Outside and Downstream of the Homeobox*

Gerald M. Edelman and Frederick S. Jones

From the Department of Neurobiology, The Scripps Research Institute, La Jolla, California 92037

There is an important puzzle at the center of modern biology that is just as demanding of biochemistry as it is of genetics. How do cells form tissues at the correct place in the body plan during embryogenesis to yield shapes characteristic of a species? Although genes constrain tissue formation, DNA does not encode specific spatiotemporal coordinates for individual cells to follow. How then can a cell at a given place have reached that location at the right time? Obviously, molecular cues and refined regulatory loops must be involved. As biochemists, we would like to understand development in terms of the regulatory biochemical signals that underlie morphogenesis.

Research in developmental biology during the last 15 years has begun to clarify the nature of this regulation and has led to the emergence of a field that deserves to be called molecular embryology. Analyses of homeobox genes and of cell adhesion molecules have made particularly significant contributions to this field. Scores of transcription factors encoded by homeobox genes have been discovered in a wide variety of animal genomes. The regional expression of these gene products is correlated with the formation of specific tissues and physiology. Somehow their expression constrains cellular activity in a place-dependent fashion, resulting in the patterning of cells within three-dimensional structures. We do not yet know how homeobox genes and their encoded homeodomain proteins carry this out. Indeed, we are just beginning to identify appropriate targets of homeobox gene products. But there is a clue that may help answer the puzzle. To convert genetic regulation into animal form, such targets must be able to affect the machinery of interacting cells and tissues during development, raising the possibility that cell adhesion molecules may be involved.

Recently putative downstream target genes for homeobox gene products have been identified. The candidates for which evidence exists include cell adhesion and substrate adhesion molecules (CAMs3 and SAMs) as well as various growth factors. Interaction of two molecular systems, one for homeobox genes controlling place-dependent transcription and another modulating cell surface signaling events and cellular mechanics through adhesion, may be essential for place-dependent pattern formation during morphogenesis.

Homeoboxes and Homeodomains

So-called homeotic genes were first discovered in Drosophila. These genes control segmentation of the animal and assign anatomical structures, such as antennae and legs, to the appropriate segment. Mutations of these genes cause homeotic transformations in which the anatomy characteristic of one segment in a given location in the anteroposterior axis is assigned incorrectly to another segment. In two of the most striking mutations, flies contain an extra thoracic segment with an additional pair of wings or have legs substituted for antennae. These mutations have been mapped respectively to two dispersed gene clusters, Bithorax (BX-C) (1) and Antennapedia (ANT-C) (2), which together are known as the homeotic gene complex. Individual genes within the clusters were found to share a 183-nucleotide DNA segment termed the homeobox (3). The homeobox encodes a 61-amino acid protein segment called the homeodomain, a helix-turn-helix motif that superficially resembles the DNA binding domains of prokaryotic repressor proteins. Homeodomains were shown to bind specific DNA sequences in vitro (4), many of which contain the motif TAAT. Nuclear magnetic resonance (NMR) spectroscopy (5, 6) and x-ray crystallographic (7, 8) analyses of homeodomains and homeodomain-DNA complexes showed that the elongated third (or recognition) helix, located near the carboxyl terminus of the homeodomain, interacts with the TAAT core sequence as well as surrounding base pairs in the major groove of DNA. Two other segments of the protein located in more amino-terminal positions fold over the DNA backbone and interact with the minor groove.

Homeodomain proteins have been shown to be transcription factors. Their effects on transcription can be either positive or negative depending on a number of variables, including the type of signal state of the cell and the composition and activity state of other transcriptional accessory proteins in the nucleus (9, 10). A search for new homeoboxes and homeodomain proteins has identified sequences that diverge considerably from those of Drosophila homeotic genes. Homeodomain proteins have been classified according to amino acid sequence similarities, DNA binding specificity, and patterns of expression during animal development (11). For example, the Antennapedia, Bicoid, and Paired proteins differ in the amino acid residue at position 9 in the recognition helix of the homeodomain. Amino acid “swapping” experiments at this position convert the binding specificity of one protein into that of another (12). Experiments involving the construction of chimeric proteins in which one homeodomain is substituted for another have demonstrated that even closely related homeodomains have distinct sites of action and different target gene specificities (13).

