Differential Gene Expression in Insects: Transcriptional Control

Lawrence G. Harshman  
University of Nebraska - Lincoln, lharshman1@unl.edu

Anthony A. James  
University of California, Irvine, aajames@uci.edu

Follow this and additional works at: https://digitalcommons.unl.edu/bioscifacpub

Part of the Life Sciences Commons

Harshman, Lawrence G. and James, Anthony A., "Differential Gene Expression in Insects: Transcriptional Control" (1998). Faculty Publications in the Biological Sciences. 87.  
https://digitalcommons.unl.edu/bioscifacpub/87

This Article is brought to you for free and open access by the Papers in the Biological Sciences at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications in the Biological Sciences by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Differential Gene Expression in Insects: 
Transcriptional Control

Lawrence G. Harshman
School of Biological Sciences, University of Nebraska–Lincoln, 
Lincoln, NE 68588-0118; email lharch@unl.edu

Anthony A. James
Department of Molecular Biology and Biochemistry, University of California, 
Irvine, CA 92697-3900; email aajames@uci.edu

Abstract
Studies on transcriptional control of gene expression play a pivotal role in many areas of biology. In non-Drosophilid insects, the cuticle, chorion, immune response, silk gland, storage proteins, and vitellogenin are foci for advances in basic research on promoter elements and transcription factors. Insects offer other advantages for gene regulation studies, including the availability of applied problems. In non-Drosophilid insects, the most serious problem for transcriptional control studies is the lack of homologous in vivo expression systems. Once this deficiency is addressed, the full impact of research on transcription control will be realized throughout the field of entomology.

Introduction
Of the tens of thousands of genes present in a eukaryote species, perhaps only 15% are expressed in any given cell at any given time (101). Differential gene expression underlies a range of biological processes, including development, reproduction, and behavior. Intense interest in this topic and the application of new technologies has produced rapid advances in the analysis of transcriptional control of gene expression in humans and a number of other species that serve as foci for study. The impact of this area of study on various fields of biology indicates that
entomology will be permeated by research on transcriptional control of gene expression.

This review focuses on gene regulation studies in insects. Emphasis is placed on studies of non-Drosophilid insects, using the *Drosophila* literature for examples, and to indicate possible future directions for studies on other insects. Categories of regulatory sequences are summarized to provide a general overview of gene regulation by *cis*-acting DNA. Hormone-mediated gene regulation and a number of well-characterized systems are discussed, including the cuticle, chorion, immune response, silk gland, storage proteins, and vitellogenin. Also discussed are the value of *in vivo* studies, transcriptional control in relationship to evolution, and physiological studies at the level of transcription. No attempt is made to cover insect viruses nor to summarize control of insect gene expression at the level of chromatin structure, splicing, translation, or post-translational processes. Regulation of genes transcribed by RNA polymerase III (150) is not addressed by this review. Unfortunately, it is necessary to omit many excellent studies from the review.

**Eukaryote Regulation of Transcription**

Scientists who wish to investigate, or exploit, some aspect of transcriptional regulation in non-Drosophilid insects usually have neither the advantage of the genetic tractability of yeast or *Drosophila*, nor homologous *in vivo* expression systems that are available for a range of organisms. However, by extrapolating from results obtained using these organisms, and by utilizing the full range of biochemical and molecular biology techniques available, this does not present insurmountable difficulties. In fact, advances in molecular and cellular biology have “democratized the genome” in the sense that *cis*-acting sequences and *trans*-acting factors can be studied in most higher organisms.

**Overview of Transcription**

Transcription is the fundamental process by which a DNA template is used to generate messenger RNA to be translated into proteins, or ribosomal, transfer, and other RNAs that have structural, regulatory, and catalytic functions. Gene regulation in eukaryotes occurs primarily at the level of transcription. This has been consistently borne out by studies on organisms such as the yeast *Saccharomyces cerevisiae*, the nematode *Caenorhabditis elegans*, sea urchins, *Drosophila melanogaster*, and a variety of vertebrates.

Eukaryote transcription is controlled by regulatory DNA sequences and transcription factors. Typically, regulatory sequences are near, or within, a gene (*cis*). Much of this *cis*-regulatory DNA is found in the 5′-end noncoding DNA associated with transcriptionally functional genes
We sometimes refer to this region as 5′-flanking DNA or upstream DNA. Regulatory DNA sequences may interact with other sequences that are not immediately adjacent. Transcription factors are proteins that serve to provide a “connection” between regulatory DNA regions that interact at a distance. These factors generally belong to one of a number of protein families defined by conserved amino acid sequence motifs that usually are DNA binding domains. Factors may interact with more than one regulatory DNA element. In addition, factors can interact with each other to form complexes with high affinity for regulatory elements. Factors are usually encoded by genes that are unlinked to the target gene and thus act in trans.

In order to understand eukaryotic gene regulation, one must be mindful of the importance of context and interaction. The capability to uniquely control the expression of each gene in an organism arises from the nature and relative position of associated regulatory sequences. These sequences interact with factors expressed in response to developmental state, physiological state, cell- or tissue-type, and the environment. Specific aspects of control mechanisms are conserved from yeast to humans (65), and yet transcriptional control of a gene is typically complex and has been characterized as consisting of “layers of regulation” (184).

Analysis of DNA Sequences and Factors that Regulate Gene Expression

Early non-Drosophilid insect gene regulation studies were largely based on the use of cloning and sequencing for the purpose of investigating promoter (regulatory) DNA of a few well-studied genes (63). Relatively recently, specialized techniques have been developed to study gene regulation. DNase I footprinting (57) and mobility shift (53) assays are two such techniques. For DNase I footprinting, prospective regulatory DNA sequences are protected from nuclease digestion by associated transcription factors derived from nuclear extracts. Mobility shift assays are based on decreased mobility of a DNA fragment migrating through a gel when interacting transcription factor protein is associated.
Transcription factors are generally amenable to study. Factors may be isolated by affinity chromatography or cloned by polymerase chain reaction (PCR) amplification using primers that correspond to conserved amino acid motifs. For the first purification of a transcription factor, Kadonaga & Tjian (88) coupled oligonucleotides that corresponded to a promoter region to a column matrix. Subsequently, numerous transcription factors have been isolated and characterized from a diverse collection of eukaryotes. The ability to identify regulatory sequences and isolate transcription factors should not obscure the need to study regulatory DNA modules using hypothesis testing and critical experiments (147). This requirement for functional investigation mandates the development of homologous in vivo expression systems for non-Drosophilid insects.

