Dissecting the regulatory switches of development: lessons from enhancer evolution in Drosophila

Matthew J. Borok*, Diana A. Tran*, Margaret C. W. Ho and Robert A. Drewell†

Summary
Cis-regulatory modules are non-protein-coding regions of DNA essential for the control of gene expression. One class of regulatory modules is embryonic enhancers, which drive gene expression during development as a result of transcription factor protein binding at the enhancer sequences. Recent comparative studies have begun to investigate the evolution of the sequence architecture within enhancers. These analyses are illuminating the way that developmental biologists think about enhancers by revealing their molecular mechanism of function.

Introduction
In order to form the tissues of the embryo, the expression patterns of genes must be tightly directed in space and time. This function is largely controlled by cis-regulatory modules (CRMs), specific regions of non-protein coding DNA that regulate genes on the same chromosome. Specifically, early patterns of gene expression regulated by CRMs are required to establish the anterior to posterior axis of the developing embryo. The best-understood CRM is the enhancer, a genetic switch that is bound by specific transcription factors (TFs) and is able to drive distinct patterns of transcription of a target gene in an orientation-independent manner (Dillon and Sabbattini, 2000; Ptashne, 1986) (see Box 1). The binding of TFs to enhancers recruits RNA polymerase II and the associated transcriptional machinery to a gene’s promoter to drive the transcription of target genes. Notably, some of the components of the machinery, such as histone acetyltransferases and nucleosome remodelers, have the ability to modify the chromatin environment in order to facilitate transcription (Kim and Dean, 2003; Madisen et al., 1998; Soutoglou and Talianidis, 2002; Utley et al., 1998).

For many years, researchers have sought to identify these modules and their respective target genes. However, recent interest has shifted towards a more complex analysis of individual enhancers in order to elucidate their molecular activity. Key questions include: which TF binding sites are responsible for functional activity; how these binding sites are organized; and what their affinities for different TFs are. The wealth of information that is emerging from the recent sequencing of many different genomes has provided answers to some of these questions, but has also brought controversy to the field. Particularly in insects, as we discuss in this Primer, evolutionary comparisons of enhancers across different species are beginning to show us that slight variations in non-coding regulatory sequences can lead to large phenotypic differences in development (Gompel et al., 2005; Jeong et al., 2006; Prud’homme et al., 2006). Such detailed analyses of enhancer evolution allow for a better understanding of the organizational and structural constraints of a CRM. In turn, the elucidation of such evolutionary constraints should reveal the nature of TF interactions with regulatory DNA sequences, as well as what is required for the transcriptional control of specific genes during development.

In this Primer, we briefly review our current understanding of classical enhancer function and then delve into recent studies that have compared enhancers in Drosophila and related insect species to reveal how evolution acts upon mechanisms of enhancer function. We focus on the even skipped stripe 2 enhancer (S2E) and neurogenic ectoderm enhancers (NEEs) in particular, owing to the lessons that these CRMs can teach us about the general principles of transcriptional control during embryonic development.

The building blocks of gene regulation: the functional characteristics of enhancers
Enhancers were first identified in the SV40 viral genome as regions of non-coding DNA that are crucial for the transcription of adjacent genes (Banerji et al., 1981; Benoist and Chambon, 1981). Functional activity of an enhancer depends upon the binding of specific TF proteins (activators, see Box 2) within the enhancer DNA sequence that help to recruit RNA polymerase II and associated protein factors to the promoter of a target gene (detailed in Box 1) (McEwan et al., 1993; Wildeman et al., 1984). In the years after these initial observations, enhancers were discovered for many genes in a wide range of model organisms (Banerji et al., 1983; Shepherd et al., 1985; Straul, 1984). In complex eukaryotes, many of the identified enhancers are responsible for directing spatiotemporally restricted patterns of gene expression in the developing embryo (Choi and Engel, 1986; Hiromi et al., 1985). Although a number of enhancers are able to activate transcription at a given eukaryotic gene, the regulation of key genes by a restricted set of tissue-specific enhancers during development must be very carefully controlled because they encode proteins that specify cellular identities in the embryo (Caplan and Ordahl, 1978; Lewis, 1978). These crucial developmental genes cannot, therefore, be expressed ubiquitously. A number of emerging studies indicate that these specific promoter-enhancer interactions are subject to subtle molecular control mechanisms, including the activity of a novel class of regulatory module – the promoter tethering element (Akbari et al., 2008; Akbari et al., 2007; Calhoun et al., 2002; Cande et al., 2009; Fujioka et al., 2009; Kwon et al., 2009).

