Annu. Rev. Entomol. 2003. 48:579–602
doi: 10.1146/annurev.ento.48.091801.112749

First published online as a Review in Advance on September 25, 2002

ANALYSIS AND FUNCTION OF TRANSCRIPTIONAL REGULATORY ELEMENTS: Insights from *Drosophila*

David N. Arnosti
Department of Biochemistry and Molecular Biology and Program in Genetics, Michigan State University, East Lansing, Michigan 48824-1319; e-mail: arnosti@msu.edu

Key Words transcription, promoter, enhancer, transcriptional regulation

■ **Abstract** Analysis of gene expression is assuming an increasingly important role in elucidating the molecular basis of insect biology. Transcriptional regulation of gene expression is directed by a variety of *cis*-acting DNA elements that control spatial and temporal patterns of expression. This review summarizes current knowledge about properties of transcriptional regulatory elements, based largely on research in *Drosophila melanogaster*, and outlines ways that new technologies are providing tools to facilitate the study of transcriptional regulatory elements in other insects.

CONTENTS

INTRODUCTION

Transcriptional regulation is woven into the entire fabric of biology; thus, knowledge of gene expression is a critical component of understanding many issues relevant to entomology, including pesticide and disease resistance, behavior, evolution, and development. These and other topics have been studied with the tools of molecular biology, allowing the identification of relevant structural genes; however, the control of these genes is often poorly understood. Transcriptional regulation is effected by *cis*-acting DNA sequences that direct the assembly of the protein machinery responsible for transcription. Alterations in these *cis*-acting regulatory sequences are thought to play driving roles in morphological diversification and evolution of developmental mechanisms (106). In addition, quantitative trait loci that contribute to adaptations often correspond to genetic alterations in putative *cis*-acting regulatory elements rather than protein-coding regions of genes (68).

Often the entomologist will be presented with sequences of genomic DNA that may contain transcriptional regulatory sequences, but functional understanding of these DNA elements can be hampered by the lack of genetic and molecular genetic tools. Functional analysis of insect transcriptional regulatory regions is based largely on work in *Drosophila melanogaster*. Fortunately, general properties of eukaryotic transcriptional regulation are highly conserved, especially among metazoans; thus, lessons learned from *Drosophila* can often be directly applied to other insects. The extensive analysis of *Drosophila* transcriptional "wiring diagrams" should provide the basis for accelerated analysis of *cis*-acting elements in other organisms.

This review focuses on the *cis*-acting DNA sequences, rather than the *trans*-acting transcriptional machinery, to assist in answering questions about subjects such as the size of a regulatory region, the possibility of coordinate regulation of closely spaced genes, the nature and importance of basal promoter sequences, and how to predict the transcriptional output of a given promoter. We are still some way from answering all these questions without the use of empirical tests, but consideration of general properties of transcriptional control regions identified in *Drosophila* will allow investigators to identify features that might be important for their system. In addition, new bioinformatic approaches promise to directly "read" transcriptional regulatory information from the genome, at least partly circumventing the need to experimentally determine the function of *cis*-acting sequences. This possibility is especially important to those working in genetically less tractable systems.

For more information on the transcriptional machinery, the reader is directed to recent reviews on RNA polymerase II transcription (58, 80, 84, 116), chromatin and chromatin remodeling machinery (90, 115), transcriptional activators

and repressors (26, 58), boundary elements (4, 36), more-specialized elements such as molecular memory modules (4, 69), and heterochromatin (42).

FEATURES OF TRANSCRIPTIONAL REGULATORY REGIONS

Three general types of *cis*-acting elements control the activity of RNA polymerase II–transcribed (i.e., protein coding) genes (Figure 1): (*a*) basal promoter sequences near the transcriptional initiation site. These elements provide a binding site for RNA polymerase II and the basal transcriptional machinery that acts on most promoters. (*b*) Enhancer elements that contain binding sites for sequence-specific transcription activators and repressors, which regulate levels of gene activity. (*c*) Boundary elements or insulators, which can functionally separate regulatory elements. Rather than acting in separate roles, the three types of elements are functionally interrelated, and these complexities contribute to the specificity of gene regulation. The term "promoter" is often used to indicate all regulatory sequences associated with a gene, including enhancer sequences and the basal promoter. More specialized elements, such as those that interact with Polycomb and Trithorax regulatory proteins, are not thought to be associated with most genes and are discussed in recent reviews (31, 69).

Transcriptional regulatory information can be located entirely within 100 bp of the transcriptional initiation site, as with testis-specific promoters, or in some

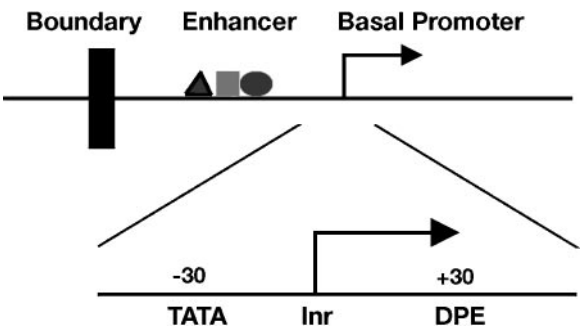

Figure 1 *Cis*-acting transcriptional control elements. Basal promoter includes ~100 bp of sequence flanking the initiation site, onto which general transcriptional machinery loads. Enhancer sequences (ranging in size from 50 bp to several hundred basepairs) are located at variable distances from +1; these bind to sequence-specific transcriptional regulators that control output levels of basal promoter. Boundary elements, found near some genes, insulate the gene from local chromatin effects and can screen the gene from the influence of distal enhancers. The inset shows the disposition of three sequence elements of basal promoters (see text).

cases distributed over a large region, as in the *bithorax* complex, which contains three HOX genes in a 300-kbp region replete with regulatory elements (34, 60). In general, larger regulatory regions are required for diverse tissue-specific and temporal patterns of expression because in most cases individual enhancers are required for separate portions of an expression pattern. Genes dedicated to a single function, such as tissue-specific structural genes, often contain all information for tissue and temporal specificity in a small regulatory element of a few hundred basepairs.

When no other information is available, the region within a few kilobasepairs 5′ of the initiation site is often the first place examined for regulatory information. For many genes, such as those encoding heat shock proteins, glue proteins, some immune-response proteins, cytochrome P450 enzymes, and chorion proteins, it appears that all necessary regulatory information is present within 1 kbp of the basal promoter (30, 57, 59, 78, 107). For a number of important regulatory genes, however, essential transcriptional information is found tens of kilobasepairs from the initiation site (33, 60, 75).

How representative are *cis*-regulatory elements in *Drosophila* for other insects and for metazoans in general? Given the high degree of conservation of the transcriptional machinery in all eukaryotes (2, 58), the ability of regulatory elements to function in heterologous systems (53, 77, 86), and the common occurrence of long-distance regulatory elements in metazoans, it is clear that in most aspects *Drosophila* is an appropriate model system. One area in which this system was not believed to be representative was in DNA methylation. Until recently, *D. melanogaster* was generally thought to lack 5-methylcytosine, an important modified nucleotide in mammalian genomes that functions in transcriptional regulatory phenomena such as imprinting and certain types of repression. DNA methylation in other insects such as aphid, cricket, and mosquito have been previously reported, leading to the question of whether *Drosophila* is atypical in this respect (37, 109). Recent genomic analyses showing the presence of genes homologous to DNA methyltransferase and methyl CpG binding proteins in *D. melanogaster* have prompted a reexamination of this issue (47, 109). Using new, more sensitive techniques, low levels of 5-methylcytosine have been identified, primarily during early stages of embryogenesis (37, 67). The possible functional role of cytosine methylation in *Drosophila* remains unknown, as in other insects; thus, we do not know whether some types of transcriptional regulation via DNA methylation will be accurately modeled in the fly.

BASAL PROMOTER ELEMENTS

A basal promoter can be defined as the ∼100 bp of sequence surrounding the transcriptional initiation site that comes into close contact with the general transcriptional machinery. This region contains TATA sequences (consensus TATAAA) centered at −30, Initiator (Inr) sequences at +1 (consensus TCAGT), and downstream promoter elements (DPE) at +30 (consensus A/GGA/TC/TGT) (Figure 1).

