Homeobox Genes and the Vertebrate Body Plan

This family of related genes determines the shape of the body. It subdivides the embryo along the head-to-tail axis into fields of cells that eventually become limbs and other structures

by Eddy M. De Robertis, Guillermo Oliver and Christopher V. E. Wright

Starting as a fertilized egg with a homogeneous appearance, an embryo made of skin, muscles, nerves and other tissues gradually arises through the division of cells. Long before most cells in the emerging body begin to specialize, however, a plan that designates major regions of the body—the head, the trunk, the tail and so on—is established. This plan helps seemingly identical combinations of tissues arrange themselves into distinctly different anatomical structures, such as arms and legs.

Recently embryologists have made great progress in uncovering the mechanisms that control this once mysterious process. In the past decade the powerful techniques of molecular biology have made it possible to isolate and characterize individual genes that mediate some of the developmental decisions involved in establishing the embryonic body plan. The key is a family of genes, known as homeobox genes, that subdivides the early embryo into fields of cells with the potential to become specific tissues and organs.

EDDY M. DE ROBERTIS, GUILLERMO OLIVER and CHRISTOPHER V. E. WRIGHT have collaborated for the past five years at the University of California, Los Angeles. De Robertis received his medical degree from the University of Uruguay in 1971 and his doctorate in biochemistry from the University of Buenos Aires in 1975. Between 1980 and 1985 he was professor of cell biology at the Biocenter of the University of Basel; since 1985 he has been the Norman Sprague Professor of Biological Chemistry at the U.C.L.A. School of Medicine. Oliver received his degree in biology from the University of Uruguay in 1980 and his degree in molecular biology from the University of Mexico in 1984. He is now associate professor of biochemistry at his alma mater in Uruguay. Wright received his doctorate in biochemistry from the University of Oxford in 1984. He is now assistant professor of cell biology at Vanderbilt University Medical School.

The development of *Xenopus laevis*, a South African clawed frog, stands as a good example of how a vertebrate takes shape. This amphibian is a favorite of modern embryologists, because at any time of year the female can be induced to lay about 1,500 large, easily fertilized eggs. Since all vertebrates develop similarly, most mechanisms of early frog embryogenesis also apply to fish, chickens, human beings and other animals.

One striking feature of early development is its rapid pace. A fertilized *Xenopus* egg cell divides into two after about 90 minutes. The cells then divide synchronously every 30 minutes until there are 4,000 of them. At that stage the embryo is called a midblastula and has the shape of a hollow sphere. To the unassisted eye, the cells all look identical, but some around the midblastula's equator are already destined to become a layer of cells called the mesoderm. The formation of the mesoderm is induced by protein growth factors released by the large yolky cells at the bottom pole of the embryo.

The entire mesodermal layer eventually moves into the interior of the embryo during a process called gastrulation. By the end of this process, three layers with distinct developmental potentials have been defined: the mesoderm, the endoderm and the ectoderm. The mesoderm gives rise to most of the body, including the vertebral column, the muscles, the bones and the body wall. The endoderm produces the epithelial layer of tissue lining the digestive tract as well as various other organs, such as the lungs, the liver and the pancreas. The ectoderm becomes the skin and the nervous system.

The ectoderm develops into the nervous system in response to chemical signals that diffuse out of the underlying mesoderm. The signals induce part of the ectoderm to thicken into a structure called the neural plate. (At this stage the embryo is termed a neurula.) The edges of the neural plate then fold toward one another while the middle sinks into the embryonic body. The edges finally fuse to form a neural tube, which becomes the basis for the brain and spinal cord.

The determination of the embryo's anteroposterior (head-to-tail) axis is a milestone in development, because it provides the major line along which later structures will develop. Ross G. Harrison of Yale University was the first to show that embryonic cells commit to become limbs and other specific anatomical structures very soon after gastrulation is complete.