In addition to activators, another class of regulatory TFs, repressors (see Box 2), are able to bind directly to enhancers, in this case to prevent target gene transcription (Dearolf et al., 1989; Stanojevic et al., 1989). Although the molecular roles of activators and repressors in promoting or preventing the recruitment of RNA polymerase II to gene promoters are not completely understood, some mechanisms for the functional activity of enhancers have been characterized. The predominant theory is that activators directly interact with components of the basal transcription machinery via specific protein domains, such as glutamine-rich, proline-rich and acidic domains or hydrophobic β sheets (McEwan et al., 1993); for a detailed review, see Kadonaga (Kadonaga, 2004). There is also...
An enhancer is an orientation-independent region of non-protein-coding DNA in the genome that is associated with a promoter and target gene. Transcription factor (TF) proteins bind to the enhancer and interact with the transcriptional machinery. The Mediator complex facilitates interactions between the enhancer and chromatin-modifying enzymes, such as histone acetyltransferases and nucleosome remodeling factors, to establish a chromatin environment that facilitates transcription of the target gene. By contrast, repressors may disrupt enhancer functional activity in one of two ways: (1) competition, where the binding sites for repressors and activators within an enhancer sequence overlap and, as a result, repressor binding excludes activators; and (2) quenching, in which repressors are able to inhibit the regulatory activity of activators bound to nearby sites within the enhancer. Histone acetyltransferases and nucleosome remodeling factors (NR, orange spheres) that facilitate changes to the chromatin environment to allow transcription to proceed. Historically, enhancers were often initially identified in genetic screens, as mutation of the enhancer module disrupts TF binding and results in an associated loss of target gene expression. Later, the development of transposon-mediated mutagenesis in Drosophila and plants allowed more sophisticated ‘enhancer trap’ reporter gene constructs to be used. Enhancers identified in these types of genetic screens include those in the bithorax complex of D. melanogaster for Kruppel (KR), Giant (GT), Bicoid (BCD) and Hunchback (HB) in metazoan species, enhancer CRMs are major players in the developmental cascade that transforms the early embryo from a mass of uniform, undifferentiated cells into a segmented and highly organized structure of differentiated cells. In Drosophila, the dynamic developmental specification of the embryonic body plan and of differentiated cell fates is accomplished by a combination of early spatiotemporal expression gradients of activator and repressor TFs that act upon downstream embryonic enhancers. Based on the timing of their expression during embryonic development, the genes in this cascade represent four basic developmental TF families: maternal, gap, pair-rule and homeotic genes (described in detail in Box 3) (for a review, see Sauer et al., 1996). At the top of the cascade, maternal mRNAs are deposited in the unfertilized egg cell during oogenesis (Berleth et al., 1988; Steward et al., 1988). Spatially restricted translation of localized maternal mRNAs in the fertilized egg establishes TF gradients in the embryo. In turn, maternal TFs bind at target embryonic enhancers for gap genes, directing gap TF expression patterns in the developing embryo (Driever and Nüsslein-Volhard, 1988; Struhl et al., 1989) (Fig. 1A). Gap TFs further regulate downstream target genes, such as those for pair-rule and homeotic TFs (Qian et al., 1991; Stanojevic et al., 1989). At each step in the cascade, gene expression patterns are controlled by specific clusters of activator and repressor binding sites within embryonic enhancers (see Box 2). Fine-tuning of transcription is mediated by the specific molecular properties of individual enhancer CRMs. Whether a given TF acts as an activator or repressor when it binds to an embryonic enhancer can be context dependent (Ip et al., 1991; Small et al., 1996). In addition, DNA sequences within embryonic enhancers may bind TFs with varying affinities (Jiang et al., 1991; Small et al., 1996). The functional consequence is that enhancers may require different threshold concentrations of interacting activators and repressors in order to regulate transcription of their target genes. The central role of enhancers in the regulatory cascade responsible for development of the Drosophila embryo makes in-depth analysis of their function crucial to the field of developmental biology.

