27. F. Mitelman in *Progress and Topics in Cytogenetics*, A. A. Sandberg, Ed. (Liss, New York, 1985), vol. 5, p. 107.
28. A. Ramaldi, D. C. Young, J. D. Griffin, *Blood* **69**, 1409 (1987); C. A. Sieff, personal communication.
29. J. R. Zucali *et al.*, *J. Clin. Invest.* 77, 1837 (1986); R. Munker, J. Gasson, M. Ogawa, H. P. Koeffler, *Nature (London)* **323**, 79 (1986); C. A. Sieff, S. Tsai, D. V. Faller, *J. Clin. Invest.* **79**, 48 (1987); V. C. Broudy *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7467 (1986); J. R. Zucali, H. E. Broxmeyer, C. A. Dinarello, M. A. Gross, R. S. Weiner, *Blood* **69**, 33 (1987); G. C. Bagby *et al.*, *J. Clin. Invest.* **18**, 1316 (1986); W. E. Fibbe *et al.*, *Blood* **68**, 1316 (1986).
30. Y.-C. Yang, S. Tsai, G. G. Wong, S. C. Clark, in preparation.
31. W. Piacibello, L. Lu, M. Wachter, B. Rubin, H. E. Broxmeyer, *Blood* **66**, 1343 (1985); F. Herrmann, S. A. Cannistra, J. D. Griffin, *J. Immunol* **136**, 2856 (1986); W. Piacibello *et al.*, *Blood* **68**, 1339 (1986); H. E. Broxmeyer, *Int. J. Cell Cloning* **4**, 378 (1986).
32. E. Spooncer, C. M. Heyworth, A. Dunn, T. M. Dexter, *Differentiation* **31**, 111 (1986); T. M. Dexter, C. Heyworth, E. Spooncer, *Prog. Clin. Biol. Res.* **184**, 235 (1985); J. Cashman, A. C. Eaves, C. J. Eaves, *Blood* **66**, 1002 (1985).
33. G. G. Wong *et al.*, *Cancer Cells 3: Growth Factors and Transformation*, J. Feramisco, B. Ozanne, C. Stiles, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1985), pp. 235–242.
34. K. Welte *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1526 (1985).
35. S. K. Das, E. R. Stanley, L. J. Guilbert, L. W. Forman, *Blood* **58**, 630 (1981).
36. R. E. Donahue, *Cold Spring Harbor Symp. Quant. Biol.* **51**, 685 (1986).
37. R. Hewick and R. Steinbrink, unpublished results.
38. E. A. Wang and K. Turner, unpublished results.
39. K. Welte *et al.*, *J. Exp. Med.* **165**, 941 (1987).
40. K. Motoyoshi *et al.*, *Blood* **60**, 1378 (1982).
41. R. Hewick and K. Ferenz, unpublished results.
42. D. D. Sabatini *et al.*, *J. Cell Biol.* **91**, 1 (1982).
43. G. G. Wong and P. Temple, unpublished results.
44. E. Platzer *et al.*, *J. Exp. Med.* **162**, 1788 (1985); E. Platzer *et al.*, *Immunobiology*, in press; A. F. Lopez *et al.*, *J. Clin. Invest.* **78**, 1220 (1986).
45. D. Metcalf *et al.*, *Blood* **67**, 257 (1986).
46. A. G. Leary *et al.*, in preparation.
47. R. H. Weisbart *et al.*, *Nature (London)* **314**, 361 (1985); A. F. Lopez *et al.*, *J. Clin. Invest.* **78**, 1220 (1986).
48. K. H. Grabstein *et al.*, *Science* **232**, 506 (1986).
49. J. D. Griffin *et al.*, *Blood* **67**, 1448 (1986); T. Hoang *et al.*, *ibid.* **68**, 313 (1986).
50. D. C. Young, K. Wagner, J. D. Griffin, *J. Clin. Invest.* **79**, 100 (1987).
51. C. A. Sieff *et al.*, in preparation.
52. A. Lopez *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2761 (1987).
53. M. K. Warren and P. Ralph, *J. Immunol.* 7, 2281 (1986).
54. R. A. Mufson, personal communication.
55. N. A. Nicola and D. Metcalf, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3765 (1984).
56. J. C. Gasson *et al.*, *ibid.* **83**, 669 (1986).
57. L. J. Guilbert and E. R. Stanley, *J. Cell Biol.* **85**, 153 (1980); *J. Biol. Chem.* **261**, 4024 (1986); R. Tushinski *et al.*, *Cell* **28**, 71 (1982).
58. R. E. Donahue *et al.*, *Nature (London)* **321**, 872 (1986); P. Mayer, C. Lam, H. Obenaus, E. Liehl, J. Besemer, *Blood*, in press.
59. D. Metcalf *et al.*, *Blood* **68**, 46 (1986); V. Kindler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1001 (1986).
60. J. Groopman *et al.*, in preparation.
61. A. W. Nienhuis *et al.*, in preparation.
62. J. C. Gasson, I. S. Y. Chen, C. A. Westbrook, D. W. Golde, in *Normal and Neoplastic Hematopoiesis*, D. W. Golde and P. A. Marks, Eds. (Liss, New York, 1983), p. 129.
63. G. P. Body, M. Buckley, Y. S. Sathe, E. Freireich, *Ann. Intern. Med.* **64**, 328 (1966).
64. L. Zon, C. Arkin, J. Groopman, *Br. J. Haematol.*, in press.
65. J. W. Eschbach *et al.*, *N. Engl. J. Med.* **316**, 73 (1987).
66. J. D. Griffin *et al.*, *Blood* **68**, 1185 (1986).
67. T. Ishibashi, J. A. Koziol, S. A. Burstein, *J. Clin. Invest.* **59**, 286 (1987).
68. K. Jacobs and R. Kriz, unpublished results.
69. Many of our colleagues at Genetics Institute, within Research and Process Development, have contributed to the work summarized in this article, especially G. Wong, Y.-C. Yang, J. Witek-Giannotti, R. Donahue, D. Stone, R. Hewick, R. Kaufman, E. Wang, K. Turner, and the members of the hematopoiesis research group. A special thanks to A. Leary, K. Jacobs, and R. Kriz for providing the data for Figs. 2 and 3. We are also grateful for the numerous productive relationships we enjoy with our academic colleagues, including C. Sieff, D. Nathan, J. Griffin, J. Groopman, M. Ogawa, A. Nienhuis, M. Vadas, M. Minden, J. Gasson, and D. Golde. Special thanks goes to Sandoz, which supported much of our work and to those at Sandoz, including D. Oette, R. Stall, D. Schoenleber, and P. Mayer, who made major contributions to the clinical development of GM-CSF. We are also grateful to U. Roy-Chowdhury for assistance in preparing the manuscript.

Regulation of Inducible and Tissue-Specific Gene Expression

TOM MANIATIS, STEPHEN GOODBOURN, JANICE A. FISCHER

Molecular genetics approaches have been used to identify and characterize cis-acting DNA sequences required for eukaryotic gene regulation. These sequences are modular in nature, consisting of arrays of short (10– to 12–base pair) recognition elements that interact with specific transcription factors. Some transcription factors have been extensively purified and the corresponding genes have been cloned, but the mechanisms by which they promote transcription are not yet understood. Positive and negative regulatory elements that function only in specific cell types or in response to extracellular inducers have been identified. A number of cases of inducible and tissue-specific gene expression involve the activation of preexisting transcription factors, rather than the synthesis of new proteins. This activation may involve covalent modification of the protein or an allosteric change in its structure. The modification of regulatory proteins may play a central role in the mechanisms of eukaryotic gene regulation.

A CENTRAL PROBLEM IN EUKARYOTIC MOLECULAR BIOLOGY is to understand the mechanisms by which specific genes are expressed in a temporal or tissue-specific manner or are activated in response to extracellular inducers. The development of methods for cloning and characterizing individual genes has provided the opportunity to study these mechanisms at the molecular level. Initially, cis-acting DNA sequences required for gene regulation were identified by introducing mutations into cloned genes and then analyzing their effects on expression in vivo. More recently, proteins that specifically bind to these regulatory DNA sequences have been identified and in some cases purified. The current challenges are to understand how specific protein-DNA interactions regulate gene expression and how these interactions are integrated into the overall pattern of gene regulation during development. In this review, we summarize the information obtained to date regarding the nature of the DNA sequence elements and protein factors required for gene regulation at the level of transcription initiation in higher eukary-

The authors are at the Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, MA 02138.

