In the Realm of the Homeodomain

When genes identified by mutant phenotypes are cloned, their sequences can sometimes turn out to be informative—and surprising. The maize gene Knotted-1 (Kn1) is one such gene. Identified 50 years ago as a dominant mutation that causes abnormal leaf development, Kn1 was cloned and sequenced last year by Sarah Hake and her collaborators (Vollbrecht et al., 1991). These investigators found that Kn1 contains a homeobox, a sequence motif found in genes from many other eukaryotes but never before in a plant gene. And Kn1 is by no means the only plant gene with a homeobox. Several other maize genes also include this sequence, Hake and her collaborators found. In addition, two groups have isolated homeobox-containing genes from Arabidopsis (Ruberti et al., 1991; Schena and Davis, 1992).

Why all the excitement about homeoboxes? The answer is simple: many genes that contain homeoboxes, 180-bp sequences that encode a stretch of 60 amino acids (the homeodomain) thought to be a DNA binding domain, play important roles in the developmental decisions that control cell specification and patterning (see Scott et al., 1989, for review). And their occurrence in plants as well as in animals and yeast confirms that the homeobox is an ancient motif that has been conserved throughout eons of evolutionary change.

A Conserved Motif in Some Drosophila Genes

The homeobox was first identified as a conserved sequence in several Drosophila genes that mutate to cause homeotic transformations (alterations in the body plan such that one segment or portion of a segment develops as if it were a different segment). Dominant alleles of the Antennapedia (Antp) and Ultrabithorax (Ubx) genes, for instance, affect adult flies. In the case of Antp, antennae are transformed into legs. Loss-of-function alleles of these genes cause homeotic transformations (and subsequent lethality) in embryos. Sequencing of the Antp, Ubx, and fushi tarazu (ftz) homeodomains revealed that they share a highly conserved sequence of about 180 bp within the coding region (McGinnis et al., 1984; Scott and Weiner, 1984). The protein sequence is even more highly conserved than the nucleotide sequence: the Antp and Ubx homeodomains are identical at 54 of 60 residues, and those of Antp and ftz are identical at 50 of 60 residues.

It has subsequently turned out that all of the homeotic genes, which are clustered in two gene complexes, have homeoboxes. However, the identification of a homeobox in the ftz gene, which is a segmentation gene rather than a homeotic gene (segmentation genes divide the embryo into repeating units instead of defining segment identity) indicated that, despite their moniker, homeoboxes are not limited to homeotic genes.

As more genes involved in anterior-posterior pattern formation in the Drosophila embryo have been cloned, this conclusion has been borne out, although only a fraction of the genes that regulate anterior-posterior patterning actually contain homeoboxes (Akam, 1987). Certain segmentation genes, such as even-skipped (eve), gooseberry, and engrailed (en), have homeoboxes, as do some of the genes that pattern larger regions of the embryo, such as the maternally expressed bicoid and caudal genes. Despite the high conservation of homeodomain sequences, they can be divided into groups of even more closely related sequences (Scott et al., 1989). The Antp, Ubx, and ftz homeodomains are all members of the Antp class of homeodomains, but the eve homeodomain, which shares 30 amino acids with that of Antp, defines a different class and the en homeodomain yet a third of the many classes.

Homeoboxes are not limited to genes that regulate pattern formation along the anterior-posterior axis. The zerknüllt gene, which is involved in embryonic dorsal-ventral pattern formation, includes a homeobox, as does cut, which specifies the fate of certain neurons; rough, which functions in eye development; and H2.0, which is expressed in gut musculature. Moreover, several segmentation genes appear to be redeployed later in development in the process of neurogenesis: in addition to its role in segmentation, eve controls the fate of a subset of neurons (Doe et al., 1988).

From Hydra to Humans

Not surprisingly, a motif shared by a slew of genes that regulate cell specification and pattern formation in Drosophila has turned up in genes from other animals—invertebrates and vertebrates, segmented and not. A remarkable finding was that not only the homeobox, but the homeotic complex itself, is conserved in other animals, among them beetles, mice, humans, and (in a more diverged form) the nematode Caenorhabditis elegans (see McGinnis and Krumlauf, 1992, for review). Recent evidence from knockout experiments suggests that the mouse genes are functionally homologous to the Drosophila genes (Chisaka and Capecci, 1991). Homeoboxes of the Antp class are found in cnidarians (Murtha et al., 1991), but it is not yet known whether they serve
homeotic functions in these primitive metazoans.

