MOLECULAR MEDICINE

TRANSCRIPTION FACTORS

ATHANASIOS G. PAPAVASSILIOU, M.D., PH.D.

A FUNDAMENTAL component of the complex instructions for building and maintaining an organism is the regulation of gene expression. The rules governing gene transcription explain why neurons and osteoclasts have completely different forms and functions even though all the nucleated cells in a person contain the same DNA and thus the same genetic information. A neuron is neither an osteoclast nor a granulocyte because of the neuron-specific pattern of gene transcription it elaborates over time and in particular anatomical spaces. The regulation of gene expression comes into play in circumstances as wide-ranging as the maturation of antibody-secreting cells, the production of intestinal epithelial cells, and the transformation of a fertilized ovum into an embryo. But differentiation and maturation are not the only processes controlled by the differential expression of genes. The cell selectively expresses certain genes in response to stress, injury, infections, hormones, growth factors, and many other external stimuli.

Clinicians encounter examples of normal and abnormal gene expression daily. Most cancers arise from abnormalities in the expression of genes concerned with growth and differentiation (proto-oncogenes). Various metabolic disorders and perturbed hormone action are the consequence of aberrant gene expression. Prolonged or incomplete healing after surgery indicates defects in the genetic signals required for the regeneration of tissue.

During the past decade, great advances have been made in uncovering the mechanisms that switch genes on and off. Of particular interest is a class of proteins, the transcription factors, with pivotal roles in regulating gene expression. As noted in a previous article in this series (see Recommended Reading), transcription factors bind to regulatory elements in DNA known as promoters and enhancers. They stimulate (or sometimes inhibit) gene transcription, and thus the formation of messenger RNA, through direct interactions with DNA. The modular arrangement of transcription factors into distinct domains serves their function well (Fig. 1). Typically, the DNA-binding domain of a transcription factor has a helical shape (α helix) within or adjacent to which reside clusters of positively charged amino acids. This domain binds specifically and tightly to short double-stranded segments of DNA in promoters or enhancers. For example, the GATA-1 transcription factor, which I shall return to later, binds specifically to the double-stranded structure made by $\frac{\text{TGATAG}}{\text{ACTATC}}$.

Transcription factors do not function in isolation. On the contrary, they form regulatory networks in which several factors interact to regulate gene transcription. A physically separate domain in transcription factors, the transactivation domain, mediates these cooperative associations. Even in synergistic combinations, transcription factors are by no means the whole story of gene regulation. A particular gene becomes accessible to transcription factors first and foremost because of the higher-order structure of the DNA in which the gene resides. This complex structure takes its shape from the reiterative coiling of the DNA double helix into supercoils and chromosomes and the association of DNA with DNA-binding and scaffolding proteins, of which the best known are the histones.

These arrangements in the cell nucleus are important, but equally notable are the extracellular or intracellular signals that lead to modifications in the DNA-binding and transactivation domains of transcription factors. Phosphorylation, triggered by biochemical signals from cell-surface receptors, is a common means of altering the function of transcription factors. Steroid hormones entering the cell can also interact with and modulate the function of transcription factors. These kinds of mechanisms form the basis of gene regulation during the cell's response to a wide variety of external stimuli, such as neurotransmitters, antigens, cytokines, and growth factors.

Recurrent structural motifs in DNA-binding domains provide the basis for the classification of transcription factors. More than 80 percent of these factors contain one of four motifs: helix–turn–helix (HTH), zinc finger, leucine zipper, or helix–loop–helix (HLH) (Fig. 1).

In all species, from microbes to humans, many different transcription factors have DNA-binding domains with the HTH motif. These proteins are important in the embryonic development of multicellular organisms. Indeed, their key role in the morphogenesis of the embryo was first identified in the larvae of the fruit fly (*Drosophila melanogaster*). HTH transcription factors have subsequently been shown to function as developmental switches in humans as well. Clinical interest in the HTH family of transcription factors has been aroused by several new findings. Some of these regulatory proteins are essential for the differentiation of hematopoietic precursor cells, and others have been implicated in the development of childhood leukemia and other neoplasms.

The so-called homeobox (or homeodomain) contains the HTH motif. This region of the HTH transcription factors consists of three closely linked α helixes, one of which interacts with DNA at multiple sites (Fig. 1). In many transcription factors of the HTH type, additional motifs flank the homeodomain and modulate its DNA-binding affinity and its specificity for particular base sequences in the DNA. The Pit-1, Oct-1 and Oct-2, and Unc-86 (POU) proteins, which provide a clear example of such combined motifs, play a critical part in the regulation of gene expression during the development of the immune and central nervous systems.