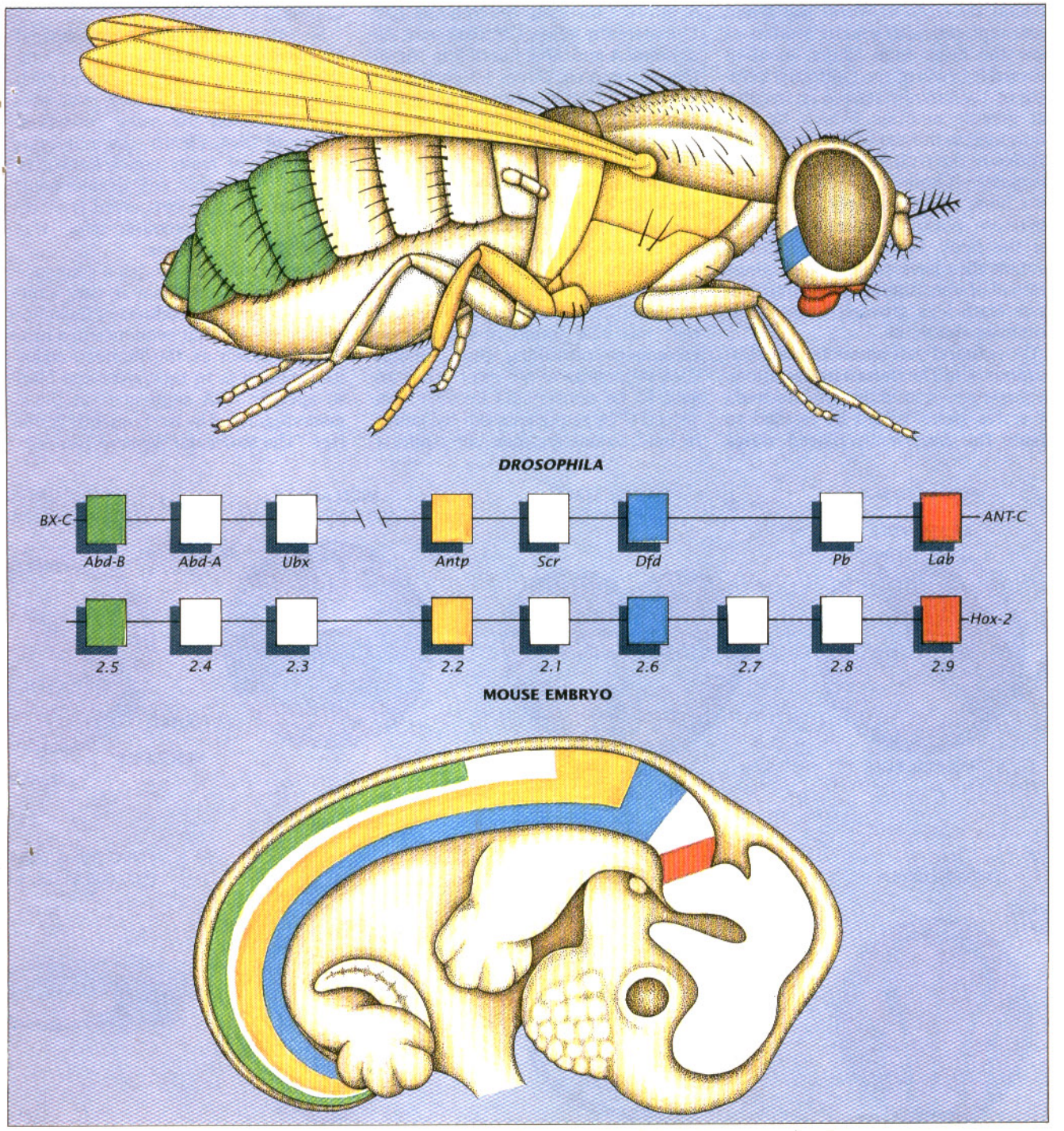
In 1918 Harrison took small fragments of mesoderm from the flanks of some amphibian neurulas and transplanted them into the sides of others. If the transplanted tissue came from a certain region of the donor, it always gave rise to an extra forelimb on the recipient. Harrison realized that although the mesoderm looked like a uniform layer in those embryos, the cells somehow already knew to which part of the body they belonged.

One peculiarity Harrison noted was that an embryo could still grow a forelimb even if he removed all the mesoderm that would normally give rise to it. He interpreted this finding to mean that the surrounding region of mesoderm also had a potential for inducing and directing limb growth. This broad

region with the potential for expressing a structure became known as a morphogenetic field.

Following Harrison's lead, many scientists conducted transplantation experiments on amphibian embryos. Those studies established that the mesoderm is the crucial cell layer that specifies which end of the embryo is the head and which is the tail. The mesoderm of the amphibian neurula was mapped, or subdivided, into morphogenetic fields for many organs: gills, balancers (structures that help tadpoles stay upright), ears, forelimbs, hind limbs, the tail and so on.

Within each morphogenetic field, the potential for forming an organ varied gradually. It was therefore proposed that each morphogenetic field contained a "gradient-field" of infor-



HOMEOBOX GENES control development in animals as different as *Drosophila melanogaster* (a fruit fly) and a mouse. These genes divide the embryo along its head-to-tail axis into bands with different developmental potentials. The location of a homeobox gene on a chromosome corresponds to where it is expressed in the body: proceeding from left to right, the genes control body areas closer to the anterior end of the animal. All homeobox genes seem to have a common evolutionary origin. In this diagram, related homeobox genes in *Drosophila* and the mouse and the body parts they control in each animal are colored similarly. The mechanism that determines the head, trunk and tail may have arisen only once during evolution.

mation for specifying an organ. As we shall explain, the behavior of these gradient-fields corresponds closely with the patterns of expression for certain sets of genes.

Transplantation studies of the mesoderm's control over the amphibian body plan ended around the close of World War II and were replaced with genetic studies of how the body took shape. In 1948 Edward B. Lewis of the California Institute of Technology started an insightful genetic analysis (which he has continued to this day) of homeotic mutations in the fruit fly *Drosophila melanogaster*. A homeotic mutation causes a body part to be replaced with a structure normally found elsewhere on the body. For example, *bithorax* mutant flies have two pairs of wings instead of one; *Antennapedia* mutants have extra legs growing where their antennae should be.