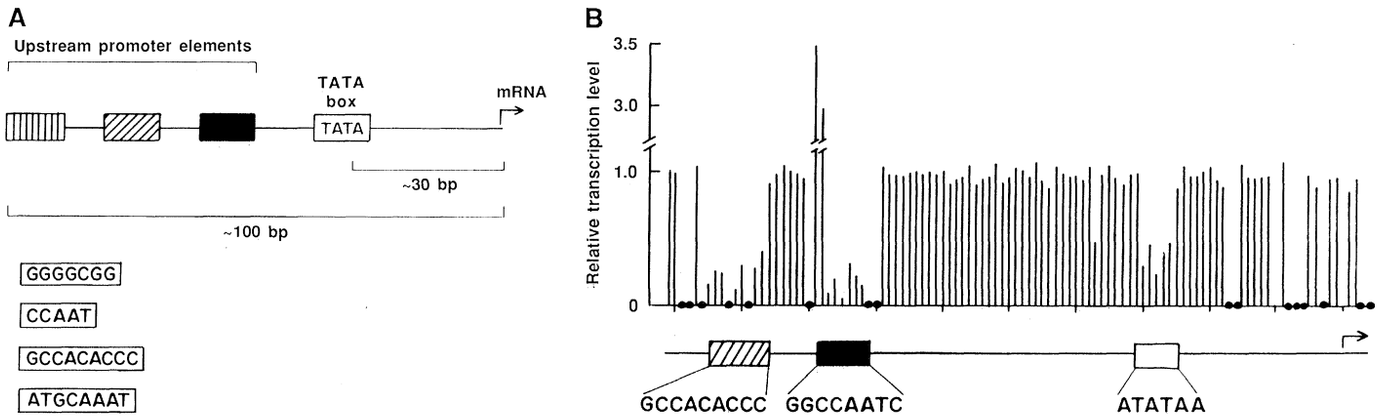

Fig. 1. (**A**) Organization of a typical higher eukaryotic promoter. The diagram shows the location of the TATA box and upstream promoter elements (UPEs) (*1*). The schematic promoter illustrated here contains three different types of UPE. Examples of the DNA sequence motifs for four well-characterized UPEs are given below the diagram. (**B**) Biological effect of point mutations in the β-globin promoter (*6*). Each line in the histogram represents the transcription level relative to a wild-type promoter observed for a particular single base mutation. The black dots represent nucleotides for which single base changes were not obtained. The diagram below the bar graph shows the position of the TATA box and two UPEs of the promoter. Only those base substitutions that lie within the three promoter elements change the level of transcription.

otes, and we discuss current views regarding the mechanisms by which these elements act.

Promoters and Enhancers

Two DNA sequence elements are required for the regulation of genes that encode messenger RNA in higher eukaryotes: promoters and enhancers. Promoters are located immediately upstream from the start site of transcription and are typically about 100 base pairs in length (*1*). The promoter is required for accurate and efficient initiation of transcription, whereas enhancers increase the rate of transcription from promoters. The distinctive characteristic of enhancers is that they can act on cis-linked promoters at great distances in an orientation-independent manner and can also function downstream from the transcription unit (*2*). However, as discussed below, the basic components of promoters and enhancers share many properties, and the mechanisms by which these components facilitate transcription may be indistinguishable.

Detailed molecular genetic analyses of a number of different promoters reveal a common pattern of organization [see (*1*) for recent reviews]. A typical promoter includes an AT-rich region designated the TATA box and one or more sequence elements of 8 to 12 base pairs designated upstream promoter elements (UPEs) (Fig. 1A). The TATA box functions primarily to ensure that transcripts are accurately initiated, whereas the UPEs increase the rate of transcription. A number of different UPEs have been identified, and some of these (for example, the CCAAT box; Fig. 1A) are found in many different promoters. Mutagenesis studies suggest that the strength of promoters is determined by the number and type of UPEs. In addition, these studies have shown that UPEs act regardless of their orientation with respect to the TATA box. However, insertion of nucleotides between the UPEs and the TATA box can decrease the level of transcription (*3, 4*). In one case, insertion of odd multiples of half a DNA turn is more detrimental to transcription than insertion of even multiples. This observation suggests that one or more proteins bound to the UPE interact with a protein (or proteins) bound to the TATA box, and this interaction requires stereospecific alignment of the proteins on the DNA helix (*3*).

UPEs have been delineated by in vitro DNA footprinting experiments (*1*) and by analysis of large numbers of promoter mutations (*5, 6*). A saturation mutagenesis study of the β-globin promoter allowed the precise localization of the TATA box and two UPEs (*6*). Single base mutations within any of these three promoter elements result in a five- to tenfold decrease in transcription in HeLa cells, whereas base substitutions elsewhere within the promoter have no effect (Fig. 1B). Two different base substitutions within the conserved CAAT box element resulted in 3.0- to 3.5-fold increase in transcription, suggesting that these changes facilitate the interactions between proteins and the CAAT box element.

Proteins that bind specifically to UPEs have been identified by DNA footprinting (*7*) and gel retardation (*8*) assays [see (*1*) for review], and in some cases these proteins have been extensively purified (*9*). Although many of these proteins are of relatively low abundance, several thousand–fold purification has been achieved by recognition-site affinity chromatography (*9*). Amino acid sequence analysis of UPE-binding proteins will provide the opportunity to clone the corresponding genes and ultimately to produce the proteins in sufficient amounts to carry out detailed biochemical studies.

Enhancers also contain discrete DNA sequence elements that specifically interact with proteins [see (*2*) for review]. This characteristic is best illustrated by the SV40 enhancer (Fig. 2). In vitro mutagenesis experiments revealed that the SV40 enhancer consists of two functional domains designated A and B (*10*). Mutations within subregions in each domain lead to a significant decrease in transcription in vivo. The A and B domains alone have weak enhancer activity, but duplication of either domain separately or in combination leads to a high level of activity (*10, 11*). The synergistic effect of A and B is independent of their relative position or their orientation. In vitro footprinting experiments have shown that the DNA sequence elements necessary for SV40 enhancer activity in vivo are specifically recognized by protein factors (*12*). Each of the binding sites shown in Fig. 2 act independently in vitro, since a mutation in one site does not affect binding to an adjacent site. In addition, alteration of the spacing between sites does not affect the in vitro footprinting pattern (*12*).

Recent evidence suggests that different sequence motifs within the SV40 enhancer are recognized by different factors present in HeLa and B cell nuclear extracts (*13*) (see Fig. 2). The functional significance of these differences is suggested by the observation that sequences specifically protected by HeLa cell factors are necessary

for in vivo enhancer function in HeLa cells but not in B cells. Similarly, mutations in the regions protected only by B cell factors alter expression in B cells but not in HeLa cells (*13*). Moreover, it has recently been shown that all of the sequence motifs within the SV40 enhancer may have different activities in different cell lines (*14*). Thus, different proteins that bind to different sites within a single enhancer can stimulate transcription in specific cell types. This complex organization of viral enhancer elements may have evolved as a means of extending the host range of animal viruses.

Not only are promoters and enhancers similarly organized, but some of their sequence elements are interchangeable. For example, a sequence found within the immunoglobulin enhancer (the "octamer") is also found in a number of different promoters (*15*). As implied by this structural similarity, enhancers and promoters appear to be functionally related. When the SV40 enhancer is located more than 100 base pairs from the β-globin promoter, UPE mutations significantly decrease the level of transcription (Fig. 1B). Deletion of both UPEs essentially abolishes transcription. However, when the SV40 enhancer is placed immediately adjacent to the deleted promoter, a high level of transcription is restored (*16*). These and similar observations with other promoters indicate that the distinction between promoters and enhancers is somewhat arbitrary. The operational difference between enhancers and promoters (that is, action at a distance) may be a consequence of the arrangement and number of transcription-factor recognition elements in each case, rather than the result of fundamental differences in the mechanisms by which these elements act. This conclusion is supported by experiments showing that a single heat-shock regulatory element cannot act at a distance, but duplication of this sequence creates an element that has all of the properties of an enhancer (*17*). The most likely explanation of this observation is that the multiple protein-protein interactions are necessary to form transcription complexes between widely separated DNA recognition elements. However, this point cannot be generalized, since multiple copies of the UPE CCAAT box do not stimulate transcription at a distance (*17*).

Several models have been proposed to explain how factors that recognize enhancers can act on promoters located many kilobases away [see (*2*) for review]. Recent observations are consistent with a model in which transcription is stimulated by interactions between proteins bound to the enhancer and the promoter, with the looping out of the intervening DNA [see (*18*) for review]. Given the modular nature of promoter and enhancer elements it is possible that factors bound to each of the recognition elements interact with each other to form transcription complexes.

