

Mark D. Biggin · Robert Tjian

Transcriptional regulation in *Drosophila*: the post-genome challenge

Received: 24 July 2000 / Accepted: 30 July 2000 / Published online: 24 October 2000
© Springer-Verlag 2000

Abstract *Drosophila melanogaster* has long been at the forefront of studies of transcriptional regulation in animals. Many fundamental ideas – such as *cis* control elements that act over long distances, the regulation of development by hierarchical cascades of transcription factors, dosage compensation, and position effect variegation– originated from studies of the fruit fly. The recent completion of the euchromatic DNA sequence of *Drosophila* is another breakthrough.

The sequence data highlight important unanswered questions. For example, only one-fifth of the 124 Mb of *Drosophila* euchromatic DNA codes for protein. The function of the remaining 100 Mb of mostly unique DNA is largely unknown. Some proportion of this non-reading frame DNA must encode the functional recognition sites targeted by the approximately 700 sequence-specific DNA binding proteins that regulate transcription in *Drosophila*, but what proportion? Most or very little? Promoter sequences by definition contain all of the *cis* information that specifies how gene transcription is regulated. However, it has been difficult to decipher this information and predict the patterns of RNA expression. How do we break this “transcriptional code”? Mechanistic studies, using simple model promoters, indicate that transcription is controlled by the coordinate action of sequence-specific DNA binding proteins interacting with the general transcriptional machinery via intermediary adapters and chromatin remodeling activities. How can we integrate this biochemical information with data from genome-wide studies to describe the generation of highly

complex patterns of transcription? Here, we discuss recent studies that may point the way ahead. We also highlight difficulties that the field faces in dissecting transcriptional control in the post-genome era.

Keywords *Drosophila* · Transcription · Protein/DNA recognition · Animal development

Why *Drosophila*?

Transcriptional regulation is vastly more complex in animals than it is in microbes. In microbes, there are 800–6,000 genes to be regulated, their promoters are generally limited to 100 bp–200 bp regions around the transcription start site, each promoter is typically controlled by only 1–4 sequence-specific DNA binding proteins, and a given microbial species will exist in only 1–3 cell or spore types. By contrast, in animals there are 14,000–80,000 genes, many promoters are spread over tens of kilobases, promoters may be regulated by 40 or more sequence-specific DNA binding proteins, and an animal generally has 80–250 distinct cell types. These differences, however, still understate the increased complexity of animals. Studies of the expression of randomly selected genes in *Drosophila* embryos suggest that even among cells of the same cell type, most mRNAs are expressed at different levels (Kopczynski et al. 1998; Liang and Biggin 1998). These differences are highly reproducible between embryos. Thus, virtually every cell in an animal may have a unique combination of gene transcription rates that is tightly regulated. In addition, there are constant changes in transcription during development and in response to alterations in the environment.

Complexity of this magnitude could be taken as a reason to avoid studying transcriptional control in animals and to focus instead on microbes. This view, however, misses the point. Truly complex multicellular life evolved only twice, giving rise to the plant and the animal kingdoms (Margulis and Schwartz 1998). The chief distinction between these organisms and the microbes

M.D. Biggin (✉)
Genome Sciences Department,
Lawrence Berkeley National Laboratory,
1 Cyclotron Road, Mailstop 84–171,
Berkeley, CA 94720, USA
e-mail: mdbiggin@lbl.gov
Fax: +1-510-4864229

R. Tjian
HHMI, Department of Molecular and Cell Biology,
401 Barker Hall 3204, University of California,
Berkeley, CA 94720–3204, USA

Most classes of sequence-specific regulatory proteins found in *Drosophila* are also present in vertebrates and nematodes. Many, however, are not found in yeast. For example, nuclear steroid receptors (20), paired box proteins (11), Hox proteins (8), POU proteins (5), T-box proteins (8), ETS factors (8), CTF-1/NF-1 (1), AP-2 (1), and runt proteins (4) are all found in fruit flies, worms, and vertebrates; none of these proteins is found in yeast (Chervitz et al. 1998; Adams et al. 2000; Rubin et al. 2000; <http://www.ebi.ac.uk/proteome>). (The numbers in brackets are the numbers of genes encoding each class in *Drosophila*, both in the above and subsequent examples.) The DNA binding domains of many animal transcription factors are homologous to those of yeast proteins— for example, C₂H₂ zinc fingers (352), HLH (61), and homeodomains (113). Very few of the yeast proteins, though, appear to be direct orthologues that serve the same biological functions as the structurally similar factors in animals. Indeed, the highly conserved heat shock transcription factor is the only example of a factor that regulates the same target genes in response to the same stimulus in yeast and animals that we are aware of (Wu 1995). Instead, the homology between yeast and animal factors appears to result largely from the duplication, divergence, and reuse of effective protein folds, which is so common in biology.

By contrast, specific orthologues of many sequence-specific regulatory factors have retained the same biological functions across animal phyla. For example, the murine and *Drosophila* orthologues of the Hox genes *Deformed*, *Labial*, *Proboscipedia*, *Antennapedia*, and *Abdominal B* are expressed in the same order along the anterior/posterior axis of mouse and fly embryos (Burglin 1994). In fact the vertebrate Hox proteins can replace the functions of their specific orthologues when expressed in the fly (e.g., McGinnis et al. 1990). It is also common to find that signal transduction pathways act via the same class of sequence-specific DNA binding proteins in different animal phyla. For example, the STAT transcription factors are downstream of the Jak kinases in flies and vertebrates (Luo et al. 1999), Su(H)/CBF1 proteins are activated by Notch signaling in flies and vertebrates (Artavanis-Tsakonas et al. 1999), Juns are phosphorylated by JNKs (Kockel et al. 1997), and LEF1-like factors are downstream of Wnts (Riese et al. 1997). At the risk of overemphasizing a point, of the 16 broad families of signal transduction pathways found in vertebrates, 15 are found in *Drosophila*, but only one in yeast (Gerhart and Kirschner 1997).

Sequence-specific transcription factors serve a broad range of functions in animals. Five of the most prominent ways that these proteins act are:

1. Tissue specific. Some factors are expressed in only one or two cell types and are often thought to bind and regulate only genes whose expression is characteristic of that tissue – for example, the muscle-specific MyoD family of proteins, which include the *Drosophila* gene *nautilus* (Keller et al. 1998 and ref-

erences therein). In effect, these tissue-specific factors determine and distinguish cell types.

2. Position specific. Many factors are expressed not in a tissue-specific manner but in localized regions within the animal. Such proteins are expressed in many cell types and generate the differences in transcription and morphology seen between cells of the same type. For example, the Hox protein Ubx is expressed in the third thoracic segment of the developing *Drosophila* embryo. Ubx instructs the cells in this region to divide and differentiate to form a small segment that bears a pair of balancing organs, halteres. In mutant animals lacking Ubx, these cells develop into a much bigger segment bearing a large pair of wings (Lawrence 1992). To a developmental biologist, such proteins specify “positional information” (Wolpert 1991).
3. Time specific. The expression of most tissue-specific and position-specific regulatory proteins varies significantly during development. Thus, these proteins also provide temporal information to the system. There are other factors, however, whose chief function is to specify time. The circadian rhythm factors are an obvious example (Darlington et al. 1998). The ecdysteroid response factors, which coordinate the timing of events such as molting and metamorphosis, are another (Segraves 1994; Thummel 1997).
4. Environmental response. Many sequence-specific DNA binding proteins alter transcription of their cognate targets in response to changes in the environment. Some factors are dedicated to this one function (e.g., the heat shock transcription factor; Wu 1995), but the activities of many tissue-, position-, and time-specific factors are also modified in response to changing environmental cues – for example, during jet-lag.
5. Ubiquitous specificity. A broad group of factors do not have a single, easily defined biological function. They tend to be expressed in most or all cells, though their levels of expression may vary somewhat among cells and over time. They bind to only a subset of genes, but the genes bound are a broad and eclectic mix. Null mutations in these proteins are often lethal, but give no specific phenotype. Examples of such factors in *Drosophila* include the GAGA factor and Adf1 (England et al. 1992; Soeller et al. 1993; Farkas et al. 1994). It is thought that these proteins increase the combinatorial possibilities available for other sequence-specific regulatory proteins. As a hypothetical example, if some of the target genes of a tissue specific factor, such as *nautilus*, are bound by GAGA, and not Adf1, and other of its targets are bound by Adf1, not GAGA, then if *nautilus* activates GAGA-bound targets more potently than Adf1-bound targets, it would have acquired greater regulatory potential.