A second motif, the zinc finger, consists of either a pair of histidines and a pair of cysteines or two pairs of

From the Differentiation Program of the European Molecular Biology Laboratory, Postfach 102209, D-69012 Heidelberg, Germany, where reprint requests should be addressed to Dr. Papavassiliou.

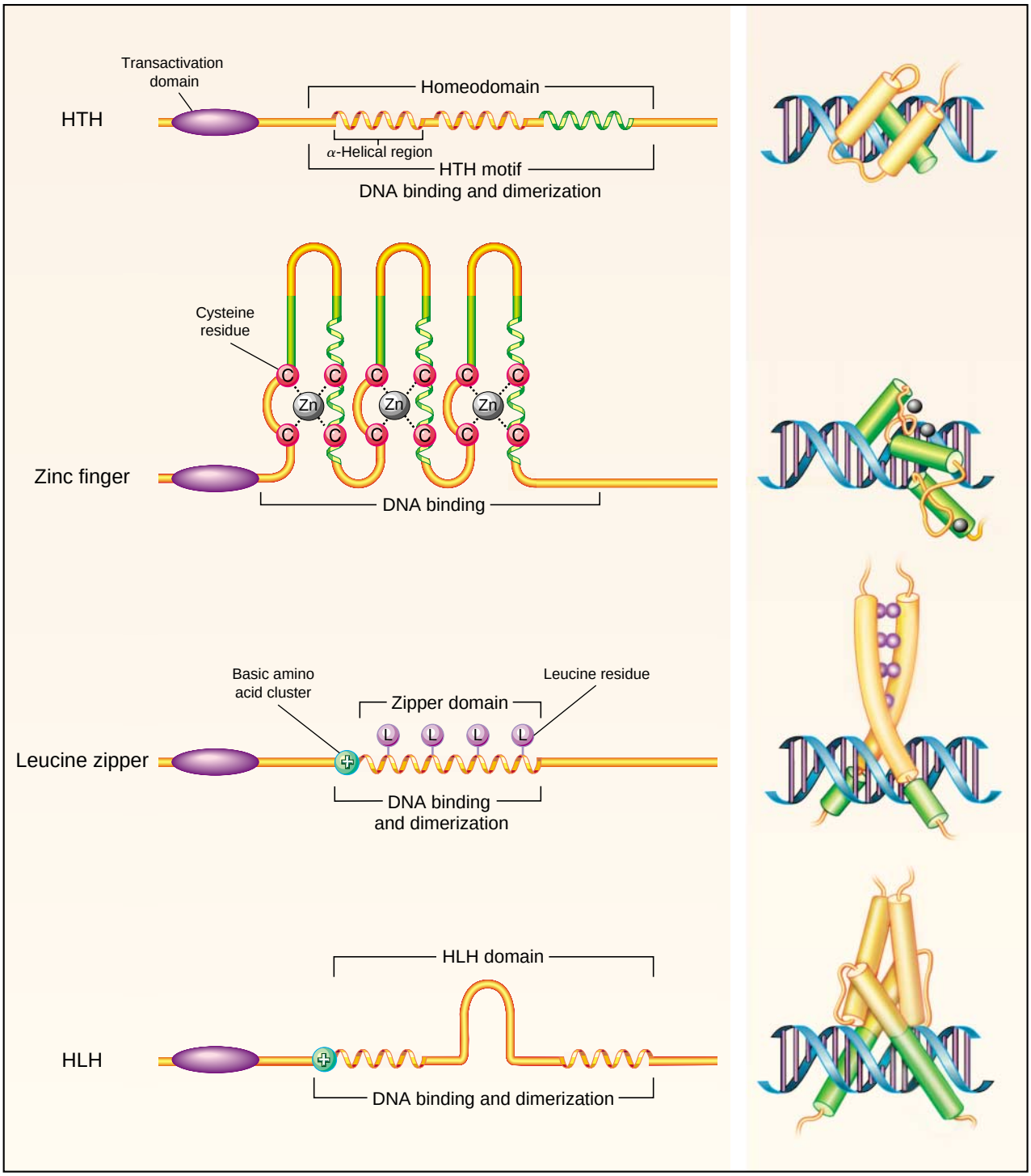

Figure 1. Transcription Factors and Their Interactions with DNA.