Not all basal promoters contain all three elements. In a survey of 205 promoters with accurately mapped start sites, approximately half had a recognizable TATA sequence, almost half had a DPE, and about one third had neither (56). Initiator-like sequences were found in approximately one quarter of arthropod promoters surveyed (24). Identification of the transcriptional initiation site requires mapping the 5′ end of a transcribed RNA by primer extension, S1 nuclease protection assays, or cloning of cDNAs, especially those isolated based on the mRNA 5′ cap structure (104). Computer identification of basal promoters in genomic sequences has met with limited success, but a current effort to map ∼2000 *Drosophila* transcription start sites from oligo-capped cDNA libraries should provide a much larger basis for development of bioinformatic methods [(81); U. Ohler, personal communication].

The elements of the basal promoter provide nucleation sites for binding by basal transcriptional machinery. The TATA sequence interacts with the TATA binding protein (TBP), a crucial part of the basal transcription machinery that helps anchor the RNA polymerase and basal transcription factors at the promoter. TBP is a subunit of the multicomponent TFIID general transcription factor, which contains about 10 TBP-associated proteins (TAFs) (2). The initiator region contacts TAFs and the RNA polymerase itself (23, 87). The DPE was identified by its contacts with TAF proteins and has been suggested to function as an alternative docking site for the TFIID factor (19).

The functional role of these basal promoter elements appears to be strongly context dependent. In vitro studies in mammalian systems indicated that TATA and Inr elements can be functionally redundant (121), but in vivo analysis of the *hsp70* promoter indicated that TATA function was indispensable and was not redundant with the Inr (117). On this promoter, Inr and DPE have important but largely redundant activities, so mutation of either element individually has little effect (117). In contrast to the *hsp70* promoter, loss of the conserved Initiator element of the TATA-less testis-specific β tubulin promoter has a stronger effect, decreasing transcription by one half (92). On the TATA-less *vermillion* promoter, a TATA element can substitute for a DPE (32), suggesting equivalence of function, but a number of embryonic enhancers tested can discriminate between otherwise identical basal promoters containing either a TATA element or a DPE (20, 83). An additional indication of functional distinctions between basal promoters is that TATA-containing promoters are repressed by the NC2 (Dr1/Drap1) basal transcription factor, whereas DPE-containing promoters are activated by the same factor in in vitro assays (113).

Basal promoter regions are also sometimes bound by sequence-specific transcription factors that play important roles in regulation of individual genes. Binding of the Zn finger Ovo protein to the initiator region of the *otu* gene is important for ovary-specific expression (64). The TATA-less *dpp* core promoter extending from -22 to $+6$ is bound by a sequence-specific factor in the embryo, and this core promoter is by itself sufficient to direct a qualitatively correct pattern of expression (96). GAGA factor binding sites located within the *even-skipped* basal promoter confer enhancer-blocking activities upon this element (82). The

temporal regulation of two promoters of Adh has been linked to direct binding of the inactive promoter by the AEF-1 repressor protein (88). For the most part, it is not known whether sequence-specific factors have the same function when bound to the basal promoter as when they bind to more distal sequences. Often sequence-specific transcription factor binding sites are found close to the basal promoter, but in contrast to the examples cited above, the exact placement of sequence-specific regulators with respect to the transcriptional initiation site is not critical; thus, these binding sites can be considered as enhancer elements.

The general conclusion to be drawn from these studies is that basal promoter sequences and elements cannot be assumed to be generic and interchangeable, a consideration to be taken into account in the design of transgenic vectors. Although in cell culture diverse basal promoters can be functionally interchangeable (51), in some cases the structure of basal promoters can be important for the function of endogenous genes (99) (see Shared or Exclusive Enhancer-Promoter Interactions, below). Alternative forms of TBP protein have been found in *Drosophila*; these proteins are expressed in tissue-specific manners and have alternative sequence specificity, suggesting that alternative basal transcriptional complexes may nucleate on distinct types of basal promoters (110).

BOUNDARY ELEMENTS

Boundary elements or insulators are regulatory regions typically several hundred basepairs in length that bind to sequence-specific factors, establishing a barrier between regulatory elements on one side and basal promoters on the other and insulating genes from repressive effects of heterochromatin-mediated silencing. These elements have been found in both *Drosophila* and vertebrates and play important roles in the regulation of HOX-complex genes in *Drosophila* and imprinted genes in vertebrates. Recent reviews discuss the functional properties of these elements and some of the proteins that interact with them (4, 36). It is clear that the *Fab* and *Mcp* boundary elements of the *bithorax* complex are critical for the proper regulation of this complex locus, separating the activation or repression signals established on distinct portions of the locus (76). However, most genes do not have the complex *cis*-regulatory design of these genes, and it appears that boundary elements do not play a decisive role in the regulation of most genes. Functional studies of these elements have therefore focused on loci into which an element has been inserted by a transposon or by the investigator (4, 36, 97).

The heterochromatic insulating properties of elements bound by the cellular Su(Hw) protein have been exploited to make transformation vectors that are less susceptible to position effects on gene expression (8). Although their effectiveness outside of *Drosophila* has not been demonstrated, boundary-element function is in some cases conserved even between *Drosophila* and vertebrates. It is therefore likely that Su(Hw) insulators will work in other insect systems.

ENHANCERS

Definition of the Term Enhancer and Functional Analysis

Enhancers were originally defined as viral, and later as cellular, DNA sequences that increased expression of a linked gene in an orientation- and distance-independent manner (7, 13). In contemporary usage, an enhancer can refer to any discrete (usually less than 1 kbp) element that binds sequence-specific transcription factors acting in a positive or negative manner. Usually more than one single type of transcription factor binds to an enhancer, but artificial multimerized binding sites for a single factor can act like an enhancer (16). For the purposes of this discussion I consider a functional cluster of transcription factor binding sites as an enhancer, regardless of the distance to the initiation site.

Enhancers have been suggested to function through two distinct pathways: through remodeling of chromatin, thereby facilitating or interfering with binding of the transcriptional machinery, and through direct interactions with the general transcriptional machinery (13). Both mechanisms appear to be important in vivo. The result of these activities can be to trip a gene from an inactive to an active state (on/off effect) or to modulate transcription levels (rheostat effect) (4).

Identification of enhancers was achieved originally through genetic approaches such as classic mutations that altered expression of HOX genes in the *bithorax* complex (60). With the advent of transgenic technology in *Drosophila*, numerous "promoter bashing" experiments allowed the direct identification of regulatory regions by functional analysis (45, 95, 102). Recent progress in bioinformatics and the completion of the *Drosophila* genome has facilitated this work by identifying potential regulatory regions based on conserved clusters of binding sites that in some cases are evolutionarily conserved. Putative factor binding sites have been traditionally identified through biochemical approaches, such as DNAseI protection or gel-mobility shift assays (122). Functional in vivo assays, in which individual elements are tested and the effects of mutating binding sites are assayed, are still required to verify these predictions, however.

Alternative approaches have been used, especially where transgenic methods are impractical, including transfection of reporter constructs in cell culture. These experiments have been useful in functional analysis for some types of compact regulatory elements, such as those associated with heat shock, cytochrome P450, and vitellogenin genes (30, 57, 72). This approach is less satisfactory with promoters that require cell-type-specific factors or input from signal transduction pathways. In vitro transcriptional assays have also been extensively employed, but as with cell culture assays, in vitro transcription assays are more successful in modeling the potential activities of a single factor rather than the activity of a complex enhancer. Less commonly, regulatory regions have also been studied using biolistic transformation of insect tissues, electroporation, and introduction of genes on viral vectors [(54) and references therein]. With the development of efficient broad-spectrum transformation systems, analysis of *cis*-regulatory regions

in transgenic organisms should be the benchmark for understanding enhancer function (5).