Most functional genomic studies that focus on gene regulation will probably concentrate on sequence-specific DNA binding proteins. As we discuss below, however,

without considerable progress in our understanding of the other transacting factors and the interaction with chromatin, we will not know how the whole system is integrated and operates.

The general transcription factors. The so-called general transcription factors play a pivotal role in directing transcription and are the smallest group of eukaryotic transcription factors. The members are usually said to be TFIIB, TFIID, TFIIE, TFIIIF, TFIIH, and RNA polymerase II, but we also include in this class elongation and termination factors such as P-TEFb, S-II, and factor 2 (reviewed by Orphanides et al. 1996; Hara et al. 1999; Price 2000). Although under certain conditions TFIIE, F, and H are not essential for initiating transcription *in vitro* (Parvin and Sharpe 1993; Tyree et al. 1993), most or all of the general factors are probably necessary for or play a significant role in the transcription of most protein-encoding genes *in vivo* (Holstege et al. 1998). The general transcription factors are highly conserved throughout the eukaryotes: specific orthologues of each show very high conservation of polypeptide sequence and subunit composition from yeast to man. TFIIB-H and RNA polymerase II form a complex around the start site of transcription – between nucleotides –39 bp to +32 bp – with different general factors recognizing distinct sequence elements within this region (Fig. 1; Lagrange et al. 1998; Kutach and Kadonaga 2000).

TFIIB, E, F, H, RNA polymerase II, and the elongation and termination factors are expressed in all animal cells, and show only moderate changes in expression between cells (e.g., Reynaud et al. 1999). Thus, these proteins may have a largely constitutive role – perhaps being used to vary the transcription rate of all protein encoding genes, but not of specific subsets of these genes.

The role of TFIID is more complicated. The canonical TFIID comprises a core polypeptide, called TBP, and a set of tightly bound TBP-associated factors (TAFs) (Tjian and Maniatis 1994; Burley and Roeder 1996; Goodrich et al. 1996). TBP is important for basal transcription and binds to the TATA box sequences found in some, but not all, promoters between –30 bp and –20 bp. The TAFs serve two main functions. One, they bind to the activation and repression domains of sequence-specific DNA binding proteins, mediating the interaction of these proteins with the general transcription factors. Two, some TAFs also make specific DNA contacts with two additional core promoter sequences: the initiator element between –2 bp and +4 bp and the downstream promoter element (DPE), between +28 bp and +32 bp (Fig. 2; Sypes and Gilmour 1994; Verrijzer et al. 1994; Kutach and Kadonaga 2000; <http://www-biology.ucsd.edu/labs/kadonaga/DCPD.html>). Further, although TBP and a core group of TAFs are expressed in all cells, other TAFs are expressed only in a subset of cells (Dikstein et al. 1996). The *Drosophila* TAFs *canonball* and *no-hitter*, for example, are only expressed in testes (M.T. Fuller, personal communication). Thus, in addition to the ubiquitously expressed canonical TFIID

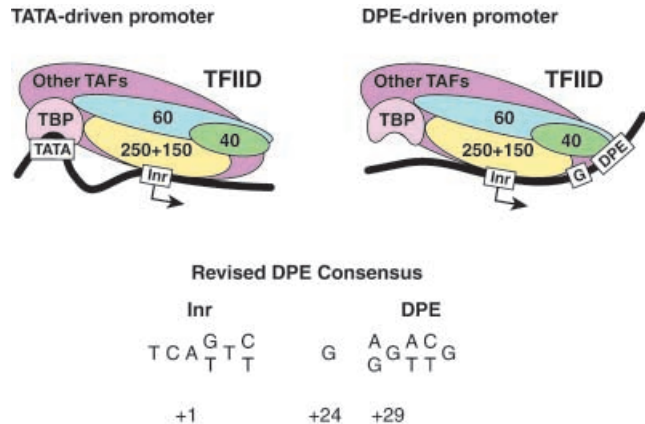


Fig. 2 A model of two distinct interactions of TFIID with TATA-versus DPE-driven core promoters taken from Kutach and Kadonaga (2000). The DPE consensus shown is 1 bp shorter than that given by Kutach and Kadonaga, based on their more recent experiments (J.T. Kadonaga, personal communication)

complex, there may be additional TFIID-like complexes containing unique collections of TAFs that are spatially restricted.

A further elaboration of the core transcriptional machinery is that there are TBP-related factors (TRFs), which are also expressed in a spatially and temporally restricted manner (Crowley et al. 1993). The TRFs are likely to be part of further, distinct TFIID-like complexes, at least one of which has preferences for core promoters that differ from those of the canonical complex (Holmes and Tjian 2000). The canonical TFIID complex may, like the other general transcription factors, provide limited gene-specific regulatory information to the system. The other TFIID-like complexes, however, appear to be hybrids since they differentially affect the expression of a subset of genes in the cells in which they are expressed – as do many sequence-specific regulatory proteins – yet they act via core promoter regions – as do the general transcription factors.

Cofactors: adapters and chromatin remodeling factors. Cofactors, or coregulators, as they are also known, do not bind to specific DNA elements and are not essential for basal transcription. Instead, they interact with sequence-specific DNA binding proteins and modify their ability to activate or repress transcription (Fig. 1; Bjorklund et al. 1999; Mannervik et al. 1999). Importantly, each cofactor interacts with a unique subset of sequence-specific regulatory factors and most act on only a subset of promoters. There are two principal means by which cofactors act: as adapters or as chromatin remodeling factors.

Adapters serve as physical bridges that help to integrate transcription signals; the activation or repression domains of sequence-specific DNA binding proteins often bind directly to adapters, which in turn bind to specific components of either the general transcriptional machinery or chromatin remodeling proteins. In the case

of activation, such interactions may recruit general factors to the promoter or make conformational changes to the macromolecular apparatus, leading in either case to an increase in the rate of transcription (Tjian and Maniatis 1994; Goodrich et al. 1996; Ptashne and Gann 1997).

