (2xPEe-Et) element, *twist* basal promoter and the GAL4 DNA-binding domain from residues 1-93.

## 2. Fly stocks

Flies expressing the full-length yeast transcriptional activator Gal4 ubiquitously throughout the embryo under the control of the actin5C enhancer, act5cGAL4/CyO (Stock # 4414) were also obtained from Bloomington. In order to obtain ubiquitous activation of the *lacZ* reporter gene in the early (2-4 hour old) embryo, act5cGAL4/CyO females were crossed to males carrying the reporter transgene.

## 3. Reporter genes

The vector **M2g3u2x-lacZ** was modified to remove the two Giant binding sites, replacing them with two sites for other transcriptional repressors with the following oligos:

**SNAIL** (Gray and Levine, 1996)

DA696: 5'GGC CGC CAG CAA GGT GGT ACT AGA CAT CAG CAA GGT GA 3' DA697: 5'AGC TTC ACC TTG CTG ATG TCT AGT ACC ACC TTG CTG GC 3' EVEN-SKIPPED (Li and Manley, 1998)

DA698: 5'GGC CGC TCA ATT AAA TGA GTA CTA GAC ATC AAT TAA ATG AA 3'

DA699: 5'AGC TT**T CAT TTA ATT GA**T GTC TAG TAC **TCA TTT AAT TGA** GC 3'

**TRAMTRACK** (Kamashev et al., 2000)

DA700: 5'GGC CGC GGT CCT GCG TAC TAG ACA GGT CCT GCA 3'

DA701: 5'AGC TTG CAG GAC CTG TCT AGT ACG CAG GAC CGC 3'

**BRINKER** (Saller and Bienz, 2001)

DA702: 5'GGC CGC GAG GCG CCA CCG TAC TAG ACA GAG GCG CCA CCA 3'

DA703: 5'AGC TTG GTG GCG CCT CTG TCT AGT ACG GTG GCG CCT CGC 3'

CUBITUS INTERRUPTUS (Muller and Basler, 2000)

DA704: 5'GGC CGC ACG GGC GGT CTG TAC TAG ACA ACG GGC GGT CTA 3'

DA705: 5'AGC TTA GAC CGC CCG TTG TCT AGT ACA GAC CGC CCG TGC 3'

SUPPRESSOR OF HAIRLESS (Guss et al., 2001)

DA706: 5'GGC CGC CGT GGG AAG TAC TAG ACA CGT GAG AAA 3'

# DA707: 5'AGC TTT TCT CAC GTG TCT AGT ACT TCC CAC GGC 3'

# **ENGRAILED** (TRANSFAC database)

DA708: 5'GGC CGC ACT AAT TAG CGT ACT AGA CAA CTA ATT AGC A 3'

DA709: 5'AGC TTG CTA ATT AGT TGT CTA GTA CGC TAA TTA GTG C 3'

RUNT (Kramer et al., 1999)

DA710: 5'GGC CGC TGC GGT CGT ACT AGA CAT GCG GTC A 3'

DA711: 5'AGC TTG ACC GCA TGT CTA GTA CGA CCG CAG C 3'

**SLOPPY PAIRED** (Yu et al., 1999)

DA714: 5'GGC CGC TCT TCG ATG TCA ACA CAC CGA CCC TCT TCG ATG TCA ACA CAC CA 3'

DA715: 5'AGC TT**G GTG TGT TGA CAT CGA AGA** GGG TC**G GTG TGT TGA** CAT CGA AGA GC 3'

4. P-element transformation, crosses to reporter genes, and whole-mount in situ hybridization of embryos.

P-element transformation vectors were introduced into the *Drosophila* germline by injection of  $yw^{67}$  embryos as described (Small et al., 1992a). Embryos were collected either directly from each transgenic reporter line or from a cross between a reporter line and a line expressing the GAL4-activator chimeric proteins in the ventral regions or ubiquitously throughout the embryo. The embryos were fixed and stained using digoxigenin-UTP labeled antisense RNA probes to either lacZ or w as described (Small et al., 1992b).