Homeobox-containing segmentation genes are ancient as well, being found in a wide variety of phyla. But, whereas it is likely that the functions of the homeotic genes have been conserved throughout much of animal evolution, the functions of the segmentation genes have probably changed. For example, enhomologues are found in annelids, arthropods, vertebrates, and echinoderms, not all of which are overtly segmented. Recent data suggest that many of the genes defined as segmentation genes in Drosophila were originally involved in neurogenesis and only later co-opted for use in segmentation (Patel et al., 1989).

In other organisms, as in Drosophila, genes that are involved in cell determination but have no obvious role in anterior-posterior patterning can also contain homeboxes (Scott et al., 1989). For example, several tissue-specific homeo-domain proteins have been found in vertebrates (Pit-1 is expressed in the pituitary, Oct-2 in B cells), and the homeo-box-containing gene mec3 controls the differentiation of certain neurons in C. elegans. Yeast, being unicellular, do not undergo pattern formation as such; nevertheless, they too have homeobox genes, among them both the a1 and a2 alleles of the master regulatory gene MAT, which controls a host of cell type specification events.

The high conservation of the homeo-domain, along with its appearance in many organisms, implies that this protein motif has an important biochemical function. A clue to this function came from comparing homeodomain sequences with the sequences of proteins of known function and, in some cases, known structure. The C-terminal portion of the homeodomain is weakly—but still significantly—similar to a stretch of amino acids in some prokaryotic DNA binding proteins, including the bacterial trp repressor and CAP protein, the bacteriophage λ repressor, and the λ Cro protein, which are known to regulate transcription (Laughon and Scott, 1984). This stretch of amino acids forms a structure known as a helix-turn-helix. Structural studies have shown that some of the residues of the second helix, the recognition helix, contact the DNA, while the first helix helps stabilize the recognition helix-DNA complex (Pabo and Sauer, 1984). Some of the most critical residues for the formation of the helix-turn-helix are highly conserved in homeodomains, and recent NMR studies demonstrate that the C-terminal portion of the homeodomain forms a helix-turn-helix structure similar to that of the prokaryote proteins (Otting et al., 1986). The prokaryotic regulatory proteins function as dimers; by analogy, homeodomain containing proteins may do so as well, although this has not yet been demonstrated.

Evidence is beginning to accumulate that homeodomain-containing proteins do indeed bind to DNA and regulate gene transcription (Scott et al., 1989). One piece of circumstantial evidence is that all homeodomain-containing proteins whose subcellular distribution has been examined are localized to the nucleus. More direct evidence has come from in vitro DNA binding and footprinting assays, which have shown that some isolated homeodomains and homeodomain-containing proteins from Drosophila bind to specific DNA sequences that differ depending on the protein, although the sequence TAAT is often included in the binding site. Additional evidence that homeodomain proteins can regulate transcription in vivo comes from cotransfection experiments. When a Ubx gene was introduced into tissue culture cells along with potential target promoters linked to reporter genes, for example, the Ubx protein repressed transcription from an Antp gene promoter and activated transcription from a Ubx gene promoter (Krasnow et al., 1989). In the case of some mammalian homeodomain proteins, such as Pit-1 and Oct-2, a role in transcriptional regulation is well established: these proteins were identified initially as transcription factors that bound to cis-acting regulatory sequences upstream of certain important genes and only later were found to contain homeodomains.

And Onward to Plants

What can the wealth of knowledge about the biochemical and developmental functions of homeodomains in animals reveal about the role of homeodomain-containing plant proteins? Although plant homeodomains are fairly divergent from animal homeodomains (Kn1 shares 12 residues with Antp and 17 with MATa2), they contain the highly conserved residues that would be necessary for DNA binding. The Arabidopsis homeobox genes (but not the maize homeobox genes, at least those that have been isolated so far) also contain a region upstream of the homeobox that appears to code for a leucine zipper. This motif, a region that forms an amphipathic helix that may function as a dimerization domain, has been found in some mammalian transcription factors but never before in combination with a homeodomain. The leucine zipper or, indeed, any other dimerization motif would probably allow for the formation of heterodimers as well as homodimers, and an intriguing possibility is that heterodimerization regulates the activity of some homeodomain-containing proteins.