Chromatin remodeling factors, on the other hand, affect transcription by altering the accessibility of DNA sites in chromatin (Kingston and Narlikar 1999; Brown et al. 2000). The general transcription factors and some sequence-specific regulatory factors bind poorly to DNA that is wrapped around an unmodified histone octamer (e.g., Workman and Roeder 1987; Taylor et al. 1991; Imbalzano et al. 1994). Other sequence-specific regulatory factors, however, bind efficiently to such nucleosomal DNA (e.g., Taylor et al. 1991). When the latter type of sequence-specific DNA binding protein recruits chromatin remodeling proteins to promoters, the accessibility of nearby sites can be altered, increasing the local interaction of those factors that otherwise bind poorly to nucleosomal DNA (Cosma et al. 1999). Some remodeling activities affect accessibility by an ATP-dependent mechanism – perhaps by creating looped DNA domains that span several nucleosomes, which may partly unwrap the DNA from the histone octamers (Bazett-Jones et al. 1999). Other chromatin remodeling activities increase accessibility by acetylating lysine residues in the N-terminal regions of the four core histones, lowering the net positive charge of the histones, which may allow partial disassociation or loosening of the DNA from the octamer. Specific acetylation of nucleosomes may also function as a molecular tag that is selectively recognized and bound by transcription factors containing a bromodomain (Dhalluin et al. 1999; Jacobson et al. 2000). Perhaps not surprisingly, histone deacetylases are used as repressive cofactors that decrease chromatin accessibility (Knoepfler and Eisenman 1999). In many animals, including some arthropods and all vertebrates, cytosine methylation and methyl-binding histone deacetylases are also used to reduce DNA site accessibility, but fruit flies lack this particular mechanism (Bird and Wolffe 1999; Colot and Rossignol 1999; Walsh and Bestor 1999).

Adapters are a heterogeneous group. They include the TAFs since although a TFIID type complex is essential for basal transcription, the TAFs themselves are not. Adapters also include other large protein complexes that associate with the general transcription factors less tightly than the TAFs, e.g., the mediator-like complexes found in yeast, flies, and vertebrates (Bjorklund et al. 1999). Yet other adapters associate with particular families of sequence-specific DNA binding protein, e.g., the Groucho family of corepressors and the retinoblastoma proteins (Rbs) (Fisher and Caudy 1998; Chen et al. 1999; Mannervick et al. 1999). Not surprisingly, such adapters tend to be conserved among animal phyla but absent from yeast, paralleling the conservation of the sequence-specific DNA binding proteins with which they interact. It is difficult to estimate the total number of adapter complexes in *Drosophila*. There are perhaps somewhere between 20 and 30 different complexes.

Chromatin remodeling proteins are also a diverse group. To date, at least three ATP-dependent remodeling complexes have been identified in *Drosophila*: dSWI/SNF, NURF, and ACF/CHRAC, each of which contains 4–8 polypeptides (Kingston and Narlikar 1999). Similar complexes are found in yeast and vertebrates, though significant changes in subunit composition have occurred across phyla. At least one, and perhaps several, large histone acetylase complexes are conserved from yeast to vertebrates, including the SAGA complex (Brown et al. 2000). TAF 250 is a histone acetylase, and the *Drosophila* homologue of the CBP histone acetylase has been shown to be a cofactor for Dpp signal transduction (Waltzer and Bienz 1999). Finally, at least two distinct histone deacetylase complexes involved in transcriptional repression are conserved from yeast to vertebrates – the Sin3-like complexes and the Mi-2/NuRD-like complexes – but other deacetylases probably exist as well (Knoepfler and Eisenman 1999).

Throughout our description of cofactors and general transcription factors, we have largely discussed complexes shown to be distinct entities by biochemical experiments. One point we have ignored until now is that a number share some polypeptides. For instance, several of the histone acetylase complexes contain a subset of the TAFs found in TFIID, while also containing other unique polypeptides (Bjorklund et al. 1999; Brown et al. 2000). Also, at least two of the *Drosophila* ATP-dependent remodeling complexes contain the ISWI polypeptide, though they share no other subunits (Kingston and Narlikar 1999). The repeated use of the same gene product in different complexes is, of course, an efficient way to build many distinct *trans*-acting factors from fewer polypeptides via combinatorial mixing and matching. This same tendency, however, makes it harder to predict the full complement of factors expressed in an organism from sequence homology alone, highlighting the importance of studying an organism for which extensive functional data are available.

In general, cofactors are ubiquitously expressed in most animal cells (e.g., Rb and Groucho), though there are exceptions, as noted above. This ubiquity does not mean, however, that their activity is constitutive. To give just one example, Rb is differentially phosphorylated in a cell-cycle-dependent manner, altering its ability to interact with E2F transcription factors and controlling a key step of the cell cycle (Sellers and Kaelin 1996). But even if some cofactors are constitutively active in all cells, they still interact with a unique subset of sequence-specific DNA binding proteins and thus are recruited to a subset of genes. Therefore, by analogy to the “ubiquitous specificity” class of sequence-specific regulatory protein discussed earlier, mixed arrays of cofactors should vastly increase the diversity of transcriptional responses. We suspect that the proliferation and elaboration of cofactors played an important role in the evolution of complex multicellular organisms.

Problems remaining

The majority of *Drosophila* transcription factors have most likely been identified. We already know the expression patterns of many of these proteins, and through the ongoing efforts of the *Drosophila* Genome Project, we will shortly know the expression profiles of them all (<http://www.fruitfly.org>). The DNA sequence of every target gene regulated by these proteins has, of course, now been determined. In principle then, if we knew enough about each transcription factor, we could predict which genes each factor binds and regulates and, eventually, the patterns of expression of all genes. Sadly, we are very far from achieving this goal. A number of problems stand in our way. Below, we present what we see as the major issues from a biochemist's perspective. We do not discuss the computational challenges faced by bioinformaticists, but we hope the points that we raise here will be useful to them. For a discussion of these additional challenges, see Crowley et al. (1997), Reinitz et al. (1998), Wasserman and Fickett (1998) and Hughes et al. (2000).

To develop predictive models, we believe that it is essential to adopt a quantitative, system-wide approach, as distinct from the qualitative, one-promoter-at-a-time approach that has dominated the last 30 years. Transcription factors in vivo are in thermodynamic equilibrium with all potential DNA binding sites in the nucleus (von Hippel et al. 1974; Lin and Riggs 1975; Yang and Nash 1995; Carr and Biggin 1999). Therefore, to predict if a transcription factor will bind to any particular promoter in vivo, a number of issues must be considered, including the concentration of the factor in nuclei and the number of potential binding sites accessible within the genome. For example, even the highest affinity sites will not be bound at functionally significant levels if the free concentration of the protein is too low because most molecules are bound throughout the genome at other high-affinity sites. Quantitative data are necessary because, as we describe below, proteins do not bind to DNA in an all or none fashion. Because the different levels at which genes are bound correlate with function, quantitation is essential.

It is not enough, therefore, to identify a handful of specific sites to which a transcription factor binds. The range of sites bound in the organism must be found. Knowing this range, it is then natural to ask what fraction of the specific sites bound are functional target sites. With this information, a picture of how the transcription factor impacts the biology of the animal can also be built. Below, we discuss in turn the problems of determining which sites are bound in vitro, which sites are bound in vivo, and which sites are functional. Additional challenges presented by the regulatory network coordinating development are also discussed. We emphasize the importance not only of enumerating all interactions and regulatory events that occur, but also of deciphering the underlying principles that determine these interactions and events.