A CRM paradigm: even skipped stripe 2 enhancer

Probably the best-characterized embryonic CRM is the even skipped (eve) S2E in Drosophila melanogaster (see Box 2). eve is a pair-rule gene that is expressed under the control of various neighboring CRMs in seven distinct stripes in the developing embryo (Macdonald et al., 1986) (Fig. 1B-D). The pattern of eve expression that S2E directs is entirely derived from the combination of transcription factor binding sites (TFBSs, see Box 2) embedded within its DNA sequence (Small et al., 1991b) (Fig. 1). Pioneering studies in the laboratory of Michael Levine twenty years ago discovered that twelve different binding sites for Kruppel (KR), Giant (GT), Bicoid (BCD) and Hunchback (HB)
TFs are located in the minimal S2E, a 480 bp region of the CRM that is sufficient to drive \textit{eve} stripe 2 expression (Fig. 1C) (Goto et al., 1989; Small et al., 1991b; Stanojevic et al., 1989). Notably, eight of the twelve TFBSs form two tight clusters at each end of the minimal S2E. Each cluster has closely spaced activator binding sites, each of which strongly overlaps with a repressor binding site (Fig. 1C). These observations suggested two important ideas regarding the molecular architecture of an enhancer: (1) clustering of TFBSs is crucial to enhancer function; and (2) the overlap between activator and repressor TFBSs enables the characteristically sharp boundaries between stripes of gene expression that are driven by many early embryonic enhancers (Fig. 1D).

Analysis of the TF gradients present in the developing \textit{Drosophila} embryo reveals how S2E directs \textit{eve} expression with such sharp boundaries. Activation of S2E-driven expression of \textit{eve} occurs through the binding of the HB and BCD activators (see Box 3), which are present at high concentrations in the anterior of the embryo (Fig. 1A). Binding of the KR and GT repressors (see Box 3), which are present in distinct non-overlapping patterns in the anterior half of the embryo, prevents activation of S2E (Small et al., 1991a) (Fig. 1C). With strong activation and no repression, S2E only drives \textit{eve} expression in a narrow 2- to 3-cell-wide region of the developing embryo (Macdonald et al., 1986) (Fig. 1D). By integrating the study of the spatial distribution of TFs in the embryo, the binding of these TFs at specific enhancers and the resulting patterns of target gene expression, we begin to see how an embryonic enhancer is analogous to a central processing unit. The enhancer must be capable of receiving input signals from TFs and of outputting a response in the form of directing a very specific spatiotemporal pattern of target gene expression.

**Box 2. Glossary of specialized terms**

**Activator.** A transcription factor that, when bound within a particular enhancer, is able to recruit the transcriptional machinery to upregulate transcription of a target gene under control of the enhancer.

**Cluster.** A group of closely spaced transcription factor binding sites thought to be the hub of enhancer activity. Clusters usually include binding sites for both activators and repressors.

**Eve stripe 2 enhancer (S2E).** An extensively studied embryonic enhancer of the even skipped gene.

**Neurogenic ectodermal enhancer (NEE).** A specific type of enhancer that directs gene expression in the neurogenic ectoderm. Spatial and temporal expression of NEE-driven genes depends upon the activators Dorsal and Twist, and the repressor Snail.

**Orthologous gene.** A gene derived from the gene of a common ancestor.

**PATSER.** A pattern search program that identifies potential transcription factor binding sites in a given DNA sequence using a position-weighted matrix.

**Position-weighted matrix (PWM).** Generated from a compilation of experimentally verified binding site sequences for a specific transcription factor (TF), it is a mathematical matrix that indicates the probability of finding a specific nucleotide at each position in the binding site.

**Repressor.** A TF that, when bound within a particular enhancer, blocks transcription of the target gene.

**Transcription factor binding site (TFBS).** A sequence of DNA known to be bound by a specific TF protein that influences transcription of nearby target genes.