Regulated Promoter and Enhancer Elements

Viral promoters and enhancers such as those of SV40 and adenovirus are active in most cell types, but the level of their activity can be regulated by viral and cellular gene products [see (*19–21*) for reviews]. The most thoroughly studied example of this type of regulation is the trans-activation and repression of transcription by viral immediate early (IE) proteins (*19, 20*). These proteins appear to act by modifying other transcription factors rather than by interacting directly with specific DNA sequences. This conclusion is based on the observation that IE proteins act on a variety of different polymerase II and III promoters with very different sequences, and on the results of in vitro transcription studies (*22*). These studies show that a partially purified IE protein can stimulate transcription in vitro, and this activity is temperature-sensitive when the factor is purified from cells infected with viruses carrying a temperature-sensitive mutation in the IE gene (*22*). More recent experiments show that the binding activity of a specific UPE factor is significantly enhanced by infection with wild-type adenovirus but not with viruses containing a mutation that inactivates the IE gene E1a (*23*). A relatively high level of the same UPE binding activity was also detected in undifferentiated embryonal carcinoma (EC) cells in the absence of adenovirus infection (*24*). Significantly, undifferentiated EC cells are unique in their ability to partially complement E1a mutations. When EC cells are induced to differentiate, a concomitant loss of UPE binding activity and E1a complementation is observed. Infection of the differentiated cells with wild-type adenovirus reactivates the binding activity of the UPE factor. These observations suggest that undifferentiated EC cells produce a cellular factor that is functionally analogous to E1a (*25*).

The functional analogy between viral and cellular gene products extends to the ability of these proteins to repress enhancer-dependent transcription (*25–27*). The activity of many viral enhancers is

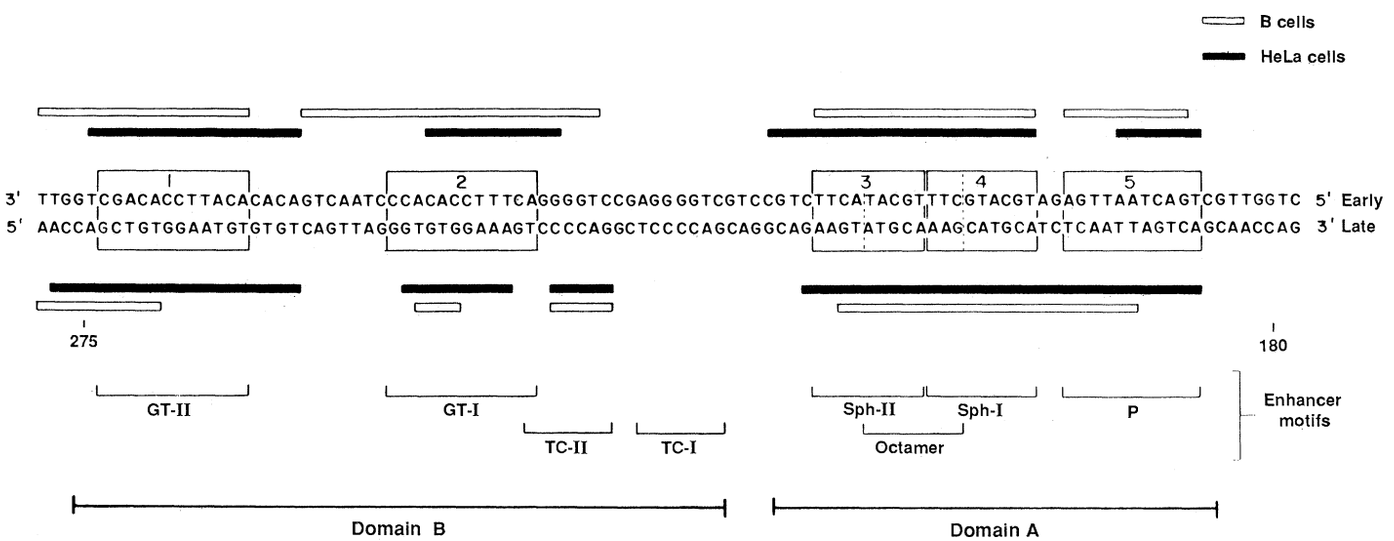

Fig. 2. Organization and protein-DNA interactions of the SV40 enhancer. The boxed sequences labeled 1 through 5 indicate the sequences that are required for maximum levels of enhancer activity in vivo and in vitro. The black and white bars indicate the sites that are protected from DNAase cleavage by extracts from HeLa and B cells, respectively. The brackets below the sequence indicate the location of sequence motifs that are repeated within the enhancer. The functional domains A and B are indicated at the bottom of the figure. [See text and (*10–13*) for further details.]

repressed by E1a and by a factor in undifferentiated mouse EC cells. In the latter case, this block is removed when the EC cells are induced to differentiate. Evidence that differentiation leads to the inactivation of a repressor that interacts directly with the enhancer is provided by the existence of polyoma enhancer mutations that circumvent the block to activity in EC cells (*28*). A functional link between the negative factor in EC cells and E1a is suggested by the observation that the same polyoma enhancer mutant that escapes repression in EC cells cannot be repressed by E1a (*29*). However, the interpretation of this result is complicated by the fact that similar polyoma enhancer mutants are fully repressed by E1a (*30*).

The general principle to emerge from the study of IE proteins is that both positive and negative regulation may be controlled by the modification of factors that interact with promoters and enhancers. Further evidence that factor modification may play a role in gene regulation is suggested by the observation that treatment of cells with phorbol esters can dramatically increase the activities of the SV40 and polyoma virus enhancers (*31*). These compounds stimulate protein kinase C activity, suggesting that enhancer binding factors may be modified by phosphorylation. As will be discussed below, there are many examples of regulated enhancers. It is thus tempting to speculate that some of these regulatory activities may be manifested by similar modifications of cellular factors.

The types of regulation observed with cellular enhancers indicate that enhancers can be divided into two categories: those that respond to changes in the environment (inducible enhancers) and those that are active only at specific times during development or only in specific tissues (temporal and tissue-specific enhancers).

Inducible Enhancers

Examples of inducible enhancers are those that respond to heat shock, exposure to heavy metals, or viral infection and those that are activated in response to growth factors or steroids. Inducible enhancers have been identified for a number of genes including genes for heat shock (*17*), metallothionein (*32*), and β-interferon (β-IFN) (*33*), as well as c-*fos* (*34*). Steroid-responsive enhancers have been identified in the long terminal repeat sequences of the mouse mammary tumor virus (*35*) and the Moloney mouse sarcoma virus (*36*) and in association with a number of cellular genes (*37–39*). In some cases, inducible enhancer elements are located at a distance from the promoters they act on, whereas others may be intimately associated with their promoters.

Fine-structure mutagenesis studies and examination of regulatory sequences conserved among different species have revealed very short DNA sequences that are necessary for regulated gene expression. For example, a 12–base pair sequence element from the human metallothionein gene is sufficient to confer metal inducibility upon a heterologous promoter (*40*), and close relatives of this element are found in metallothionein gene regulatory regions from different species (*41*). Each gene contains multiple copies of this sequence, and the level of heavy metal induction depends on the number of copies present. The additive effects of tandemly reiterated regulatory elements appears to be a general phenomenon, since similar findings have been reported for heat-shock (*42*) and β-IFN (*33*) genes and for glucocorticoid regulatory elements (*43*). There is also some evidence that the number of reiterated elements needed for efficient induction can vary among different cell types (*33, 42, 44*), which presumably reflects differences in the levels or kinds of trans-acting factors in these cells.

Inducible enhancer elements are capable of activating heterologous promoters and therefore by definition act positively. However, this positive activation could be achieved by a number of different mechanisms (Fig. 3), including the synthesis or activation of a positive transcription factor or the inactivation of a negative regulatory protein. Alternatively, enhancer activation could be a consequence of both the inactivation of a repressor and the activation of a positive regulatory factor. Temperature induction of the *Drosophila* heat-shock hsp70 gene appears to be an example of the simple positive regulatory mechanism (see below). In contrast, the human β-IFN gene is regulated primarily by a negative mechanism. The interferon gene regulatory element (IRE) consists of two regulatory domains—a negative element and an adjacent or overlapping positive transcription element. The negative element was initially identified by the analysis of deletion mutants that resulted in high levels of transcription in the absence of inducer. Subsequently, the negative element was shown to repress transcription from a heterologous promoter, and the positive element was shown to act as a constitutive enhancer (*45*). The negative regulatory mechanism was further supported by genomic footprinting experiments (*46*). The negative regulatory region was protected from deoxyribonuclease (DNAse) digestion before induction but not after. Conversely, the positive regulatory domain was not protected from DNAse I digestion prior to induction but was protected after induction (*46*). None of these observations rule out the possibility that the positive transcription element also responds to induction. In fact, some evidence indicates that this is indeed the case. First, deletion of the negative regulatory element from the IRE increases the basal level of transcription, but a low level of induction is still observed (*45*). In addition, when the positive regulatory element of the IRE is placed upstream from a heterologous promoter, a low but reproducible level of induction is detected (*45*). Second, when multiple copies of a conserved hexanucleotide sequence located within the positive regulatory element of the IRE is placed upstream from a heterologous promoter, a low level of virus induction is observed (*47*). Thus, activation of the interferon promoter may involve both the removal of a repressor (or repressors) and the stimulation of a positive regulatory factor.