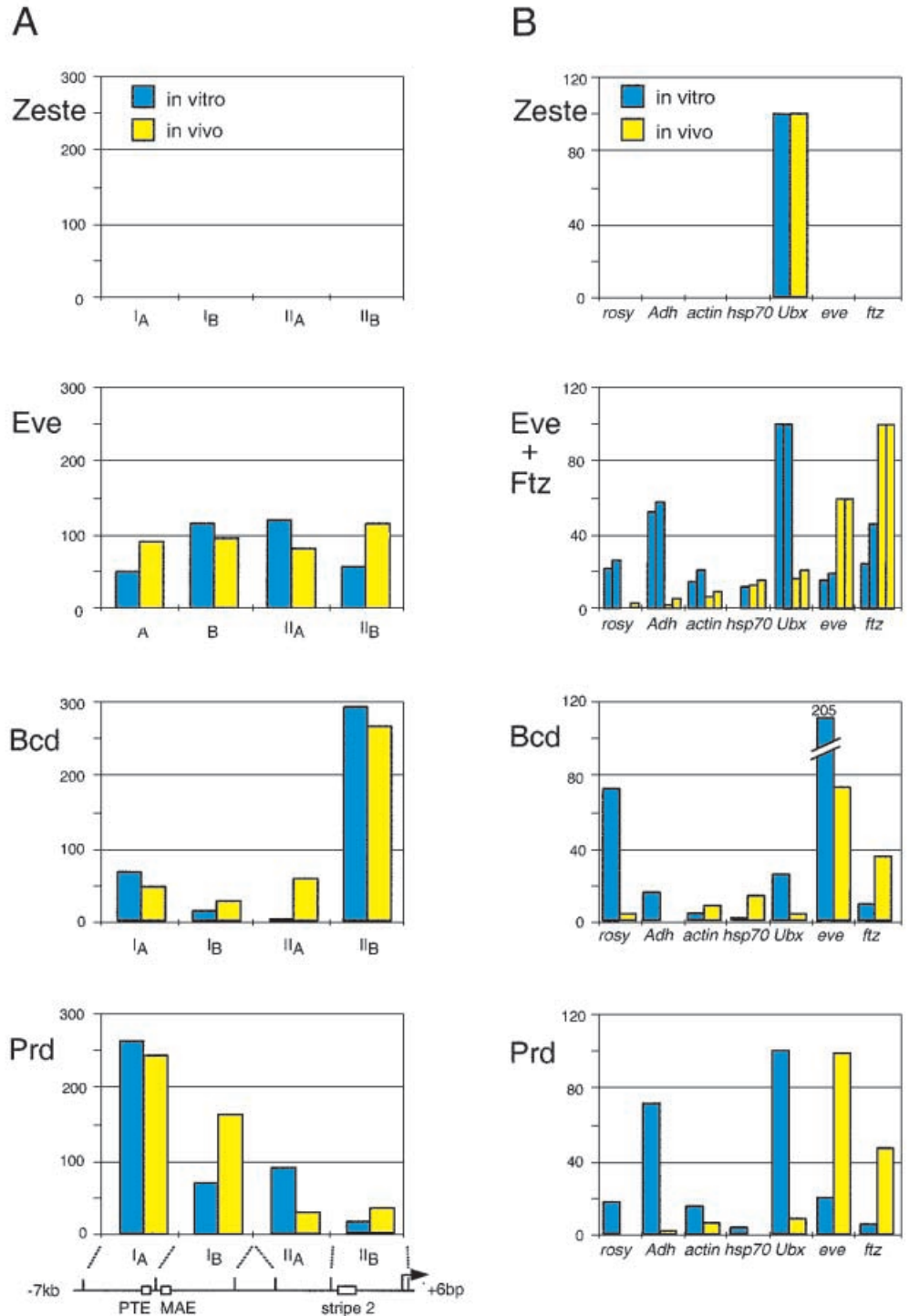
The range of DNA sequences bound in vitro. Given the affinities of a sequence-specific DNA binding protein for all variants of its recognition site, the degree to which any DNA fragment is bound in vitro by the purified protein should be predictable from sequence data alone. It might be thought that such predictions are already possible for many transcription factors, but very few proteins have, in fact, been studied in sufficient detail to allow such calculations. Consensus binding sites have been proposed for many factors (Faisst and Meyer 1992; Heinemeyer et al. 1998; <http://transfac.gbf.de/TRANSFAC>), but most analyses have been qualitative, seeking to identify the highest affinity sites and largely ignoring sites bound only several fold more weakly.

A further problem results from homomeric cooperative interactions between molecules. Frequently, DNA binding preferences are determined by selection of sites from pools of random oligonucleotides using PCR-based methods (e.g., Pollock and Treisman 1990; Ekker et al. 1991). The randomized portion of the oligonucleotides used tends to be only 20 bp–30 bp long. For technical reasons, it is not uncommon for sites to be selected using only the DNA binding domain, rather than the full-length protein. Unfortunately, many transcription factors bind as multimers to DNA sites separated by more than 20 bp, and domains that mediate these homomeric protein/protein interactions often lie outside of the DNA binding domain. To give one illustration of how misleading a typical PCR binding site selection experiment might be, the isolated DNA binding domain of the *Drosophila* transcription factor Zeste binds strongly to sites of sequence (C/T)GAG(C/T)G. The full-length Zeste protein, by contrast, does not bind well to such isolated consensus sites but requires at least three consensus sites for efficient binding, each separated by 18 bp–60 bp (Chen and Pirrotta 1993). In other words, intact Zeste protein binds in vitro with high affinity to a much smaller percentage of gene fragments than does its DNA binding domain alone, and the sites it interacts with are longer than those detectable in typical binding site-selection experiments.

Homomeric cooperative interactions are actually a double-edged sword. Not only can they increase the selectivity of DNA binding, as in the above example, but they can also have the opposite effect. The homeoprotein Eve, for instance, binds efficiently to DNA sites that poorly match the consensus, providing that these “weak” sites are within a few hundred base pairs of high affinity consensus sites (TenHarmsel et al. 1993). Perhaps because the homomeric interactions between Eve molecules are relatively weak, they do not significantly increase binding to consensus sites. In effect, homomeric cooperativity decreases the selectivity of binding of Eve by increasing the total range of DNA sequences bound. Therefore, it seems that a thorough quantitative analysis of in vitro DNA binding specificities requires several experimental approaches to ensure a complete understanding.

Predicting transcription factor DNA binding in vitro is particularly difficult in eukaryotes, both multicellular

Fig. 3A, B Comparison of the in vitro (blue) and in vivo (yellow) DNA binding preferences of Zeste and four homeoproteins: Eve, Ftz, Paired (*Prd*), and Bicoid (*Bcd*). **A** The relative in vitro and in vivo DNA binding specificities of Paired, Bicoid, and Eve for four fragments of the *eve* promoter are broadly similar. Zeste does not bind to this promoter and Ftz binds similarly to Eve. At the bottom is a diagram of the *eve* gene, indicating the four restriction fragments, the paired target element (PTE), the minimal autoregulatory element activated by Eve (MAE), the Bicoid responsive stripe 2 element (*stripe 2*), and the mRNA start site (arrow). **B** The in vitro and in vivo DNA binding preferences of four homeoproteins differ when compared across differently transcribed genes. DNA binding per kb of DNA to a series of DNA fragments from several different gene loci are shown. The fragments range in size from 1.5 kb to 8 kb. The *eve* and *ftz* genes are strongly transcribed at the stage of development at which in vivo binding was assayed; the other genes are either not transcribed or are only weakly transcribed at this stage. Binding of Zeste has not been assayed in vivo to the *hsp 70* and *Adh* genes. This figure was adapted from Carr and Biggin (1999)



and microbial, because factors in these organisms generally bind to shorter more degenerate sequences than do typical prokaryotic proteins. It is not uncommon for prokaryotic factors to recognize sites of 14 bp–18 bp (Robinson et al. 1998; <http://arep.med.harvard.edu/dpinteract>), whereas many eukaryotic factors bind 5 bp–8 bp sequences (Faisst and Meyer 1992; Heinemeyer et al. 1998; <http://transfac.gbf.de/transfac>). The sheer frequency of these shorter sites in the genome enhances the pos-

sibility of error. In animals, including *Drosophila*, this problem is compounded by the much larger size of the genome, a greater proportion of which is non-protein coding, substantially increasing the number of potential sites to be predicted.