**Two-dimensional similarity plots.** A graphical method that identifies regions of similarity between two sequences.

Evolution of the eve stripe 2 embryonic enhancer

Recent evolutionary analyses of TFBSs within enhancer sequences and their relative rate of evolutionary turnover are beginning to provide a useful perspective on the molecular mechanisms of enhancer function. Early comparative studies in the Kreitman laboratory concentrated on evolutionary analysis of the sequence of S2E across several \textit{Drosophila} species, including \textit{D. melanogaster}, \textit{D. yakuba}, \textit{D. erecta} and \textit{D. pseudoobscura} (see Fig. 2A for phylogenetic relationship). Their studies revealed that all of the \textit{Drosophila} species investigated possess an orthologous early embryonic enhancer (see Box 2) capable of driving \textit{eve} expression.
Early in Drosophila embryonic development, the initial anterior-to-posterior morphogen gradient comprises maternally deposited (MAT) mRNAs in the unfertilized egg, including bicoid and hunchback. Translation produces the MAT transcription factors (TFs), which can activate or repress spatiotemporally restricted patterns of expression of gap genes, such as Kruppel, knirps and giant. In turn, these gap TFs regulate downstream expression of pair-rule genes, such as eve, in stripes in the developing Drosophila embryo. Pair-rule TFs regulate segment polarity genes further downstream in the cascade. In addition, gap and pair-rule TFs both regulate the expression of the homeotic genes, which ultimately define segmental identity.

**Maternal genes**

**Bicoid** (BCD). The unfertilized egg contains a broad, maternally deposited distribution of bicoid mRNA, which is transported to the anterior pole of the developing embryo prior to translation. After translation, BCD protein diffuses from a high concentration at the anterior end across the embryo and acts as an activator of hunchback.

**Hunchback** (HB). The hunchback gene is activated by BCD in the anterior half of the embryo. Maternally deposited hunchback mRNAs also help to produce this protein in the embryo. HB protein functions as a regulator of other gap genes and as a threshold level-dependent activator or repressor at different enhancers of the pair-rule genes, including even skipped (eve).

**Gap genes**

**Kruppel** (KR). The Kruppel gene is repressed by high levels of HB at the anterior end of the embryo, leading to its localization in the middle of the embryo, where HB levels begin to diminish. KR protein functions as a repressor of downstream target genes, including eve.

**Knirps** (KNI). The knirps gene is repressed by intermediate levels of HB, and thus KNI protein is present in a peak posterior to that of KR. KNI functions as a repressor of eve.

**Giant** (GT). The giant gene is repressed by low levels of HB, and is therefore only expressed in one anterior and one posterior region of the embryo, where HB is not present. GT protein functions as a repressor of eve.

**Pair-rule genes**

**Even skipped** (EVE). The eve gene is expressed in seven stripes in the developing embryo. eve enhancers are activated by Caudal and HB, and repressed by HB, KR, KNI and GT. EVE protein establishes segmentation in the embryo.

Box 3. Key players in anterio posterior patterning in early Drosophila development

A posterior shift or expansion (Ludwig et al., 2000), suggesting that nuanced enhancer architecture might be encoding important functional information. Bioinformatic analysis revealed that the TFBSs contained within the S2E are not well-conserved between orthologs (Ludwig and Kreitman, 1995; Ludwig et al., 1998). There is, in fact, significant turnover of TFBSs between the Drosophila S2E orthologs studied: of the seventeen known binding sites analyzed in the 798 bp full-length S2E from D. melanogaster, only three are completely conserved at the sequence level in the other three Drosophila species studied (Ludwig et al., 1998) (Fig. 2B). In addition, none of the sixteen binding sites that were surveyed from the full-length D. melanogaster S2E are conserved in all thirteen sequenced Drosophila species (Ludwig et al., 2000). Detailed sequence alignment revealed that binding sites for BCD, HB, KR and GT identified in D. melanogaster are subject to extensive nucleotide substitutions in other Drosophila species (Ludwig et al., 2000; Ludwig et al., 1998). The key question therefore becomes: how is the functional activity of the enhancer preserved at all, despite the evolutionary turnover of TFBSs within the CRM?