**Insects as Subjects for Transcriptional Control Studies**

Unique characteristics of insects provide favorable targets for gene regulation studies. The silk gland, as described below, is one such example. In addition, advances in technology have made it feasible to study gene regulation in various pest insect species.

*The Silk Gland as the First Insect Model System for Gene Regulation Studies*

The silk glands of the silk moth, Bombyx mori, are modified salivary glands. Fibroins are silk fiber proteins that are synthesized and secreted by the posterior silk gland (PSG). Sericins, proteins that act to adhere silk fibers, are synthesized and secreted by the middle silk gland (MSG). Active transcription from the fibroin gene was determined to occur during the last four days of the fifth instar, in parallel with a surge of silk production for production of the cocoon (120, 158). The mRNA from a different silk fiber protein gene, P25, is synthesized in the PSG and accumulates simultaneously with fibroin gene message (37). In the MSG, mRNA from different sericin genes appears during the same period, but expression is gene-dependent rather than tightly coordinated (36, 80). Silk gland genes have various attributes that contributed to their utility for the study of gene expression, including abundant mRNA production and tissue region–specific control of gene expression. In addition, in vitro transcription methods were developed to study B. mori silk gland gene promoters (167). Although there are no completely defined in vitro transcription systems for eukaryotes (94), the cell-free transcription assays in B. mori pioneered an emphasis on functional analyses of insect gene expression.
Emerging Areas of Research Focus

Insects as agents of agricultural or medical problems are prime targets for gene regulation studies. A systematic investigation of olfaction genes is in progress in *D. melanogaster* (22), and a similar effort in other insects is encouraged by the prospect of manipulating the mechanism that pest species use to find food. In general, it is important to investigate the regulation of insect genes that are expressed in the process of food location or in the course of feeding.

An ancillary advantage of working on applied problems is that the research is typically conducted in a defined environmental context. Thus, the opportunity exists to meaningfully interpret a particular mechanism of gene regulation in terms of ecological response or evolutionary adaptation.

An important goal in medical entomology is understanding transcriptional control of gene expression in disease vector salivary glands (81). Salivary gland secretions are the first vector gene products to interact with a blood meal, and salivary glands are the last site of interaction between vectors and the majority of human disease-causing microbes. Studies of salivary gland gene expression in the yellow fever mosquito, *Aedes aegypti*, show that maltase-like gene mRNA is expressed in females only in the proximal regions of the lateral lobes in females (82). Salivary gland gene promoters could be used to effectively target the expression of recombinant genes that are toxic to human pathogens carried by insect vectors.

**CIS-Acting DNA Sequences**

This section briefly describes *cis*-acting DNA elements that regulate gene expression. Transcription factors are also discussed, although they typically are not well characterized in non-Drosophilid insects. Gene expression in the silk gland of *Bombyx mori* is featured as a source of examples of *cis*-elements in recognition of the important contributions derived from this system.

**Core Promoter**

The core promoter consists of regulatory DNA in immediate 5′-end proximity of the transcription start site of a gene (Figure 1). This regulatory DNA tends to be highly conserved in sequence, position, and function. The core promoter typically confers low-level constitutive gene expression.

**TATA BOX** — The core promoter often includes a sequence with a TATA base sequence motif. The TATA “box” (sequence) is functionally constrained in sequence, orientation, and position relative to the start of transcription. It typically interacts with the TATA binding protein (TBP),
which serves as a point of interaction with other transcription factors and RNA polymerase.

An upstream sequence comparison of the *B. mori* fibroin gene (168) and P25 gene (35) was conducted, and the TATA box region was compared. Specifically, the TATA box region of the fibroin gene ends at −30 compared to −32 for P25 (base + 1 is defined as the transcription start site). Moreover, the TATA box is identical in sequence and nearly identical in position for fibroin and sericin silk gland genes (35, 121).

The effect of point mutations on the fibroin gene core promoter was investigated by *in vitro* mutagenesis (70). Mutagenesis and the cell-free transcription assay were used to infer the actual point of transcription factor contact with specific core promoter bases (69). As has also been observed for *Drosophila* and mammalian genes, the fibroin gene core promoter possesses basal (constitutive) and modulatory functions (159).

**INITIATOR REGION** – In vertebrates, conserved sequences in the immediate vicinity of the transcription start site form a regulatory unit called the initiator region (Inr) (83). There is now evidence for an arthropod-specific Inr. A comparative DNA sequence analysis by Monte Carlo simulation based on data from the arthropod promoter database (*Nucleic Acids Research* 1996:24) has revealed that a high percentage of arthropod promoters contain a consensus sequence, TCAGT, in immediate proximity to the transcription start site (26).

**Upstream Promoter**

Characteristic regulatory sequences are often found in the 5′-end proximal region of genes transcribed by RNA polymerase II (100). They consist of conserved sequences that identify them as CAAT, OCT, or GC-rich promoter elements, and they may be referred to collectively as upstream promoter elements (Figure 1). As an example, the octamer sequence (OCT), ATGCAAAT, is associated with developmentally defined and tissue-specific gene expression.

Upstream promoter elements tend to be constrained in position but may function unconstrained by orientation. Prediction is not always possible of how the presence of any one of these elements will influence the activity of a specific gene, but the effect of interaction with other factors is often transcription stimulation. In general, upstream promoter elements are essential for efficient promoter activity.

Sequences that correspond to several classes of upstream promoter elements have been identified immediately upstream of non-Drosophilid insect genes. For example, a CAAT sequence is found in an upstream interval (−93 to −83) of the *B. mori* fibroin and sericin genes (63). As another example, the canonical octamer sequence was found upstream of a cuticular protein gene in the saturniid moth, *Hyalophora cecropia* (96). This se-
quence may be transcriptionally functional, as suggested by the fact that cell lysates from various tissues retarded the mobility of an upstream fragment from this H. cecropia gene. Moreover, interacting protein was displaced using an oligonucleotide that matched the octamer sequence but not with an oligonucleotide that had a slightly different sequence.