The range of DNA sequences bound in vivo. Even when accurate predictions of DNA binding in vitro are available, it is likely that conditions in vivo modify the DNA

binding specificities of most transcription factors. Predicting which target genes a factor binds *in vivo*, therefore, requires that we understand the forces modifying binding and the degree of their effect. To provide a framework for our discussion of this issue, we first describe the results of a unique comparison of the *in vitro* and *in vivo* DNA binding specificities of five different *Drosophila* transcription factors (Fig. 3; Walter et al. 1994; Laney and Biggin 1996; Carr and Biggin 1999).

Zeste binds *in vivo* with high specificity only to those promoter regions to which it binds strongly *in vitro* (described earlier). The relative occupancy of these sites *in vivo*, however, differs up to tenfold from the relative levels of occupancy *in vitro*. The homeoproteins Eve, Ftz, Paired, and Bicoid have much broader recognition properties. They bind to DNA sites throughout the length of most genes, both in *Drosophila* embryos and *in vitro*. There is a good correlation between the *in vivo* and *in vitro* DNA binding specificities of each of these homeoproteins on actively transcribed genes (Fig. 3A), but there is no such correlation on inactive genes. These genes are bound at least tenfold more weakly *in vivo* than would be predicted from *in vitro* affinities (Fig. 3B).

What do the above results suggest about the forces influencing transcription factor DNA binding *in vivo*? First, they imply that conditions in cells do not radically alter DNA binding specificity, at least for these five proteins, but instead they quantitatively modify it. Second, in the case of the homeoproteins, a possible explanation for the data is that chromatin structure may specifically reduce binding on inactive genes. This view is supported by considerable evidence that chromatin structure is modified as genes become active and by data showing that in isolated nuclei, homeoprotein DNA sites are less accessible to restriction enzyme digestion in inactive genes than they are in active genes (Wallrath and Elgin 1995; Beato and Eisefeld 1997; Wolffe 1998; Carr and Biggin 2000). Unfortunately, the precise degree to which chromatin structure affects homeoprotein/DNA interactions *in vivo* is difficult to assess – estimates vary from a mean effect of two- to tenfold.

It is also difficult to predict how chromatin affects DNA binding of other classes of factor. As we described previously, some proteins bind efficiently to nucleosomal DNA without the aid of chromatin remodeling factors, but these data do not preclude the possibility that higher order structures may influence interactions of these proteins at some sites. Zeste binds *in vivo* with equal efficiency to promoter constructs whether they are transcriptionally active or inactive (Laney and Biggin 1997). Additionally, restriction enzyme accessibility assays indicate that unlike homeoprotein recognition sites, sites recognized by other sequence-specific regulatory proteins are not differentially inhibited at most sites on inactive genes (Carr and Biggin 2000). Thus, for many proteins, chromatin structure may modify binding at only a minority of sites and predicting which will require a better understanding of chromatin remodeling than we currently possess.

The other major factor modifying DNA binding *in vivo* is heteromeric cooperative association between proteins. Interactions between different transcription factors frequently increase the affinity of one or both proteins to individual DNA sites *in vitro*, and in a few instances, binding to some sites *in vivo* does depend on these interactions (Johnson 1992). Based on these data, it is frequently assumed that the DNA binding specificity of many transcription factors will be much more selective *in vivo* than *in vitro*. In other words, it is assumed that factors are expressed at such low levels in cells that they cannot bind efficiently to *any* specific sites without heteromeric cooperativity – even sites that are not blocked by chromatin structure – and that, as a consequence, only recognition sites adjacent to those of heterodimer partners will be bound efficiently. There probably are proteins for which this is true, but measurements of *in vivo* DNA binding have not borne out these predictions to date. For example, homeoproteins such as Ftz and Eve have been proposed to bind highly specifically *in vivo*; each protein being thought to bind to different sets of genes (Mann and Chan 1996; Yu et al. 1997), but direct measurement of DNA binding by these proteins *in vivo* suggests that they bind with similar specificity to the same targets (Walter et al. 1994).

Predicting which genes are bound in cells will require, therefore, that any increase in affinity resulting from heteromeric cooperativity be taken into account, in addition to the concentrations of proteins in nuclei, the total number of specific sites in the genome, and the degree of accessibility of each site in each cell. We will also need more experimental data about the range of sites bound *in vivo*, including higher resolution mapping of sites, data for more sequence-specific factors, and measurements on many more genes.

However difficult it is to predict genes targeted by a sequence-specific DNA binding protein *in vivo*, the problem is worse for cofactors and the general transcription factors. The recruitment of these proteins is dependent on multiple protein/protein interactions with a constellation of promoter-bound factors (Fig. 1). Since we do not understand all of the players in these complexes, any success at predicting full promoter activity is likely to be limited at present.

The broad DNA binding by some sequence-specific regulatory proteins in *Drosophila* has implications for the function of the majority of DNA that does not code for protein. Because many promoter deletion experiments have implicated short discrete regions within genes as important transcriptional regulatory elements, it is often assumed that the sequences in between these enhancer and silencer elements serve no function. Homeoproteins, however, bind to long regions between enhancers to the same extent that they bind to their known response elements (Fig. 3). If a large proportion of these sites are functional, and if other classes of factor also bind broadly, a much larger percent of the nonprotein coding portion of the genome could encode transcriptional *cis* elements than currently believed.

The range of genes directly regulated in vivo. Even if we knew which genes a transcription factor binds in vivo, we cannot assume that all are regulated by that factor, as some binding sites may be nonfunctional. For example, the activities of individual transcription factors within enhancosomes are dependent upon the presence of other factors precisely positioned at nearby sites (Thanos and Maniatis 1995; Ptashne and Gann 1997). On their own, individual promoter-bound factors may not be significantly active.

Because of the extreme complexity of the developmental network in *Drosophila*, it is useful to break the problem of determining which genes are regulated into two steps: one, to ascertain the range of all genes that are regulated directly or indirectly; two, to determine what fraction of genes are only directly regulated. By directly, we mean genes that are regulated as a consequence of being bound by a factor. By indirectly, we mean genes that are not direct targets, but whose transcription is altered by subsequent changes in the regulatory network that result from the changed expression of direct targets, some of which may be transcription factors. The reasons for first enumerating direct and indirect targets together is because this is simpler and it will always be useful to know the complete impact a transcription factor has on the network.

Whole genome microarray analysis of mRNA expression has proven to be a powerful method to identify genes regulated directly or indirectly by transcription factors in micro-organisms (e.g., DeRisi et al. 1997; Chu et al. 1998). Comparing mRNAs from wild-type cells with mRNA from cells either lacking or over-expressing a specific regulator shows the total range of variation of gene expression. Microarrays have also been successful for examining gene expression in animal tissue culture cells (Iyer et al. 1999; Ross et al. 2000; Zhao et al. 2000). This approach is less effective, however, for whole animals. The problem is that array analysis compares two populations of cells. As described earlier, virtually every cell in *Drosophila* has a unique profile of mRNA expression. Thus, to establish a thorough catalogue of gene expression would require pairwise comparison of mRNAs from every cell in the animal – clearly an impossible task at present.

To overcome this difficulty, a different strategy has been used. The expression patterns of a random sample of several hundred genes have been analyzed in wild-type and mutant *Drosophila* by whole-mount in situ analysis (Liang and Biggin 1998). In the fly, these experiments are cheap and take only a few weeks to perform. They do not identify all genes regulated by a transcription factor because it is unrealistic to examine all genes, but they do provide a good estimate of the percentage of genes regulated. For example, each of the Hox genes were shown to regulate the expression of at least 80% of all genes directly or indirectly (Liang and Biggin 1998). If instead mRNAs from whole embryos, mutant and wild-type, had been compared by array analysis, it is doubtful that more than 2% of the genes would have

been found to be regulated. Indeed, microarray analysis of whole flies suggest that only 10% of genes are regulated during metamorphosis, a time when many tissues are histolyzed, presumably eliminating the expression of all mRNAs in these cells (White et al. 1999). A combination of microarray analysis – preferably comparing RNA from small groups of cells from specific regions of the body – and whole-mount in situ experiments will probably provide the best overview of the total fraction of genes directly and indirectly regulated by a given transcription factor.