Models of Enhancer Activity

Two distinct models of enhancer action have been proposed to ascribe different computational functional roles to the enhancer (Figure 2). In the first, the "enhanceosome" model, the arrangement of binding sites within the enhancer is critical to dictating the correct output of the element, so the enhancer acts as a molecular computer, leading to a single output directed to the general machinery (25, 103). In the best-studied example, the regulatory element controlling the human β interferon promoter, cooperative interactions between architectural proteins and sequence-specific activators assemble a monolithic complex that subsequently recruits coactivators required to fire the promoter (74) (Figure 2*A*). In the second model the enhancer acts as an information display, or "billboard," which is then read and interpreted by consecutive interactions with the basal machinery. In the case of a billboard enhancer, exact binding site locations are less critical, and both activating and repressing states can be represented at the same time within an enhancer (55). It is not cooperative assembly of a higher-order structure, but successive interactions with the basal transcriptional machinery, and the biochemical consequences of these multiple interactions, that dictates the output of the enhancer

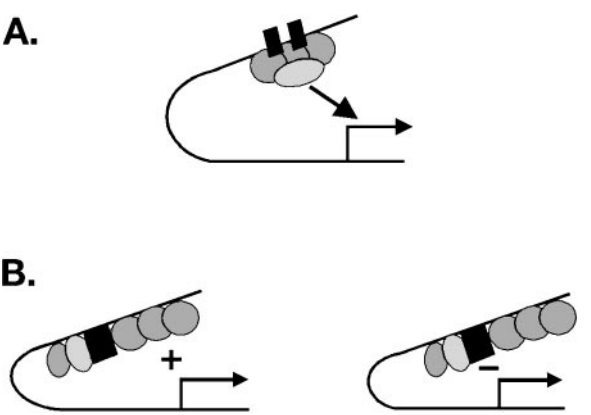

Figure 2 Alternative models of enhancer action. (*A*) "Enhanceosome" complex of carefully positioned, cooperatively bound transcriptional regulators provides a single output to the basal promoter. (*B*) "Billboard" or information display enhancer contains loosely constrained transcription factor sites binding proteins that interact with the basal machinery in a variety of conformations. The subelements of the enhancer can represent conflicting signals, which are deciphered by consecutive interactions with the basal machinery. (*Left*) Activators send positive output to promoter, resulting in transcriptional initiation; (*right*) negative signal output from a repressor.

(Figure 2*B*). The billboard enhancer model appears to more accurately describe many developmentally active enhancers, whose exact composition of binding sites is subject to rapid change in evolutionary time, even as the overall output remains constant (65, 66). Considering these functional differences, a distinction should be made between the terms *enhanceosome*, which implies important cooperative assembly processes within the enhancer-bound proteins, and *enhancer*, which may or may not function in this manner.

Many enhancers activated by signal transduction pathways require more than one type of activator to achieve correct tissue specificity. The requirement for multiple signal inputs is suggested to reflect a common enhancer design scheme, whereby a field of cells is made competent for expression by the presence of a widely expressed activator (e.g., field-selector gene) that by itself is a weak activator (9). Only in those nuclei receiving an input from a signal transduction cascade are additional activating proteins bound to the enhancer, allowing the gene to fire. To achieve greater specificity, in the absence of this signal corepressors bind in place of the coactivators, suppressing the weak activity of the widely expressed general activator. In the case of the Notch signaling pathway, the DNA binding protein Su(H) can interact either with a fragment of Notch as a coactivator or with Hairless as a corepressor (9). Synergistic interactions with the widely expressed Scalloped/Vestigial activator proteins provide a strong transcriptional output. The exact positioning of these transcription factors within a native enhancer is not critical, for an arbitrary arrangement of two binding sites each for Su(H) and Sd is sufficient to mimic the expression pattern of a native element (39).

Pattern Generation by Enhancers

Enhancers can act as information integrators or simply passively replicate a particular pattern of gene expression established previously (4). It is difficult to predict a priori whether tissue and temporal specificity of transcription is associated with a large, complex regulatory element or a simple element. In the case of the *eve* pair ruled gene, the iterated seven-stripe blastoderm expression pattern reflects the integration of positional and temporal information by five separable elements comprising 4 kbp of DNA bound by over seven distinct types of transcriptional regulator (33, 101). The large amount of *cis*-regulatory sequence is required to program cells as either *eve* expressing or *eve* nonexpressing, an essential step in the establishment of segmentation. In contrast, after cells have differentiated, expression of tissue-specific factors can recapitulate this tissue-specific pattern by binding to simple *cis* elements. For example, small autoregulatory elements in *eve* and *ftz* are sufficient to drive the multistripe pattern in later embryos, in response to earlier stripe patterning, and a short element containing multimerized Pax6/eyeless factor sites is sufficient to drive gene expression in an eye-specific pattern (49, 94, 95). In experimental settings, precise tissue-specific gene expression can be achieved using transgenes containing five binding sites for the yeast Gal4 activator, which is expressed under the control of tissue-specific enhancers (16).

The biochemistry of gene activation does not require more than one single type of transcription factor; Pax6/eyeless and Gal4 alone can drive robust expression. More typically, though, several proteins are involved in regulation, even on small tissue-specific transcriptional switches. A minimal regulatory element from the *yp1/2* ovarian-specific enhancer contains binding sites for four distinct proteins within ~40 bp: a tissue-specific activator that synergizes with a ubiquitously expressed activator, a constitutive repressor that represses the gene in nonovarian tissues, and a female-specific activator that displaces the repressor (1). A 125-bp larval fat body–specific element for the *Fbp1* gene contains a timing element in the form of an ecdysone receptor binding site that permits expression in response to proper levels of hormone, an AEF-1 repressor binding site that keeps the gene off in nonfat body tissues, and redundant GATAb protein sites that provide both activation and antirepression activities (17).

These minimal elements often represent only the pattern-specific part of the entire control region, with other sequences playing a role in overall signal amplification. In one case, the proximal element of the mosquito vitellogenin regulatory region binds to four distinct activators and is sufficient to drive low-level expression of the element in the correct stage and tissue-specific manner, but upstream sequences with additional factor binding sites are required to achieve wild-type levels of expression (53). Similar regulatory designs with separate specificity/signal amplification modules have been described in the sea urchin (3).

Shared or Exclusive Enhancer-Promoter Interactions

Enhancers can activate basal promoters over long distances, thus raising the question of how regulatory information is targeted to the correct gene. In some cases, regulatory sequences lie between divergently transcribed genes, which may share *cis*-regulatory elements. In other cases, regulation may be selective. An example of shared regulation is seen in the case of the divergent yolk protein 1 and 2 genes (*yp1* and *yp2*), which are separated by 1225-bp-containing fat body and ovary-specific enhancers (15). Both genes are simultaneously fired in each tissue by the same regulatory elements. In some cases in which an enhancer engages multiple promoters, promoter competition occurs, but in the case of the *yp1* and *yp2* genes, access to enhancers does not appear to be rate limiting; rather, a poorly understood promoter dependence was noted, in which damage to one basal promoter reduced the activity of the other (97).

Divergently transcribed genes may also share regulatory sequences, but in a temporally separated fashion. The divergently transcribed *Pig1* and *Sgs4* genes are expressed in the larval salivary gland in a sequential fashion, with shut-off of *Pig1* transcription accompanying upregulation of the *Sgs4* gene (59, 79). These genes are separated by a 838-bp region that contains multiple factor binding sites for activators and at least one inhibitory factor. One cluster of activators is capable of activating both genes simultaneously, but an additional element allows transcriptional activity to be directed to the Pig1 promoter until mid third instar (79).

Then transcriptional activity is focused on *Sgs4* until an ecdysone-regulated loss of one activator protein inhibits *Sgs4* expression (89).

Some divergently transcribed genes do not share regulatory information at all. Such enhancer-promoter specificity can be dictated by the elements of the basal promoter, including TATA, DPE, and Inr elements. Distinct types of basal promoters allow enhancers near the *dpp* and *gsb* genes to selectively activate the correct promoter and not neighboring genes (62, 75). A molecular basis of such specificity has been suggested from studies of the AE-1 enhancer, which normally activates only the *ftz* promoter and not the divergent *scr* promoter. This enhancer's specificity appears to be dependent on the presence of a TATA sequence within the activated gene (83). A recent transgenic comparison of DPE- or TATA-containing basal promoters identified endogenous enhancers that activated only DPE-containing promoters, as well as enhancers that activated only a TATA-containing promoter. A majority of the enhancers tested activated both genes, however, suggesting that most enhancers activate disparate promoters (20).