**Enhancers**

Enhancer sequences (Figure 1) by definition have a stimulatory effect on transcription. Whereas core and upstream promoter elements have relatively defined locations relative to the transcription start site, enhancers may function tens of thousands of nucleotides away from the target gene. They may be found upstream, downstream (beyond the 3′ noncoding end), or within a gene. Enhancer sequences function in either orientation, and their action may be tissue specific. For example, yolk protein genes of D. melanogaster are known to have fat body and ovarian follicle cell enhancers (16, 58, 59). Categories of DNA also exist that act to inhibit transcription, which are known as silencers or repressors (17, 86).

Enhancers in non-Drosophilid insects are poorly understood undoubtedly because such sequences cannot usually be defined on the basis of functional characteristics in the general absence of in vitro or in vivo transcription assays. Nevertheless, from in vitro studies on fibroin gene expression, enhancers were determined to exist in two sequence intervals approximately 50–250 bases upstream from the transcription start site (165, 166). Experimental inversion of the region (bases −234 to −66) that contained the enhancers did not alter the stimulatory effect on transcription, as is characteristic for these elements.

**Temporal Control, Cell Type, and Tissue-Specific Regulatory Sequences**

These promoter sequences interact with factors that are differentially expressed in time or location within the organism (105). Temporal control factors are expressed at defined stages of development. As an example, the concentration of two factors, Sp1 and GATA-1, can account for temporal and tissue-defined activation of B-hemoglobin gene expression in embryos (114). Examples of temporal and tissue-specific promoters are also available from work on D. melanogaster. Temporal expression of the alcohol dehydrogenase (Adh) gene is controlled by stage-specific promoters; one promoter mediates larval expression, and a different promoter mediates adult expression (141, 142, 147). An example of tissue-specific expression is the promoter that restricts the expression of certain genes to the D. melanogaster male accessory gland (25).

In non-Drosophilid insects, only a limited number of DNA sequences and factors have been clearly implicated in temporal or tissue-specific control of gene expression. An example of temporal and tissue specific-
ity is provided by work on a GATA sequence in a chorion gene promoter in *B. mori*. This sequence interacts with a GATA-like factor in conjunction with the activation of late chorion gene expression in *B. mori* follicle cells (44, 45). An example of temporal and tissue region specificity is provided by work on the expression of the P25 gene in the PSG of *B. mori* (46, 73). This tissue region has approximately 500 cells, and expression of the P25 gene is restricted to PSG at a specific time in development. Transcription factors are known to bind 5′-flanking DNA of the P25 gene. One factor, PSGF, acts to stabilize binding of a second factor, SGFB, at its recognition sequence, and apparently this interaction mediates PSG-specific gene expression.

**Sex Determination and the Homeobox**

Specific regulatory DNA sequences control fundamental processes in organisms including cell division, sex determination, and development. These sequences may also control the expression of genes that are not directly associated with these fundamental processes.

**SEX-DETERMINATION FACTOR BINDING SITES** — The sex-determination pathway in *D. melanogaster* consists of a hierarchy of regulatory gene expression (7). At the bottom of the hierarchy, an mRNA for a transcription factor (Doublesex protein) is differentially spliced to produce female- or male-specific factors that act to determine gender and define sex-specific somatic gene expression. Specific target genes are known for this transcription factor. For example, between yolk protein genes 1 and 2 of *D. melanogaster*, doublesex binding sites have been established to exist (2, 19), and accordingly, yolk protein gene expression is known to depend on a functional sex-determination hierarchy (9). The vitellogenin A1 gene from the yellow fever mosquito, *A. aegypti*, has putative doublesex binding sites in the noncoding regulatory DNA region (139).

**THE HOMEBOX** — The products of homebox genes are transcription factors that recognize highly conserved DNA sequences known as homeoboxes. Homebox gene transcription factors and homeoboxes were first described in *D. melanogaster*, where they are now known to play an overarching organizing role in development. These genes encode transcription factors that determine embryonic polarity, segmentation boundaries, and developmental commitment within segments. The homeobox genes that regulate early development in *D. melanogaster* are partially conserved in sequence, gene organization, and function in mammals (64).

As revealed by a comparison of *Tribolium castaneum* and *D. melanogaster*, features of certain homeobox genes are highly conserved between the two taxa, but differences are also present. Specifically, *T. castaneum* has a
single homeotic gene complex (153) rather than the anterior and posterior
determining clusters of *D. melanogaster*. *B. mori* appears to be more like *D.
 melanogaster* in that its homeotic genes are organized into two complexes
(169, 170). Homeobox genes may have played an important role in insect
evolution. For example, the suppression of homeotic gene expression in
specific abdominal cells in Lepidoptera is associated with the generation
of abdominal prolegs in the order (171).

Homeobox sequences and genes appear to be involved in the regulation
of a diverse set of genes. An example is provided by the homeobox
sequences studied in the upstream region of the *B. mori* silk gland fibroin
gene. Upstream fragments from the fibroin gene, including the fragment
from −234 to −66, were used in mobility shift assays to detect DNA bind-
ing proteins in PSG and MSG extracts (157). Hui *et al.* (75) demonstrated
that upstream regions of the fibroin gene bind different silk gland tran-
scription factors (SGFs). SGF-1 binds a site found in the 5′-flanking DNA
of sericin-1 and fibroin genes. SGF-1, and its binding site may play a more
general role in regulation of silk gland genes (107, 108). It is interesting to
note that sequences with similarity to the SGF-1 site exist in *Drosophila*
salivary gland genes (75). Three of the SGFs recognize a DNA sequence
that corresponds to the homeobox (74). Clusters of homeodomain bind-
ing sites are in the promoters of silk gland genes, and multiple homeobox
genes are expressed in the PSG (75). Homeobox genes and homeoboxes
appear to play a major role in transcriptional control of silk gland genes
that is not yet understood in detail.

**Response Elements**

Many genes have associated DNA sequences that serve as specific or
global response elements (Figure 1). These sequences typically are in-
volved in the induction of gene expression in response to environmental
factors or hormonal signals.