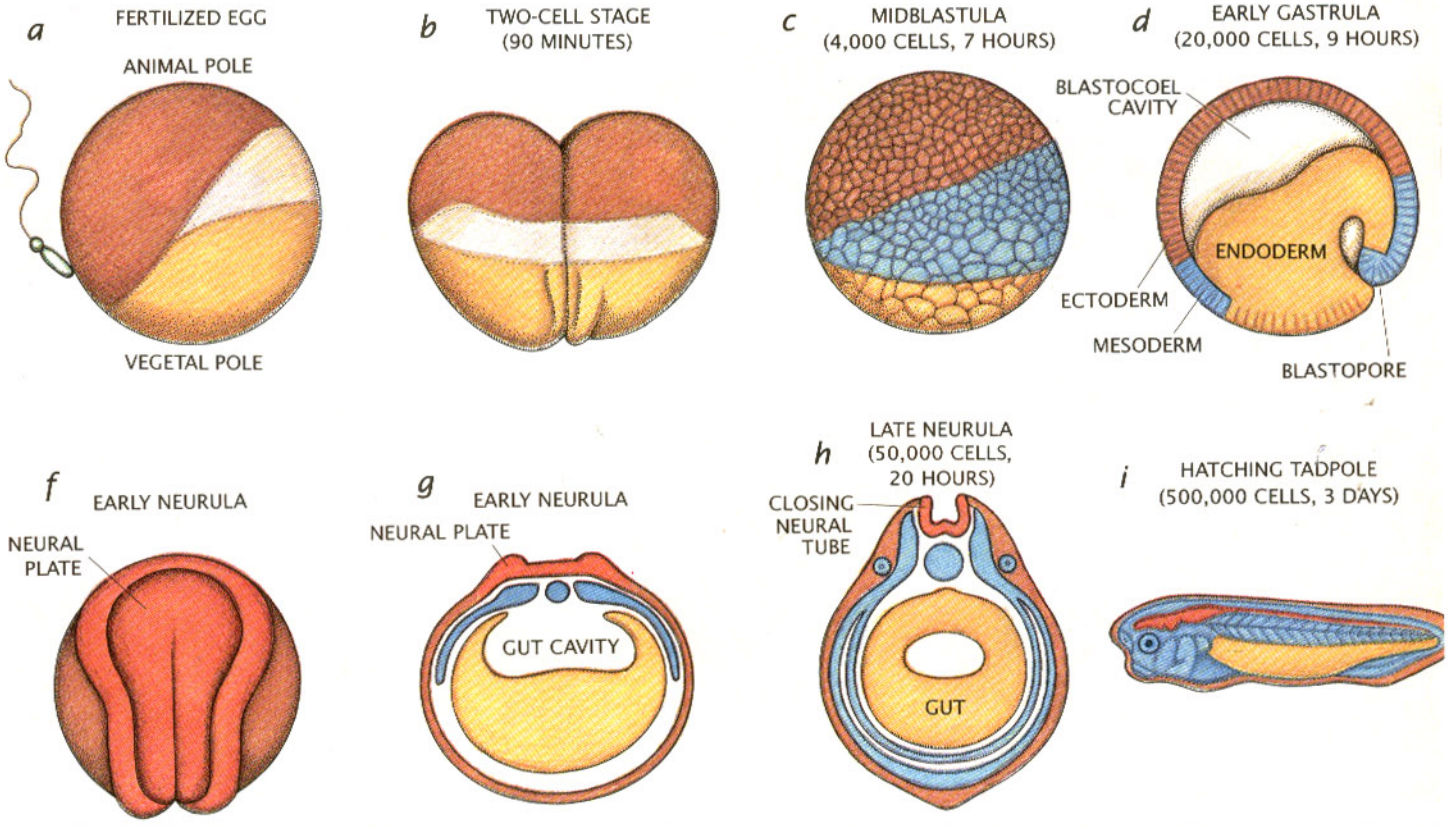
Lewis found that homeotic transformations could be caused by mutations in single genes, even though hundreds of active genes would be needed to create the abnormally placed wings and legs. It was reasonable to assume, then, that the mutations were affecting master genes that controlled the activity of many subordinate genes.

Once genetic engineering enabled scientists to isolate genes, the race to find and study the homeotic genes was on. During the early 1980's David S. Hogness and Welcome Bender of Stanford University became the first to isolate the genes *Ultrabithorax*, *Abdominal-A* and *Abdominal-B* in the *bithorax* complex. Walter J. Gehring and Richard L. Garber of the Biocenter at the University of Basel and Matthew P. Scott and Thomas C. Kaufman of Indiana University isolated the genes of the *Antennapedia* complex, including ones called *Labial*, *Proboscapedia*, *Deformed* and *Antennapedia*.

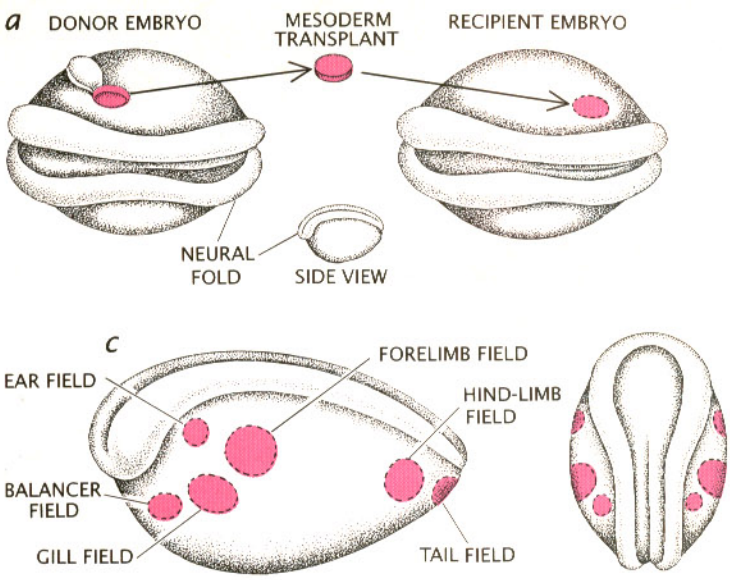
A crucial discovery came in 1983, when Gehring and his colleague William J. McGinnis found the *Antennapedia* gene contained a DNA sequence that was also found in another development-controlling gene. (Similar DNA sequences in different genes are said to be conserved.) Because conserved DNA sequences can hybridize, or bind, to one another, one could label the conserved DNA sequence from *Antennapedia* radioactively and use it as a probe to locate other genes containing the same region. In this way, Gehring and McGinnis isolated *Ultrabithorax*, *Deformed* and other homeotic genes. The conserved DNA region was identified independently by Scott, who was then at the University of Colorado at Boulder.