# REFERENCES

Andrioli, L. P., Vasisht, V., Theodosopoulou, E., Oberstein, A. and Small, S. (2002). Anterior repression of a Drosophila stripe enhancer requires three position-specific mechanisms. *Development* 129, 4931-40.

Arnosti, D. N., Barolo, S., Levine, M. and Small, S. (1996a). The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* 122, 205-14.

Arnosti, D. N., Gray, S., Barolo, S., Zhou, J. and Levine, M. (1996b). The gap protein knirps mediates both quenching and direct repression in the Drosophila embryo. *Embo J* 15, 3659-66.

Arnosti, D. N., Gray, S., Barolo, S., Zhou, J. and Levine, M. (1996c). The gap protein knirps mediates both quenching and direct repression in the Drosophila embryo. *Embo J* 15, 3659-66.

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.

**Bang, A. G., Bailey, A. M. and Posakony, J. W.** (1995). Hairless promotes stable commitment to the sensory organ precursor cell fate by negatively regulating the activity of the Notch signaling pathway. *Dev Biol* **172**, 479-94.

Bang, A. G., Hartenstein, V. and Posakony, J. W. (1991). Hairless is required for the development of adult sensory organ precursor cells in Drosophila. *Development* 111, 89-104.

**Bang, A. G. and Posakony, J. W.** (1992). The Drosophila gene Hairless encodes a novel basic protein that controls alternative cell fates in adult sensory organ development. *Genes Dev* **6**, 1752-69.

**Barolo, S. and Levine, M.** (1997). hairy mediates dominant repression in the Drosophila embryo. *Embo J* 16, 2883-91.

**Barolo, S., Stone, T., Bang, A. G. and Posakony, J. W.** (2002). Default repression and Notch signaling: Hairless acts as an adaptor to recruit the corepressors Groucho and dCtBP to Suppressor of Hairless. *Genes Dev* 16, 1964-76.

Barolo, S., Walker, R. G., Polyanovsky, A. D., Freschi, G., Keil, T. and Posakony, J.
W. (2000). A notch-independent activity of suppressor of hairless is required for normal mechanoreceptor physiology. *Cell* 103, 957-69.

Cadigan, K. M., Grossniklaus, U. and Gehring, W. J. (1994a). Functional redundancy: the respective roles of the two sloppy paired genes in Drosophila segmentation. *Proc Natl Acad Sci U S A* 91, 6324-8.

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

**Campbell, G. and Tomlinson, A.** (1999). Transducing the Dpp morphogen gradient in the wing of Drosophila: regulation of Dpp targets by brinker. *Cell* **96**, 553-62.

**DiNardo, S. and O'Farrell, P. H.** (1987). Establishment and refinement of segmental pattern in the Drosophila embryo: spatial control of engrailed expression by pair-rule genes. *Genes Dev* 1, 1212-25.

Dou, S., Zeng, X., Cortes, P., Erdjument-Bromage, H., Tempst, P., Honjo, T. and Vales, L. D. (1994). The recombination signal sequence-binding protein RBP-2N functions as a transcriptional repressor. *Mol Cell Biol* 14, 3310-9.

Fisher, A. L. and Caudy, M. (1998). Groucho proteins: transcriptional corepressors for specific subsets of DNA-binding transcription factors in vertebrates and invertebrates. *Genes Dev* 12, 1931-40.

Flores-Saaib, R. D. and Courey, A. J. (2000). Analysis of Groucho-histone interactions suggests mechanistic similarities between Groucho- and Tup1-mediated repression. *Nucleic Acids Res* 28, 4189-96.

Fujioka, M., Jaynes, J. B. and Goto, T. (1995). Early even-skipped stripes act as morphogenetic gradients at the single cell level to establish engrailed expression. *Development* 121, 4371-82.