Colinearity and the Hox Gene Family of Vertebrates

In his genetic experiments with Drosophila, Lewis (1) noted a correlation that he termed colinearity between the position of a homeobox gene within the chromosome and the position in the anteroposterior axis at which morphological defects would occur. More recently, cross-genome hybridization assays with sequences of the Drosophila ANT-C and BX-C genes led to the discovery of a vertebrate set of genes related to the Drosophila homeotic genes, the Hox genes (14). A total of 38 different Hox genes have been isolated in vertebrate species. There are four separate clusters of genes, Hox A, B, C, and D (previously designated Hox 1, 2, 3, and 4) (15), located on four different chromosomes. As with Drosophila, the vertebrate Hox genes are expressed in a colinear fashion; during development, each gene is sequentially activated in a defined spatiotemporal order consistent with its position within the cluster. The gene expressed earliest is located at the 3’-end of the cluster and expressed latest is located at the 5’-end of the cluster.
pression proceeds caudal as more 5′-genes in the cluster are activated. Genes expressed at later times yield structures whose anterior borders start in progressively more posterior positions in the embryo (16). Overexpression or deletion of particular genes in a cluster results in transformations; a striking example is that in which posterior vertebrae are transformed into ones with more anterior characteristics (17).

**Linkage of Homeobox Genes to Pattern**

The existence, function, and expression patterns of homeobox genes and homeodomains do not in themselves explain the complex cell patterning mechanisms that drive the formation of tissues during morphogenesis. The problem that needs to be solved is easy to state. What kinds of molecules (structural gene products) are regulated by homeoboxes to alter the dynamics of cell grouping into tissues? During tissue patterning in vertebrates, cells of different histories are brought into contact with each other through complex cellular movements.Signals are exchanged across borders of different cell collectives, leading to new programs of gene expression and to patterns of movement, proliferation, or milieu-dependent differentiation (embryonic induction) (18, 19).

Unlike Drosophila, vertebrates show no outward manifestations of segmentation. It has been suggested that Hox genes collectively constitute a "code" that specifies the appropriate anatomy at a particular place in the body plan of the animal. While the idea of a code appears to be attractive at first glance, what is definitely required is to determine how particular target molecules regulated by homeobox genes can affect the mechanics of cell-cell interaction. Identification of such targets downstream of the homeobox genes may help to explain how a particular morphology is associated with a particular place.

Analysis of the regulatory pathways governing homeodomains themselves is also of critical importance, particularly in studying cell behavior as a function of place. Recent work has supported the notion that both the Drosophila homeotic selector (20, 21) and vertebrate Hox genes (22) autoregulate and cross-regulate each other. An additional factor that may influence the sequential nature of the expression of such genes is how susceptible they are to activation or repression by growth factors such as retinoic acid (23). Signaling between homeobox genes and growth factors may occur in a loop; for example the Drosophila genes ubx and abdominal-A control the expression of a gene called decapentaplegic (dpp), a homologue of genes encoding proteins of the transforming growth factor-β family. Expression of dpp is, in turn, required for transient synthesis of the homeotic gene labial in a subset of endodermal cells (24, 25).

As exciting as these findings are, however, they do not directly address how tissues are mechanically assembled in a heritable fashion. What is required is to show a linkage between control by homeobox genes and the expression of molecules that actually control cell-cell assembly and cell movement.