**ENVIRONMENTAL RESPONSE ELEMENTS** — Elements are known
that stimulate gene expression in response to a range of environmental
states. For example, heat shock elements and metal response elements are
seemingly universally present in organisms. Other elements mediate re-
sponse to cyclic AMP, oxidative stress, toxins, microbial infection, and star-
vation. Few such elements are identified in non-Drosophilid insects. As one
possible example from work on the blow fly, *Calliphora vicina*, a sequence
has been identified in the 5′-end noncoding DNA of a storage protein gene
that is essentially the same as one known to mediate starvation condition
control of amino acid biosynthesis gene expression in yeast (52).

Work by Berenbaum and colleagues on swallowtail butterflies, *Papilio*
spp., and their host plants illustrates the value of insects for the study
of environment-dependent gene expression (10, 11). In *Papilio polyxenes*, and related species, the pattern of host plant usage correlates with the inducibility of cytochrome P450s that detoxify plant secondary compounds such as xanthotoxin (33, 126). A cDNA cloned from *P. polyxenes* encodes a cytochrome P450 that inactivates xanthotoxin (33, 34). In *Papilio glaucus* and *P. polyxenes*, a series of related P450 genes are induced by furanocoumarins including xanthotoxin (77). In the 5′-flanking region of some of these genes, prospective regulatory sequences are found that correspond to vertebrate response elements such as a dioxin receptor element, a barbituate response element, and an antioxidant response element. In the process of defining the function of these sequences, a fragment from the upstream promoter region of one of these genes has been found to be sufficient to drive xanthotoxin-induced reporter gene expression in a heterologous lepidopteran cell line (126).

**HORMONE RESPONSE ELEMENTS** — An ecdysone response element (EcRE) is a DNA sequence that binds the ecdysone receptor (EcR), which is usually complexed with the hormone 20-hydroxyecdysone (20HE). The first EcRE described was found upstream of the *D. melanogaster* heat shock protein 27 gene (138). This element is similar in sequence to the mammalian estrogen, thyroid, and retinoic acid response elements. Remarkably, a two-base increase in the spacing within the ecdysone element converts it into a functional estrogen response element (106). EcREs have been functionally investigated in *D. melanogaster*. Some noteworthy studies are on *Eip28/29*, a gene that is rapidly activated in the presence of ecdysteroids (27). Within 10 kb around *Eip28/29*, there are 38 sequences that partially match the expectation of an EcRE. However, only 5–6 of these sequences have the projected symmetrical pattern of base sequence necessary for interaction with both arms of the receptor. Of importance, only three functional EcREs are present; one in the 5′-end noncoding region and two in the 3′-end noncoding region of the gene. The functional EcREs were rigorously identified by the ability of a putative EcRE cloned into an unrelated DNA sequence to bind EcR and the ability of a putative element to confer ecdysone responsiveness to an unrelated reporter gene. Numerous EcRE-like sequences may exist in proximity to a gene, but only a few of them may actually possess response element function.

**Hormonal Control of Gene Expression**

One of the distinctive features of insect gene regulation studies is the support provided by the rich tradition of insect endocrinology. Insect endocrinology is based on a well-established experimental logic that uses extirpation, ligation, transplantation, and administration of hormonally active compounds. This powerful experimental framework is used to in-
vestigate regulatory elements and factors. The recent cloning of the ecdysone receptor (EcR) establishes a bridge between classical insect endocrinology and molecular biology.

Ecdysone Receptor (EcR)

The recent cloning of the ecdysone receptor (EcR) establishes a bridge between classical insect endocrinology and molecular biology. The first demonstration that ecdysteroids can activate transcription was made using polytene chromosome puffs of the midge Chironomus tentans (31, 32). An important conceptual advance was made by Ashburner et al. (6), who proposed a model in which 20HE, interacting with a receptor at the time of the D. melanogaster larval-prepupal transition, acts as a transcription regulator by blocking expression of a small class of repressor genes and at the same time activating a larger set of regulatory genes that control target gene expression. The Ashburner model for the indirect action of 20HE may often apply to ecdysone modulation of gene expression (39, 68). However, it is now clear that the regulatory hierarchy controlled by ecdysteroids is quite complicated (164). Moreover, there is evidence, derived from a study of the Fbp1 gene in D. melanogaster, that the ecdysone receptor can have a direct role in controlling gene expression in response to a change in ecdysteroid titer. Specifically, in vivo footprinting was used to demonstrate occupancy of an EcRE by the ecdysone receptor complex during the time of active transcription of the Fbp1 gene (5).

The ecdysone receptor (EcR) was first cloned in D. melanogaster (92). Multiple isoforms of EcR characteristically differ in the N-terminal region (160). The EcR-B1 isoform is found in larval tissues, and the EcR-A isoform in imaginal discs. The product of another gene, ultraspiracle (USP), forms a heterodimer with EcR to produce the functional receptor (185). In addition, at least a dozen steroid-like receptors are now known to exist, whose endogenous ligand and function are unknown (68).

The developmental pattern of the EcR gene expression is noteworthy because in conjunction with 20HE, the receptor controls a battery of genes at each molt. In Manduca sexta, analogs to EcR-B1, EcR-A, and USP have been cloned (55, 84, 85). The mRNA abundance of EcR-B1 and EcR-A analogs have been monitored in the epidermis of M. sexta during development. EcR-B1 mRNA rises at the beginning of a molt, whereas the peak of EcR-A mRNA coincides with the peak of ecdysteroid titer during the molt (55, 84). Possibly, EcR-B1 may induce the molt, and EcR-A may control initiation of cuticle deposition (134). The level of USP message is relatively high and invariant except for a decrease at the end of the pupal molt (85). In M. sexta, JH modulates expression of EcR and USP mRNAs during pre-differentiation events (134). In this context, it may be relevant that thyroxin and prolactin act antagonistically on gene expression during amphibian metamorphosis (163) reminiscent of the often antagonistic action of JH and 20HE.
Genes similar to *D. melanogaster* EcRs are under investigation in non-Drosophilid Diptera. In *C. tentans*, EcR mRNA abundance parallels changes in ecdysteroid titer (78). In this species, the EcR mRNA has been observed to be relatively uncommon during diapause (68). An EcR cDNA probe has been used to assess mRNA abundance during pre-vitellogenesis and vitellogenesis in *A. aegypti* (29). In pre-vitellogenic mosquitoes, EcR mRNA abundance drops at 2–3 days post-eclosion, but a rapid accumulation of mRNA occurs in response to a blood meal followed by a decline at 12–24 h post-feeding.

