Neuron, Vol. 7, 177–181, August, 1991, Copyright © 1991 by Cell Press

Mechanisms for Diversity in Gene Expression Patterns

Review

Kevin Struhl
Department of Biological Chemistry
and Molecular Pharmacology
Harvard Medical School
Boston, Massachusetts 02115

The human central nervous system contains about 10^{12} cells whose actions define the world as we know it. Although the number of classically defined cell types is rather small, the regulatory complexity displayed by individual genes indicates that many, and perhaps nearly all, of the cells in the central nervous system are distinct with respect to which genes are expressed. In addition to this cellular specificity, gene regulatory patterns are constantly changing throughout development and in response to extracellular signals. As a result, transcriptional regulatory patterns in the central nervous system are extraordinarily complex. As a rough estimate, probably between 10^4 and 10^5 genes are expressed, many of them in unique and unexpectedly complicated cellular patterns (Bier et al., 1989; McKay and Hockfield, 1982; Sutcliffe, 1988).

This enormous diversity and flexibility in gene expression patterns is accomplished with a relatively small number of transcription factors. There are undoubtedly hundreds of transcription factors and possibly as many as a few thousand, but more than this seems very unlikely. It is evident, therefore, that a "one regulatory protein per gene" model, such as frequently applies to prokaryotic organisms, is grossly inadequate. Instead, combinatorial action of transcriptional regulatory proteins is necessary for multicellular organisms to generate the requisite diversity in gene expression patterns. This review will discuss the molecular mechanisms involved in generating diversity.

Combinatorial Activation of Transcription

The most important mechanism for achieving diversity, combinatorial activation, relies on the basic properties of the eukaryotic transcriptional machinery. For protein-coding genes, this machinery consists of RNA polymerase II and several auxiliary factors including TFIID, which binds to the conserved TATA element found in most eukaryotic promoters (Sawadogo and Sentenac, 1990). By itself, RNA polymerase II is transcriptionally inactive on normal DNA templates. However, after binding of TFIID to the TATA element and subsequent assembly of the other factors into an active transcription complex, RNA polymerase II can initiate synthesis at a site 25–30 bp downstream of the TATA element. However, this "basic transcriptional machinery" is not sufficient to promote transcription in vivo because promoters containing only the TATA element and initiation region are essentially inactive. Thus, the eukaryotic RNA polymerase II transcriptional machinery is qualitatively different from prokaryotic RNA polymerase holoenzymes that are sufficient for efficient transcriptional initiation.

In eukaryotic organisms, gene expression requires activator proteins that bind to specific promoter sequences and stimulate the basic transcriptional machinery (Mitchell and Tjian, 1989; Ptashne, 1988; Struhl, 1989). Thus, a first-order description of a particular transcriptional regulatory pattern is simply a matter of which specific activator proteins can interact at the promoter. In this view, a set of genes can be coordinately regulated if their promoters contain related DNA sequence elements that can interact with a common activator protein. In general, the related promoter elements are not identical, but strongly resemble a consensus sequence, which is often functionally optimal. This ability to interact efficiently with a range of related sequences allows for regulatory and evolutionary flexibility. However, the number of distinct DNA-binding specificities is far too limited to account for the diversity of transcriptional regulatory patterns. More importantly, generating complex expression patterns would be impossible if a single activator protein were sufficient to enhance transcription, because all genes containing a common promoter element would be coordinately expressed in a given cell type.

The fundamental aspect of the RNA polymerase II machinery that addresses the diversity problem is that efficient transcription requires the combinatorial action of activator proteins. A single activator protein bound at one site in the promoter typically confers a very low level of gene expression. In contrast, transcription is stimulated much more efficiently (factors of 5–1000) by the combination of multiple activator proteins bound at distinct promoter sites. Most importantly, such transcriptional synergy is frequently observed even when the multiple binding sites are recognized by distinct, and even evolutionarily distant, proteins. As an example of such promiscuity, the combination of the mammalian glucocorticoid receptor and the yeast GAL4 protein is much more effective than either protein alone. Although the mechanism(s) of synergistic and promiscuous activation remains to be elucidated, the requirement for multiple activator proteins at a promoter permits a very large number of possible combinations, each of which might be biologically distinct.