The general protein structures of the four major classes of transcription factors described in the text are shown on the left. The structures include a transactivation domain linked to a DNA-binding domain and, in certain cases, a dimerization domain. The types of transcription factor take their names from the characteristic motifs (helix–turn–helix [HTH], zinc finger, leucine zipper, and helix–loop–helix [HLH]) involved in DNA binding and protein dimerization and are shown on the right interacting with DNA. The cylinders represent α-helical regions, and the areas that contact DNA directly are green. Zn denotes zinc, C cysteine, and L leucine. The plus sign indicates a positive charge.

cysteines that together grasp a zinc ion. The polypeptide in between these pairs of amino acids loops out and folds into a finger-shaped configuration that is usually organized as a single series of tandem repeats (Fig. 1). The specificity for a DNA sequence resides in the amino acid residues at the base and body, rather than the tip, of each finger. Zinc-finger transcription factors function by binding as dimers to response ele-

ments in the upstream region of the gene under control. Thus far, a broad spectrum of transcription factors has been shown to possess the zinc finger or related motifs.

Steroid hormone receptors are a special class of ligand-activated zinc-finger proteins. Thyroid hormone receptors also have this general structure. The zinc-finger family includes genes for retinoic acid and vitamin D receptors as well. A single amino acid replacement within one of the zinc fingers of a vitamin D receptor was found to be the cause of hereditary vitamin D–resistant rickets in one family. Recent investigations suggest that steroid hormone receptors are involved in malignant growth. These receptors may activate the hormone-regulated gene in a constitutive manner that is independent of the hormone. Mutant estrogen receptors, which have been reported in some breast cancers, may contribute to the loss of estrogen dependence. Transcription factors with the zinc-finger motif are also implicated in other cancers. The oncogene *bcl*-6, which has an important role in large-cell lymphoma, is a zinc-finger protein. Another clinically relevant zinc-finger protein is WT1, a tumor-suppressor protein that is inactivated in children with Wilms' tumor. The GATA-1 protein, mentioned previously, also belongs to the zinc-finger family. This transcription factor has an important role in the regulation of gene expression in differentiating erythroid cells, megakaryocytes, and mast cells. In the erythroid line, GATA-1 participates in the control of globin gene expression. Other kinds of zinc-finger proteins regulate the transcription of genes for the gastric proton pump (H^+/K^+-ATPase genes); these proteins have been identified in gastric parietal cells.

Leucine-zipper proteins consist of an extended α helix in which the amino acid leucine occupies every seventh position (Fig. 1). These periodic leucine residues along one face of the α helix of the protein interact with the leucine side chains of the corresponding helix of its partner. The result is the formation of an intermolecular coil that stabilizes (or "zips up") two proteins containing the leucine-zipper motif. These dimers form the correct configuration for DNA binding and consist of either two identical proteins (homodimers) or two different proteins, both containing leucine zippers (heterodimers).

The leucine-zipper motif underlies the interactions between the proto-oncogene products Jun and Fos, which constitute the heterodimeric transcription factor AP-1 and are implicated in signal-dependent processes that determine cell lineage and growth, and hence carcinogenesis. This motif also governs the activity of CREB (cyclic AMP response element–binding protein), a transcriptional transducer of changes in intracellular cyclic AMP involved in a host of biologic functions, including gluconeogenesis, neuronal excitation, establishment of circadian rhythms, pituitary proliferation, and opiate tolerance.

Members of the HLH group include proteins that bind to immunoglobulin gene enhancers, the muscle-specific transcription factors MyoD and myogenin (controlling myogenesis), and the *myc* genes (which are the cellular counterparts of viral oncogenes and are involved in growth regulation). The HLH transcription factors have some similarities to the leucine-zipper family. Both leucine-zipper proteins and HLH transcription factors have a region of positively charged amino acids that contacts the DNA and an adjoining region that mediates dimer formation (Fig. 1). Both families of transcription factors have key roles in differentiation and development, and in both the activity is regulated by heterodimer formation.

Transcription factors are essential for normal growth and development, whether in the embryo or the child, for normal cell differentiation, and for a large number of cellular responses to external stimuli. We can thus expect to find increasing numbers of examples of abnormal transcription factors in many areas of medicine, but especially in oncology, endocrinology, pediatrics, and immunology. Inevitably, medical research on transcription factors will soon shift from the essential first steps of identification and cataloguing to the development of treatments that act on the transcription factors involved in the misexpression of specific proteins.

Recommended Reading

Latchman DS. Eukaryotic transcription factors. London: Academic Press, 1991.

Rosenthal N. Regulation of gene expression. N Engl J Med 1994;331: 931-3.

Travers A. DNA–protein interactions. London: Chapman & Hall, 1993.