Fujioka, M., Yusibova, G. L., Patel, N. H., Brown, S. J. and Jaynes, J. B. (2002). The repressor activity of Even-skipped is highly conserved, and is sufficient to activate engrailed and to regulate both the spacing and stability of parasegment boundaries. *Development* 129, 4411-21.

Gray, S. and Levine, M. (1996). Short-range transcriptional repressors mediate both quenching and direct repression within complex loci in Drosophila. *Genes Dev* 10, 700-10.

Gray, S., Szymanski, P. and Levine, M. (1994). Short-range repression permits multiple enhancers to function autonomously within a complex promoter. *Genes Dev* 8, 1829-38.

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.

Guss, K. A., Nelson, C. E., Hudson, A., Kraus, M. E. and Carroll, S. B. (2001). Control of a genetic regulatory network by a selector gene. *Science* 292, 1164-7. Hasson, P., Muller, B., Basler, K. and Paroush, Z. (2001). Brinker requires two corepressors for maximal and versatile repression in Dpp signalling. *Embo J* 20, 5725-36.

Hewitt, G. F., Strunk, B. S., Margulies, C., Priputin, T., Wang, X. D., Amey, R., Pabst, B. A., Kosman, D., Reinitz, J. and Arnosti, D. N. (1999). Transcriptional repression by the Drosophila giant protein: cis element positioning provides an alternative means of interpreting an effector gradient. *Development* 126, 1201-10.

Howard, K. and Ingham, P. (1986). Regulatory interactions between the segmentation genes fushi tarazu, hairy, and engrailed in the Drosophila blastoderm. *Cell* 44, 949-57.

Jazwinska, A., Kirov, N., Wieschaus, E., Roth, S. and Rushlow, C. (1999a). The Drosophila gene brinker reveals a novel mechanism of Dpp target gene regulation. *Cell* 96, 563-73.

Jazwinska, A., Rushlow, C. and Roth, S. (1999b). The role of brinker in mediating the graded response to Dpp in early Drosophila embryos. *Development* 126, 3323-34.

Jimenez, G., Paroush, Z. and Ish-Horowicz, D. (1997). Groucho acts as a corepressor for a subset of negative regulators, including Hairy and Engrailed. *Genes Dev* 11, 3072-82.

Jimenez, G., Pinchin, S. M. and Ish-Horowicz, D. (1996). In vivo interactions of the Drosophila Hairy and Runt transcriptional repressors with target promoters. *Embo J* 15, 7088-98.

Kamashev, D. E., Balandina, A. V. and Karpov, V. L. (2000). Tramtrack protein-DNA interactions. A cross-linking study. *J Biol Chem* 275, 36056-61.

Keller, S. A., Mao, Y., Struffi, P., Margulies, C., Yurk, C. E., Anderson, A. R., Amey, R. L., Moore, S., Ebels, J. M., Foley, K. et al. (2000). dCtBP-dependent and - independent repression activities of the Drosophila Knirps protein. *Mol Cell Biol* 20, 7247-58.

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.

Lecourtois, M. and Schweisguth, F. (1998). Indirect evidence for Delta-dependent intracellular processing of notch in Drosophila embryos. *Curr Biol* 8, 771-4.

Li, C. and Manley, J. L. (1998). Even-skipped represses transcription by binding TATA binding protein and blocking the TFIID-TATA box interaction. *Mol Cell Biol* 18, 3771-81.

Mannervik, M., Nibu, Y., Zhang, H. and Levine, M. (1999). Transcriptional coregulators in development. *Science* 284, 606-9.

Manoukian, A. S. and Krause, H. M. (1992). Concentration-dependent activities of the even-skipped protein in Drosophila embryos. *Genes Dev* 6, 1740-51.

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

Minami, M., Kinoshita, N., Kamoshida, Y., Tanimoto, H. and Tabata, T. (1999). brinker is a target of Dpp in Drosophila that negatively regulates Dpp-dependent genes. *Nature* **398**, 242-6.