Significantly, McGinnis also showed that other invertebrates—such as centipedes and earthworms, from which insects are thought to have evolved—also contained the same conserved region of DNA. Clearly, the molecular structures of many genes known to control embryonic cell development were related. The conserved DNA region in each homeotic gene was dubbed the homeobox.

The homeobox encodes a sequence of 60 amino acids that is very similar in the protein products of most homeotic genes. That sequence in a protein is known as the homeodomain. Its function is to recognize and bind to specific DNA sequences in those genes regulated by the homeotic genes [see "The Molecular Basis of Develop-



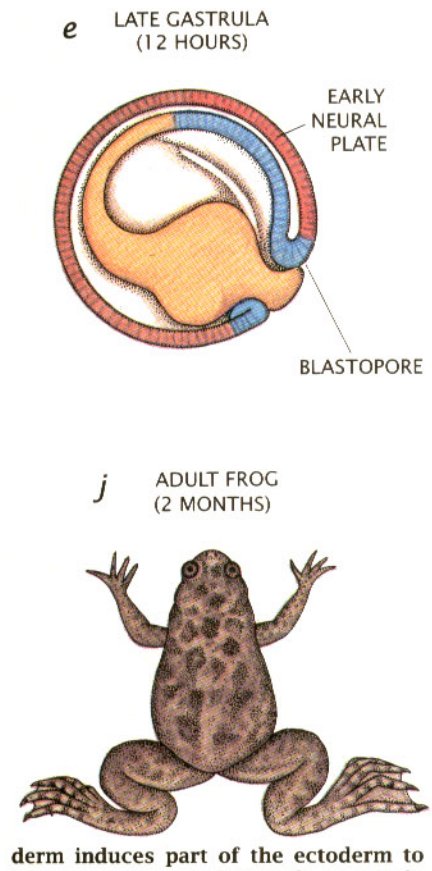
VERTEBRATE BODY PLAN is generated through the chemically induced formation and movement of cell layers, as seen in the development of *Xenopus laevis*, a South African frog. Through rapid cell divisions (*a–c*), a fertilized egg becomes a hollow ball of cells. The large yolky cells at the bottom pole of the embryo release protein growth factors that induce the overlying cells to become the mesoderm layer (*blue*). The mesoderm is the critical layer that determines the embryo's anteroposterior polarity. Two other layers of cells—the ectoderm (*brown*) and the endoderm (*yellow*)—are established during gastrulation, the process by which the mesoderm migrates into the interior of the embryo (*d–e*). During the neurula stage (*f–g*), the meso-



TRANSPLANTS in which mesoderm from one neurula-stage embryo is grafted onto another will induce the formation of additional limbs or organs. If mesoderm from the forelimb area is transplanted, for example (*a*), the recipient embryo will have an extra forelimb (*b*). Through such experiments on amphibians, embryologists have identified morphogenetic fields that specify the development of various structures (*c*).

ment," by Walter J. Gehring; SCIENTIFIC AMERICAN, October, 1985].

The polypeptide chain in the homeodomain consists of four helixes, one of which is responsible for recognizing a specific DNA sequence. Because this helix is nearly the same in all homeodomain proteins, the proteins all bind to fairly similar DNA sequences. When they bind to genes in a cell, homeodomain proteins activate or repress the expression of those subordinate genes.



derm induces part of the ectoderm to become the neural plate (*red*). As seen in cross section (*h–i*), the neural plate closes on itself and becomes a neural tube, which is the forerunner of the brain and the spinal cord in the mature animal (*j*).

We began research on the homeobox in 1983, when one of us (De Robertis) had his laboratory at the biocenter on the same floor as Gehring. We had been interested for some time in the development of *Xenopus laevis* [see "Gene Transplantation and the Analysis of Development," by Eddy M. De Robertis and J. B. Gurdon; SCIENTIFIC AMERICAN, December, 1979]. As we followed the great advances being made in studies of *Drosophila*, it became evident that we would have to identify master genes in vertebrates if we were ever to gain a comparable understanding of their development.