In addition to basal promoter specificity, restriction of enhancer activity to one promoter over another might also be effected by placement of a boundary element between the two genes. Experimentally, this has been observed for the *yp1* and *yp2* genes using Su(Hw) elements (97). In another case, boundary element–like activity mediated by the *eve* basal promoter was also observed (82). Aside from these experimental situations, it is not known whether boundary elements are frequently utilized to screen promoters from unwanted enhancer attention. In the best understood cases of biological activity of boundary elements in the *bithorax* complex, the activity of these elements seems critical to separate differentially regulated enhancers (76).

Enhancer Redundancy

The billboard model of enhancers suggests that there are multiple configurations in which enhancer-bound factors can interact with the basal transcriptional machinery. Consistent with this model, functional analysis of many *cis*-regulatory elements suggests that redundant regulatory information is represented in discrete, separable sequences. In the case of yolk protein genes, for example, the 125-bp enhancer that gives fat body activity can be deleted from the upstream region without measurably affecting expression, and a deletion that removes most of the intergenic region only had subtle effects on expression (86, 97). *Krüppel* expression in the central domain of the embryo is directed by two adjacent enhancers that each independently drive the correct pattern of gene expression (45). In the case of the *eve* stripe 2 enhancer, removal of activator binding sites in the context of a 5.2-kbp element had weak and variable effects in contrast to the more drastic effects noted when a minimal stripe 2 element was assayed (101, 102). A similar redundancy in enhancer design was noted with enhancers that drive the blastoderm stripe pattern of expression of *ftz*, where mutations in individual activator sites had little measurable effect on activity, although mutation of multiple

sites abolished activity (40). Repressor binding sites also function in a redundant manner, as in the case of multiple Brinker binding sites within the *zen* promoter (91). Regulatory mutations that involve a single protein binding site are in general rarely detected, suggesting that most enhancers are built to withstand a fair degree of change without suffering catastrophic loss of activity. In addition to the likely functional redundancy built into many systems, our inability to measure subtle changes in transcriptional output of altered elements may also explain the apparent robustness of *cis*-regulatory elements to alteration (105).

In contrast to this redundancy, some complex regulatory regions do show strong dependence on a few sites. The regulatory region of the *diptericin* antibacterial gene, for instance, contains conserved binding sites for Rel-type transcriptional activators (73). Although the sites are not sufficient to confer activation, loss of two of these sites in a 2.2-kbp enhancer abolished induction of a reporter in response to bacterial challenge, consistent with an enhanceosome type of model (73). Both insect and vertebrate genes utilize Rel domain proteins in innate immune responses, and enhancers such as this that are designed for sudden inducibility might tend to function as enhanceosomes (103).

SOME SPECIAL PROMOTERS

Modular Enhancers and Short-Range Repression

Many developmentally regulated genes have independently acting modular control elements. An initial indication of the modularity of transcriptional regulatory regions came from the identification of mutant alleles of the *hairy* gene, which is normally expressed in a seven-stripe pattern in the embryo (46). Deletions upstream of the structural gene caused a loss of individual stripes rather than a general effect on all regions of expression. Subsequent dissection of *hairy* regulatory regions revealed the presence of individual modular enhancer elements, each of which is subject to the action of activators and repressor proteins, a finding that has been replicated with many other genes, including *even-skipped* (33, 101). One surprising finding from studies of *eve* enhancers is that in an individual nucleus at a given moment, one enhancer can be repressed while an adjacent enhancer actively signals the promoter. The reason for this functional independence is that the silenced enhancers are repressed by proteins that work over a short range ($<$100 bp), so distances between enhancers of a few hundred bases are sufficient to prevent improper cross-regulation (38). Artificially juxtaposing two normally distant enhancers causes improper repression of one element by repressors bound to the other (100). Repressors that act over long distances have also been identified, but it is not yet clear how their activities are integrated into modular control regions (26, 38).

Compact Testis-Specific Promoters

Unlike the uniformly compact promoters associated with RNA Pol I and III transcription units, RNA Pol II–regulated promoters active in most tissues are variable

in size. An exception to this observation is formed by a number of testis-specific genes with unusually small transcriptional control elements, often extending no more than 100 nt 5′ of the initiation site. The expression of the testis-specific $\beta 2$ tubulin gene depends on a 14-bp element at -40 that contributes tissue specificity and on two small elements just 5′ and 3′ of $+1$ that are important for overall activity of the promoter (92). As is the case with other testis-specific genes, the nature of the proteins that might interact with these functional motifs is not known, although binding activities have been identified in gel-shift assays (92). Like other testis-specific promoters, the $\beta 2$ tubulin gene contains a conserved Inr but lacks a TATA box. The 14-bp element does not activate a heterologous TATA-containing basal promoter, suggesting that a distinct type of basal machinery architecture might be used on this class of gene. This notion is supported by the finding that the testis-specific TAFII80 and TAFII110 homologs encoded by *cannonball* and *no-hitter*, respectively, are required for wild-type expression of a number of testis-specific genes (2, 44). Cannonball and No-hitter may function as part of a tissue-specific form of TFIID or, like other TAFII's, they may function as part of a histone acetylation complex (90, 115).

In some cases, the close spacing of testis-specific genes would indicate small *cis*-acting regions. For example, a cluster of four adenylyl cyclase genes had intergenic spacings of only 166, 84, and 39 bp (21). However, it is important to consider that some regulatory elements map within transcribed regions or within the sequence of upstream genes. In $\beta 2$ tubulin and the H1-like *don juan* genes, there are transcriptional control elements immediately downstream of the initiation site, overlapping sequences that play a role in translational control of the mRNA (14, 92). In some cases, control elements are located within transcribed regions of upstream genes, such as the 53-bp element responsible for male-specific transcription of *gonadal* found at -330 bp within the coding sequence of an upstream gene not expressed in male germline and a control element for male-specific transcription of *Janus B* located within an exon of the upstream *Janus A* gene (93, 119).

Promoters in Heterochromatin

Most protein-coding genes are located in euchromatic regions, but a small number of genes such as *light*, *rolled*, and *concertina* are embedded in heterochromatin, chromosomal regions with a more compact structure and rich in repetitive sequences (42, 111). Little is known about the transcriptional regulation of such genes, except that in contrast to most genes, which are repressed by proximity to heterochromatin, these are fully functional, and in the case of *light* are positively regulated by heterochromatin (41). The *D. melanogaster light* gene has an unusual pattern of transcriptional initiation, with start sites spread over a region of $\sim$200 nucleotides. In *D. virilis*, by contrast, in which the *light* gene is located in euchromatin, transcription initiates at a single site. When introduced into *D. melanogaster*, the *D. virilis* gene can rescue a *light* mutant, which indicates that the overall regulation of the genes has been conserved despite the differences in

chromatin context and basal promoter function (120). Future work should identify the specific regulatory features that allow this gene to function in a heterochromatic context.

runt: A Disperse Regulatory Element

Similar to *eve* and *ftz*, initial embryonic patterning of the *runt* gene is under control of individual, modular stripe enhancers. Unlike *eve* and *ftz*, however, later expression of *runt* is not regulated by a compact autoregulatory element, but by a large region, over 5 kbp in size, that cannot be subdivided into independent elements (52). This large regulatory element appears to be unusual, but traditional approaches to promoter analysis have concentrated on those elements that are compact (<1 kbp); thus, the true frequency of disperse elements may be higher than generally appreciated. Evolutionary conservation of this disperse *runt* element is low, suggesting that interacting functional subelements might be widely distributed within this regulatory region (114).