Morel, V., Lecourtois, M., Massiani, O., Maier, D., Preiss, A. and Schweisguth, F. (2001). Transcriptional repression by suppressor of hairless involves the binding of a hairless-dCtBP complex in Drosophila. *Curr Biol* 11, 789-92.

Morel, V. and Schweisguth, F. (2000). Repression by suppressor of hairless and activation by Notch are required to define a single row of single-minded expressing cells in the Drosophila embryo. *Genes Dev* 14, 377-88.

Muller, B. and Basler, K. (2000). The repressor and activator forms of Cubitus interruptus control Hedgehog target genes through common generic gli-binding sites. *Development* 127, 2999-3007.

Nibu, Y., Zhang, H., Bajor, E., Barolo, S., Small, S. and Levine, M. (1998a). dCtBP mediates transcriptional repression by Knirps, Kruppel and Snail in the Drosophila embryo. *Embo J* 17, 7009-20.

Nibu, Y., Zhang, H. and Levine, M. (1998b). Interaction of short-range repressors with Drosophila CtBP in the embryo. *Science* 280, 101-4.

Nibu, Y., Zhang, H. and Levine, M. (2001). Local action of long-range repressors in the Drosophila embryo. *Embo J* 20, 2246-53.

**Pagans, S., Ortiz-Lombardia, M., Espinas, M. L., Bernues, J. and Azorin, F.** (2002). The Drosophila transcription factor tramtrack (TTK) interacts with Trithorax-like (GAGA) and represses GAGA-mediated activation. *Nucleic Acids Res* **30**, 4406-13.

Petcherski, A. G. and Kimble, J. (2000). Mastermind is a putative activator for Notch. *Curr Biol* 10, R471-3.

**Poortinga, G., Watanabe, M. and Parkhurst, S. M.** (1998). Drosophila CtBP: a Hairyinteracting protein required for embryonic segmentation and hairy-mediated transcriptional repression. *Embo J* 17, 2067-78.

Qi, H., Rand, M. D., Wu, X., Sestan, N., Wang, W., Rakic, P., Xu, T. and Artavanis-Tsakonas, S. (1999). Processing of the notch ligand delta by the metalloprotease Kuzbanian. *Science* 283, 91-4. Saller, E. and Bienz, M. (2001). Direct competition between Brinker and Drosophila Mad in Dpp target gene transcription. *EMBO Rep* 2, 298-305.

Schweisguth, F. and Posakony, J. W. (1994). Antagonistic activities of Suppressor of Hairless and Hairless control alternative cell fates in the Drosophila adult epidermis. *Development* 120, 1433-41.

Seipel, K., Georgiev, O. and Schaffner, W. (1992). Different activation domains stimulate transcription from remote ('enhancer') and proximal ('promoter') positions. *Embo J* 11, 4961-8.

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

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

Struhl, G. and Adachi, A. (1998). Nuclear access and action of notch in vivo. Cell 93, 649-60.

Struhl, G. and Greenwald, I. (1999). Presenilin is required for activity and nuclear access of Notch in Drosophila. *Nature* 398, 522-5.

Sundqvist, A., Sollerbrant, K. and Svensson, C. (1998). The carboxy-terminal region of adenovirus E1A activates transcription through targeting of a C-terminal binding protein-histone deacetylase complex. *FEBS Lett* **429**, 183-8.

Tolkunova, E. N., Fujioka, M., Kobayashi, M., Deka, D. and Jaynes, J. B. (1998). Two distinct types of repression domain in engrailed: one interacts with the groucho corepressor and is preferentially active on integrated target genes. *Mol Cell Biol* 18, 2804-14.

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.

Yu, Y., Yussa, M., Song, J., Hirsch, J. and Pick, L. (1999). A double interaction screen identifies positive and negative ftz gene regulators and ftz-interacting proteins. *Mech Dev* 83, 95-105.

Zhang, H. and Levine, M. (1999). Groucho and dCtBP mediate separate pathways of transcriptional repression in the Drosophila embryo. *Proc Natl Acad Sci U S A* 96, 535-40.