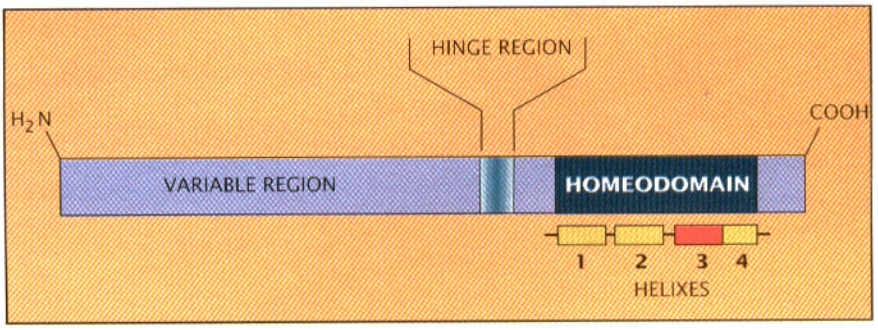
Yet the lack of knowledge about frog genetics seemed to be an insurmountable barrier to further progress. Even though the genetics of mice had been reasonably well studied, there were no real candidates for master genes controlling embryogenesis.

We decided to try what seemed, at the time, a crazy experiment: to isolate a gene similar to *Antennapedia* from frog DNA with McGinnis and Gehring's fruit fly homeobox probes. There was little reason to believe that the frog DNA contained such a gene or that the genes of such unrelated species would be significantly similar. Still, we felt it was worth the attempt. Some of our colleagues were skeptical that such an experiment could ever work, and two of our students declined to help on those grounds.

We were soon celebrating with a bottle of champagne. On our very first attempt, while working with Andrés E. Carrasco, a postdoctoral student in our laboratory, we succeeded. We analyzed the DNA sequence in the frog gene that our experiment had isolated, which is now called *XlHbox 1*, and confirmed that it contained the homeobox region. That finding strongly suggested that a gene directly controlling vertebrate development might at last be at hand.

Little did we imagine after our first experiment that it would take six more years and the efforts of laboratories throughout the world before it was certain that vertebrate homeobox genes were directly involved in the control of development. Even so, initial progress was swift in studies of mammals. Working with mice, Frank H. Ruddle of Yale University (who was then on sabbatical at the biocenter) and Peter Gruss of the Max Planck Institute for Experimental Medicine in Göttingen, West Germany, isolated many genes containing homeoboxes. Dado Boncinelli of the University of Naples had similar success with human genes. The proteins encoded by all these homeobox genes differ greatly from one another except at the highly conserved homeodomain.

The roster of proteins known to contain homeodomains grew in 1988, when researchers purified transcription factors for the first time. These factors are proteins that increase the expression of particular target genes. When the transcription factors were sequenced, some were found to contain homeodomains, which indicated that they were products of genes with homeobox regions. These biochemical



HOMEODOMAIN PROTEINS bind to DNA and regulate gene expression. They are composed of a variable region, which determines a protein's specific activity, a small connective hinge region and a homeodomain, a 60 amino acid sequence that is similar in all proteins of this type. This sequence is encoded by the homeobox regions of genes. The homeodomain consists of four alpha helixes (*1-4*), one of which (*red*) recognizes and binds to a specific DNA sequence in the target genes.

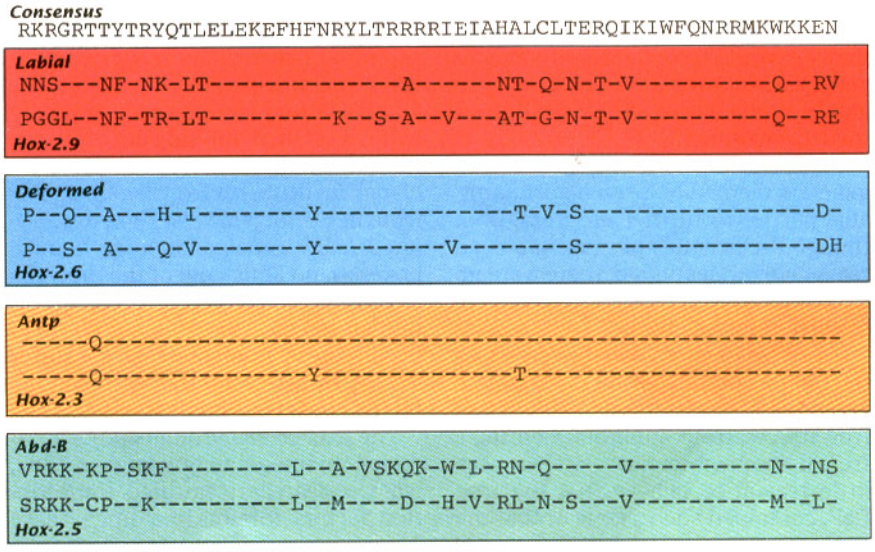
studies independently confirmed that homeobox genes regulate the activity of other genes.

But how do homeobox genes orchestrate cellular differentiation during development? An inkling of the answer comes from observing the regions in which the proteins made by homeobox genes are located in the embryonic body at various developmental stages. The *Xenopus* XlHbox 1 protein, for example, is found in a narrow band of cells just behind the frog embryo's head. This band consists of both the mesoderm and the anterior spinal cord and neural crest.

The anterior and posterior boundaries of *XlHbox 1* expression in these tissue layers are neatly aligned. Because mesoderm is known to induce the anteroposterior characteristics of neural tissue, it seems possible that the mesoderm expressing *XlHbox 1* also induces cells in the overlying neural plate to express the gene as well.

Other homeobox genes are active in different regions. On the basis of homeobox gene expression patterns, therefore, one can view the vertebrate embryo as subdivided into anteroposterior fields of cells with different developmental capacities. This subdivision of the embryonic body precedes the formation of specific organs or structures.