EVOLUTION OF *CIS*-REGULATORY ELEMENTS

The rates of change in transcriptional regulatory regions during evolution can differ markedly from those in protein-coding sequences. In some cases the sequences of transcriptional regulatory regions are considerably altered, yet the transcriptional output remains the same, indicating that different arrangements of transcription factors can have the same function (65, 66). The constraints on regulatory regions are considerably more relaxed than those on protein-coding regions, probably owing in part to the redundant, flexible nature of transcriptional regions. In other cases, transcriptional regulatory regions themselves appear to be the driving force behind evolutionary changes, as for instance in the altered expression patterns of HOX genes correlated with diverse limb development schemes in arthropods (6). Duplicated genes can diverge in function solely on the basis of transcriptional regulation, as demonstrated in the case of the homologous transcription factors Paired, Gooseberry, and Gooseberry neuro. These proteins have essentially identical activities but differ in developmental roles because of their genes' distinct transcriptional control regions (63).

Functional Analysis

Conservation or divergence in transcriptional control regions has been tested both functionally and by sequence analysis. One of the first demonstrations of conservation of transcriptional regulation between different insect orders was the finding that a chorion gene cluster from *Bombyx mori* is correctly regulated in *Drosophila* in a sex- and tissue-specific manner, despite the great evolutionary distance between lepidopterans and dipterans, the lack of homology between the chorion

genes of the two species, and the distinct gene arrangement (77). Subsequent tests have shown that the vitellogenin gene from the mosquito *Aedes aegyti* is regulated in the correct stage and tissue-specific manner in *Drosophila* (although not in a female-specific manner), despite the great differences in yolk proteins and vitellogenesis between these organisms (53). Other heterologous genes correctly regulated in *Drosophila* include gut-specific protease genes from the mosquito *Anopheles gambiae* and the black fly *Simulium vittatum* [(98, 118) and references therein].

Sequence Analysis

Given these often remarkable similarities in the function of regulatory regions from divergent organisms, is it possible to identify similar sequences by computer alignment? The picture is mixed. A comparison of yolk-protein regulatory regions among *D. melanogaster* and Hawaiian drosophilids demonstrated that there have been extensive changes in the short region between the divergently transcribed genes, so much so that there was no clear overall alignment between previously mapped enhancer elements in *D. melanogaster* and similar areas in *D. grimshawi* (86). Only short, interspersed sequences were found, but transgenic analysis confirmed that these sequences were functionally conserved (86). The extent of divergence between yolk protein genes may be an extreme, however, because of selective evolutionary pressures on vitellogenesis, which leads to subtle changes in regulation that are not readily evident in an experimental setting.

Much closer similarities were noted in an analysis of enhancers of the conserved pair rule gene *ftz* in *D. hydei*. In sequence alignments with previously characterized enhancers from *D. melanogaster*, conserved blocks hundreds of basepairs in size with 50%–65% overall identity were noted. These enhancers gave a *ftz*-type expression pattern in *D. melanogaster* (50, 70). A recent comparative analysis of 100 kbp of known or suspected regulatory DNA from 40 loci between *D. melanogaster* and *D. virilis* (∼40 million years of separation, similar to *D. hydei*) found that overall about 25% of the sequence was conserved, with most of the conserved sequences present as small blocks of median length of 19 bp, and a substantial fraction up to 40–60 bp in length (11). Thus, even between divergent drosophilids, there appears to be sufficient conservation to identify regulatory blocks in many cases and to identify divergent features of enhancer design. For instance, comparison of the *even-skipped* stripe 2 enhancer from an array of *Drosophila* species revealed that overall organization and general size of this enhancer was maintained and many binding sites were conserved (66). A specific binding site for the Bicoid activator that played an important role in *D. melanogaster* was not conserved, however. A recent deletion appears to have brought a nearby Giant repressor site closer to this recently acquired Bicoid activator site, an example of the "tuning" of enhancer output (43). Thus, whereas both enhancers have the same overall output when

tested in *D. melanogaster*, a chimeric enhancer containing half *D. pseudoobscura* and half *D. melanogaster* sequences gave improper regulatory output, evidence of changes in enhancer design (65).

The greater degree of divergence between regulatory regions in more distantly related species complicates efforts to locate enhancers by sequence alignment alone, even when functional aspects are similar. In some cases, lack of conservation reflects a fundamental change in regulatory strategy. An 8.7-kbp fragment 5′ of *hairy* from the beetle *Tribolium casteneum* drives a portion of the conserved striped pattern in central regions of *D. melanogaster*, but there are no obvious conserved regulatory regions. A more detailed functional analysis of individual conserved sequence clusters might reveal conserved functional elements, but the failure to recapitulate some portions of the stripe pattern in the fly is probably due to fundamental differences in anterior and terminal patterning in *Tribolium* and *Drosophila* (29). Preliminary analysis of putative regulatory elements in the *Tribolium* HOX complex reveals a similar lack of conservation between these two species, suggesting that with current approaches direct functional testing will remain critical for identification and analysis of possibly conserved regulatory regions (18).

NEW TECHNOLOGIES

Genomics

Our current understanding of transcriptional switch elements derives from empirical experiments involving the dissection of basal promoters and enhancers and from in vivo tests. The most powerful studies, such as those conducted on complex regulatory regions such as the *bithorax* complex and *even-skipped*, are initiated with genetic characterization of *trans*-activating factors, identification of *cis*-regulatory elements, mapping of factor binding sites within those regions, mutagenesis of those putative binding sites, and testing of the elements in transgenic organisms. This approach requires good genetic and transgenic tools; thus, most studies of this type have been restricted to *Drosophila*.

Applications of genomics and bioinformatics associated with the *Drosophila* genomic project, as well as more extensive phylogenetic comparisons and advances in transgenesis, will facilitate the analysis of *cis*-regulatory elements, enabling work to proceed in other, less genetically tractable systems. Comprehensive analysis of gene expression in *Drosophila* is being conducted by gene array analysis and in situ hybridization. Gene arrays have been used to track expression on an organismal level and more recently in a tissue-specific basis in both wild-type and mutant backgrounds (48, 61). As examples, recent gene array analysis of the *Drosophila* immune response, mesoderm-specific genes, and metamorphosis will provide a baseline set of data for comparison in other species (27, 35, 112). A systematic analysis of embryonic expression patterns by in situ hybridization now underway at the Berkeley Drosophila Genome Project will add a spatial component

to the temporal information gained from gene array studies (108). From these studies we can expect to develop a comprehensive knowledge of the *Drosophila* transcriptome, including genetic pathways. This large body of empirical data will give researchers models for gene regulation by which to measure homologs in other insect species, possibly revealing conserved or divergent features of regulatory elements or expression patterns.

Bioinformatics

Bioinformatic analysis can be used to identify binding sites of known or novel factors. This identification is carried out in two ways. For factors with well-characterized binding sites, a consensus binding site, or a position weight matrix representing a population of known binding sites can be used to search genomic sequences. Relatively simple search paradigms, such as searching for clusters of a binding motif within 400 bp, have been successful in producing useful information. A genomic search for clustered Dorsal protein binding sites identified 15 putative Dorsal-regulated enhancers, some of which were assayed by transgenic analysis (71). A search for clusters of activator and repressor binding sites in the 1-Mbp region surrounding the *even-skipped* gene successfully identified previously functionally mapped transcriptional control elements flanking *eve* and identified a novel enhancer that proved to be a control element for the *giant* gene (12).

When the *trans*-acting factors for genes are not known, one can search for overrepresented motifs near genes that show similar expression patterns. Such an approach was successfully used to identify repeated motifs common to eight odorant receptor proteins expressed primarily in maxilla. Mutagenesis of the regulatory regions confirmed that the elements are important for the tissue-specific regulation of these genes (22). Alternatively, one can search a known regulatory region using statistical methods for motifs that are overrepresented; a recent study showed that this technique located sites of known regulatory importance in *Drosophila* enhancers (85).

Phylogenetic Analysis

Phylogenetic comparisons are powerful tools to identify regions of genomic DNA that likely contain regulatory information. The imminent or completed sequencing of additional dipteran genomes, such as that of *D. pseudoobscura* and *An. gambiae* will facilitate analysis of *cis*-acting control regions on a global scale, so the putative regulatory regions of any conserved gene will be available for bioinformatic and functional analysis. It remains to be seen how close the conservation of regulatory regions will be in more evolutionarily distant insects and whether the genetic control networks are sufficiently similar to allow direct modeling of entire enhancers from dipterans. At least at the level of individual factor binding sites, genes with similar regulatory profiles, such as those encoding rhodopsins, antimicrobial

peptides, and cytochrome P450s, appear to use the same transcription factors; thus, at least short motifs should be conserved (10, 27, 30).