Evidence that inducible regulatory elements specifically interact with cellular factors has been provided by a number of DNA binding and footprinting studies. Significantly, gene activation occurs in the absence of protein synthesis with all of the inducible genes thus far analyzed. Thus, the activation of gene expression must involve the modification of preexisting cellular factors. This modifi-

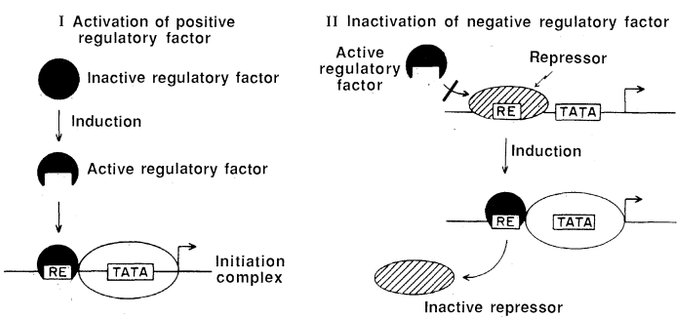

Fig. 3. Models for the activation of inducible enhancer elements. In model I the regulatory factor is present in an inactive configuration. Treatment with inducer leads to a modification that allows the factor to bind to the regulatory element (RE). In model II, access to the RE is blocked in uninduced cells by a negatively acting factor. Induction leads to the displacement of the repressor and binding of the regulatory factor to the RE. In the simplest situation the repressor would be the target for modification and would lose affinity for the template to be replaced by a constitutively active transcription factor. However, an alternative possibility is that the transcription factor could be modified with an increase in template affinity and displace the repressor by simple competition. In a third model (not shown) induction leads to both the activation of a regulatory factor and the inactivation of a repressor.

cation might increase the affinity of the factor for DNA, or it might change the conformation of a factor already bound to the regulatory element. These alternatives have recently been investigated in the heat-shock, c-*fos*, and steroid-responsive genes.

A factor that is capable of binding heat-shock response elements (HSEs) has been identified in *Drosophila* (*48, 49*), yeast (*50*), and man (*51*). Although it was initially thought that this factor is active in extracts from uninduced cells (*48*), subsequent studies have revealed that induction leads to a dramatic increase in the HSE-binding activity (*49, 51, 52*). The HSE-binding factor has been highly purified and shown to stimulate transcription from the hsp70 promoter in vitro (*48*) and therefore has been designated the heat-shock transcription factor (HSTF). Occupation of more than one HSE within the hsp70 promoter is essential for both in vivo (*42*) and in vitro (*53*) transcription activity, and there is evidence that HSTF can achieve this by binding in a cooperative fashion. The binding of the first molecule of HSTF [which appears to bind as a dimer to the bilaterally symmetrical HSE (*54*)] causes a bend in the DNA, possibly facilitating the binding of the second HSTF molecule (*55*). Binding the second molecule of HSTF allows an additional conformational change, as evidenced by a change in the contacts of the DNA to HSTF (*54*). This second conformational change may be important in triggering transcription.

The cellular oncogene c-*fos* is highly inducible by a variety of factors, including serum growth factors such as platelet-derived growth factor (PDGF), epidermal growth factor (EGF), and nerve growth factor. In addition, the gene is inducible by phorbol esters, cyclic adenosine monophosphate, and calcium ionophores. All of these stimuli lead to the activation of cellular kinases. Both endogenous and transfected c-*fos* genes are highly inducible in human HeLa and mouse 3T3 cells, but a c-*fos* enhancer–specific DNA-binding activity has been detected in nuclear extracts prepared from either uninduced or induced HeLa cells (*56, 57*). In contrast, a c-*fos* enhancer–binding activity is not detected in nuclear extracts from mouse A431 cells unless they are induced with EGF (*57*). Finally, a DNA-binding protein that binds to a different region of the c-*fos* 5′ flanking region has been detected in nuclear extracts from BALB/3T3 cells (*58*), and this activity is detected only when the cells have been treated with PDGF. These results suggest that the regulation of c-*fos* expression may involve different cis-acting sequences and transcription factors in different cell types.

The best understood trans-acting regulatory factors in higher eukaryotes are the steroid receptor proteins that bind to and activate the steroid-inducible enhancers (*59*). Several laboratories have shown that the binding site for the receptor-ligand complex coincides with the DNA sequences required for steroid-dependent gene regulation (*35, 37, 38, 60*). More recently, the genes encoding the estrogen (*61*), glucocorticoid (*62, 63*), and progesterone receptors (*64*) have been cloned and sequenced. Direct evidence that binding of the steroid receptor–ligand complex to the enhancer results in the activation of transcription was provided by experiments in which plasmids expressing receptor proteins and a plasmid containing a steroid-inducible enhancer were cotransfected into receptor-negative cell lines (*65, 66*). However, the mechanism by which this activation occurs is controversial. Early studies suggested that the steroid binds to receptor in the cytoplasm, and the hormone-receptor complex migrates to the nucleus where it binds to the gene and activates transcription (*59*). However, the evidence relevant to this model is conflicting. First, in some studies the estrogen receptor has been found in the cytoplasm in the absence of hormone, whereas in others it is found in the nucleus [see (*59*) and (*67*) for a discussion of these conflicting observations]. Second, whereas immunoaffinity-purified ligand-free glucocorticoid receptors bind specifically to steroid regulatory elements (*68*), ligand-free receptor synthesized in vitro (*69*) or present in crude cytoplasmic extracts that have not been activated by heat treatment (*68, 70*) will not bind to target DNA sequences. One possible explanation for this discrepancy could be that the ligand-free receptor is bound to a cytoplasmic protein and that steroid binding releases the receptor and allows nuclear translocation. A candidate for such a protein is a 90-kD heat-shock protein, which binds the receptor in untreated cells (*71*). Third, genomic footprinting experiments with the putative glucocorticoid response element of the tyrosine transaminase gene reveal only minor differences in the presence and absence of hormone (*72*). These differences could mean that the receptor is not bound in the absence of hormone. However, these data are also consistent with the possibility that the receptor is bound to the enhancer in the absence of hormone but that steroid binding induces a conformational change in the receptor that alters its contacts with DNA. Thus, in spite of intensive efforts, the question of whether hormone receptors are localized in the nucleus and bound to DNA in the absence of hormone has yet to be answered.

Temporal and Tissue-Specific Enhancers

The lymphoid cell–specific expression of immunoglobulin (Ig) genes is the best characterized example of tissue-specific gene expression (*73*). Immunoglobulin gene expression is not only limited to B cells; the genes are also activated at a specific stage during lymphoid cell differentiation. Analysis of the DNA sequence requirements for Ig gene regulation shows that a minimum of three distinct sequences are needed. In addition to a B cell–specific enhancer, Ig promoters are activated specifically in lymphoid cells, and intragenic sequences appear to be involved in a cell-specific post-transcriptional control (*74*). Specific interaction between Ig enhancers and promoters is indicated by the observation that these two elements act synergistically to control the level of Ig gene expression (*75*). This synergy could be the consequence of interactions between proteins bound to the two sequence elements, since a highly conserved sequence element designated the Ig octamer is found within both Ig enhancers and promoters, and this sequence interacts specifically with factors present in nuclear extracts (*76*).

B cell–specific enhancer elements have been localized within the introns of both the mu heavy-chain and kappa light-chain genes. Deletion studies with the heavy-chain gene enhancer showed a minimal functional region of approximately 140 base pairs (*77*). This region includes a number of conserved enhancer core sequences, but mutagenesis of these elements individually resulted in only minor effects on transcription (*78*). This result suggests that the enhancer is composed of several recognition elements that act together to establish maximum levels of transcription. The inactivation of any one of these elements is not sufficient to inactivate the enhancer. Certain regions of the enhancer appear to be negative regulatory elements, since their deletion leads to higher levels of transcription in non-B cells (*78, 79*).