The regulatory flexibility due to transcriptional synergy is greatly enhanced by the ability of activator proteins to function bidirectionally at long and variable distances either upstream or downstream from the mRNA initiation site. Such action at a distance is believed to reflect interactions between distantly bound proteins that are brought into close proximity by looping out of the intervening DNA. In general, an activator protein becomes less efficient when bound

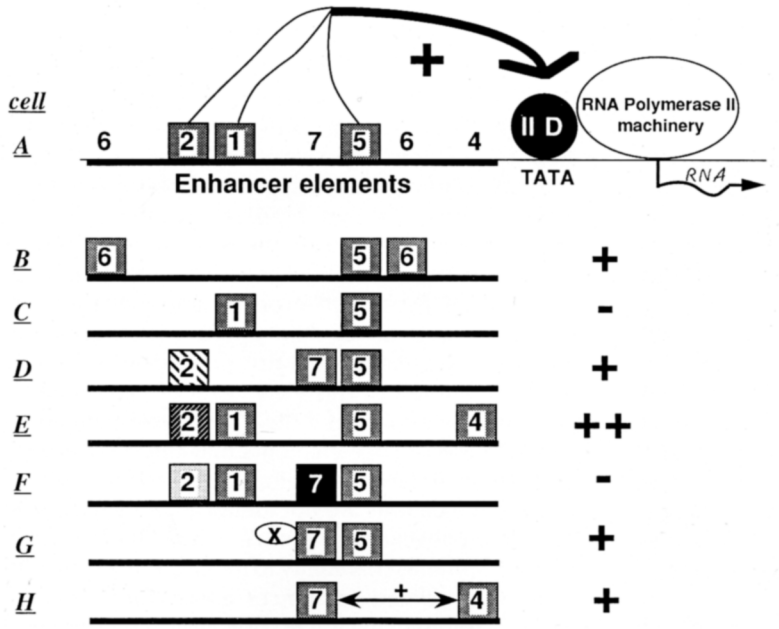

Figure 1. Generation of a Complex Expression Pattern for a Hypothetical Gene Containing Seven Enhancer Elements Upstream of the TATA Element and mRNA Initiation Site

Each of the eight cells (A–H) contains a particular array of activator proteins (1–7) bound directly to the promoter region (closed boxes, with distinctive shadings indicating individual members of a multiprotein family as in 2 and 7) or indirectly via a protein–protein interaction (between X and 7 in cell G); protein 6 is cell type specific, whereas protein 5 is nearly ubiquitous. As a simple arbitrary rule, any three activator proteins can stimulate the basic transcriptional machinery unless a repressor protein is also bound to the promoter (at site 7 in cell F). Four activator proteins permit higher expression levels (cell E), and two activator proteins are insufficient (cell C) unless there is a synergistic protein–protein interaction (cell H).

at increasing distances from the initiation site, but individual proteins display considerable variability (in this regard, the common distinction between "promoter" and "enhancer" binding proteins is artificial). Whatever the precise molecular mechanisms involved, the important principle is that a promoter can be subject to the action of numerous proteins whose target sequences can be spread out over a large chromosomal region. Indeed, there are already examples in flies and mammals in which sequences 30–50 kb from the initiation site play an important regulatory role (Grosveld et al., 1987; Karch et al., 1990). Protein-binding sites are often tightly clustered into enhancers that can be moved as a functionally autonomous unit, but such a genomic organization is not essential.

Given these properties, an enormous diversity of transcriptional regulatory patterns can be generated (see Figure 1). In simple cases, dedicated promoters responding to a single activator protein can be arranged by having multiple copies of a common binding site. More typically, genes whose promoters contain multiple distinct sites could be efficiently expressed only when certain developmental or environmental conditions are met simultaneously. Redundant promoters that contain more elements than necessary can permit expression under several different, but specific, circumstances. For the most complex expression patterns, such as observed for genes that determine cell fate or that are responsible for the synthesis of neurotransmitters, numerous protein-binding sites are scattered over large regions of DNA such that a wide variety of different protein combinations can activate transcription. Genetic experiments in Drosophila often reveal that individual elements are required for strikingly discrete portions of the overall pattern. Finally, the principle of combinatorial activation results in regulatory networks in which sets of genes are coordinately controlled by specific environmental or developmental signals, yet the individual genes can be members of many different sets.

Families of Transcription Factors

Although the combinatorial activation process clearly generates an impressive amount of diversity, it is limited by the number of distinct DNA-binding specificities of the activator proteins. The number of possible recognition sequences is limited by the small number of base pairs (typically 6–8) that are involved in high affinity protein–DNA interactions. Moreover, it is very likely that the inherent chemistries of nucleotides and amino acids severely restrict which DNA sequences can serve as protein-binding sites. This restriction is compounded by structural and evolutionary constraints on the number of DNA-binding motifs (e.g., helix-turn-helix, zinc finger, bZIP, and helix-loop-helix).