Some resolution to this intriguing issue is being provided by exciting new studies that have expanded the evolutionary scope of S2E analysis. These studies analyze evolutionarily divergent species outside of the Drosophila genus (Fig. 2). Drosophila species generally have small, compacted genomes with relatively uniformly conserved non-coding DNA. In comparison, species of the true fruit flies (Tephritidae family, Sepsid species) have genomes that are between 4- and 6-times larger than that of D. melanogaster and contain blocks of conserved non-coding sequence flanked by regions that are poorly conserved (Peterson et al., 2009). The increased evolutionary divergence between these species, combined with the differences in overall genome structure, enable a more accurate measure of TFBS turnover and of the functional significance of sequence conservation.

Recent studies have compared the eve regulatory locus in Drosophila with that of Sepsid species, which diverged approximately 100 million years ago (Hare et al., 2008a) (Fig. 2A). Nevertheless, eve is expressed in the same seven transverse stripes in Sepsid embryos as in Drosophila embryos (Hare et al., 2008a). In addition, many of the TFs upstream in the developmental cascade, including HB, GT and KR, are also expressed in conserved embryonic patterns in the Sepsid species Themira minor. Despite the fact that eve regulatory regions are found to be only minimally conserved between these groups, Sepsid eve enhancers identified through clusters of HB, Caudal (CAD), Knirps (KNI), KR and BCD binding sites are able to drive conserved gene expression patterns in transgenic D. melanogaster.

Functional conservation of the Sepsid and Drosophila S2E orthologs despite relatively low sequence conservation indicates that some other shared molecular property is responsible for their activity. The identity of this common molecular mechanism is currently the focus of active research and debate. Bioinformatic analysis in the Eisen laboratory demonstrates that there has been large-scale reorganization of the TFBSs in the eve enhancers from different Drosophila and Sepsid species (Fig. 2B), indicating that the spatial organization of the binding sites within enhancer regions may not be crucial (Hare et al., 2008a). However, detailed analysis of the architecture of S2E orthologs reveals the existence of 20-30 bp blocks of highly conserved sequence enriched in pairs of neighboring or overlapping TFBSs. This suggests that the relative position of binding sites to one another might be more important than their overall spatial arrangement within an enhancer (Hare et al., 2008a) (Fig. 2B).
Fig. 2. Phylogenetic relationship and eve stripe 2 enhancer architecture of Drosophila and Sepsid species. (A) The Drosophila (green box) diverged from the Sepsid family (blue box) approximately 100 million years ago (Mya). The distantly related Drosophila species, D. melanogaster and D. virilis, diverged 60 Mya (Tamura et al., 2004). (B) The organization of a subset of bioinformatically predicted binding sites for the transcription factors Bicoid (BCD, blue), Hunchback (HB, purple), Giant (GT, green), Kruppel (KR, red) and Sloppy paired 1 (SLP1, yellow) within the minimal eve stripe 2 enhancer [summarized from Hare (Hare et al., 2008a)] are shown for eight species: D. melanogaster, D. simulans, D. yakuba, D. erecta, D. pseudoobscura, D. virilis, Sepsis cynipsea and Themira minor.

Several crucial issues were raised following the conclusions drawn from the initial studies of the orthologous insect S2Es (Hare et al., 2008a). A primary concern is that the S2E sequences in Drosophila and Themira are not as diverged as originally indicated, and therefore the lack of sequence homology does not indicate a lack of conserved TF organization. Using two-dimensional similarity plots (see Box 2), Crocker and Erives aligned the S2Es from Drosophila melanogaster and Themira putris and found extensive homology between a series of specific 14-41 bp stretches of sequence. The order of these sequence blocks is conserved across the entire length of the enhancer, suggesting that their position relative to each other is crucial. As a result, these authors argue that these sequences harbor binding sites for KR, BCD and GT and thus that the TFBS architecture is largely unchanged between these insect families (Crocker and Erives, 2008a). In response, Hare et al. note that the plots only align half of the minimal S2E region; by contrast, the other half of S2E exhibits little or no conservation despite its necessity for proper functional activity of the S2E module (Hare et al., 2008b). Furthermore, computational studies using PATSER [a bioinformatics tool that can be used to scan a given DNA sequence with a position-weighted matrix (PWM, see Box 2) representing a consensus binding site for a specific transcription factor] to predict TFBSs within S2E orthologs reveals that an extensive genomic reorganization of binding sites has occurred between Drosophila and Sepsid families. Indeed, 24 of the TFBSs analyzed in this study are not conserved between Drosophila and Sepsids, despite the fact that many of these sites bind conserved proteins known to regulate eve expression through S2E, such as HB, BCD, GT, KR and an additional pair-rule repressor TF, Sloppy paired 1 (SLP1) (Andrioli et al., 2002; Hare et al., 2008b). The studies by Hare et al. suggest that TFBSs are the essential molecular components within enhancer regions that, even when extensively reorganized, can still function to produce the same patterns of expression (Hare et al., 2008a; Hare et al., 2008b).