Transgenesis in Other Systems

Genomic and bioinformatic tools can accelerate identification and characterization of *cis* elements, but models of transcriptional regulatory regions still require validation with functional tests. Recent advances in insect transgenesis (5) offer the possibility of testing the function of putative regulatory regions in most species. The lepidopteran PiggyBac vector has been reported to successfully transform many different insect orders, including Coleoptera, Lepidoptera, and Diptera (10, 29). The use of eye-specific green fluorescent protein marker genes to identify transgenic progeny allows transformation of wild-type animals, obviating the need to identify mutant strains with phenotypes that can be rescued. In addition to studying regulatory elements, transgenesis has the potential to generate mutant phenotypes using RNA interference (28), allowing genetic analysis in nonmodel system insects. A future paradigm for characterizing the regulation of a novel gene in a nonmodel insect system would be identification of possible homologs in a completed genome, examination of possible evolutionarily conserved regulatory regions and binding sites for *trans*-acting factors, experimental testing of *cis* elements on transformation vectors, and disruption of putative regulatory factors through RNA interference.

CONCLUDING REMARKS

Genome-wide analysis of transcriptional programs in *Drosophila* will contribute greatly to our understanding of basic transcriptional mechanisms used in all metazoans by providing for the first time a comprehensive overview of all transcription factors, a genome-wide comparison of transcriptional elements, and their conservation in disparate species. Comparisons with *Drosophila* do not eliminate the requirement for direct functional analysis of transcriptional elements in species of interest, but new advances will facilitate the design of these experiments, and work in previously nonmodel organisms will undoubtedly make important contributions to basic science. A better understanding of insect *cis*-acting regulatory regions will also contribute to many areas relevant to entomology, including identification of useful promoters for engineering resistance in disease vectors, identifying the molecular basis of pesticide resistance, and understanding evolution mediated by changes in transcriptional control elements.

ACKNOWLEDGMENTS

Special thanks to S. Brown, T. Ip, B. Lemaitre, U. Ohler, A. Raikhel, D. Tautz, and B. Wakimoto for helpful suggestions and sharing information prior to publication.

The *Annual Review of Entomology* is online at http://ento.annualreviews.org

LITERATURE CITED

1. An W, Wensink PC. 1995. Integrating sex- and tissue-specific regulation within a single *Drosophila* enhancer. *Genes Dev.* 9:256–66
2. Aoyagi N, Wassarman DA. 2000. Genes encoding *Drosophila melanogaster* RNA polymerase II general transcription factors: diversity in TFIIA and TFIID components contributes to gene-specific transcriptional regulation. *J. Cell Biol.* 150: F45–50
3. Arnone MI, Davidson EH. 1997. The hardwiring of development: organization and function of genomic regulatory systems. *Development* 124:1851–64
4. Arnosti DN. 2002. Design and function of transcriptional switches in *Drosophila*. *Insect Biochem. Mol. Biol.* 32(10):1257–73
5. Atkinson PW, Pinkerton AC, O'Brochta DA. 2001. Genetic transformation systems in insects. *Annu. Rev. Entomol.* 46: 317–46
6. Averof M, Patel NH. 1997. Crustacean appendage evolution associated with changes in *Hox* gene expression. *Nature* 388:682–86
7. Banerji J, Rusconi S, Schaffner W. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27:299–308
8. Barolo S, Carver LA, Posakony JW. 2000. GFP and beta-galactosidase transformation vectors for promoter/enhancer analysis in *Drosophila*. *Biotechniques* 29:726–32
9. Barolo S, Posakony JW. 2002. Three habits of highly effective signaling pathways; principles of transcriptional control by developmental cell signaling. *Genes Dev.* 16:1167–81
10. Berghammer AJ, Klingler M, Wimmer EA. 1999. A universal marker for transgenic insects. *Nature* 402:370–71
11. Bergman CM, Kreitman M. 2001. Analysis of conserved noncoding DNA in *Drosophila* reveals similar constraints in intergenic and intronic sequences. *Genome Res.* 11:1335–45
12. Berman BP, Nibu Y, Pfeiffer BD, Tomancak P, Celniker SE, et al. 2002. Exploiting transcription factor binding site clustering to identify *cis*-regulatory modules involved in pattern formation in the *Drosophila* genome. *Proc. Natl. Acad. Sci. USA* 99:757–62
13. Blackwood EM, Kadonaga JT. 1998. Going the distance: a current view of enhancer action. *Science* 281:60–63
14. Blümer N, Schreiter K, Hempel L, Santel A, Hollmann M, et al. 2002. A new translational repression element and unusual transcriptional control regulate expression of *don juan* during *Drosophila* spermatogenesis. *Mech. Dev.* 110:97–112
15. Bownes M. 1994. The regulation of the yolk protein genes, a family of sex differentiation genes in *Drosophila melanogaster*. *BioEssays* 16:745–52
16. Brand AH, Perrimon N. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118:401–15
17. Brodu V, Mugat B, Fichelson P, Lepesant JA, Antoniewski C. 2001. A UAS site substitution approach to the in vivo dissection of promoters: interplay between the GATAb activator and the AEF-1 repressor at a *Drosophila* ecdysone response unit. *Development* 128:2593–602
18. Brown SJ, Fellers JP, Shippy TD, Richardson EA, Maxwell M, et al. 2002. Sequence of the *Tribolium castaneum* homeotic complex. The region corresponding to the *Drosophila melanogaster* Antennapedia complex. *Genetics* 160:1067–74
19. Burke TW, Willy PJ, Kutach AK, Butler JE, Kadonaga JT. 1998. The DPE,

a conserved downstream core promoter element that is functionally analogous to the TATA box. *Cold Spring Harbor Symp. Quant. Biol.* 63:75–82
20. Butler JE, Kadonaga JT. 2001. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev.* 15:2515–19
21. Cann MJ, Chung E, Levin LR. 2000. A new family of adenylyl cyclase genes in the male germline of *Drosophila melanogaster*. *Dev. Genes Evol.* 210:200–6
22. Carlson J. 2002. *Odor and taste receptors: genetics and e-genetics.* Presented at Annu. Drosophila Res. Conf., 43rd, San Diego, CA
23. Chalkley GE, Verrijzer CP. 1999. DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *EMBO J.* 18:4835–45
24. Cherbas L, Cherbas P. 1993. The arthropod initiator: the capsite consensus plays an important role in transcription. *Insect Biochem. Mol. Biol.* 23:81–90
25. Courey AJ. 2001. Cooperativity in transcriptional control. *Curr. Biol.* 11:R250–52
26. Courey AJ, Jia S. 2001. Transcriptional repression: the long and the short of it. *Genes Dev.* 15:2786–96
27. De Gregorio E, Spellman PT, Rubin GM, Lemaitre B. 2001. Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. *Proc. Natl. Acad. Sci. USA* 98:12590–95
28. Denell R, Shippy T. 2001. Comparative insect developmental genetics: phenotypes without mutants. *BioEssays* 23: 379–82
29. Eckert C, Wolff C, Wimmer E, Tautz D. 2002. Functional analysis of the regulatory region of the pair-rule gene *hairy* in *Tribolium* suggests regulatory divergence in spite of conserved expression. Submitted
30. Feyereisen R. 1999. Insect P450 enzymes. *Annu. Rev. Entomol.* 44:507–33
31. Francis NJ, Kingston RE. 2001. Mechanisms of transcriptional memory. *Nat. Rev. Mol. Cell. Biol.* 2:409–21
32. Fridell YW, Searles LL. 1992. In vivo transcriptional analysis of the TATA-less promoter of the *Drosophila melanogaster vermilion* gene. *Mol. Cell. Biol.* 12:4571–77
33. Fujioka M, Emi-Sarker Y, Yusibova GL, Goto T, Jaynes JB. 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
34. Fuller MT. 1993. Spermatogenesis. In *The Development of* Drosophila melanogaster, ed. M Bate, A Martinez Arias, 1:71–147. New York: Cold Spring Harbor Press
35. Furlong EE, Andersen EC, Null B, White KP, Scott MP. 2001. Patterns of gene expression during *Drosophila* mesoderm development. *Science* 293:1629–33
36. Gerasimova TI, Corces VG. 2001. Chromatin insulators and boundaries: effects on transcription and nuclear organization. *Annu. Rev. Genet.* 35:193–208
37. Gowher H, Leismann O, Jeltsch A. 2000. DNA of *Drosophila melanogaster* contains 5-methylcytosine. *EMBO J.* 19: 6918–23
38. Gray S, Levine M. 1996. Transcriptional repression in development. *Curr. Opin. Cell. Biol.* 8:358–64
39. Guss KA, Nelson CE, Hudson A, Kraus ME, Carroll SB. 2001. Control of a genetic regulatory network by a selector gene. *Science* 292:1164–67
40. Han W, Yu Y, Su K, Kohanski RA, 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
41. Hearn MG, Hedrick A, Grigliatti TA, Wakimoto BT. 1991. The effect of modifiers of position-effect variegation on