Multiprotein families of transcription factors that recognize related DNA sequences constitute an important diversity mechanism for overcoming some of the above constraints. Examples of such families are homeodomain proteins that control key developmental decisions (Levine and Hoey, 1988); steroid hormone receptors (Evans and Hollenberg, 1988); the AP-1 and ATF/CREB proteins, which utilize the bZIP structural motif (Curran and Franza, 1988; Hai et al., 1989); and the helix-loop-helix proteins such as myoD, E12/E47, and achaete-scute (Murre et al., 1989). Such protein families are likely to regulate a core group of genes in a variety of cell types and developmental stages or in response to extracellular signals. How-

ever, the individual proteins in each of these families, though structurally and functionally related, do not necessarily have identical DNA-binding specificities. Thus, the precise DNA sequence of a promoter element can determine which particular members of a multiprotein family will regulate the expression of the gene. Moreover, the spectra of genes affected by individual family members could differ dramatically, especially in the common situation in vivo in which the relatively large number of potential binding sites are competing for the limited amounts of protein.

Heterodimer formation between individual members of a protein family can provide an additional diversity mechanism that increases the number of DNA-binding transcription factors. Such dimerization interactions can be mediated by the leucine zipper (Landschulz et al., 1988) or helix-loop-helix (Murre et al., 1989) motifs, and the resulting heterodimers can be functionally distinct from the parental homodimers with respect to their DNA-binding or their transcriptional activation properties (i.e., inherent strength or regulated activity). In a given cell type or under particular environmental circumstances, the constellation of transcription factors of a particular family will depend on the amounts of the individual proteins present and on the relative strengths of the dimerization interactions.

Protein–Protein Interactions

The diversity mechanisms described above make the simplifying assumption that a particular pattern of gene expression reflects synergistic activation that occurs in the absence of direct interactions between the specific transcription factors bound at the promoter. However, such protein–protein interactions clearly occur, and they can result in dramatic transcriptional effects. Although relatively few such protein–protein interactions have been characterized in detail at the present time, it is very likely that they serve as an important source of regulatory diversity.

Protein–protein interactions between transcription factors can influence gene regulation by a variety of distinct molecular mechanisms. First, a DNA-binding protein with low transcriptional activity can be converted to a potent activator by interacting with a separate non-DNA-binding protein that contains a strong acidic activation region. For example, the acidic activation domain of herpesvirus VP16 interacts with the homeodomain of Oct-1 (Stern et al., 1989). Second, interactions between two proteins can result in the cooperative binding of both proteins to target DNA sequences in the promoter under conditions in which neither protein can bind alone. Cooperative DNA binding can involve two molecules of the same protein, as is the case for steroid hormone receptors (Schmid et al., 1989; Tsai et al., 1989), or two distinct protein species, as in the MCM1/α2 interaction (Keleher et al., 1988). As initially described for developmental decisions of bacteriophage λ (Johnson et al., 1981), and as is likely for the initial response to the *bicoid* gradient morphogen in early Drosophila embryos (Driever et al., 1989; Struhl et al., 1989), cooperative DNA binding provides a means by which the level of gene expression is extremely sensitive to small changes in protein concentration. Third, interactions between two DNA-binding transcription factors can either augment or inhibit gene expression, as observed for the steroid hormone receptors and the AP-1 protein family (Diamond et al., 1990; Schüle et al., 1990; Yang-Yen et al., 1990) as well as for yeast MCM1 and the cell type regulators α1 and α2 (Bender and Sprague, 1987; Keleher et al., 1988). Fourth, heteromeric protein complexes such as CTF1 (Chodosh et al., 1988) and HAP2/3/4 (Olesen and Guarente, 1990) can be necessary for a single DNA-binding event, whereas complexes such as a1/α2 can have DNA sequence specificities that differ from either of the individual components (Goutte and Johnson, 1988). Whatever the particular molecular mechanism, the regulatory combinations mediated by protein–protein interactions add a new level of diversity beyond combinatorial activation and multiprotein families.