Overall, there exist caveats to the use of sequence alignments, including similarity plots, which can overemphasize weak homology across short stretches of sequence that are sometimes near the threshold for statistical significance. Although sequence analysis and alignment approaches can be useful in the identification of potential regions of conservation, the biological relevance of such alignments is not easily deciphered. To date, no functional synthetic S2E has been built, which could support the idea that this CRM is a loose cluster of TFBSs with as yet undefined organizational requirements. Further analysis of S2E will continue to shed light on the molecular mechanisms of enhancer regulation. In the meantime, studies of additional Drosophila CRMs may be beginning to elucidate this issue.

Molecular motifs in neurogenic ectodermal enhancer architecture

The functional significance of TFBSs has also been investigated in NEEs (see Box 2), which are crucial for dorsoventral patterning of the neurogenic ectoderm in the developing Drosophila embryo (Erives and Levine, 2004). The gene expression patterns directed by this class of CRMs are mediated largely by three TFs: the activators Dorsal (DL) and Twist (TWI) and the repressor Snail (SNA) (reviewed by Rusch and Levine, 1996). The genomes of D. melanogaster, D. pseudoobscura and D. virilis were analyzed for the presence of a consensus NEE motif comprising several sequences, including closely spaced binding sites for DL and TWI and a SNA binding site that overlapped with the TWI binding site (Crocker et al., 2008b). Using their consensus motif, the authors identified five NEEs in D. melanogaster, five in D. virilis and four in D. pseudoobscura. In agreement with the findings of Hare et al. for the eve gene (Hare et al., 2008a), orthologous NEE-driven genes from different Drosophila species show nearly identical embryonic expression patterns. However, when the bioinformatically identified NEEs were tested in transgenic D. melanogaster, they produced different patterns of expression. Although the predicted NEEs directed expression in a band on the ventral side of the embryo extending along the entire anterioposterior axis (as is normal), the number of nuclei in the band was not consistent. As the computational criteria by which the NEEs were identified depended on homology to a consensus sequence, the orthologous NEEs were very similar in their organization of DL, TWI and SNA TFBSs. These results raise a fundamental question: how are the endogenous NEE-driven patterns of gene expression the same in all species, whereas the identified NEE orthologs direct different patterns of reporter gene transcription in transgenic D. melanogaster?

A crucial difference between NEE orthologs that appears to regulate the distinct functional activity of these enhancers from different Drosophila species is the spacing between DL and TWI
Fig. 3. Changing regulatory environments can drive compensatory changes in enhancer organization. In (A) D. melanogaster and (B) D. virilis, the Dorsal (orange) and Twist (gray) activators cooperatively bind to neurogenic ectoderm enhancer (NEE) sequences (green) to upregulate the transcription of target genes (arrow). This interaction is dependent upon compatibility between the transcription factors and their respective binding sites. In the ~60 million years of evolution separating D. virilis and D. melanogaster (Tamura et al., 2004), protein sequences have diverged only slightly, but enough to affect the DNA-binding domains of the proteins. As a result, in the vnd NEEs of (C) D. melanogaster and (D) D. virilis, the spacing between a conserved Dorsal binding site (orange boxes) and a conserved Twist binding site (gray boxes) is different. The intervening sequence between the binding sites is only 37.5% conserved, but there is also a deletion of two nucleotides (white boxes) in the D. virilis NEE. Even small changes in transcription factor protein domains between different species can drive the architectural reorganization of DNA binding sites. In this case, a decrease in the spacing between sites preserves enhancer functional activity.