Even though the homeodomains encoded by different homeobox genes are very similar to one another, characteristic differences in their amino acid sequences can be used to identify them. Some homeodomains resemble one another much more closely than others. Interestingly enough, on the basis of these similarities and differences, some mammalian homeodomains strongly resemble those produced by particular fruit fly genes.



ALL HOMEODOMAINS are fundamentally similar, but homeodomains made by some insect and mammalian genes are particularly alike. Amino acid sequences in several homeodomains are shown here. The genes *Labial, Deformed, Antp* and *Abd-B* are from *Drosophila* fruit flies, and the four analogous *Hox* genes are from mice. In each sequence the letters stand for amino acids. A hyphen indicates that the amino acid is the same as in the consensus string, an average of all homeodomain sequences.

When patterns of expression for many homeobox genes in mouse embryos were analyzed, a remarkable observation was made independently by Robb Krumlauf of the Medical Research Council in London and Denis Duboule of the European Molecular Biology Laboratory in Heidelberg. Investigators had previously shown that in both vertebrates and invertebrates, homeobox genes cluster in complexes, or groups, on a chromosome. In other words, the homeobox genes are arranged in a precise order, left to right, on the linear DNA molecule that makes up a chromosome.

Krumlauf and Duboule made the unexpected discovery that in mice the order of the homeobox genes in a cluster corresponds directly to where the genes are expressed. Homeobox genes located near the left end of a complex are expressed in posterior parts of the body and genes to the right are expressed closer to the head. Lewis had noticed the same pattern in *Drosophila* many years earlier.

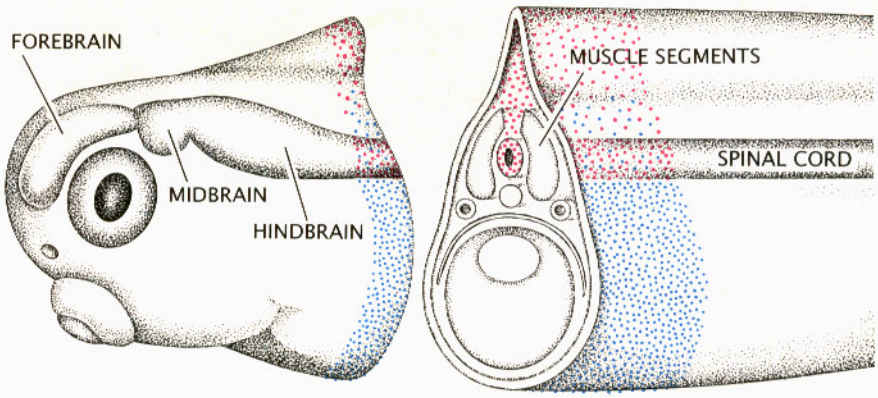
All vertebrates have four homeobox complexes, each located on separate chromosomes. These complexes probably arose during evolution through duplications of the single cluster of homeobox genes in invertebrates. Consequently, every human being has four genes that resemble the fruit fly gene *Abdominal-B,* for example, and four others that resemble *Deformed.*

One unifying principle applies to all homeobox complexes: genes expressed posteriorly are located at the left, and those expressed anteriorly are at the right. Homeobox genes are therefore arranged in the chromosomal DNA in the same order in which they are expressed along the anteroposterior body axis. This extraordinary arrangement may have come about because homeobox genes must be activated in a particular order.

Evidence for how this sequential deployment of homeobox genes occurs is accumulating. In vertebrate embryos, retinoic acid (a compound related to vitamin A that can sometimes cause severe birth defects) and peptide growth factors are good candidates for providing such positional clues. They could convey such information

by activating homeobox genes selectively in the mesoderm, the key element in determining the body plan.

By adding retinoic acid to cultured embryonic cells, Boncinelli's research group has shown that the compound can activate many homeobox genes. In frog embryos, Douglas A. Melton of Harvard University has proved that fibroblast growth factor (which induces the formation of the mesoderm in early embryos) can activate posterior homeobox genes selectively. In our laboratory at the University of California at Los Angeles, Ken W. Y. Cho has shown that a protein resembling transforming growth factor β activates only anterior genes.



HOMEOBOX GENES are expressed in discrete bands along the anteroposterior axis of an embryo. In a *Xenopus laevis* tadpole, for example, the *XlHbox 1* gene is expressed in a region in the anterior trunk of the body. The protein produced by the gene is found in the cell nuclei of both mesodermal (*blue*) and ectodermal (*red*) tissues. The forelimb grows entirely from mesoderm cells expressing the XlHbox 1 protein.

Once activated, do homeobox genes directly specify the identities and fates of embryonic cells, thereby shaping the body and guiding the formation of organs? Or are their effects indirect? The results of two experiments argue for their having a direct role.

In the first experiment, we injected antibodies directed against the XlHbox 1 protein into single-cell *Xenopus* embryos. The antibodies bound to the protein and inactivated it during the crucial period in which the body plan is established. When we examined the tadpoles that developed, we discovered the tissues that normally expressed *XlHbox 1* and that should have become a section of the anterior spinal cord had instead become hindbrain structures. In effect, the "loss-of-function" of *XlHbox 1* changed part of the spinal cord into a more anterior structure.