As with the SV40 enhancer, DNA binding studies with the heavy-chain gene enhancer have revealed a complex pattern of protein-DNA interactions. Initially, genomic footprinting experiments showed five discrete binding sites within the enhancer, and the footprints were observed only in lymphoid cells (*80*). Four of these binding sites were designated E1 through E4 (Fig. 4). Each of these sites contains a sequence related to 5′-CAGGTGGC-3′. The presence of this sequence initially suggested that all four sites may be recognized by the same B cell–specific factor. However, subsequent in vitro footprinting experiments strongly suggest that each site binds to a different factor (*81–84*). The fifth site, designated "O," contains the Ig octamer sequence. Deletion of any one of the five

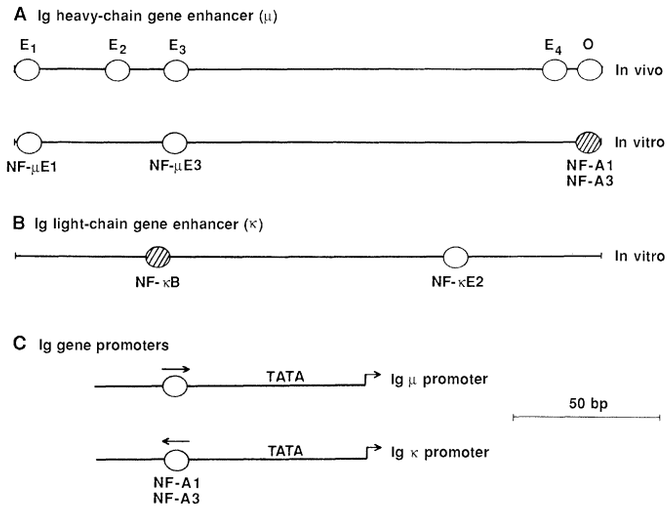

Fig. 4. The location of specific protein binding sites in the μ and κ enhancer elements. (**A**) Binding sites in the heavy-chain gene enhancer detected in vivo and in vitro. The binding sites are designated E1 through E4 and O (octamer). The names below the line representing the in vitro binding sites designate the different factors detected in cell extracts. (**B**) Binding sites in the kappa light-chain gene enhancer detected in vitro. (**C**) Binding sites in the μ and κ gene promoter detected in vitro. See text for further discussion.

recognition sites reduces, but does not eliminate, enhancer activity (*73*).

In vitro binding studies with the heavy-chain gene enhancer revealed specific binding to the E1, E3, and O sites (Fig. 4) and to additional sites that were not detected by genomic footprinting experiments (*81–84*). The functional significance of binding to these additional sites, if any, remains to be established. Failure to observe in vitro binding to the E2 and E4 sites is not understood, but may be the result of the inactivation of the binding proteins in the crude extracts.

An important point to emerge from the in vitro binding studies is that the factors binding to the E1, E3, and O sites are not limited to B cell extracts. The apparent discrepancy between this observation and the results of the genomic footprinting experiments, which show that these sites are occupied only in B cells, can be explained by assuming that the binding factors are ubiquitous, but the binding sites within the mu gene enhancer are accessible only in B cells. In addition to the ubiquitous factor that binds to the Ig octamer sequence, a B cell–specific factor that binds to the same sequence has been identified (*85, 86*). Similar studies with the kappa-chain gene enhancer have revealed two factors that bind to different sites (*87*). One of these factors is found in all cell types examined, while the other is detected only in mature B cell lines (*87*). Initially, this factor, called NF-κB, was thought to be present only in B cells, but more recent experiments have shown that it is present in an inactive form in many cell types but must be modified in some way before it can bind specifically to the kappa-chain gene enhancer (*88*). For example NF-κB was not detected in extracts prepared from early pre-B cells or from a variety of nonlymphoid cell types. However, when pre-B cells are treated with lipopolysaccharides (LPS), phorbol esters, or cycloheximide, which are known to induce kappa gene expression, the factor can be detected in nuclear extracts. Surprisingly, the binding activity of NF-κB is also induced by phorbol esters in T cells and HeLa cells, neither of which express the kappa-chain immunoglobulin gene (*88*).

The existence of a post-translational mechanism for activating regulatory factors involved in tissue-specific gene expression suggests one solution to the "regulatory regression" problem during development (*88*)—that is, how is the expression of regulatory genes regulated? If this is achieved entirely by regulation at the level of transcription, there must be regulatory cascades involving the sequential activation of genes encoding trans-acting regulatory factors. On the other hand, in a post-translational mechanism, all of the factors necessary for tissue-specific expression are present in immature precursor cells but are inactive. Inducers of cellular differentiation would then act in the absence of new transcription to convert the appropriate set of regulatory factors from an inactive to an active form.

An interesting role for NF-κB in Ig enhancer activity was recently suggested by the analysis of a plasmacytoma cell line that expresses the endogenous kappa-chain gene but fails to express transfected genes (*89*). These cells do not express NF-κB. Other plasmacytoma lines do express NF-κB and allow enhancer-dependent transcription of transfected kappa-chain genes. Pre-B cells also lack this factor, but treatment with LPS stimulates the appearance of NF-κB with the concomitant activation of the endogenous kappa gene enhancer. On the basis of the difference in behavior of the endogenous and transfected genes, and their apparent requirement for NF-κB, the authors speculate that the kappa gene enhancer is required for the establishment of kappa chain transcription, but not for its maintenance. Similar conclusions were reached on the basis of the analysis of mutant B cell lines that express high levels of heavy-chain immunoglobulins from genes in which the enhancer was deleted (*90*) and on competition experiments with the SV40 enhancer (*91*). Although the results of each of these studies are consistent with the existence of different mechanisms for the establishment and maintenance of gene activity, none of them provide direct evidence, and alternative explanations of the data are possible. The validity of this interesting and potentially important insight into the role of enhancers in gene regulation must await further studies.

In contrast to immunoglobulin genes, which are expressed only in one cell type, many genes are transcribed in several tissues or at several different times during development (or both). The fruit fly *Drosophila* is particularly well suited for studies of these more complex examples of gene regulation. Several *Drosophila* genes that are expressed in multiple tissues and at different developmental stages have been cloned, and the development of the P-element germ-line transformation method (*92*) has provided the opportunity to identify regulatory DNA sequences. With this method, single copies of cloned genes can be stably introduced into random locations in germ-line chromosomes. In most cases, *Drosophila* genes require only a few kilobases or less of their flanking DNA for normal regulation (*93*), and unlike the case in cultured mammalian cells and transgenic mice, the site of integration usually has little effect on the expression of the introduced gene.

On the basis of the immunoglobulin and SV40 paradigms, two simple models for the regulation of genes expressed in multiple cell types (or at different times) can be considered. In one model, expression of a single gene is controlled by different trans-acting factors present in different tissues. These tissue-specific factors could interact either with separate enhancers (one per cell type) or with the same enhancer. In the latter case, different factors could recognize different sequence features of the single enhancer element, similar to the case with the SV40 enhancer. Such tissue-specific trans-acting factors could control the expression of many genes transcribed in a particular tissue. In an alternative model, common trans-acting factors present in all of the cell types in which a gene is expressed control the expression of the gene. Such factors would interact, with a single enhancer in an identical manner in each cell type.

Deletion studies of the regulatory regions of several *Drosophila* genes have demonstrated that there are different DNA sequence requirements for expression of single genes in different tissues (*94–*

96), thus providing evidence for the first model described above in which different factors in different tissues regulate a single gene. The clearest example of separate DNA sequences that regulate the same promoter in different tissues is in the 5′ regulatory regions of the divergently transcribed yolk protein genes, *yp1* and *yp2*, which are expressed in the fat body and ovaries of adult females. The expression of both yolk protein genes in the fat body and ovaries is controlled by two distinct cis-acting elements (*94, 97*), each of which acts on the promoters of both genes. The adult female fat body element has been localized to a 125–base pair DNA fragment, which has all the characteristics of a stage-, sex-, and tissue-specific enhancer (*97*).

A much more complex example of tissue-specific gene regulation is seen in the larval alcohol dehydrogenase gene, *Adh-1*, of *Drosophila mulleri*. *Adh-1* is expressed in the larval fat body, Malpighian tubules, anterior midgut, and middle midgut of P-element–transformed *D. melanogaster* (*96*). Maximum levels of expression of *Adh-1* in all four tissues require at least three DNA elements, two within 300 base pairs on the 5′ side of the gene and another 3′ to the start of transcription. However, lower levels of expression can be achieved in the fat body, Malpighian tubules, and middle midgut, or in all four tissues, with the 3′ element plus one or the other of the two elements 5′ to the gene (*98*). The different sequence requirements in different cell types could reflect the presence of different concentrations of common factors in all four tissues or could indicate that different factors in each tissue interact in different ways with the three cis-acting DNA sequences. However, a series of clustered point mutations in one of the 5′ elements did not reveal any tissue-specific sequence requirements (*98*).

A similar phenomenon has also been observed in the mouse α-fetoprotein gene, which is expressed in the fetal yolk sac, liver, and gastrointestinal tract (*99*). Experiments in transgenic mice suggest that factors present in all of these tissues interact with three separate regions of an enhancer upstream of the α-fetoprotein gene, since any one of these enhancer regions can, along with the α-fetoprotein promoter, activate transcription in all three tissues. However, these experiments also suggest that there may be tissue-specific factors that interact more effectively with some regions of the enhancer than others, because the different enhancer regions were more efficient in activating transcription in some of the tissues examined.