**CAMs and SAMs Are Downstream Targets of Vertebrate Homeobox Genes**

A great number of cell surface adhesion molecules and extracellular matrix glycoproteins have been found to affect morphogenetic movements and the formation of cell collectives (26). The partitioning of such collectives by cell sorting into the three germ layers (27) is a fundamental theme in vertebrate development. This process of gastrulation has been shown to be affected by a number of adhesion proteins (28). During development, CAMs and SAMs have been found to modulate cellular primary processes such as proliferation, movement, interaction, differentiation, and death (19). It is striking that adhesion molecules are expressed in a highly place-dependent fashion that often reflects temporal growth and differentiation gradients appearing during the formation of tissues (28). What determines the place-dependent expression of the genes encoding these morphoregulatory proteins?

Recently, we have demonstrated a link between the expression of homeobox genes and the expression of specific morphoregulatory molecules. In studies aimed at identifying the sequences of DNA (particularly promoters and enhancers) that control the expression of different adhesion molecule genes, we isolated the promoter region for the cytotactin/tenascin gene (29). Cytotactin is an extracellular matrix glycoprotein (30) that regulates cell adhesion and migration during morphogenesis of the brain and other embryonic structures (31, 32). The promoter region of the cytotactin gene (29) exhibits many potential transcription factor binding motifs including those resembling homeodomain binding sites (HBs).

The gene promoter for the mouse neural cell adhesion molecule (N-CAM) has also been found to contain potential homeodomain binding sites (33). In experiments prompted by these observations, we found that homeobox gene products affected the activity of both the cytotactin/tenascin and N-CAM gene promoters in vitro. To show this, NIH 3T3 cells were transfected with plasmids driving the expression of various homeobox-containing genes and, at the same time, with plasmids containing a chloramphenicol acetyltransferase gene driven by either the cytotactin/tenascin or N-CAM promoter sequences.

In the earliest work, homeodomain binding sites and a 67-base pair region critical for activation of the cytotactin/tenascin promoter by the homeobox gene even-1 were identified (34). even-1 is a vertebrate homolog of the Drosophila pair-rule segmentation gene (known as even-skipped) (35). The promoter region responding to even-1 contained a TRE/AP-1 motif, a DNA element known to bind products of the fos and jun gene family of transcription factors and to mediate responsiveness to growth factor signals (36, 37). Mutation of this element abolished activation of the promoter by even-1.

An intriguing correlation was found between even-1 activation and the growth factor stimulation of promoter activity. Transfection of cells with even-1 mimicked the effects of growth factors and could substitute for the high serum conditions normally required to get comparable activation levels of the cytotactin/tenascin promoter (34). These data suggest that the activity of the cytotactin promoter may be regulated combinatorially at the TRE/AP-1 site by convergent inputs both from growth factor signals and homeodomain proteins. In accord with these results, recent evidence from other laboratories has suggested that expression of homeodomain proteins can mimic growth factor effects by regulating the transcriptional activity of serum-inducible genes. A human homeobox gene product, Phox1, and other related Paired homeodomain proteins were found to enhance binding of the serum response factor (SRF) to its DNA target sequence (38), thereby activating SRF target genes. Another homeodomain protein, encoded by the murine gtx gene and quite unrelated to Paired class proteins, bound to a DNA sequence similar to that bound by SRF and repressed serum-induced activation of a reporter gene controlled by this DNA sequence (39).

Cytotactin is an extracellular matrix protein. What about cell-cell adhesion molecules present on the cell surface? In simultaneous studies, we have demonstrated that N-CAM gene expression is controlled by different Hox gene products. Expression of the HoxB9 (Hox 2.5) gene activated N-CAM promoter activity. Co-transfection with a neighboring gene, HoxB8 (Hox 2.4), mitigated this effect, i.e. co-expression of these two Hox genes controlled N-CAM promoter activity in a switch-like
duration depended on the concentration ratio of the Hox plasmids used, and showed a thresholded response. A 47-base pair DNA segment in the N-CAM promoter containing two potential HBSs with TAAT motifs was found to be sufficient for promoter modulation by these homeobox gene products. Mutation of the HBS motifs completely abolished these effects (40). The HBS regions of the N-CAM promoter not only affected responsiveness to homeobox genes in transfected cells but also bound to homeodomain proteins in a DNA sequence-specific manner. For example, a homeoprotein encoded by the HoxC6 (Hox 3.3) gene was bound to a specific TAAT-containing sequence CCTATTATTAA in the N-CAM promoter, designated HBS-I (41).