The most common way that individual direct targets have been identified involves the mutation of specific binding sites in transgenic promoter constructs (e.g., Schier and Gehring 1992; Small et al. 1992; Simpson-Brose et al. 1994). This approach is too laborious to allow more than a few potential target genes to be determined, but the data are extremely useful: the elements identified can be used to establish what levels of binding in vivo are functionally significant. The most promising way to identify quickly many direct targets uses fusion proteins that contain the ligand binding domain of a heterologous steroid nuclear receptor and the transcription factor under study. These fusion proteins are activated upon addition of hormone and if protein synthesis is inhibited at the same time that hormone is administered, only the expression of direct targets should be affected. To date, this approach has been used to examine a small number of potential target genes in frogs and *Arabidopsis* (Gammill and Sive 1997; Sablowski and Meyerowitz 1998). If this method is used to analyze the expression either of a random sample of potential direct targets using whole-mount in situs or of all genes using an array, the number of direct targets can be estimated. Although not all genes bound in vivo may be direct targets, at least all direct targets should be bound in the test cells.

Finally, no discussion of animal gene regulatory networks is complete without an acknowledgement of redundancy. It is well known that deletion of some transcription factors has little or no obvious effect on the organism or on gene expression, despite the fact that other experiments indicate that the same proteins can directly regulate important target genes in vivo (e.g., Goldberg et al. 1989; Rudnicki et al. 1992). Such functional redundancy has been amply discussed by others (Tautz 1992; Thomas 1993; Pickett and Meeks-Wagner 1995). Suffice it to say that any attempt to determine all the direct targets of a transcription factor will be forever haunted by this issue. All estimates are likely to be minimum estimates, and distinguishing between nonfunctional binding and functionally redundant binding will require detailed analysis (e.g., Laney and Biggin 1996). Certainly, it is best to study an animal, such as the fruit fly, in which redundancy is less prevalent.

Additional challenges. Beyond the difficulties mentioned above, several other issues must be resolved. One problem is to understand why enhancers and silencer elements interact with a particular core promoter region and

not with other nearby promoters. There are three features of the *Drosophila* genome that make this a challenge. One, around 7% of genes lie within the transcription unit of another gene, typically within an intron (Ashburner et al. 1999). Two, many genes initiate transcription from several discrete sites – for example, the *Adh* gene has two separate core promoters that are active at different stages of development (Hansen and Tjian 1995). Three, the enhancer and silencer elements of neighboring genes are frequently close together. Insulator elements have been discovered that block the ability of regulatory elements to act on neighboring promoters, but our understanding of insulator proteins is too rudimentary to predict insulators from DNA sequence data alone (Cai and Levine 1995; Gaszner et al. 1999; Scott et al. 1999). We will need more information about how distant regulatory elements communicate their effect to the general transcription factors before we can predict enhancer/core promoter selectivity (for further discussion see Bulger and Groudine 1999; Rollins et al. 1999).

The highly dynamic, interacting nature of the network will also have to be taken in to account. Development is not regulated by a strict temporal cascade of transcription factors. One group of factors does not first activate another group of factors, which after a time then activate another group, etc., etc. Instead, the process is a continuous network of cascades. Some transcription factors are expressed earlier than others; but as soon as their targets are affected, the initially expressed factors together with those of their targets that are themselves transcription factors begin to mutually affect new targets and also modify the expression of genes previously regulated. In multicellular organisms, there is extensive crosstalk between neighboring cells using various signal transduction pathways. Direct targets will quickly also become indirect targets of the same transcription factor. There is no such thing as *the* pattern of a gene's expression: mRNA expression changes constantly throughout development. This continuous variation of many parameters imposes a further significant challenge.

Given these difficulties, a fully quantitative modeling of animal transcriptional control may appear unrealistic, but the progress made over the last 30 years has been nothing short of remarkable. We optimistically conclude that the next 30 years may bring us much closer to our goal than currently seems possible. With a complete understanding of animal transcription systems, it should be possible to predict the physiological effects on an organism of mutating or misexpressing a transcription factor. Also, it should allow the design of novel heterologous promoters that express selected gene products in a specified group of cells at a controlled level. Such progress would be important both for human gene therapy and for the development of improved transgenic animals. Quantitative genome-wide analyses are already changing the way we think of transcriptional control and will continue to do so in the future.

Acknowledgements We thank Al Courey, Michael Holmes, Kevin Peet, and Trevor Williams for helpful comments on this manuscript. The work in our laboratories was funded in part by grants from the National Institutes of Health (M.D.B. and R.T.) and the Howard Hughes Medical Institute (R.T.). Work at the Lawrence Berkeley National Laboratory was performed under Department of Energy Contract DE-AC03-76SF00098, University of California.

References

- Adams MD, et al (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287:2185–2195
- Artavanis-Tsakonas S, Rand MD, Lake RJ (1999) Notch signaling: cell fate control and signal integration in development. *Science* 284:770–776
- Ashburner M (1989) *Drosophila*: a laboratory handbook. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Ashburner M, et al (1999) An exploration of the sequence of a 29 Mb region of the genome of *Drosophila melanogaster*: the *Adh* region. *Genetics* 153:179–219
- Bazett-Jones DP, Cote J, Landel CC, Peterson CL, Workman JL (1999) The SWI/SNF complex creates loop domains in DNA and polynucleosomal arrays and can disrupt DNA histone contacts within these domains. *Mol Cell Biol* 19:1470–1487
- Beato M, Eisefeld K (1997) Transcription factor access to chromatin. *Nucleic Acids Res* 25:3559–3563
- Bird AP, Wolffe AP (1999) Methylation-induced repression – belts, braces, and chromatin. *Cell* 99:451–454
- Bjorklund S, Almouzni G, Davidson I, Nightingale KP, Weiss K (1999) Global transcription regulators of eukaryotes. *Cell* 96:759–767
- Brown CE, Lechner T, Howe LA, Workman JL (2000) The many HATs of transcription coactivators. *Trends Biochem Sci* 25:15–19
- Bulger M, Groudine M (1999) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev* 13:2465–2477
- Burglin TR (1994) A comprehensive classification of homeobox genes. In: Duboule D (ed) *Guidebook to the homeobox genes*. Oxford University Press, Oxford, pp 27–269
- Burley SK, Roeder RG (1996) Biochemistry and structural biology of transcription factor IID (TFIID). *Annu Rev Biochem* 65:769–799
- Cai H, Levine M (1995) Modulation of enhancer-promoter interactions by insulators in the *Drosophila* embryo. *Nature* 376:533–536
- Carr A, Biggin MD (1999) A comparison of in vivo and in vitro DNA binding specificities suggests a new model for homeoprotein DNA binding in *Drosophila*. *EMBO J* 18:1598–1608
- Carr A, Biggin MD (2000) Accessibility of transcriptionally inactive genes is specifically reduced at homeoprotein-DNA binding sites. *Nucleic Acids Res* (in press)
- Chen G, Fernandez J, Mische S, Courey AJ (1999) A functional interaction between the histone deacetylase Rpd3 and the corepressor Groucho in *Drosophila* development. *Genes Dev* 13:2218–2230
- Chen JD, Pirrotta V (1993) Multimerization of the *Drosophila* zeste protein is required for efficient DNA binding. *EMBO J* 12:2075–2083
- Chervitz SA, et al (1998) Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science* 282:2022–2028
- Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, Brown PO, Herskowitz I (1998) The transcriptional program of sporulation in budding yeast. *Science* 282:699–705
- Colot V, Rostingol JL (1999) Eukaryotic DNA methylation as an evolutionary device. *Bioessays* 21:402–411
- Cosma MP, Tanaka T, Naysmyth K (1999) Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle and developmentally regulated factor. *Cell* 97:299–311