the variegation of heterochromatic genes of *Drosophila melanogaster*. *Genetics* 128:785–97
42. Henikoff S. 2000. Heterochromatin function in complex genomes. *Biochim. Biophys. Acta* 1470:1–8
43. Hewitt GF, Strunk BS, Margulies C, Priputin T, Wang X-D, et al. 1999. Transcriptional repression by the *Drosophila* giant protein: *cis* element positioning provides an alternative means of interpreting an effector gradient. *Development* 126:1201–10
44. Hiller MA, Lin TY, Wood C, Fuller MT. 2001. Developmental regulation of transcription by a tissue-specific TAF homolog. *Genes Dev.* 15:1021–30
45. Hoch M, Schröder C, Seifert E, Jäckle H. 1990. *cis*-acting control elements for *Krüppel* expression in the *Drosophila* embryo. *EMBO J.* 9:2587–95
46. Hooper KL, Parkhurst SM, Ish-Horowicz D. 1989. Spatial control of hairy protein expression during embryogenesis. *Development* 107:489–504
47. Hung MS, Karthikeyan N, Huang B, Koo HC, Kiger J, Shen CJ. 1999. *Drosophila* proteins related to vertebrate DNA (5-cytosine) methyltransferases. *Proc. Natl. Acad. Sci. USA* 96:11940–45
48. Imam FB, Johnson EJ, Arbeitman M, Furlong E, Null B, et al. 2002. *A transcriptional view of* Drosophila *development.* Presented at Annu. Drosophila Res. Conf., 43rd, San Diego, CA
49. Jiang J, Hoey T, Levine M. 1991. Autoregulation of a segmentation gene in *Drosophila*: combinatorial interaction of the *even-skipped* homeo box protein with a distal enhancer element. *Genes Dev.* 5: 265–77
50. Jost W, Yu Y, Pick L, Preiss A, Maier D. 1995. Structure and regulation of the *fushi tarazu* gene from *Drosophila hydei*. *Roux's Arch. Dev. Biol.* 205:160–70
51. Kermekchiev M, Pettersson M, Matthias P, Schaffner W. 1991. Every enhancer works with every promoter for all the combinations tested: could new regulatory pathways evolve by enhancer shuffling? *Gene Exp.* 1:71–81
52. Klingler M, Soong J, Butler B, Gergen JP. 1996. Disperse versus compact elements for the regulation of *runt* stripes in *Drosophila*. *Dev. Biol.* 177:73–84
53. Kokoza VA, Martin D, Mienaltowski MJ, Ahmed A, Morton CM, Raikhel AS. 2001. Transcriptional regulation of the mosquito vitellogenin gene via a blood meal-triggered cascade. *Gene* 274:47–65
54. Kravariti L, Thomas J, Sourmeli S, Rodakis GC, Mauchamp B, et al. 2001. The biolistic method as a tool for testing the differential activity of putative silkmoth chorion gene promoters. *Insect Biochem. Mol. Biol.* 31:473–79
55. Kulkarni M, Arnosti DN. 2002. Parallel signaling by transcriptional enhancers. Submitted
56. Kutach AK, Kadonaga JT. 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. *Mol. Cell. Biol.* 20:4754–64
57. Lawson R, Mestril R, Schiller P, Voellmy R. 1984. Expression of heat shock-beta-galactosidase hybrid genes in cultured *Drosophila* cells. *Mol. Gen. Genet.* 198: 116–24
58. Lee TI, Young RA. 2000. Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* 34:77–137
59. Lehmann M. 1996. Drosophila *Sgs* genes: stage and tissue specificity of hormone responsiveness. *BioEssays* 18:47–54
60. Lewis EB. 1998. The bithorax complex: the first fifty years. *Int. J. Dev. Biol.* 42: 403–15
61. Li TR, White KP. 2002. *Developmental gene expression profiles of individual tissues and organs in* Drosophila. Presented at Annu. Drosophila Res. Conf., 43rd, San Diego, CA
62. Li X, Noll M. 1994. Compatibility between enhancers and promoters determines the transcriptional specificity of

gooseberry and *gooseberry neuro* in the *Drosophila* embryo. *EMBO J.* 13:400–6
63. Li X, Noll M. 1994. Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different *cis*-regulatory regions. *Nature* 367: 83–87
64. Lu J, Oliver B. 2001. Drosophila OVO regulates ovarian tumor transcription by binding unusually near the transcription start site. *Development* 128:1671–86
65. Ludwig MZ, Bergman C, Patel NH, Kreitman M. 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403:564–67
66. Ludwig MZ, Kreitman M. 1998. Functional analysis of *eve* stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. *Development* 125:949–58
67. Lyko F, Ramsahoye BH, Jaenisch R. 2000. DNA methylation in *Drosophila melanogaster*. *Nature* 408:538–40
68. Mackay TF. 2001. Quantitative trait loci in Drosophila. *Nat. Rev. Genet.* 2:11–20
69. Mahmoudi T, Verrijzer CP. 2001. Chromatin silencing and activation by Polycomb and trithorax group proteins. *Oncogene* 20:3055–66
70. Maier D, Preiss A, Powell JR. 1990. Regulation of the segmentation gene *fushi tarazu* has been functionally conserved in *Drosophila*. *EMBO J.* 9:3957–66
71. Markstein M, Markstein P, Markstein V, Levine MS. 2002. Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. USA* 99:763–68
72. Martin D, Wang SF, Raikhel AS. 2001. The vitellogenin gene of the mosquito *Aedes aegypti* is a direct target of ecdysteroid receptor. *Mol. Cell. Endocrinol.* 173:75–86
73. Meister M, Braun A, Kappler C, Reichhart JM, Hoffmann JA. 1994. Insect immunity. A transgenic analysis in *Drosophila* defines several functional domains in the *diptericin* promoter. *EMBO J.* 13:5958–66
74. Merika M, Thanos D. 2001. Enhanceosomes. *Curr. Opin. Genet. Dev.* 11:205–8
75. Merli C, Bergstrom DE, Cygan JA, Blackman RK. 1996. Promoter specificity mediates the independent regulation of neighboring genes. *Genes Dev.* 10:1260–70
76. Mihaly J, Hogga I, Barges S, Galloni M, Mishra RK, et al. 1998. Chromatin domain boundaries in the Bithorax complex. *Cell. Mol. Life Sci.* 54:60–70
77. Mitsialis SA, Kafatos FC. 1985. Regulatory elements controlling chorion gene expression are conserved between flies and moths. *Nature* 317:453–56
78. Mitsialis SA, Spoerel N, Leviten M, Kafatos FC. 1987. A short 5'-flanking DNA region is sufficient for developmentally correct expression of moth chorion genes in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 84:7987–91
79. Mougneau E, von Seggern D, Fowler T, Rosenblatt J, Jongens T, et al. 1993. A transcriptional switch between the *Pig-1* and *Sgs-4* genes of *Drosophila melanogaster*. *Mol. Cell. Biol.* 13:184–95
80. Näär AM, Lemon BD, Tjian R. 2001. Transcriptional coactivator complexes. *Annu. Rev. Biochem.* 70:475–501
81. Ohler U. 2000. Promoter prediction on a genomic scale: the *Adh* experience. *Genome Res.* 10:539–42
82. Ohtsuki S, Levine M. 1998. GAGA mediates the enhancer blocking activity of the *eve* promoter in the *Drosophila* embryo. *Genes Dev.* 12:3325–30
83. Ohtsuki S, Levine M, Cai HN. 1998. Different core promoters possess distinct regulatory activities in the *Drosophila* embryo. *Genes Dev.* 12:547–56
84. Orphanides G, Reinberg D. 2002. A unified theory of gene expression. *Cell* 108:439–51
85. Papatsenko DA, Makeev VJ, Lifanov AP, Regnier M, Nazina AG, Desplan C. 2002. Extraction of functional binding sites from unique regulatory regions:

the *Drosophila* early developmental enhancers. *Genome Res.* 12:470–81
86. Piano F, Parisi MJ, Karess R, Kambysellis MP. 1999. Evidence for redundancy but not *trans* factor-*cis* element coevolution in the regulation of *Drosophila* Yp genes. *Genetics* 152:605–16
87. Purnell BA, Emanuel PA, Gilmour DS. 1994. TFIID sequence recognition of the initiator and sequences farther downstream in *Drosophila* class II genes. *Genes Dev.* 8:830–42
88. Ren B, Maniatis T. 1998. Regulation of *Drosophila Adh* promoter switching by an initiator-targeted repression mechanism. *EMBO J.* 17:1076–86
89. Renault N, King-Jones K, Lehmann M. 2001. Downregulation of the tissue-specific transcription factor Fork head by *Broad-Complex* mediates a stage-specific hormone response. *Development* 128: 3729–37
90. Roth SY, Denu JM, Allis CD. 2001. Histone acetyltransferases. *Annu. Rev. Biochem.* 70:81–120
91. Rushlow C, Colosimo PF, Lin MC, Xu M, Kirov N. 2001. Transcriptional regulation of the *Drosophila* gene *zen* by competing Smad and Brinker inputs. *Genes Dev.* 15:340–51
92. Santel A, Kaufmann J, Hyland R, Renkawitz-Pohl R. 2000. The initiator element of the *Drosophila* $\beta 2$ tubulin gene core promoter contributes to gene expression in vivo but is not required for male germ-cell specific expression. *Nucleic Acids Res.* 28:1439–46
93. Schulz RA, Xie XL, Miksch JL. 1990. *cis*-acting sequences required for the germ line expression of the *Drosophila gonadal* gene. *Dev. Biol.* 140:455–58
94. Sheng G, Thouvenot E, Schmucker D, Wilson DS, Desplan C. 1997. Direct regulation of rhodopsin 1 by *Pax-6/eyeless* in *Drosophila*: evidence for a conserved function in photoreceptors. *Genes Dev.* 11:1122–31
95. Schier AF, Gehring WJ. 1993. Analysis of a *fushi tarazu* autoregulatory element: multiple sequence elements contribute to enhancer activity. *EMBO J.* 12:1111–19
96. Schwyter DH, Huang JD, Dubnicoff T, Courey AJ. 1995. The *decapentaplegic* core promoter region plays an integral role, in the spatial control of transcription. *Mol. Cell. Biol.* 15:3960–68
97. Scott KS, Geyer PK. 1995. Effects of the su(Hw) insulator protein on the expression of the divergently transcribed *Drosophila* yolk protein genes. *EMBO J.* 14:6258–67
98. Skavdis G, Siden-Kiamos I, Muller HM, Crisanti A, Louis C. 1996. Conserved function of *Anopheles gambiae* midgut-specific promoters in the fruitfly. *EMBO J.* 15:344–50
99. Smale ST. 2001. Core promoters: active contributors to combinatorial gene regulation. *Genes Dev.* 15:2503–8
100. Small S, Arnosti DN, Levine M. 1993. Spacing ensures autonomous expression of different stripe enhancers in the *even-skipped* promoter. *Development* 119:767–72
101. Small S, Blair A, Levine M. 1992. Regulation of *even-skipped* stripe 2 in the *Drosophila* embryo. *EMBO J.* 11:4047–57
102. Stanojevic D, Small S, Levine M. 1991. Regulation of a segmentation stripe by overlapping activators and repressors in the *Drosophila* embryo. *Science* 254: 1385–87
103. Struhl K. 2001. Gene regulation. A paradigm for precision. *Science* 293:1054–55
104. Suzuki Y, Tsunoda T, Sese J, Taira H, Mizushima-Sugano J, et al. 2001. Identification and characterization of the potential promoter regions of 1031 kinds of human genes. *Genome Res.* 11:677–84
105. Tautz D. 2000. A genetic uncertainty problem. *Trends Genet.* 16:475–77
106. Tautz D. 2000. Evolution of transcriptional regulation. *Curr. Opin. Genet. Dev.* 10:575–79

107. Tingvall TO, Roos E, Engstrom Y. 2001. The GATA factor Serpent is required for the onset of the humoral immune response in *Drosophila* embryos. *Proc. Natl. Acad. Sci. USA* 98:3884–88
108. Tomancak P, Beaton A, Weiszmann R, Kwan E, Celniker S, Rubin GM. 2002. *Patterns of gene expression in development-embryology meets bioinformatics.* Presented at the Annu. Drosophila Res. Conf., 43rd, San Diego, CA
109. Tweedie S, Ng HH, Barlow AL, Turner BM, Hendrich B, Bird A. 1999. Vestiges of a DNA methylation system in *Drosophila melanogaster*? *Nat. Genet.* 23:389–90
110. Verrijzer CP. 2001. Transcription factor IID: not so basal after all. *Science* 293: 2010–11
111. Weiler KS, Wakimoto BT. 1995. Heterochromatin and gene expression in *Drosophila*. *Annu. Rev. Genet.* 29:577–605
112. White KP, Rifkin SA, Hurban P, Hogness DS. 1999. Microarray analysis of *Drosophila* development during metamorphosis. *Science* 286:2179–84
113. Willy PJ, Kobayashi R, Kadonaga JT. 2000. A basal transcription factor that activates or represses transcription. *Science* 290:982–85
114. Wolff C, Pepling M, Gergen P, Klingler M. 1999. Structure and evolution of a pair-rule interaction element: *runt* regulatory sequences in *D. melanogaster* and *D. virilis*. *Mech. Dev.* 80:87–99
115. Workman JL, Kingston RE. 1998. Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu. Rev. Biochem.* 67:545–79
116. Woychik NA, Hampsey M. 2002. The RNA polymerase II machinery: structure illuminates function. *Cell* 108:453–63
117. Wu CH, Madabusi L, Nishioka H, Emanuel P, Sypes M, et al. 2001. Analysis of core promoter sequences located downstream from the TATA element in the *hsp70* promoter from *Drosophila melanogaster*. *Mol. Cell. Biol.* 21:1593–602
118. Xiong B, Jacobs-Lorena M. 1995. Gut-specific transcriptional regulatory elements of the carboxypeptidase gene are conserved between black flies and *Drosophila*. *Proc. Natl. Acad. Sci. USA* 92:9313–17
119. Yanicostas C, Lepesant JA. 1990. Transcriptional and translational *cis*-regulatory sequences of the spermatocyte-specific *Drosophila janusB* gene are located in the 3′ exonic region of the overlapping *janusA* gene. *Mol. Gen. Genet.* 224:450–58
120. Yasuhara JC, DeCrease C, Slade D, Wakimoto BT. 2002. *A comparative study of heterochromatic and euchromatic* light *genes in* Drosophila *species*. Presented at the 43rd Annu. Drosophila Res. Conf., San Diego, CA
121. Zenzie-Gregory B, O'Shea-Greenfield A, Smale ST. 1992. Similar mechanisms for transcription initiation mediated through a TATA box or an initiator element. *J. Biol. Chem.* 267:2823–30
122. Zhu J, Kokoza V, Raikhel A. 2002. Analysis of stage- and tissue-specific gene expression in insect vectors. In *The Biology of Disease Vectors*, pp. xx–xx, ed. WC Marquardt. Niwot: Univ. Colorado Press. 2nd ed.