These data support the notion that the N-CAM gene may be a downstream target of Hox genes. Recently, we have found that genes encoding more divergent homeodomains, including those from the paired class, also affect N-CAM promoter activity. Paired homeobox genes are expressed widely by cells undergoing neural development (42). They may provide important means to control the activity of promoters for the N-CAM gene and other immunoglobulin-like cell adhesion molecule genes during formation of the nervous system.

With the exception of the homeobox genes themselves (22), the cytotactin/tenascin and the N-CAM genes are the first in vivo transcriptional targets of proteins encoded by vertebrate homeobox-containing genes. These findings buttress the hypothesis that place-dependent expression of cell and substrate adhesion molecules may turn out to depend on the local expression patterns of different homeobox genes (43). Definitive support for this hypothesis will require demonstration that the putative target sequences are actually involved in place-dependent expression in vivo.

We would expect that regulatory sequences of a number of other cell adhesion molecule genes would also be targets. Recently, for example, the promoter for the gene encoding the mouse immunoglobulin-like adhesion molecule L1 has been isolated, and its sequence showed potential binding sites for homeoproteins (44). Connectin, a cell adhesion molecule, was found to be an in vivo target for the ultrabithorax homeotic gene in Drosophila (45). Moreover, evidence is mounting to suggest a regulatory link between homeobox genes and neuronal cell morphology and migration, phenomena that are affected by a number of adhesion molecules. Mutations in the Caenorhabditis elegans mab-5 gene, the product of which is an antennapedia class homeoprotein, lead to abnormal neural pathfinding during development (46). It has also been suggested that cellular uptake of antennapedia homeodomain peptide and binding to target DNA sequences in primary cultures of neurons can alter their neurite extension properties (47).

**Prospects and Synthesis**

Place-dependent signaling by specific homeodomain proteins may program precise expression levels of the genes that encode morphoregulatory molecules such as CAMs and SAMs. Central to this hypothesis is the fulfillment of the prediction that, during development, the activities of CAM and SAM gene promoters will be found to be regulated in vivo by homeodomains in precise patterns. One would expect that particular combinations of CAMs and SAMs would be synthesized at discrete locations of the body plan. It is also reasonable to expect that promoters for CAM and SAM genes would have different threshold sensitivities to a given concentration of homeoprotein. Homeobox gene products may affect responses to appropriate growth factors as well. If this turns out to be the case, then the activities of homeodomains, growth factors, and cell adhesion molecules would be more intimately intertwined than had been previously imagined. As cell collectives of different

**Fig. 1. Hypothesized components of a morphoregulatory control loop governing place dependence**. This scheme links place-dependent Hox gene expression to the mechanochemical effect of morphoregulatory molecules on cells. A particular place in the developing vertebrate organism shows differential gene expression, and in the case of Hox genes, colinearities. Cells at a particular place in the embryo adopt shapes and movement and proliferation patterns on the basis of place-dependent integration of three kinds of signaling components, growth factors (GF), homeodomain proteins (HD), and CAMs and SAMs. Cells having specific patterning properties characteristic of a given place and time denoted by (Place and Time), would express specific growth factor, homeodomain protein, and CAM and SAM repertoires characteristic of that particular place. Different components of these repertoires are indicated by subscripts. Interplay between cell groups and tissues during morphogenesis would lead to qualitative and quantitative changes in each of the components of the signal loop. The result would be a different pattern with different signaling components at (Place and Time)₂. Dynamic reciprocity is indicated by arrows drawn between different types of gene products. The scheme suggests that changes in one type of component would lead to correlative changes in others.

**Acknowledgment**—We thank Dr. Kathryn Crossin for helpful comments and suggestions during the preparation of the manuscript.

**REFERENCES**

Minireview: Outside and Downstream of the Homeobox