- Crowley TE, Hoey T, Liu JK, Jan YN, Jan LY, Tjian R (1993) A new factor related to TATA-binding protein has highly restricted expression patterns in *Drosophila*. *Nature* 361:557–561
- Crowley EM, Roeder K, Bina M (1997) A statistical model for locating regulatory regions in genomic DNA. *J Mol Biol* 268: 8–14
- Darlington TK, Wager-Smith K, Ceriani MF, Staknis D, Gekakis N, Steeves TDL, Weitz CJ, Takahashi JS, Kay SA (1998) Closing the circadian loop: CLOCK-induced transcription of its own inhibitors per and tim. *Science* 280:1599–1603
- DeRisi JL, Iyer VR, Brown PO (1997) Exploring metabolic and genetic control of gene expression on a genomic scale. *Science* 278:680–686
- Dhalluin C, Carlson JE, Zeng L, He C, Aggarwal AK, Zhou M-M (1999) Structure and ligand of a histone acetyltransferase bromodomain. *Nature* 399:491–496
- Dikstein R, Zhou S, Tjian R (1996) Human TAF_{II}105 is a cell type-specific TFIID subunit related to hTAF_{II}130. *Cell* 87: 137–146
- Ekker SC, Young KE, von Kessler DP, Beachy PA (1991) Optimal DNA sequence recognition by the Ultrabithorax homeodomain of *Drosophila*. *EMBO J* 10:1179–1186
- England B, Admon A, Tjian R (1992) Cloning of *Drosophila* transcription factor Adf-1 reveals homology to Myb oncoproteins. *Proc Natl Acad Sci USA* 89:683–687
- Faisst S, Meyer S (1992) Compilation of vertebrate-encoded transcription factors. *Nucleic Acids Res* 20:3–26
- Farkas G, Gausz J, Galloni M, Reuter G, Gyurkovics H, Karch F (1994) The Trithorax-like gene encodes the *Drosophila* GAGA factor. *Nature* 371:806–808
- Fisher AL, Caudy M (1998) Groucho proteins: transcriptional corepressors for specific subsets of DNA binding transcription factors in vertebrates and invertebrates. *Genes Dev* 12:1931–1940
- Gammill LS, Sive H (1997) Identification of otx2 target genes and restrictions in ectodermal competence during *Xenopus* cement gland formation. *Development* 124:471–481
- Gaszner M, Vazquez J, Schedl P (1999) The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interactions. *Genes Dev* 13:2098–2107
- Gerhart J, Kirschner M (1997) *Cells, embryos, and evolution*. Blackwell, Oxon
- Goldberg ML, Colvin RA, Mellon AF (1989) The *Drosophila* zeste locus is nonessential. *Genetics* 123:145–155
- Goodrich JA, Cutler G, Tjian R (1996) Contacts in context: promoter specificity and macromolecular interactions in transcription. *Cell* 84:825–830
- Hansen SK, Tjian R (1995) TAFs and TFIIA mediate differential utilization of the tandem Adh promoters. *Cell* 82:565–575
- Hara R, Selby CP, Liu M, Price DH, Sancar A (1999) Human transcription release factor 2 dissociates RNA polymerase I and II stalled as a cyclobutane thymine dimer. *J Biol Chem* 274: 24779–24789
- Heinemeyer T, et al (1998) Databases on transcriptional regulation: transfac, trrd and compel. *Nucleic Acids Res* 26:362–367
- Holmes MC, Tjian R (2000) Promoter-selective properties of the TBP-related factor TRF1. *Science* 288:867–870
- Holstege FC, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, Green MR, Golub TR, Lander ES, Young RA (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95: 717–728
- Hughes JD, Estep PW, Tavazoie S, Church GM (2000) Computational identification of cis-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. *J Mol Biol* 296:1205–1214
- Imbalzano AN, Kwon H, Green MR, Kingston RE (1994) Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature* 370:481–485
- Iyer VR, et al (1999) The transcriptional program in the response of human fibroblasts to serum. *Science* 283:83–87
- Jacobson RH, Ladurner AG, King DS, Tjian R (2000) Structure and function of a human TAF_{II}250 double bromodomain module. *Science* 288:1422–1425
- Johnson A (1992) A combinatorial regulatory circuit in budding yeast. In: McKnight SL, Yamamoto KR (eds) *Transcriptional Regulation*. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY
- Keller CA, Grill MA, Abmayr SM (1998) A role for nautilus in the differentiation of muscle precursors. *Dev Biol* 202:157–171
- Kingston RE, Narlikar GJ (1999) ATP dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes Dev* 13:2339–2352
- Knoepfler PS, Eisenman RN (1999) Sin meets NuRD and other tails of repression. *Cell* 99:447–450
- Kockel L, Zeitlinger J, Staszewski LM, Mlodzik M, Bohmann D (1997) Jun in *Drosophila* development: redundant and nonredundant functions and regulation by two MAPK signal transduction pathways. *Genes Dev* 11:1748–1758
- Kopczynski CC, Noordermeer JN, Serano TL, Chen WY, Pendleton JD, Lewis S, Goodman CS, Rubin GM (1998) A high throughput screen to identify secreted and transmembrane proteins involved in *Drosophila* embryogenesis. *Proc Natl Acad Sci USA* 95:9973–9978
- 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* 20:4754–4764
- Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH (1998) New core promoter element in RNA polymerase II dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev* 12:34–44
- Laney JD, Biggin MD (1996) Redundant control of Ultrabithorax by zeste involves functional levels of zeste binding at the Ultrabithorax promoter. *Development* 122:2303–2311
- Laney JD, Biggin MD (1997) zeste mediated activation by an enhancer is independent of cooperative DNA binding in vivo. *Proc Natl Acad Sci USA* 94:3602–3604
- Lawrence PA (1992) *The making of a fly*. Blackwell, Oxford
- Liang Z, Biggin MD (1998) Eve and Ftz regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in *Drosophila*. *Development* 125:4471–4482
- Lin S, Riggs AD (1975) The general affinity of lac repressor for E coli DNA: implications for gene regulation in prokaryotes and eukaryotes. *Cell* 4:107–111
- Luo H, Asha H, Kockel L, Parke T, Mlodzik M, Dearolf CR (1999) The *Drosophila* Jak kinase Hopscotch is required for multiple developmental processes in the eye. *Dev Biol* 213: 432–441
- Mann RS, Chan S-K (1996) Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. *Trends Genet* 12:258–262
- Mannervik M, Nibu Y, Levine M (1999) Transcriptional coregulators in development. *Science* 284:606–609
- Margulis L, Schwartz KV (1998) *The five kingdoms*. Freeman, New York
- McGinnis N, Kuziora MA, McGinnis W (1990) Human *Hox-42* and *Drosophila Deformed* encode similar regulatory specificities in *Drosophila* embryos and larvae. *Cell* 63:969–976
- Orphanides G, Lagrange T, Reinberg D (1996) The general transcription factors of RNA polymerase II. *Genes Dev* 10:2657–2683
- Parvin JD, Sharpe PA (1993) DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell* 73:533–540
- Pickett FB, Meeks-Wagner DR (1995) Seeing double: appreciating genetic redundancy. *Plant Cell* 7:1347–1356
- Pollock R, Treisman R (1990) A sensitive method for the determination of protein-DNA-binding specificities. *Nucleic Acids Res* 18:6197–6204
- Price DH (2000) P-TEFb: a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol Cell Biol* 20:2629–2634
- Ptashe M, Gann A (1997) Transcriptional activation by recruitment *Nature* 386:569–577