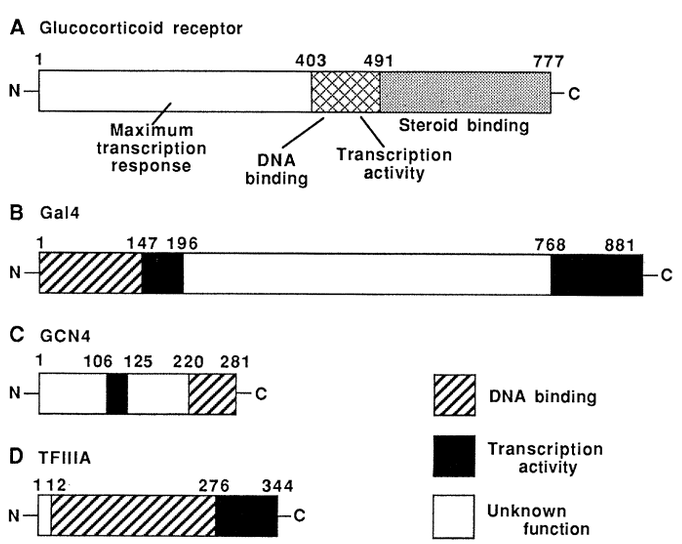

Fig. 5. Functional domains of eukaryotic transcription factors. A diagram showing the location of amino acid sequences required for specific DNA binding, steroid binding, or transcriptional activity of eukaryotic trans-acting regulatory factors. The various functional domains are represented by boxes with the number of amino acids from the NH_2 terminus of the protein indicated. (A) Glucocorticoid receptor; (B) Gal4 regulatory factor; (C) GCN regulatory factor; and (D) TFIIIA, 5*S* gene transcription factor.

Transcription Factors

Molecular cloning of the genes that encode trans-acting factors allows a functional analysis of the proteins themselves. In the case of the steroid receptors, these studies have led to the identification of domains of the proteins involved in DNA and steroid binding and in the stimulation of transcription (*63, 65, 66, 69, 100*) (Fig. 5A). The DNA binding domain has a cysteine-rich "finger" region, a structural feature that has been found in several other proteins with transcriptional regulatory function, most notably the DNA-binding transcription factor TFIIIA (*101*). A graphic illustration of the modular organization of steroid receptors was provided by experiments in which the DNA-binding region of the progesterone receptor was replaced by the corresponding region of the glucocorticoid receptor. The hybrid protein binds specifically to glucocorticoid regulatory regions and leads to the progesterone-dependent activation of linked genes (*102*). Thus, the DNA-binding and transcriptional activation domain can be physically separated from the domain responsible for hormone binding (Fig. 5A).

Linker insertion studies of the human glucocorticoid receptor initially suggested that sequences required for transcriptional activation could be separated from the DNA-binding domain (*66*). However, subsequent studies of the rat (*103*) and human (*104*) receptors showed that deletion of the regions thought to be required for transcription decreases, but does not inactivate, receptor-dependent transcription. In fact, only 86 to 88 of the 795 amino acids of the glucocorticoid receptor are necessary and sufficient for low levels of activity of the glucocorticoid regulatory element (*103, 104*). When truncated receptor molecules contain the hormone-binding domain, DNA binding and transcription occur only when the steroid is bound. However, deletion of the steroid binding domain results in the constitutive activation of the receptor (*69, 103, 104*). Thus, steroid binding may induce a conformational change in the receptor that unmasks the DNA binding and transcriptional activation potential.

Further studies are necessary to determine whether the DNA binding and transcriptional activity functions of the glucocorticoid receptor are separable. A naturally occurring receptor mutant nt^i, which is missing 417 NH_2-terminal amino acids, binds to hormone and to DNA but fails to function. Since deletion of the corresponding region of the wild-type receptor gene does not inactivate transcription, the nt^i mutant may contain additional mutations that uncouple DNA binding and enhancer activation (*103*). In contrast to the glucocorticoid receptor (Fig. 5A), DNA-binding and transcriptional activation domains are separable in the yeast transcriptional regulatory proteins Gal4 (*105*) and GCN4 (*106*), and in the *Xenopus* 5*S* gene transcription factor TFIIIA (*107*) (Fig. 5, B–D). Remarkably, most of the amino acid sequences of these transcription factors can be removed without effect. For example, as much as 80 percent of the Gal4 protein can be deleted and significant levels of transcription activity remain (*108*). A common feature of the amino acid sequences required for GCN4 (*106*) and Gal4 (*108*) transcriptional activity is the presence of a short stretch of acidic residues. How these residues interact with other proteins to stimulate transcription is not known.

The Future

The application of recombinant DNA techniques to the study of eukaryotic gene regulation has led to the identification and characterization of regulatory DNA sequences, and significant advances

are being made in the characterization of the protein factors that specifically interact with these sequences to promote or repress transcription. Information regarding the nature of these regulatory components will undoubtedly continue to accumulate at a rapid pace. However, the complexity of the transcriptional apparatus may prove to be a major obstacle to understanding mechanisms at the molecular level. The recent development of cell-specific in vitro transcription systems is encouraging in this regard (*109*). If inducible or tissue-specific gene activation involves the modification of one or a few limiting factors that are required for the assembly of a transcription complex, it may be difficult to establish conditions that accurately reflect those that exist in vivo.