activator binding sites. For example, in the D. melanogaster ventral nervous system defective (vnd) NEE, the DL and TWI TFBSs are separated by 10 bp (Fig. 3A,C). In the orthologous D. virilis vnd NEE, the distance between the sites is shortened to 8 bp (Crocker et al., 2008b) (Fig. 3B,D). Could this relatively minor modulation of TFBS architecture have a pronounced influence on functional activity? In several experiments, Crocker and colleagues found that adding or deleting just a few base pairs between the two TFBSs in an NEE from one species to mimic the spacing in a second species (or another NEE from the same species) is in fact enough to recreate the pattern of gene expression directed by the second NEE (Crocker et al., 2008a; Crocker et al., 2008b). What could be directing these subtle modifications in CRM architecture? Previous studies suggest that slight differences in the trans environment of different species can drive evolutionary changes at the sequence level in CRMs (Gasch et al., 2004; Ronshaugen et al., 2002). A close investigation of DL and TWI protein sequences has revealed that several changes have occurred between the species, including one mutation in D. melanogaster DL that is known to affect DL-TWI gene activation (Jia et al., 2002). Therefore, the way in which these two proteins interact with their DNA binding sites, and thus with each other indirectly, has probably slightly changed between D. virilis and D. melanogaster (Fig. 3). These modifications at the protein level in turn lead to the co-evolution of the regulatory sequences bound by these TFs (Fig. 3). The pattern of expression of NEE-driven genes is crucial for the development of the fly body plan, so we can assume that any change in expression of NEE-target genes would probably be strongly selected against in evolution. Thus, as the trans environment changes, natural selection also demands compensatory changes in the cis-regulatory sequences. In the case of the NEEs, cis-evolution appears to act by modulating the spacing of binding sites to ensure that functional activity of the CRM is maintained in a given species (Fig. 3). Accordingly, when NEEs are tested in a different species the trans environment will be different and thus the NEEs will not respond to the upstream TFs with the same transcriptional outputs.

Is transcription factor binding site organization crucial for enhancer function?

The current debate over the functional significance of TFBS organization in CRMs highlights a key mystery in developmental biology: how exactly does an enhancer function at the molecular level? Without a doubt, the TF binding sites within an enhancer are crucially important. However, studies of the eve stripe 2 regulatory module suggest that binding site turnover occurs frequently in evolution, resulting in changes to the overall number and architecture of TFBSs within orthologous enhancers. Surprisingly, these pronounced differences appear to have no significant effect on the regulation of target gene transcription, suggesting that enhancers are under selective pressure to be functionally robust to evolutionary changes at the sequence level. Our own studies on early embryonic enhancers from the homeotic genes in Drosophila (Ho et al., 2009) have corroborated these discoveries. In general, enhancers from different species maintain their conserved functional activity despite significant modulation of their sequence architecture during evolution.

Can we use these observations of enhancer activity to develop a clearer idea of the molecular function of CRMs? Recent studies do in fact lend support to the information display/billboard model of enhancer function (Fig. 4A). The information display/billboard model suggests that an enhancer functions not as a single large processor, but as a series of autonomous signals, each of which can have an effect on target gene transcription (Arnosti et al., 1996; Kulkarni and Arnosti, 2003). In this model, as long as there is never a net decrease in activator and repressor TFBSs, the enhancer can continue to function normally. The only requirement is for the presence of the correct combination of TFBSs somewhere in the enhancer sequence (Fig. 4A). Could this loose requirement be all that is needed to establish a precise response to the complex binding of multiple TFs at a CRM? Hare and colleagues concede that certain binding site clusters may not be so flexible (Hare et al., 2008a). Functionally linked sites, such as the tight clusters at either end of the minimal S2E, may not be subject to an equivalent rate of evolutionary turnover as more independent sites within the CRM. This is certainly the case with the NEEs, where even slight changes in the spacing between binding sites can produce large differences in target gene expression in the same trans environment (Crocker et al., 2008b).