Modification of Protein Activity

Although much of the diversity in multicellular organisms depends simply upon which transcription factors are present in the various cell types, variations in the activities of the proteins also make a major contribution. Differences in protein activity can occur at the level of DNA binding, inherent transcriptional activation potential, or protein–protein interactions; hence they amplify all the diversity mechanisms described above. One standard means by which protein activity can be altered is by phosphorylation or by other covalent modifications. In the case of phosphorylation, the major protein kinases are activated by second messengers (cAMP, inositol phosphates, diacylglycerol, and calcium) that are generated by signal transduction pathways; however, other protein kinases are almost certainly involved as well. Another classic way to affect protein activity is by allosteric interaction with small molecules (e.g., hormones, amino acids, and cAMP). Both of these mechanisms for altering the activity of specific transcription factors are utilized extensively in prokaryotic organisms, and they provide the major basis for modulating gene expression in response to extracellular signals.

Eukaryotic cells have a novel way to modify protein activity effectively, namely, regulation of nuclear localization. In the case of NF-κB, the protein is translocated to the nucleus only under particular conditions that inactivate a specific inhibitor protein (IκB) which otherwise sequesters NF-κB in the cytoplasm (Baeuerle and Baltimore, 1988). Other members of the NF-κB family, the *rel* oncoprotein and the *dorsal* morphogen of Drosophila, presumably function in a similar manner (Gilmore, 1990). A different mechanism for regulating nuclear localization is exemplified by the

glucocorticoid receptor, which in the absence of hormone is excluded from the nucleus by virtue of an interaction with a heat shock protein (Picard et al., 1990).

Negative Regulation

By counterbalancing the actions of activator proteins, transcriptional repressors provide another fundamental mechanism for achieving diversity. Repressors inhibit gene expression by a variety of molecular mechanisms, including competitive DNA binding to coincident or overlapping promoter elements, inactivation of a bound activator protein, or direct repression (silencing) of the basic transcriptional machinery (Levine and Manley, 1989). Regardless of the particular molecular mechanism, repressors contribute to diversity by using the basic principles of combinatorial action, multiprotein families, heterodimerization, protein–protein interactions, and modification of protein activity. Moreover, multiprotein families often include both activators and repressors, and protein–protein interactions can have synergistic or antagonistic consequences for gene expression.

Summary

Despite the relatively low number of transcriptional regulatory proteins, the number of possible combinations that act in particular cell types at specific times and in response to appropriate extracellular stimuli is enormous. In considering the regulatory patterns of a particular gene, the critical determinants of diversity are the specific promoter sequences that govern the potential DNA-binding proteins which function either directly or indirectly in association with other proteins; constellations of proteins in the nucleus and their transcriptional activities; and synergistic or antagonistic protein–protein interactions. Although some of these regulatory principles operate in prokaryotes, the combinatorial nature of the transcriptional activation process, the existence of multiprotein families, and the prevalance of heteromeric protein complexes are characteristic of eukaryotic cells and are essential for the extraordinary complexity of gene expression patterns in multicellular organisms.

References

Baeuerle, P. A., and Baltimore, D. (1988). IκB: a specific inhibitor of the NF-κB transcription factor. Science *242*, 540–546.

Bender, A., and Sprague, G. F., Jr. (1987). MATα1 protein, a yeast transcription activator, binds synergistically with a second protein to a set of cell-type-specific genes. Cell *50*, 681–691.

Bier, E., Vaessin, H., Shepherd, S., Lee, K., McCall, K., Barbel, S., Ackerman, L., Carretto, R., Uemura, T., Grell, E., Jan, L. Y. and Jan, Y. N. (1989). Searching for pattern and mutation in the *Drosophila* genome with a P-*lacZ* vector. Genes Dev. *3*, 1273–1287.

Chodosh, L. A., Baldwin, A. S., Carthew, R. W., and Sharp, P. A. (1988). Human CCAAT-binding proteins have heterologous subunits. Cell *53*, 11–24.

Curran, T., and Franza, B. R., Jr. (1988). Fos and Jun: the AP-1 connection. Cell *55*, 395–397.

Diamond, M., Miller, J. N., Yoshinaga, S. K., and Yamamoto, K. R. (1990). c-*jun* and c-*fos* levels specify positive or negative glucocorticoid regulation from a composite GRE. Science *249*, 1266–1272.

Driever, W., Thoma, G., and Nusslein-Volhard, C. (1989). Determination of spatial domains of zygotic gene expression in the *Drosophila* embryo by the affinity of binding sites for the bicoid morphogen. Nature *340*, 363–367.

Evans, R. M., and Hollenberg, S. M. (1988). Zinc fingers: gilt by association. Cell *52*, 1–3.

Gilmore, T. D. (1990). NF-κB, KBF1, *dorsal*, and *rel*ated matters. Cell *62*, 841–843.