REFERENCES AND NOTES

1. W. S. Dynan and R. Tjian, *Nature (London)* **316**, 774 (1985); S. McKnight and R. Tjian, *Cell* **46**, 795 (1986).
2. E. Serfling, M. Jasin, W. Schaffner, *Trends Genet.* **1**, 224 (1985).
3. K. Takahashi *et al.*, *Nature (London)* **319**, 121 (1986).
4. S. L. McKnight, *Cell* **31**, 355 (1982).
5. P. Dierks *et al.*, *ibid.* **32**, 695 (1983); D. Baty, H. A. Barrera-Saldana, R. D. Everett, M. Vigneron, P. Chambon, *Nucleic Acids Res.* **12**, 915 (1984); S. L. McKnight, R. C. Kingsbury, A. Spence, M. Smith, *Cell* **37**, 253 (1984); B. J. Graves, P. F. Johnson, S. L. McKnight, *ibid.* **44**, 565 (1986).
6. R. M. Myers, K. Tilly, T. Maniatis, *Science* **232**, 613 (1986).
7. D. J. Galas and A. Schmitz, *Nucleic Acids Res.* **5**, 3156 (1978).
8. M. Fried and D. M. Crothers, *ibid.* **9**, 6505 (1981); M. M. Garner and A. Rezvin, *ibid.*, p. 3047.
9. M. R. Briggs, J. T. Kadonaga, S. P. Bell, R. Tjian, *Science* **234**, 47 (1986); K. A. Jones *et al.*, *Cell* **48**, 79 (1986); J. T. Kadonaga and R. Tjian, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5889 (1986); P. J. Rosenfeld and T. J. Kelly, *J. Biol. Chem.* **261**, 1398 (1986); M. S. Kasher, D. Pintel, D. C. Ward, *Mol. Cell. Biol.* **6**, 3117 (1986); L. A. Chodosh, R. W. Carthew, P. A. Sharp, *ibid.*, p. 4723.
10. M. Zenke *et al.*, *EMBO J.* **5**, 387 (1986).
11. W. Herr and Y. Gluzman, *Nature (London)* **313**, 711 (1985); W. Herr and J. Clarke, *Cell* **45**, 461 (1986).
12. A. G. Wildeman *et al.*, *Mol. Cell. Biol.* **6**, 2098 (1986).
13. I. Davidson *et al.*, *Nature (London)* **323**, 544 (1986).
14. S. Schirm, J. Jiricny, W. Schaffner, *Genes Dev.* **1**, 65 (1987).
15. D. Bohmann *et al.*, *Nature (London)* **325**, 268 (1987); T. G. Parslow, S. D. Jones, B. Bond, K. R. Yamamoto, *Science* **235**, 1498 (1987).
16. R. Treisman and T. Maniatis, *Nature (London)* **315**, 72 (1985).
17. M. Bienz and H. R. B. Pelham, *Cell* **45**, 753 (1986).
18. M. Ptashne, *Nature (London)* **322**, 697 (1986).
19. R. E. Kingston, A. S. Baldwin, P. A. Sharp, *Cell* **41**, 3 (1985).
20. N. C. Jones, *Nature (London)* **321**, 202 (1986).
21. I. S. Y. Chen, *Cell* **47**, 1 (1986).
22. S. M. Abmayr, L. D. Feldman, R. G. Roeder, *ibid.* **43**, 821 (1985).
23. I. Kovesdi, R. Reichel, J. Nevins, *ibid.* **45**, 219 (1986).
24. R. Reichel, I. Kovesdi, J. Nevins, *ibid.* **48**, 501 (1987).
25. A. Velcich and E. Ziff, *ibid.* **40**, 705 (1985); R. Hen, E. Borrelli, P. Chambon, *Science* **230**, 1391 (1985).
26. C. M. Gorman, P. W. J. Rigby, D. P. Lane, *Cell* **42**, 519 (1985).
27. R. W. Stein and E. B. Ziff, *Mol. Cell. Biol.* **7**, 1164 (1987).
28. M. Katinka *et al.*, *Nature (London)* **290**, 720 (1981); K. Sekikawa and A. J. Levine, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1100 (1981); F. K. Fujimura and E. Linney, *ibid.* **79**, 1479 (1982); F. Melin *et al.*, *EMBO J.* **4**, 1799 (1985).
29. R. Hen, E. Borrelli, C. Fromental, P. Sassone-Corsi, P. Chambon, *Nature (London)* **321**, 249 (1986).
30. A. Velcich, F. G. Kern, C. Basilico, E. B. Ziff, *Mol. Cell. Biol.* **6**, 4019 (1986).
31. R. J. Imbra and M. Karin, *Nature (London)* **323**, 555 (1987); C. Wasylyk *et al.*, *Cell* **48**, 525 (1987).
32. E. Serfling, A. Lubbe, K. Dorsch-Hasler, W. Schaffner, *EMBO J.* **4**, 3851 (1985).
33. S. Goodbourn, K. Zinn, T. Maniatis, *Cell* **41**, 509 (1985); T. Fujita, S. Ohno, H. Yasumitsu, T. Taniguchi, *ibid.*, p. 489.
34. R. Treisman, *ibid.* **42**, 889 (1985).
35. V. L. Chandler, B. A. Maler, K. R. Yamamoto, *ibid.* **33**, 489 (1983); H. Ponta, N. Kennedy, P. Skroch, N. E. Hynes, B. Groner, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1020 (1985).
36. D. DeFranco and K. R. Yamamoto, *Mol. Cell. Biol.* **6**, 993 (1986); R. Miksicek *et al.*, *Cell* **46**, 283 (1986).
37. R. Renkawitz, G. Schutz, D. von der Ahe, M. Beato, *Cell* **37**, 503 (1984); M. Karin *et al.*, *Nature (London)* **308**, 513 (1984); E. P. Slater, O. Rabenau, M. Karin, J. D. Baxter, M. Beato, *Mol. Cell. Biol.* **5**, 2984 (1985); H. M. Jantzen *et al.*, *Cell* **49**, 29 (1987).
38. L. Klein-Hitpass, M. Schorpp, U. Wagner, G. U. Ryffel, *Cell* **46**, 1053 (1986).
39. A. Seiler-Tuyns *et al.*, *Nucleic Acids Res.* **14**, 8755 (1986).
40. P. F. Searle, G. W. Stuart, R. D. Palmiter, *Mol. Cell. Biol.* **5**, 1480 (1985).
41. G. W. Stuart, P. F. Searle, R. D. Palmiter, *Nature (London)* **317**, 828 (1985).
42. R. Dudler and A. Travers, *Cell* **38**, 391 (1984).
43. D. DeFranco, O. Wrange, J. Merryweather, K. R. Yamamoto, in *U.C.L.A. Symposium on Molecular and Cellular Biology 20*, I. Herskowitz and M. Simon, Eds. (Liss, New York, 1985), p. 305.
44. H. R. B. Pelham, *Cell* **30**, 517 (1982).
45. S. Goodbourn, H. Burstein, T. Maniatis, *ibid.* **45**, 601 (1986).
46. K. Zinn and T. Maniatis, *ibid.*, p. 611.
47. T. Fujita *et al.*, *ibid.* **49**, 357 (1987).
48. C. S. Parker and J. Topol, *ibid.* **37**, 273 (1984).
49. C. Wu, *Nature (London)* **309**, 229 (1984).
50. G. Wiederrecht, D. J. Shuey, W. A. Kibble, C. S. Parker, *Cell* **48**, 507 (1987).
51. R. E. Kingston, T. J. Schuetz, Z. K. M. B. Larin, *Mol. Cell. Biol.* **7**, 1530 (1987).
52. C. Wu, C. Parker, H. Pelham, personal communications.
53. J. Topol, D. M. Ruden, C. S. Parker, *Cell* **42**, 527 (1985).
54. D. J. Shuey and C. S. Parker, *J. Biol. Chem.* **261**, 7934 (1986).
55. ———, *Nature (London)* **323**, 459 (1986).
56. R. Treisman, *Cell* **46**, 567 (1986); M. Z. Gilman, R. N. Wilson, R. A. Weinberg, *Mol. Cell. Biol.* **6**, 4305 (1986); M. E. Greenberg, Z. Siegfried, E. B. Ziff, *ibid.* **7**, 1217 (1987).
57. R. Prywes and R. G. Roeder, *Cell* **47**, 777 (1986).
58. T. E. Hayes, A. M. Kitchen, B.H. Cochran, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1272 (1987).
59. K. R. Yamamoto, *Annu. Rev. Genet.* **19**, 209 (1985).
60. F. Payvar *et al.*, *Cell* **35**, 381 (1983); D. von der Ahe *et al.*, *Nature (London)* **313**, 706 (1985); A. C. B. Cato, R. Miksicek, G. Schutz, J. Arnemann, M. Beato, *EMBO J.* **5**, 2237 (1986); J.-P. Jost, M. Seldran, M. Geiser, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 429 (1984).
61. S. Green *et al.*, *Nature (London)* **320**, 134 (1986); G. L. Greene *et al.*, *Science* **231**, 1150 (1986); A. Krust *et al.*, *EMBO J.* **5**, 891 (1986).
62. R. Miesfield *et al.*, *Nature (London)* **312**, 779 (1984); S. M. Hollenberg *et al.*, *ibid.* **318**, 635 (1985).
63. M. Danielsen, J. P. Northrop, G. M. Ringold, *EMBO J.* **5**, 2513 (1986).
64. J. M. Jeltsch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5424 (1986); O. M. Conneely *et al.*, *Science* **233**, 767 (1986); H. Loosfelt *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9045 (1986).
65. R. Miesfield *et al.*, *Cell* **46**, 389 (1986); P. M. Druege *et al.*, *Nucleic Acids Res.* **14**, 9329 (1986).
66. V. Giguere, S. M. Hollenberg, M. G. Rosenfeld, R. M. Evans, *Cell* **46**, 645 (1986).
67. C. M. Szego and R. J. Pietras, *Nature (London)* **317**, 88 (1985).
68. A. Bailly, C. LePage, M. Rauch, E. Milgrom, *EMBO J.* **5**, 3235 (1986).
69. P. J. Godowski, S. Rusconi, R. Miesfeld, K. R. Yamamoto, *Nature (London)* **325**, 365 (1987); S. Rusconi and K. R. Yamamoto, *EMBO J.* **6**, 1309 (1987).
70. T. Willmann and M. Beato, *Nature (London)* **324**, 688 (1986).
71. E. R. Sanchez, D. O. Toft, M. J. Schlesinger, W. B. Pratt, *J. Biol. Chem.* **260**, 12398 (1985); M. G. Catelli *et al.*, *EMBO J.* **4**, 3131 (1985); I. Joab *et al.*, *Nature (London)* **308**, 850 (1984).
72. P. B. Becker, B. Gloss, W. Schmid, U. Strahle, G. Schutz, *Nature (London)* **324**, 686 (1986).
73. P. Matthias, T. Gerster, W. Schaffner, in *Nucleic Acids and Molecular Biology*, F. Eckstein and D. M. J. Lilley, Eds. (Springer-Verlag, Heidelberg, in press); T. Gerster, P. Matthias, M. Thali, J. Jiricny, W. Schaffner, *EMBO J.* **6**, 1323 (1987).
74. R. Grosschedl and D. Baltimore, *Cell* **41**, 885 (1985); J. O. Mason, G. T. Williams, M. S. Neuberger, *ibid.*, p. 479; T. V. Gopal, T. Shimada, A. W. Baur, A. W. Nienhuis, *Science* **229**, 1102 (1985); D. E. Kelley and R. P. Perry, *Nucleic Acids Res.* **14**, 5431 (1986); T. Gerster, D. Picard, W. Schaffner, *Cell* **45**, 45 (1986).
75. J. V. Garcia, J. Stafford, C. Queen, *Nature (London)* **322**, 383 (1986).
76. F. G. Falkner and H. F. Zachau, *ibid.* **310**, 71 (1984); T. G. Parslow, D. L. Blair, W. J. Murphy, D. K. Granner, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2650 (1984).
77. S. D. Gillies, S. L. Morrison, V. T. Oi, S. Tonegawa, *Cell* **33**, 717 (1983); J. Banerji, L. Olson, W. Schaffner, *ibid.*, p. 729.
78. T. Kadesch, P. Zervos, D. Ruezinsky, *Nucleic Acids Res.* **14**, 8209 (1986).
79. C. Wasylyk and B. Wasylyk, *EMBO J.* **5**, 553 (1986).
80. A. Ephrussi, G. M. Church, S. Tonegawa, W. Gilbert, *Science* **227**, 134 (1985); G. Church, A. Ephrussi, W. Gilbert, S. Tonegawa, *Nature (London)* **313**, 798 (1985).
81. H. Singh, R. Sen, D. Baltimore, P. Sharp, *Nature (London)* **319**, 154 (1986); J. Weinberger, D. Baltimore, P. Sharp, *ibid.* **322**, 846 (1986).
82. U. Schlokat, D. Bohmann, H. Scholer, P. Gruss, *EMBO J.* **5**, 3258 (1986).
83. C. Peterson, K. Orthe, K. L. Calame, *Mol. Cell. Biol.* **6**, 4168 (1986).
84. P. Augereau and P. Chambon, *EMBO J.* **5**, 1791 (1986).
85. N. F. Landolfi, J. D. Capra, P. W. Tucker, *Nature (London)* **323**, 548 (1986).
86. L. M. Staudt *et al.*, *ibid.*, p. 640.
87. R. Sen and D. Baltimore, *Cell* **46**, 705 (1986).
88. ———, *ibid.* **47**, 921 (1986).
89. M. L. Atchison and R. P. Perry, *ibid.* **48**, 121 (1987).
90. M. R. Wabl and P. D. Burrows, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2452 (1984); D. M. Zaller and L. A. Eckhardt, *ibid.* **82**, 5088 (1985); R. J. Aguilera, T. J. Hope, H. Sakano, *EMBO J.* **4**, 3689 (1985); I. Klein *et al.*, *Nucleic Acids Res.* **13**, 8901 (1986).
91. X. F. Wang and K. Calame, *Cell* **47**, 241 (1986).
92. A. C. Spradling and G. M. Rubin, *Science* **218**, 341 (1982); G. M. Rubin and A. C. Spradling, *ibid.*, p. 348.
93. S. B. Scholnick, B. A. Morgan, J. Hirsh, *Cell* **34**, 37 (1983); A. C. Spradling and G. M. Rubin, *ibid.*, p. 47; D. A. Goldberg, J. W. Posakony, T. Maniatis, *ibid.*, p. 59; T. Hazelrigg, R. Levis, G. M. Rubin, *ibid.* **36**, 469 (1984).
94. M. J. Garabedian, M. C. Hung, P. C. Wensink, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1396 (1985).
95. Y. Hiromi, A. Kuroiwa, W. J. Gehring, *Cell* **43**, 603 (1985); R. L. Levis, T. Hazelrigg, G. M. Rubin, *EMBO J.* **4**, 3489 (1985); V. Pirotta, H. Steller, M. P. Bozzetti, *ibid.*, p. 3501; S. B. Scholnick, S. J. Bray, B. A. Morgan, C. A. McCormick, J. Hirsh, *Science* **234**, 998 (1986).
96. J. A. Fischer and T. Maniatis, *EMBO J.* **5**, 1275 (1986).
97. M. J. Garabedian, B. M. Shepherd, P. C. Wensink, *Cell* **45**, 859 (1986).
98. J. A. Fischer and T. Maniatis, in preparation.
99. R. E. Hammer, R. Krumlauf, S. A. Camper, R. L. Brinster, S. M. Tilghman, *Science* **235**, 53 (1987).
100. V. Kumar, S. Green, A. Staub, P. Chambon, *EMBO J.* **5**, 2231 (1986).
101. J. Miller, A. D. McLachlan, A. Klug, *ibid.* **4**, 1609 (1985); J. Berg, *Science* **232**, 485 (1986); *Nature (London)* **319**, 264 (1986); S. C. Harrison, *ibid.* **322**, 597 (1986).