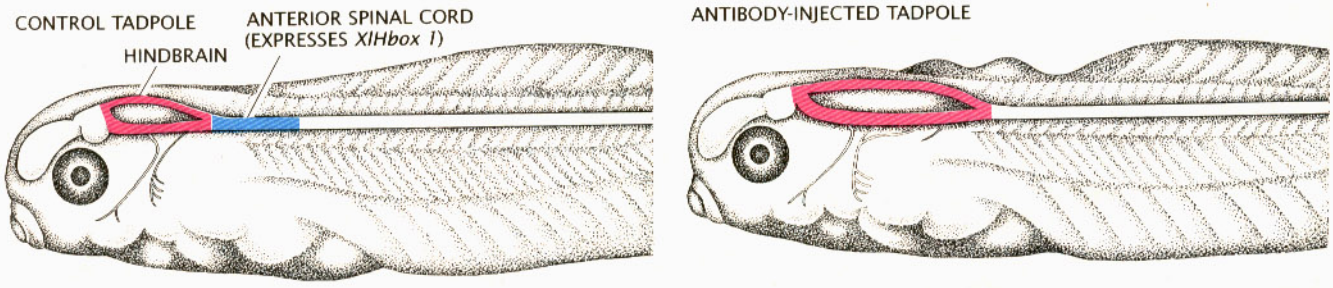
In the second experiment, Gruss and Michael Kessel of the Max Planck Institute injected DNA containing a mouse homeobox gene into mouse embryos. The piece of DNA was designed so that the homeobox gene would be expressed throughout the body, even in regions where it normally would not, such as the head and neck. The resulting mice frequently had severe head defects, such as cleft palates. Even more interesting, they also had an extra vertebra and intervertebral disk at the base of the cranium, and some had an extra pair of ribs in the neck region. In this way, the "gain-of-function" of a homeobox gene induced homeotic transformations precisely like those observed in fruit fly mutations.

Other work also suggests a role for homeobox genes in specifying cell identity. As previously described, homeobox genes are strongly expressed in bands along the anteroposterior axis early in development. Later, when organs are forming in these regions, the same homeobox genes are once again expressed intensely. At these later stages, homeobox genes seem to provide molecular tags that remind cells of where in the embryo they originated.

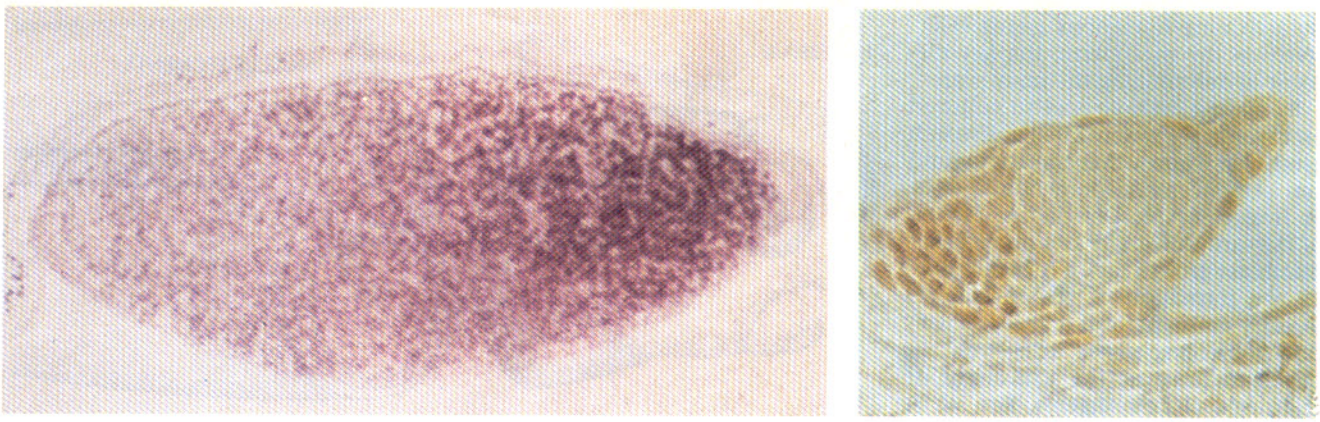
The development of the forelimb is a particularly informative case. The entire forelimb field is derived from the band of mesoderm that expresses *XlHbox 1*. Cells proliferate in the band and form a small forelimb bud that appears on *Xenopus* by the third week after fertilization.

At this stage the forelimb bud mesoderm appears uniform, but it contains a gradient of XlHbox 1 protein. That is, the protein is most abundant in cell nuclei along the anterior side of the limb bud—the side that gives rise to the thumb—and least abundant in nuclei on the posterior side, which gives rise to the smallest digit. As the limb extends and takes shape, the concentration of XlHbox 1 protein stays highest near the shoulder, at the proximal end of the arm. In contrast, the protein from another gene, *Hox 5.2*, establishes a gradient that is highest along the posterior side and the distal end of the limb—a pattern precisely the reverse of that for *XlHbox 1*.

Gradients of XlHbox 1 and Hox 5.2 proteins can be detected in frog, chicken and mouse embryos. Other homeobox genes are also involved in forelimb development: Duboule has identified three other homeobox genes adjacent to *Hox 5.2* that turn on sequentially as the limb tip extends. The order in which the genes are ac-



INHIBITION of a homeodomain protein can alter the developmental fate of embryonic tissues. In normal tadpoles (*left*), XlHbox 1 protein is expressed in a defined region of the anterior spinal cord (*blue*). If antibodies against XlHbox 1 protein are injected into a one-celled embryo, then in the resulting tadpole (*right*), that region is transformed into hindbrain (*red*).