Goutte, C., and Johnson, A. D. (1988). a1 protein alters the DNA binding specificity of α2 repressor. Cell *52*, 875–882.

Grosveld, F., van Assendelft, G. B., Greaves, D. R., and Kollias, B. (1987). Position-independent, high-level expression of the human β-globin gene in transgenic mice. Cell *51*, 975–985.

Hai, T., Liu, F., Coukos, W. J., and Green, M. R. (1989). Transcription factor ATF cDNA clones: an extensive family of leucine zipper proteins able to selectively form DNA-binding heterodimers. Genes Dev. *3*, 2083–2090.

Johnson, A. D., Poteete, A. R., Lauer, G., Sauer, R. T., Ackers, G. R., and Ptashne, M. (1981). λ repressor and cro-components of an efficient molecular switch. Nature *294*, 217–223.

Karch, F., Bender, W., and Weiffenbach, B. (1990). *abdA* expression in *Drosophila* embryos. Genes Dev. *4*, 1573–1587.

Keleher, C. A., Goutte, C., and Johnson, A. D. (1988). The yeast cell-type-specific repressor α2 acts cooperatively with a non-cell-type-specific protein. Cell *53*, 927–936.

Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science *240*, 1759–1764.

Levine, M., and Hoey, T. (1988). Homeobox proteins as sequence-specific transcription factors. Cell *55*, 537–540.

Levine, M., and Manley, J. L. (1989). Transcriptional repression of eukaryotic promoters. Cell *59*, 405–408.

McKay, R. D. G., and Hockfield, S. J. (1982). Monoclonal antibodies distinguish antigenically discrete neuronal types in the vertebrate central nervous system. Proc. Natl. Acad. Sci. USA *79*, 6747–6751.

Mitchell, P., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science *245*, 371–378.

Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C. V., Buskin, J. N., Hauschka, S. D., Lassar, A. B., Weintraub, H., and Baltimore, D. (1989). Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell *58*, 537–544.

Olesen, J. T., and Guarente, L. (1990). The HAP2 subunit of yeast CCAAT transcriptional activator contains adjacent domains for subunit association and DNA recognition: model for the HAP2/3/4 complex. Genes Dev. *4*, 1714–1729.

Picard, D., Khursheed, B., Garabedian, M. J., Fortin, M. G., Lindquist, S., and Yamamoto, K. R. (1990). Reduced levels of hsp90 compromise steroid receptor action *in vivo*. Nature *348*, 166–168.

Ptashne, M. (1988). How eukaryotic transcriptional activators work. Nature *335*, 683–689.

Sawadogo, M., and Sentenac, A. (1990). RNA polymerase B (II) and general transcription factors. Annu. Rev. Biochem. *59*, 711–754.

Schmid, W., Strahle, U., Schutz, G., Schmitt, J., and Stunnenberg, H. (1989). Glucocorticoid receptor binds cooperatively to adjacent recognition sites. EMBO J. *8*, 2257–2263.

Schüle, R., Rangarajan, P., Kliewer, S., Ransone, L. J., Bolado, J., Yang, N., Verma, I. M., and Evans, R. M. (1990). Functional

antagonism between oncoprotein c-Jun and the glucocorticoid receptor. Cell *62*, 1217–1226.

Stern, S., Tanaka, M., and Herr, W. (1989). The Oct-1 homeodomain directs formation of a multiprotein-DNA complex with the HSV transactivator VP16. Nature *341*, 624–630.

Struhl, G., Struhl, K., and Macdonald, P. M. (1989). The gradient morphogen *bicoid* is a concentration-dependent transcriptional activator. Cell *57*, 1259–1273.

Struhl, K. (1989). Molecular mechanisms of transcriptional regulation in yeast. Annu. Rev. Biochem. *58*, 1051–1077.

Sutcliffe, J. G. (1988). Messenger RNA in the mammalian central nervous system. Annu. Rev. Neurosci. *11*, 157–198.

Tsai, S. Y., Tsai, M.-J., and O'Malley, B. W. (1989). Cooperative binding of steroid hormone receptors contributes to transcriptional synergism at target enhancer elements. Cell *57*, 443–448.

Yang-Yen, H.-F., Chambard, J.-C., Sun, Y.-L., Smeal, T., Schmidt, T. J., Drouin, J., and Karin, M. (1990). Transcriptional interference between c-Jun and the glucocorticoid receptor: mutual inhibition of DNA binding due to direct protein–protein interaction. Cell *62*, 1205–1215.