_Juvenile Hormone Control of Transcription_

JH apparently controls gene expression by multiple mechanisms (87, 135, 173). JH may act by induction of transcription factors, and evidence from work on *D. melanogaster* (183) and *Rhodnius prolixus* (143) shows that JH may operate by stimulation of a second messenger system. One approach to study the mode of action of JH is to isolate a set of genes whose transcription is controlled by the hormone and compare the sequence of the 5′-end flanking DNA of each gene to determine what sequences are held in common as prospective JH responsive elements (87, 122). A comparative sequence analysis of the upstream region of the JH-suppressible storage protein genes identified several DNA sequences held in common but not present in the promoter region of a JH-insensitive storage protein gene. A mutation approach may prove to be especially useful to investigate JH transcriptional control mechanisms. Wilson (175) has obtained mutants (Met) of *D. melanogaster* that can tolerate high concentrations of a JH analog (methoprene). The sequence of a Met cDNA suggests that it may encode a protein that belongs to the PAS family of transcription regulators (TG Wilson, M Ashok, G Thomas, personal communication).

The range of biological activities of JH and considerable background knowledge about the agents that control JH titer provide an interesting context for gene regulation studies. JH titer is known to be partially controlled by juvenile hormone esterase (JHE), and reciprocally JH can control expression of this enzyme (66). In order to understand this regulatory relationship, it will be essential to investigate JH regulation of JHE gene expression (67, 176). Some progress has been made in this endeavor. Specifically, lysates from a tissue expressing JHE produce a gel mobility shift of an upstream fragment from a *Trichoplusia ni* JHE gene (87).

_Model Systems_

In addition to *B. mori* silk glands, a few systems are the focus of non-Drosophilid gene regulation studies. The advantage of working on model
Differential Gene Expression in Insects: Transcriptional Control

systems is clear, and each can be expected to provide information on transcriptional control of gene expression well into the future. However, the concept of model systems can have a mixed effect on research. Although it has served to push ahead aspects of basic research, it may also have the effect of restricting work on other organisms and applied problems.

**Chorion**

*B. mori* chorion proteins are the diverse constituents of eggshell that provide an opportunity for study of sharply defined tissue and temporal coordinate regulation of gene families. The characteristics of chorion genes have been previously reviewed (63, 179). Two chorion gene families (A and B) can be subdivided into three temporal expression classes: early (ErA, ErB), middle (A, B), and high-cysteine late genes (HcA, HcB). In each class, A and B genes are paired and simultaneously bidirectionally transcribed. The information necessary for correct temporal and tissue gene expression is contained in 300 bp of noncoding DNA that separate each A and B pair of genes (115, 116).

A conserved hexamer, TCACGT, found in each 5′-flanking region of a chorion gene is identical in sequence to an element found upstream of *D. melanogaster* chorion genes (149). Mutation analysis of this chorion-specific hexamer was conducted on a promoter fragment from a silk moth, *Antereaea pernyi*, chorion gene (50). Using heterologous expression in *D. melanogaster*, single-base substitutions at positions 2 and 4 of the hexamer were determined to suppress reporter gene expression driven by the silk moth hexamer, which experimentally verifies the critical nature of the sequence for chorion gene function. In a heterologous gene expression study using *D. melanogaster*, deletion of the chorion gene-specific conserved element, TCACGT, resulted in loss of expression control and thereby revealed a functional regulatory sequence shared between flies and moths (116).

Skeiky & Iatrou (146) used mobility shift and footprinting assays to analyze the upstream region of the Hc chorion genes and identified protein-binding DNA sequences and two transcription factors, BCFI and BCFII. One of the protein-binding sequences, AGATAA, is the consensus binding site of the GATA transcription factor, which was first characterized in vertebrates (123). GATA is a transcription factor that controls expression of vertebrate globin genes, which are a multigene family expressed in a specific tissue in a temporally defined pattern (132).

Conserved amino acid sequences among GATA factors were used to design primers for PCR amplification of GATA factor gene fragments (44). Amplified gene fragments were used to probe cDNA expression libraries. Two single-copy genes were isolated and shown to encode GATA-like factors. The expressed product of one gene retarded mobility of an upstream fragment that contained the GATA consensus binding
site (44). This gene, BmGATAB, produces three isoforms by alternative mRNA splicing, and changes in splice site preference occur during the late stages of choriogenesis (45). In this system, the abundance of the active transcription factor is regulated by protein modification. Specifically, an inactive phosphorylated form of BCF1 is sequestered in the cytoplasm of specific cells until it appears dephosphorylated in the nucleus coincident with activation of high-cysteine chorion genes (145).

Evidence exists for positive and negative regulation of chorion gene expression. In vitro mutagenesis of a well-studied B. mori chorion gene promoter (A/B.L12) followed by heterologous expression in D. melanogaster revealed the presence of an element that suppresses gene expression (150). Current understanding of the transcriptional control of B. mori chorion genes is summarized by Kafatos et al. (89). They noted that factors binding promoters control gene expression by quantitative and temporal regulation and that the interaction of factors may be “complex, synergistic and hierarchical.”

Cuticle

The cuticle is a tissue that facilitates the study of sequential gene expression (137). Several Lepidopteran species, especially H. cecropia and M. sexta, serve as a focus for transcription regulation studies on the cuticle.

Ecdysteroid and juvenile hormone (JH) titers have been carefully determined throughout M. sexta growth and metamorphosis, which provides a valuable context to investigate the hormonal control of cuticular gene expression (133, 137). Within this framework, cuticular genes have been identified that are differentially expressed during the larval intermolt, at the molt, at the inception of metamorphosis, or during the pupal stage. These genes may be hormonally activated, transiently repressed and then activated by a hormonal signal, or deeply repressed. In vitro culture of the larval cuticle permits isolation of this tissue from the influence of endogenous 20HE and JH, and thus control of specific cuticular gene expression can be studied by in vitro administration of 20HE and JH (71).