GRADIENTS of homeodomain proteins can be seen in these stained limb buds. In the chick wing bud (*left*), the concentration of Hox 5.2 protein is highest in the cell nuclei near the posterior region (*right edge*). In the pectoral fin bud of a zebrafish (*right*), the concentration of XlHbox 1 protein is highest in the anterior region (*left edge*). The fish pectoral fin is the evolutionary precursor of the tetrapod forelimb. Gradients are efficient mechanisms for conveying positional information.

tivated corresponds to their order in the DNA.

Gradients of proteins or other molecules are good signposts for specifying the positions of cells and efficiently directing their fates. For example, cells in a limb bud can form separate digits by responding differently to varying amounts of a single protein. It would be less economical if a distinct protein had to specify each digit.

In conclusion, analysis of homeodomain protein gradients during limb development reveals that the same set of homeobox genes that establishes the head-to-tail axis is used again later to specify the positions of cells during limb development. Homeodomain proteins are found in cell nuclei, as would be expected of DNA-binding proteins that turn genes on and off. How gradients of nuclear proteins are established in limb buds is not yet known. Intercellular communication signals, similar to those involved in axis formation and perhaps mediated by growth factors or retinoic acid, are probably involved.

In addition to what the study of homeobox genes has explained about embryonic development, it has provided insights into evolution. Because the order of homeobox genes is similar in vertebrates and invertebrates, the first homeobox complexes must have evolved eons ago in flatworms or other primitive organisms that were the common ancestors of both human beings and insects. It would be interesting to know whether the most primitive multicellular organisms with anteroposterior polarity, such as rotifers, also have homeobox gene complexes. The amazing conservation of the complexes throughout evolution suggests that once an efficient way of specifying the anteroposterior axis was found, it was easier to produce new body shapes by modifying that system than to develop entirely new strategies.

Homeobox gene activity also offers clues about how specific anatomical structures might have evolved. Scientists have long wondered, for example, about the origin of the arm, or forelimb, in tetrapods. Because the primitive fish called coelacanths have pectoral fins with bony joints, investigators assumed for many years that the arm evolved from pectoral fins.

Support for this theory has come out of work by Anders Molven and Charles B. Kimmel of the University of Oregon. In zebrafish embryos, *XlHbox 1* is first expressed in a circular region of the lateral mesoderm corresponding to the pectoral fin field. At this stage the expression of the gene corresponds exactly with a morphogenetic field defined by Harrison in 1918.

As the cells proliferate, XlHbox 1 protein forms a steep gradient in the pectoral fin bud, similar to the gradients in frog, chicken and mouse forelimb buds. This pattern suggests that *XlHbox 1* is an ancient gene, whose function in the limb gradient-field antedates the appearance of tetrapod structures such as digits. Much can probably be learned by reexamining the comparative embryology of vertebrates at the level of gene expression.

Although it will take a long time to understand exactly how genes cooperate to organize cells from an apparently homogeneous egg into a swimming tadpole, molecular analysis of vertebrate development has already made a great leap forward. The expression of homeobox genes may provide a molecular explanation for the gradient-fields recognized by experimental embryologists many decades ago. The genes that control the anteroposterior axis are conserved in the zoological spectrum to a degree beyond investigators' wildest expectations. Molecules that may be involved in transmitting positional information, such as retinoic acid and growth factors, have been identified. Possibilities are now open for analyzing how body shape changes during the course of evolution. Students starting work in laboratories today may one day be able to answer simple questions, such as what makes an arm different from a leg. What a good time to begin!

FURTHER READING

CLONING OF A X. LAEVIS GENE EXPRESSED DURING EARLY EMBRYOGENESIS CODING FOR A PEPTIDE REGION HOMOLOGOUS TO DROSOPHILA HOMEOTIC GENES. Andrés E. Carrasco et al. in *Cell*, Vol. 37, No. 2, pages 409–414; June, 1984.

A GRADIENT OF HOMEODOMAIN PROTEIN IN DEVELOPING FORELIMBS OF XENOPUS AND MOUSE EMBRYOS. Guillermo Oliver et al. in *Cell*, Vol. 55, No. 6, pages 1017–1024; December 23, 1988.

THE MURINE AND DROSOPHILA HOMEOBOX GENE COMPLEXES HAVE COMMON FEATURES OF ORGANIZATION AND EXPRESSION. Anthony Graham et al. in *Cell*, Vol. 57, No. 3, pages 367–378; May 5, 1989.

INTERFERENCE WITH FUNCTION OF A HOMEOBOX GENE IN XENOPUS EMBRYOS PRODUCES MALFORMATIONS OF THE ANTERIOR SPINAL CORD. Christopher V. E. Wright et al. in *Cell*, Vol. 59, No. 1, pages 81–93; October 6, 1989.

VARIATIONS OF CERVICAL VERTEBRAE AFTER EXPRESSION OF A *HOX 1.1* TRANSGENE IN MICE. Michael Kesset et al. in *Cell*, Vol. 61, No. 2, pages 301–308; April 20, 1990.