- Reinitz J, Kosman D, Vanario-Alonso CE, Sharp DH (1998) Stripe forming architecture of the gap gene system. *Dev Genet* 23: 11–27
- Reynaud E, Lomelf H, Vazquez M, Zurita M (1999) The *Drosophila melanogaster* homologue of the xeroderma pigmentosum D gene product is located in euchromatic regions and has a dynamic response to UV light-induced lesions in polytene chromosomes. *Mol Biol Cell* 10:1191–1203
- Riese J, Yu X, Munnerlyn A, Eresh S, Hsu SC, Grosschedl R, Bienz M (1997) LEF-1, a nuclear factor coordinating signaling inputs from wingless and decapentaplegic. *Cell* 88:777–787
- Robinson K, McGuire AM, Church GM (1998) A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete *Escherichia coli* K-12 genome. *J Mol Biol* 284: 241–254
- Rollins RA, Morcillo PA, Dorsett D (1999) Nipped-B, a *Drosophila* homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics* 152:577–593
- Rong YS, Golic KG (2000) Gene targeting by homologous recombination in *Drosophila*. *Science* 288:2013–2018
- Ross DT, et al (2000) Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 24:227–235
- Rubin GM, et al (2000) Comparative genomics of the eukaryotes. *Science* 287:2204–2215
- Rudnicki MA, Braun T, Hinuma S, Jaenisch R (1992) Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. *Cell* 71:383–390
- Sablowski RWM, Meyerowitz E (1998) A homolog of NO APICAL MERISTEM is an immediate target of the floral homeotic genes APETALA3/PISTILLATA. *Cell* 92:93–103
- Schier A, Gehring WJ (1992) Direct homeodomain-DNA interaction in the autoregulation of the fushi tarazu gene. *Nature* 356:804–807
- Scott KC, Taubman AD, Geyer PK (1999) Enhancer blocking by the *Drosophila* gypsy insulator depends upon insulator anatomy and enhancer strength. *Genetics* 153:787–798
- Segraves WA (1994) Steroid receptors and orphan receptors in *Drosophila* development. *Semin Cell Biol* 5:105–113
- Sellers WR, Kaelin WG (1996) RB as a modulator of transcription. *Biochim Biophys Acta* 1288:M1–M5
- Simpson-Brose M, Treisman J, Desplan C (1994) Synergy between the Hunchback and Bicoid morphogens is required for anterior patterning in *Drosophila*. *Cell* 78:855–865
- Small S, Blair A, Levine M (1992) Regulation of even-skipped stripe 2 in the *Drosophila* embryo. *EMBO J* 11:4047–4057
- Soeller WC, Poole SJ, Kornberg T (1988) In vitro transcription of the *Drosophila* engrailed gene. *Genes Dev* 2:68–81
- Soeller WC, Oh CE, Kornberg TB (1993) Isolation of cDNAs encoding the *Drosophila* GAGA transcription factor. *Mol Cell Biol* 13:7961–7970
- Spradling AC (1986) P element-mediated transformation. In: Roberts DB (ed) *Drosophila: a practical approach*. IRL Press, Washington, D.C., pp 175–197
- Sypes MA, Gilmour DS (1994) Protein/DNA crosslinking of a TFIID complex reveals novel interactions downstream of the transcription start. *Nucleic Acids Res* 22:807–814
- Tautz D (1992) Redundancies, development and the flow of information. *BioEssays* 14:263–266
- Taylor IC, Workman JL, Schuetz TJ, Kingston RE (1991) Facilitated binding of GAL4 and heat shock factor to nucleosomal templates: differential function of DNA-binding domains. *Genes Dev* 5:1285–1298
- TenHarmsel A, Austin RJ, Savenelli N, Biggin MD (1993) Cooperative binding at a distance by even-skipped protein correlates with repression and suggests a mechanism of silencing. *Mol Cell Biol* 13:2742–2752
- Thanos D, Maniatis T (1995) Virus induction of human IFN β gene expression requires the assembly of an enhancersome. *Cell* 83:1091–1100
- The FlyBase Consortium (1999) The FlyBase Database of the *Drosophila* Genome Projects and community literature. *Nucleic Acids Res* 27:85–88
- Thomas JT (1993) Thinking about genetic redundancy. *Trends Genet* 9:395–399
- Thummel CS (1997) Dueling orphans—interacting nuclear receptors coordinate *Drosophila* metamorphosis. *Bioessays* 19:669–672
- Tjian R, Maniatis T (1994) Transcriptional activation: a complex puzzle with few easy pieces. *Cell* 77:5–8
- Tyree CM, George CP, Lira-DeVito LM, Wampler SL, Dahmus ME, Zawel L, Kadonaga JT (1993) Identification of a minimal set of proteins that is sufficient for accurate initiation of transcription by RNA polymerase II. *Genes Dev* 13:1254–1265
- Verrijzer CP, Yokomori K, Chen JL, Tjian R (1994) *Drosophila* TAFII 150: similar to yeast TSM-1 and specific binding to core promoter. *DNA Sci* 264:933–941
- von Hippel PH, Revzin A, Gross CA, Wang AC (1974) Nonspecific DNA binding of genome regulating proteins as a biological control mechanism: 1. The lac operon: equilibrium aspects. *Proc Natl Acad Sci USA* 71:4808–4812
- Wallrath LL, Elgin SCR (1995) Position effect variegation in *Drosophila* is associated with an altered chromatin structure. *Genes Dev* 9:1263–1277
- Walsh CP, Bestor TH (1999) Cytosine methylation and mammalian development. *Genes Dev* 13:26–34
- Walter J, Dever C, Biggin MD (1994) Two homeodomain proteins bind with similar specificity to a wide range of DNA sites in *Drosophila* embryos. *Genes Dev* 8:1678–1692
- Waltzer L, Bienz M (1999) A function of CBP as a transcriptional co-activator during Dpp signalling. *EMBO J* 18:1630–1641
- Wasserman WW, Fickett JW (1998) Identification of regulatory regions which confer muscle-specific gene expression. *J Mol Biol* 278:167–181
- White KP, Rifkin SA, Hurban P, Hogness DS (1999) Microarray analysis of *Drosophila* development during metamorphosis. *Science* 286:2179–2184
- Wolffe A (1998) Chromatin structure and function, 3rd edn. Academic Press, San Diego
- Wolpert L (1991) The triumph of the embryo. Oxford University Press, Oxford
- Workman JL, Roeder RG (1987) Binding of transcription factor TFIID to the major late promoter during in vitro nucleosome assembly potentiates subsequent initiation by RNA polymerase II. *Cell* 51:613–622
- Wu C (1995) Heat shock transcription factors: structure and regulation. *Annu Rev Cell Dev Biol* 11:441–469
- Yang S-W, Nash HA (1995) Comparison of protein binding to DNA in vivo and in vitro: defining an effective intracellular target. *EMBO J* 14:6292–6300
- Yu Y, Li W, Su K, Yussa M, Han W, Perrimon N, Pick L (1997) The nuclear hormone receptor Ftz-F1 is a cofactor for the *Drosophila* homeodomain protein Ftz. *Nature* 385:552–555
- Zhao R, Gish K, Murphy M, Yin Y, Notterman D, Hoffman WH, Tom E, Mack DH, Levine AJ (2000) Analysis of p53-regulated gene expression patterns using oligonucleotide arrays. *Genes Dev* 14:981–993