Hiruma & Riddiford (72) present a model for JH and ecdysteroid regulatory interaction based on the study of dopa decarboxylase (DDC) gene expression in the epidermis of M. sexta. The hypothesis was advanced that JH induces a longer-lived positive transcription factor and ecdysone induces a shorter-lived negative-effect factor. In this scenario, when the ecdysone-induced factor degrades, a temporal burst of DDC activity is expected to occur until the JH-induced factor also degrades. As a general perspective, negative control of transcription may play an important role in determining the temporal or spatial boundaries of gene expression (97).

In H. cecropia, genes encoding proteins from the flexible cuticle (HCCP12) and rigid cuticle (HCCP66) have been cloned and sequenced,
including their 5′-end noncoding regions (12, 96). A computer homology search for possible cis-acting elements revealed an imperfect palindromic EcRE at +20 to +34 (96). However, mobility shift assays did not detect protein binding to this sequence. A perfect match to the canonical octamer sequence was found between −74 to −81 of the HCCP66 gene. A mobility shift assay detected interacting protein from cell lysates, but the binding was not tissue specific (96). In vertebrates, Oct-1 protein is broadly expressed throughout the body, whereas Oct-2 is restricted to specific cell types. Also in vertebrates, a related transcription factor, Pit-1, binds a sequence similar to the octamer and is instrumental in pituitary-specific gene expression where it controls expression of the prolactin and growth hormone genes (79, 119). A Pit-like recognition sequence has been detected in the cis-upstream region of a B. mori cuticular gene (117).

**Immune Response**

The insect immune response is a paradigm for the study of inducible gene expression. The insect acute-phase (injury or pathogen-mediated) response includes induction of antibacterial proteins and peptides (76). This peptide response is increasingly well understood at the level of transcriptional control of gene expression. Interest in the mechanism of the insect immune response partially stems from the compelling argument that “biological control strategies have not realized their promise because, at least in part, insects elaborate efficacious immune responses” (152, p. 1).

In *H. cecropia*, Sun *et al.* (156) cloned two genes that were found to encode a major group of antibacterial peptides called attacins. The two genes respond in concert to bacterial infection but otherwise are not expressed in the same pattern. An analysis of the 5′ upstream region of each gene revealed the presence of a κB-like sequence previously known from the promoters of vertebrate immune system genes. Sun *et al.* (156) suggested that the rapid insect immune response may be analogous to the vertebrate acute-phase response to infection typified by the appearance of a specific transcription factor, NF-κB (99).

Mobility shift assays identified a prospective *H. cecropia* acute-phase transcription factor, CIF, which is a protein present in bacterial-induced tissue extracts (154). The behavior of CIF is reminiscent of the vertebrate factor NF-κB; in particular, the lack of gene expression inhibition after addition of cycloheximide suggests the factor is activated from an existing precursor protein. Also, administration of phorbol-ester induces CIF, which implies the involvement of a second-messenger signal transduction pathway. Sun & Faye (155) used a κB affinity column to purify CIF.

Another immune response peptide, cecropin B, was cloned from *B. mori*. A probe from cecropin B was used to detect induced mRNA in fat body and hemocytes and to assess the effect of lipopolysaccharide (LPS)
kinetics on cecropin B gene expression (90, 91, 162). In the upstream gene region of the cecropin B gene, two consensus κB binding sites and an interleukin-6 (IL6) response element were identified (161).

Antibacterial peptide genes have been investigated in various Dipteran species (42, 109, 130). In the flesh fly, Sarcophaga peregrina, expression of a gene encoding an antibacterial peptide was investigated in naive (no acute-phase induction) individuals (109). It was observed that transient mRNA accumulated at both embryonic and pupal stages (109). Accumulation of mRNA from genes that encode two other defense peptides (sarcotoxin I and sarcotoxin II) supported the observation of expression of this class of genes during development in naive (non acute-phase) insects (4, 118).

Diptericin is an anti-bacterial peptide produced as part of the acute-phase response in D. melanogaster. As is characteristic of insect acute-phase response genes, the 5′-end noncoding region of the diptericin gene has DNA sequences that correspond to vertebrate κB response elements (62). In addition, IL6-like response elements and sequences that confer a response to interferon were found in the 5′-flanking DNA of the diptericin gene. Quantitative transgenic analyses of the 5′-end–flanking DNA region of the gene demonstrated that the two κB-like sequences are responsible for inducibility and tissue-specific expression (110). Other elements are required for a response beyond minimal stimulation. Specifically, the upstream sequence with similarity to an IL6 response element has been shown to significantly contribute to inducible gene expression (61). Moreover, it is increasingly clear that genes that control formation of the embryonic dorsal-ventral axis in D. melanogaster are integrally involved in the expression of immune response anti-microbial peptides (98, 131) in a manner that is analogous to the vertebrate immune response. In general, the immune systems of invertebrates and vertebrates are strikingly similar in the nature of the mechanisms employed for control of gene expression.

Storage Proteins

Storage proteins allow the study of genes that are proximally controlled by hormones and distally regulated by photoperiod, gender, and nutrition. The classes of hexameric storage proteins include JH suppressible, methionine rich, and arylphorins. The principle roles of storage proteins are to provide for cuticle formation and metamorphosis, as well as provision for adult non-feeding stages.

Storage protein genes have been investigated in the well-studied lepidopterans B. mori and M. sexta. In B. mori, the upstream region of a sex-specific storage protein (SP1) gene has regions of sequence similarity to an EcRE, and the first intron has sequences similar to a viral (SV40) enhancer (140). The upstream region of an arylphorin-type storage protein (SP2) gene also has a sequence similar to the SV40 enhancer and
another sequence associated with fat body–specific expression in *D. melanogaster* (54). A larva-specific *B. mori* storage protein (BmLSP) gene has a TGATAAA heptamer (56) that is typically found in the upstream region of storage protein genes (174). The late larval mRNA developmental profiles from a female-specific storage protein gene and a gender-un-differentiated storage protein gene were determined in *M. sexta* (136). Allatectomy, used in conjunction with administration of a JH analog, demonstrated that female-specific storage protein gene expression was specifically suppressed by JH (172).