102. S. Green and P. Chambon, *Nature (London)* **325**, 75 (1987).
103. R. Miesfeld, P. J. Godowski, B. A. Maler, K. R. Yamamoto, *Science* **236**, 423 (1987).
104. S. M. Hollenberg, V. Giguere, P. Segui, R. M. Evans, *Cell* **49**, 39 (1987).
105. R. Brent and M. Ptashne, *ibid.* **43**, 729 (1985).
106. I. A. Hope and K. Struhl, *ibid.*, p. 177.
107. D. R. Smith, I. J. Jackson, D. D. Brown, *ibid.* **37**, 645 (1984).
108. J. Ma and M. Ptashne, *ibid.* **48**, 847 (1987).
109. J. M. Sugano and R. G. Roeder, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8511 (1986); Y. Suzuki *et al.*, *ibid.*, p. 9522; K. Gorski, M. Carneiro, U. Schibler, *Cell* **47**, 767 (1986).

Homeo Boxes in the Study of Development

WALTER J. GEHRING

The body plan of *Drosophila* is determined to a large extent by homeotic genes, which specify the identity and spatial arrangement of the body segments. Homeotic genes share a characteristic DNA segment, the homeo box, which encodes a defined domain of the homeotic proteins. The homeo domain seems to mediate the binding to specific DNA sequences, whereby the homeotic proteins exert a gene regulatory function. By isolating the normal *Antennapedia* gene, fusing its protein-coding sequences to an inducible promoter, and reintroducing this fusion gene into the germline of flies, it has been possible to transform head structures into thoracic structures and to alter the body plan in a predicted way. Sequence homologies suggest that similar genetic mechanisms may control development in higher organisms.

ORGANISMS DEVELOP ACCORDING TO A PRECISE DEVELOPmental program that specifies their body plan in great detail and also determines the sequence and timing of the developmental events. This developmental information is stored in the nucleotide sequences of the DNA. The question of how the one-dimensional sequence information stored in the DNA is converted into the three-dimensional structure of an embryo, or four-dimensional formation if we also include time, is the fundamental problem of developmental biology. Structural genes have been identified that specify the molecular building blocks from which the organism is constructed. The developmental program consists of a precise spatial and temporal pattern of expression of these structural genes that forms the basis of development. Normal development requires the coordinate expression of thousands of structural genes in a concerted fashion. Since independent control of the individual structural genes would lead to chaotic development, we might predict that there are controlling genes that regulate the activity of groups of structural genes coordinately. Such genes would presumably be arranged hierarchically or form a controlling network that ensures the proper timing of the developmental events and generates the proper spatial pattern. However, it proved to be difficult to find the controlling genes that specify the architecture, the body plan. Candidates for such developmental controlling genes were first identified as homeotic mutations in *Drosophila* as early as 1915 (*1*), but their molecular analysis had to await the advent of DNA technology. Homeotic mutations transform certain parts or an entire body segment into the corresponding structures of another body segment, thereby changing the architecture of the organism. *Drosophila* belongs to the dipteran insects that have only one pair of wings. However, in certain homeotic mutants, like those of the bithorax complex, the third thoracic segment becomes transformed into a second thoracic segment with a second pair of wings. This dramatic change in the architecture also may reflect evolutionary history, since the diptera evolved from more primitive insects that had four wings. Such homeotic mutations were found mainly in insects and other arthropods whose body is subdivided into typical segments along the anteroposterior body axis. However, they may also exist in vertebrates including humans.

The first homeotic genes were cloned in the absence of any biochemical information about their gene products by "chromosome walking" and by microdissection of bands from giant polytene chromosomes (*2*). The structural analysis of the *Antennapedia* (*Antp*) gene led to the discovery of the homeo box (Fig. 1), a small DNA segment of approximately 180 bp, that is characteristic for homeotic genes (*3, 4*). The significance of the homeo box homology was demonstrated by isolating previously unknown homeotic genes from *Drosophila* with the homeo box as a probe (*3*); perhaps more importantly, sequences homologous to the homeo box have been isolated from higher organisms including vertebrates (*5*), mammals, and humans (*6*). This might provide an entry point to cloning the genes that control development in higher organisms, on the basis of their partial homology to the *Drosophila* homeo box.

Comparative Anatomy of the Homeo Box

Analysis of the DNA sequences of the various homeo boxes shows that these sequences are highly conserved during evolution, whereas the flanking sequences differ considerably among different genes. The various homeo boxes share the same open reading frame, which extends into the flanking sequences and so indicates that the homeo box encodes a particular domain of the homeotic proteins, the homeo domain. A first hint with regard to the function of the homeo domain came from comparative protein sequence analysis, which revealed a small but significant degree of homology to the yeast mating–type proteins MAT a1 and MAT α 2 (*7*). These proteins are known to control cellular differentiation into mating-types a or α, or into spores, that is, into the three cell types that yeasts can form (*8*). They are sequence-specific DNA-binding

The author is professor of Developmental Biology and Genetics, Biozentrum, University of Basel, Switzerland.