The most parsimonious explanation for these conflicting discoveries is that individual CRMs may in fact be very idiosyncratic. Enhancers may function differently depending upon the TFs by which they are bound and the genes they regulate. This might reflect the unique properties of specific TFs. Different TFs will use distinct molecular interactions to bind to DNA and the mechanisms by which they promote transcription will also vary. Innovative work from the Carroll laboratory on CRMs that regulate yellow gene expression in wing pigmentation in different Drosophila species has demonstrated that enhancers with novel regulatory activity can in fact be generated by modifications to TF binding (Gompel et al., 2005) (reviewed by Prud’homme et al., 2007). In a recent study investigating the evolution of Dorsal target enhancers in Drosophila, the mosquito Anopheles gambiae and the flour beetle Tribolium castaneum, putative orthologous enhancers were found to lack sequence conservation. However, what does appear to be under
evolutionary constraint is the position of these enhancers relative to their target gene. As a result, as these enhancers are modified over evolutionary time, they become capable of directing divergent patterns of gene expression (Cande et al., 2009).

In support of the idea of idiosyncratic CRMs, there are examples of enhancers that do not appear to fit well with the information display/billboard model. In vertebrates, the number of well-characterized enhancers is low and few examples support the information display model perfectly. For example, mammalian sequences with multiple binding sites for the GATA1 TF will continue to function as an enhancer if only a single binding site is intact. However, a very particular site must remain intact: eliminating two of the three GATA1 sites in an enhancer may not have any effect on its function, but the single loss of the third crucial site can completely abolish enhancer activity in mice (Cheng et al., 2008). This example stresses the functional importance of a subset of possible binding sites, although the issue of overall TFBS architecture at mammalian enhancers remains largely unexplored.

Perhaps the best-studied counter-example to the information display/billboard model in vertebrates is the interferon β (IFNB1) enhanceosome. The IFNB1 gene is upregulated in mammals in response to viral infection and the enhancer that directs this expression requires a very specific set of TFs to carry out this regulatory activity (Thanos and Maniatis, 1995). Every TF component is required for the enhanceosome to successfully form and to mediate the upregulation of the IFNB1 gene (Panne et al., 2003) (Fig. 4B). As a result of this highly structured enhancer-TF complex, there has been virtually no evolutionary binding site turnover at this CRM throughout 100 million years of mammalian evolution. The architectural constraints on the Drosophila NEEs may be somewhere between the two extremes presented by the IFNB1 enhanceosome and the S2E orthologs. Although the NEEs do not appear to have significant binding site turnover, these CRMs are able to tolerate slight shifts in TFBS position within the module. Indeed, it seems that evolution may have favored these shifts in response to a changing trans environment between different species. In the case of the NEEs, the binding site architecture of a CRM is able to co-evolve with a changing regulatory environment, demonstrating that molecular mechanisms of enhancer function can be plastic. These recent studies demonstrate that evolution can impact enhancers via a number of distinct mechanisms including, but not limited to, reorganization of the total number, binding affinity and spatial arrangement of TFBSs. These discoveries have increased our understanding of the precise molecular mechanisms of enhancer-mediated transcriptional control for a number of key genes during development.

Conclusions

Future studies now need to ask precisely what it is within the enhancer sequence, or the genomic context, or the TFs themselves, that contributes to the differences in functional activity between enhancers. These questions will be answered largely through a combination of detailed bioinformatic analyses of genomic sequences and functional tests across related eukaryotic species. As our knowledge of TF binding activity during embryogenesis increases, computational tools, including position-weight matrices, can be used to find binding sites that are conserved between species and to compare their distribution in genomic DNA sequences. In vivo functional studies should follow such investigations, guided by rapid cell-based assays to screen potential sequences. Although these types of experiments have been performed on a few enhancers, the next step is to apply the same techniques to the ever-growing array of other known enhancers. In addition, expanding our current catalog of known enhancer CRM orthologs to other insects and mammals/vertebrates would greatly contribute to our understanding.
of the evolution of CRM function. Only by exploring the evolution of a wide range of different enhancers can we hope to understand exactly how these CRMs function and direct the precise patterns of gene expression that give rise to the elegant complexity of embryonic development.

Acknowledgements
Research in our laboratory is supported by funding to R.A.D. from the National Science Foundation and National Institutes of Health and a Howard Hughes Medical Institute Undergraduate Science Education Program grant to the Biology department at Harvey Mudd College. M.J.B. was supported by an Engman award. D.A.T. was supported as an Arnold and Mabel Beckman Foundation Scholar. M.C.W.H. received support from the Merrick-American Association for the Advancement of Science (AAAS) Undergraduate Science Research Program. Deposited in PMC for release after 12 months.

References