Expression of storage protein genes and their endocrine control have been investigated in the waxmoth, *Galleria mellonella* (95, 111, 128, 129). No exact *D. melanogaster* EcRE sequences were found in the 5′-end sequence of an arylphorin gene (Lhp76), even though ecdysteroids have been shown to suppress Lhp76 gene expression (113). Lhp82, another storage protein gene, is suppressed by JH in the first four larval instars and by ecdysteroids when JH is absent in pupae. Nuclear run-off assays on debrained pupae demonstrated that the Lhp82 gene is transcribed in the absence of JH and 20HE (112). As in the case of the Lhp76 gene, no EcRE-like sequence was detected upstream of the Lhp82 gene (112).

Storage protein genes may have associated EcRE-like sequences. In the Colorado potato beetle, *Leptinotarsa decemlineata*, the mRNA from the diapause protein 1 gene disappears during metamorphosis and reappears during short-day conditions, which indicates that gene expression is under control of JH (40). Upstream of this gene is a sequence with similarity to an EcRE that also includes the storage protein gene consensus heptamer (93).

**Vitellogenin**

Vitellogenin gene expression provides a focus for the investigation of hormone transcriptional control. Vitellogenin is the principle protein used to provision the insect egg. The vitellogenin synthesized by most insects is a member of a conserved class of proteins that is present in organisms as divergent as nematodes and vertebrates (13, 148). *D. melanogaster* produces yolk proteins that are unrelated to vitellogenin (15). The *D. melanogaster* yolk protein genes are expressed in the fat body and ovarian follicle cells and are under control of regulatory sequences responsive to tissue-specific factors, 20HE, and nutrition-mediated signals (16).

In a study of vitellogenin in the migratory locust, *Locusta migratoria*, it was established that a JH analog stimulates vitellogenin mRNA accumulation in reproductive-age females whose corpora allata were destroyed (28). Two vitellogenin genes were cloned from *L. migratoria* (102). The genes were found to be coordinately expressed after injection of methoprene, a JH analog, into allotetomized females or into female fifth instars (41).
Comparison of 5′-end upstream sequences from *L. migratoria* vitellogenin genes (177, 178) and JH-controlled *Periplaneta americana* oothecin genes (124, 125) suggested the presence of a possible JH-response element sequence. It shares an octanucleotide sequence, AAGGGTTC, found in the *D. melanogaster* hsp27 gene and also bears similarity to half of a consensus steroid response element. Braun & Wyatt (18) tested upstream vitellogenin gene fragments for mobility shift with nuclear lysates and found that a fragment with the conserved octanucleotide sequence did not bind protein. However, mobility retardation of a different fragment suggested the presence of a juvenile hormone–induced factor (JHF) binding site in the vicinity of −610 to −549 (18).

An appreciable delay has been observed between the time of administration of a JH analog to *L. migratoria* and the appearance of vitellogenin mRNA. This time lag is increased when cycloheximide is co-administered, which suggests the possibility that activation of vitellogenin gene expression requires JH to first stimulate the synthesis of a transcription factor (47). Experiments conducted with an *in vitro* transcription system support the hypothesis that JH controls *L. migratoria* vitellogenin gene expression by transcription factor induction (180).

Additional evidence for an indirect mechanism of JH control of vitellogenin gene expression comes from the work of Agui *et al.* (3). In one set of experiments, decapitated female houseflies, *M. domestica*, were injected with increasing concentrations of 20HE, followed by Northern analysis that indicated vitellogenin mRNA accumulated in a dose-dependent manner. The response to hormone administration was maximal within 8 h after the time of initial exposure. When JHIII, or a JH-analog, was applied topically to decapitated females, vitellogenin gene expression was similarly stimulated, but it required 24 h for maximum response in the ovary.

Both 20HE (49) and JH (14) are capable of stimulating *A. aegypti* vitellogenin synthesis (14, 49). 20HE injected into decapitated females that had not blood-fed increased mRNA production to a lesser degree than hormone injection into blood-fed females, which implies that blood feeding has an additional effect on gene expression (60, 127). Using cDNA probes that correspond to vitellogenic carboxy peptidase (VCP) and vitellogenin genes, Deitsch *et al.* (39) tested whether the modulatory effect of 20HE on gene expression was direct or indirect. They found that cycloheximide suppressed VCP and vitellogenin gene expression after hormone administration. The implication is that 20HE indirectly acts to control expression of these genes.

**Expression Systems**

The most serious limitation for non-Drosophilid insect gene regulation studies is the absence of homologous *in vivo* expression systems. Homolo-
Heterologous transgenic expression using *D. melanogaster* has been useful for expression of genes from other insects, but more than occasionally the results are problematic. For example, no detectable expression existed in *D. melanogaster* after P-element transformation of a portion of the locust vitellogenin gene along with 5′-end–flanking DNA from the gene (181). Transfected *D. melanogaster* cells have also been used for non-Drosophilid insect gene expression. Abraham *et al.* (1) transfected the *B. mori* cytoplasmic actin gene and 2480 bp of sequence upstream from the actin gene transcription start site. After transfection, *B. mori* actin mRNA accumulated, but the relative amount of heterologous actin mRNA decreased after administration of the hormone 20HE, which is known to stimulate expression of the actin gene in *B. mori*. Abraham *et al.* (1) speculated that this decrease in heterologous mRNA may reflect the greater competitive ability of host-cell actin gene regulatory sequences for available transcription factors.

In spite of the evident problems, successes have occurred using heterologous expression undoubtedly because DNA regulatory elements and factors are often highly conserved in evolution. For example, Mitsialis *et al.* (116) used P-element transformation to introduce the 5′-end region of a *B. mori* chorion gene into the *Drosophila* genome in a construct that expressed bacterial chloramphenical transferase as a reporter. *B. mori* upstream sequences expressed CAT in the correct gender, tissue (ovary), and temporal pattern. Bello & Couble (8) used upstream sequences from the P25 silk gland gene to drive reporter gene expression after P-element transformation of *D. melanogaster*. The regulatory sequences expressed the reporter enzyme only in *D. melanogaster* larval salivary glands. Gut-specific fidelity of heterologous expression in *D. melanogaster* was demonstrated using a black fly, *Simulium vittatum*, carboxypeptidase gene (182) and a midgut trypsin gene from the vector mosquito, *Anopheles gambiae* (144). Although heterologous expression can be a valuable tool, it should be noted that *cis*-elements can be present and yet undetected simply because certain factors are missing in the heterologous organism (89). Clearly, a primary goal for non-Drosophilid insect gene regulation studies is the development of homologous whole-organism *in vivo* expression systems. This is an attainable goal, as indicated by the recently achieved success in germline transformation of the medfly, *Ceratitis capitata* (103).

