Homeobox Linked to Gene Control

Researchers have discovered a homeobox, a gene sequence of known developmental importance, in proteins that regulate mammalian gene activities

FOR MANY YEARS developmental biologists have thought that control of development actually comes down to control of gene expression. A flurry of reports now coming out of several laboratories is putting that hypothesis on even firmer footing.

Researchers have found a genetic link between three proteins that serve as transcription factors for initiating gene expression in mammals and a protein that regulates the development of the simple roundworm *Caenorhabditis elegans*. The genes for the four proteins contain a highly conserved sequence that apparently encodes a DNA-binding domain. This suggests that the proteins may all work the same way, by interacting with DNA and altering gene expression.

But especially interesting is the finding that the proposed DNA-binding domain includes a previously undiscovered type of "homeobox." Homeoboxes are gene sequences that were originally identified about 5 years ago in the homeotic genes of the fruit fly. These are genes that help to specify the development of the different cell types that form the various segments of the fruit-fly body.

Since their discovery, homeoboxes have attracted a great deal of attention, to the point of engendering what some researchers have called "homeo-madness." The gene sequences are widely distributed, occurring in species ranging from sea urchins to mammals, including man. The hypothesis is that the mammalian homeobox-containing genes, like their fruit-fly counterparts, are developmental control genes.

Before the current work, however, all the mammalian homeoboxes had been identified only on the basis of their structural similarity to the fruit-fly sequences. The functions of the genes with which these mammalian homeoboxes are associated are unknown. The discovery of homeodomains in the transcription factors has for the first time provided a biochemical function for mammalian homeobox-containing proteins.

Moreover, that function, as a component of the DNA-binding regions of the transcription factors, is in full accord with the mode of action predicted for homeobox-containing genes in development. In the fruit fly and presumably in mammals as well, the proteins produced by the homeobox genes are supposed to work by interacting with other genes and regulating their expression. The specific binding of the proteins to their target genes is thought to be mediated by the homeodomains, which are some 60 amino acids long. However, there has been little direct proof that homeodomains bind to specific DNA sites—at least until now.

The mammalian transcription factors in which the new homeobox has been found are the two called Oct-1 and Oct-2, and a third that acts specifically in pituitary cells. Their isolation is the result of the continuing effort by molecular biologists to dissect the mechanisms by which genes are controlled. This work has shown that cells produce several different kinds of factors that, by binding to specific control sites on genes, turn the genes on (or off, as the case may be).

The proposed DNA-binding domain includes a previously undiscovered homeobox.

Because genes are easier to handle than proteins, researchers generally work backward, starting by first identifying the control sites on genes and then looking for the protein transcription factors that bind to the sites. That is how Oct-1 and -2 and the pituitary factor were isolated.

Oct-1 and -2 bind to the same DNA sequence, which contains eight nucleotides and is sometimes called the OCTA (for octanucleotide) element. The discovery of the OCTA element presented molecular biologists with a paradox. It was originally detected in antibody genes as the control element that restricts antibody production to the B lymphocytes of the immune system. However, the OCTA element is also a positive control site in a large number of additional genes that are expressed in many different cell types. "The question," says Phillip Sharp of the Massachusetts Institute of Technology, "is how can the sequence be B cell–specific and non-B cell–specific at the same time?"

Researchers, including Sharp, David Baltimore of the Whitehead Institute for Biomedical Research at MIT, Winship Herr of Cold Spring Harbor Laboratory on Long Island, and Robert Roeder of Rockefeller University in New York City, found that this puzzle could be partly explained by the existence of two transcription factors needed for activating OCTA-element genes. One, now called Oct-2, is made only in B cells where it activates antibody gene expression. The other, Oct-1, is produced in all cell types, and is therefore responsible for the nonspecific expression of OCTA-element genes.

Meanwhile, Michael Rosenfeld of the Howard Hughes Medical Institute at the University of California, San Diego (UCSD), and his colleagues identified a pituitary-specific transcription factor that they called Pit-1. It activates the expression of the genes for prolactin and growth hormone in pituitary cells. The regulatory site recognized by Pit-1 is structurally related to the OCTA element.

The genes for the three factors have now been cloned and their nucleotide sequences determined. The first published report (in the 7 October issue of *Cell*) came from Louis Staudt who had previously worked on Oct-1 and -2 in Baltimore's laboratory but is now at the National Cancer Institute in Bethesda, Maryland. Staudt, with his NCI colleagues Hon-Sum Ko, Patricia Fast, and Wesley McBride, found that Oct-2 contains an identifiable homeobox that is nevertheless structurally distinct from previously identified homeoboxes. They also showed that mutations that alter the homeobox prevent Oct-2 from binding to DNA.

Meanwhile, researchers from the Baltimore and Sharp laboratories determined the complete nucleotide sequence of the *oct-2* gene, and the Herr group has done the same for the *oct-1* gene. The Roeder group has also obtained a clone and sequence for the *oct-2* gene.

The situation regarding the pituitary transcription factor is somewhat confused at the moment. The Rosenfeld group reports in the 4 November issue of *Cell* that they have

cloned and sequenced the gene encoding Pit-1. In the same issue, Michael Karin of the UCSD School of Medicine and his colleagues describe the sequence of the gene for a pituitary-specific transcription factor that they have identified.