It is too early to predict the developmental functions of plant homeobox genes except in a general way. By analogy with homeobox genes in animals, the plant genes probably serve as control points in cell specification. That is, when a group of cells is being "patterned" such that different subsets of cells adopt different fates, homeodomain proteins may step in to regulate the transcription of cell type-specific genes. Because all Kn1 alleles are dominant mutations, it is not possible to deduce with certainty the function of the wild-type Kn1 gene. The expression
of dominant alleles of Kn1 in internal layers of the leaf apparently leads to the production of a signal that causes adjacent cells to proliferate, forming "knots" along lateral veins (Sinha and Hake, 1990). The Kn1 protein may thus control the transcription of genes that specify a signal that determines the behavior of nearby cells.

Given the very different structures of plants and animals, it is also possible that the functions of plant homeobox genes are not strictly analogous to those of animals. The plasmodesmata connect the cytoplasm of plant cells, so that, whereas an animal is made up of individual cells, a plant can be thought of as constituting a cytoplasmic continuum. This architecture has profound implications for the functions of plant homeobox genes. The plasmodesmata connect the cytoplasm of plant cells, so that, whereas an animal is made up of individual cells, a plant can be thought of as constituting a cytoplasmic continuum. This architecture has profound implications for the organization of plant form and for many other aspects of plant growth and development (Kaplan and Hagemann, 1991).

For example, although homeodomain proteins themselves are unlikely to move freely from cell to cell, substances that regulate them—or that they regulate—may do so. This mode of cell-cell interaction could mean that, in at least some cases, plant homeobox genes regulate the expression of different kinds of genes than those regulated by animal homeobox genes.

The recent discovery that plants contain homeobox genes raises the issue of why they were not found earlier. Given the extent to which the plant and animal homeodomain sequences have diverged, it is now clear that a gene probe for the entire homeobox region would be unlike-ly to hybridize to a plant homeobox. The same holds true for C. elegans as well: when degenerate oligonucleotides spanning just the most conserved region of the homeobox were used as probes, numerous nematode homeobox genes were identified (Bürglin et al., 1989).

A further explanation for the belated unearthing of homeobox-containing plant genes may lie in the fact that fewer genes known to be involved in pattern formation and cell specification have been cloned from plants than from animals such as Drosophila. As genes involved in developmental processes from embryogenesis and leaf development to flower commitment and fruit development are identified genetically and then cloned and sequenced, it will not be surprising if some of them turn out to have homeodomains. Many developmental events in plants are regulated at the level of gene transcription, after all, and homeodomain proteins are, presumably, transcriptional regulators. Some genes implicated in the regulation of plant development have been sequenced, and several have turned out to encode other sorts of transcription factors. The protein encoded by the Arabidopsis floral homeotic gene agamous, for example, contains a region (the MADS box) that resembles the DNA binding domains of yeast and human transcription factors (Yanofsky et al., 1990), and the gl-1 gene, which is required for the initiation of leaf trichomes in Arabidopsis, contains a Myb DNA binding motif (Oppenheimer et al., 1991).

The finding that plant genes with important roles in pattern formation and cell specification can have Myb domains or MADS boxes illustrates an important point: in plants, as well as in animals, homeodomain proteins are just one class of transcriptional regulators that play important roles in development (Katagiri and Chua, 1992). Like all proteins that regulate transcription in subsets of cells, they must activate or repress the transcription of downstream effector genes that direct-ly or indirectly cause metabolic or structural changes in the cells in which they are active or lead to the production of diffusible signals that influence neighboring cells. They must also, themselves, be regulated by upstream events—such as transcriptional activation or post-transcriptional modification in response to external signals—that limit their activity to restricted sets of cells.

The discovery of recognizable homeoboxes in plants confirms that plants, like other eukaryotes, use homeodomain proteins—and, in at least one case, in a process whereby cell fates may be determined. But this finding raises many more puzzles than it solves: what kinds of processes do plant homeodomain proteins regulate? Do plant homeodomain proteins have the key regulatory roles that animal homeodomain proteins often seem to have? Do regulatory networks of plant homeodomain proteins, like the network involved in anterior-posterior patterning in Drosophila, control any plant pattern forming processes? Do differences in the modes of plant and animal development lead to different strategies for the use of homeodomain proteins? The solutions to these mysteries may begin to unfold as more plant homeobox genes are isolated and their functions probed by genetic analysis, studies of their patterns of expression, overexpression experiments, and gene disruption experiments.

Rebecca Chasan

REFERENCES


In the realm of the homeodomain.
R Chasan
DOI 10.1105/tpc.4.3.237

This information is current as of October 20, 2017