**Perspectives**

Here we advocate three points of view. 1. Transcriptional control of gene expression should be viewed as a quantitative process. 2. Physiol-
ogy may be usefully conceived in terms of transcriptional control of gene expression. 3. An evolutionary perspective is valuable for understanding control of gene expression.

**Qualitative vs Quantitative Transcriptional Control**

The regulation of transcription is often thought of in terms of discreet processes. However, transcriptional control will often be revealed to be a quantitative process when gene regulation is studied in sufficient depth. For example, in the study of sea urchin early development, Calzone et al. (20) used gel shift assays for quantification of factor affinity for a DNA fragment, for measurement of the stability of factor-DNA association, and to estimate the number of factors present in a given cell type. Results from the study reveal that transcriptional control of gene regulation is quantitative in nature. The following quote encapsulates this perspective: “*cis*-regulatory modules, it is now clear, come in many different designs. Although as a field we tend to regard them qualitatively as switches, they are of course actually quantitative mechanisms for which the decisive variables are the intranuclear concentrations of the relevant active transcription factors” (38, p. 612).

**Physiology at the Level of Transcriptional Control**

Simplistically, two fundamental tenets of physiology are response to stimuli and homeostasis. Physiology can be reformulated in terms of response at the level of transcription (response to stimuli) and feedback that modulates transcriptional control of gene expression (homeostasis). Clearly, some physiological responses occur on a time scale that precludes gene expression from playing a major role in physiological responses, and obviously, tissue and organism levels of biological organization must be included to understand physiological processes. Nevertheless, much may be gained from introducing the conceptual framework of “transcriptional physiology” and the expedient inclination to work “backward” from patterns of gene expression for the purpose of studying of physiological processes.

Transcriptional physiology should be based in part on transgenic experiments. For example, in humans and mice, the regulatory enzymes in glycolysis and gluconeogenesis have been cloned and promoters characterized. As a result, it is possible to investigate hormonal and dietary regulation of metabolic pathways in terms of gene expression in transgenic mice (104). Another focus of transcription physiology will be a general understanding of what regulates the regulatory factors (97). Regulation of transcription factor abundance could result from control of factor gene expression, differential splicing of factor mRNA, formation of factor complexes, and factor activation (or inactivation) by chemical modification. A logic underlies transcription regulatory circuits, and eventually generalization about homeostatic processes may be possible in terms of this logic.
Once we conceive of a transcriptional physiology, numerous questions can be asked. Does response to stimuli usually involve a hierarchical pattern of gene regulation, or do genes respond relatively directly to stimuli? Does co-expression of genes that are responding to a particular environmental stimulus identify the genes that are involved in physiological adaptation to that environmental state? Is feedback regulation of transcription a prevalent basis for homeostasis? Does long-term acclimation result from regulation of transcription factor abundance? Does a rapid response to stimuli, such as the acute-phase immune response, typically result from the activation of pre-existing inactive transcription factors? Is physiological adaptation to environmental change often the result of an evolved inducible response? The process of extending the study of physiology to the level of transcriptional control of gene expression is exciting and inevitable.

The Value of an Evolutionary Perspective

An evolutionary perspective provides a basis for understanding between species patterns of transcriptional control of a gene. For example, the expression of glucose dehydrogenase (Gld) has been studied in D. melanogaster and other Drosophila species (23, 24). Gld is instrumental in modifying the puparium to permit eclosion. The enzyme is synthesized at each larval molt in the epidermis of both sexes under control of ecdysone and tissue factors. In addition, Gld is expressed in various tissues at the adult stage, and the pattern of gene expression varies markedly among Drosophila species. Experimental alteration of gene expression of the sex determination hierarchy in D. melanogaster changes the tissue-specific pattern of Gld expression in a manner that mirrors the tissue variation in expression among species of Drosophila (51). The argument is thus made that variation in adult Gld expression among Drosophila species merely reflects variation in expression of the sex-determination hierarchy genes. The pattern of adult Gld gene expression is viewed as incidental rather than adaptive.

In general, interpretation of patterns of gene expression may be very difficult without recourse to an evolutionary interpretation. Another example that supports this perspective is derived from a comparison of gene expression in humans and Drosophila. The alcohol dehydrogenase gene (Adh) of D. melanogaster is activated by the transcription factor C/EBP, and another factor, AEF-1, is capable of specifically suppressing the effect of C/EBP in the fat body (48). The insect fat body is sometimes described as functionally analogous to the vertebrate liver, and seemingly accordingly, C/EBP and AEF-1 control expression of Adh in human liver. However, these human and Drosophila Adhs are evolutionarily unrelated. Moreover, the fat body arises from mesoderm, whereas the liver is derived from endoderm. It is known that C/EBP is typically present in adipose tissue (21, 30), and perhaps in the course of evolution unrelated Adhs were “recruited” by C/
EBP and AEF-1 into different adipose tissues (fat body and liver). The suggestion that investigators include an evolutionary perspective as one of their research tools is best advocated by the argument that “nothing in biology makes sense except in the light of evolution” (43).

Acknowledgments
We thank K Kumaran, M Goldsmith, C Brown, C Nelson, S Quisenberry, and J Wingrove for reading various drafts of the manuscript. We also thank L Riddiford, K Iatrou, L Miller, B Rybczynski, and M Learned for helpful conversations. This review had its inception at a Molecular Insect Science Meeting symposium organized with generous assistance from H Hagedorn. General and genial advice from L Gilbert is also appreciated. This research was supported by grants from the John D and Catherine T MacArthur Foundation, Burroughs Wellcome Fund, and National Institutes of Health NIAID to AA James.

Literature Cited


Differential Gene Expression in Insects: Transcriptional Control