According to Karin, this factor, called GHF-1, activates only the growth hormone gene and should not be the same as Pit-1. However, the Pit-1 and GHF-1 sequences are identical. One group may have inadvertently cloned the gene for the other's transcription factor. It remains for the researchers to work out what has happened, although it seems clear that the two groups have the gene for a pituitary-specific transcription factor of some kind.

In any event, comparison of the amino acid sequences of Oct-1 and -2 and Pit-1 revealed, Herr says, "that they all have homeoboxes that are more related to each other than to other homeoboxes." In addition, they all have another region of even greater sequence similarity, which contains some 75 amino acids and is located about 25 amino acids in front of the homeobox in the protein sequence. Since the transcription factors bind to similar DNA sequences, it is perhaps not surprising that they contain related amino acid sequences, although the identification of the homeobox as part of their putative DNA-binding site was no doubt a gratifying bonus.

An even more startling result came, however, with the sequencing of the *unc-86* gene of *C. elegans* by H. Robert Horvitz of the Howard Hughes Medical Institute at MIT and Michael Finney and Gary Ruvkun of Massachusetts General Hospital in Boston.* This gene helps to regulate neuronal development in the roundworm. The Unc-86 protein also carries the new homeobox in conjunction with the 75–amino acid conserved sequence. "Unc-86 is almost certainly a transcription factor," Horvitz concludes.

The members of the Baltimore-Sharp, Herr, Horvitz, and Rosenfeld groups are suggesting that the 160–amino acid sequence that includes the homeobox and the 75–amino acid sequence be called the POU (pronounced "pow") domain because it was found in *P*it-1, *O*ct-1 and -2, and *U*nc-86.

In addition to indicating that the Unc-86 protein is a transcription factor, the new findings also raise the possibility that the mammalian transcription factors, especially Oct-2 and the pituitary factor, which are made in specific cell types, might themselves help to determine cell fates. Staudt notes, for example, that the *oct-2* gene is active very early in the development of B cells, even before the antibody genes undergo the rearrangements needed to make the genes functional.

A great many questions still need to be answered about the POU domain and the proteins that carry it. Researchers will be looking, for example, for definitive proof that Unc-86 is a transcription factor and that the transcription factors are developmental control proteins. They also want to find out what the 75–amino acid conserved sequence does. Its high degree of conservation throughout evolution implies an important function.

Another question concerns whether the POU domain will be found in still more proteins. In addition, the investigators would like to get their hands on the genes that regulate the expression of *unc-86*, *oct-1* and *-2*, and the pituitary factor, as well as any additional genes regulated by the transcription factors and the Unc-86 protein. Homeo-madness may soon be compounded by POU-madness. ■ **JEAN L. MARX**

*The Horvitz paper is in press in *Cell*. The Baltimore-Sharp paper on Oct-2 and the Herr paper on Oct-1 will be published in the December *Genes and Development*. That issue will also feature a joint publication of the four groups (including Rosenfeld's) concerning the sequence similarity of the proteins they are studying.
In addition, Baltimore, Sharp, and Roeder described their results at the symposium "Gene Regulation and Oncogenes," which was held from 23 to 27 October in Chatham, Massachusetts, under the aegis of the American Association for Cancer Research.

Schizophrenia Genetics a Mixed Bag

Two groups of researchers have just reported new information on the genetics of schizophrenia. The results of the first group indicate that seven families appear to carry a single dominant gene for schizophrenia on chromosome 5. The second research team finds no such evidence for a genetic defect on the long arm of chromosome 5 in a large, multigenerational family. Taken together, the two reports, which appear in the 10 November issue of *Nature*, confirm that the disease does not follow a simple pattern of inheritance in every family that is affected.

Both reports are studded with cautions about interpreting the new results too broadly. Nevertheless, both groups say they have "strong evidence" to support their conclusions and emphasize that the conflicting results are not surprising.

"We have a linkage marker that identifies the approximate location of the gene for schizophrenia," says Hugh Gurling of the University of London. "It is monogenic [caused by a single gene] in seven families that we have studied. That is probably our most important finding." Neither the defective gene itself nor its precise position on chromosome 5 has been identified, however. Instead, Gurling, Robin Sherrington, also of the University of London, and their colleagues find it is linked to and inherited with other genetic "markers"—two identified pieces of DNA on the long arm of chromosome 5 that are located in about the same region as the proposed defective gene.

A second research group used more markers and did not find any evidence of an abnormal gene for schizophrenia on the long arm of chromosome 5. "We would not have published these particular results had it not been for the contrasting positive data from Gurling's study," says Kenneth Kidd of Yale University School of Medicine in New Haven, Connecticut. But, he adds, "We fully expect to find a marker. It's just that in this family, it's not on chromosome 5."

The study populations differ in several respects. Kidd, James Kennedy, also of Yale, and their colleagues studied one large Swedish family with 157 members. Thirty-one were strictly diagnosed as schizophrenic and most had an early, severe onset of the disease. Gurling, in contrast, purposely selected seven smaller families—two from England and five from Iceland—because they include an unusually large percentage of affected individuals. Thirty-nine of 104 people were diagnosed as schizophrenic, five as having schizoid personality types, and ten as having "fringe" mental disorders. Given the broad range of psychiatric illnesses in Gurling's study, Kidd expresses surprise that all of the families that Gurling studied could be linked to the same chromosome 5 markers. Gurling sees the different illnesses as "unusual expressions of the schizophrenia mutation."

While the new results are an important advance in the genetics of schizophrenia, they are far from being the final word. "We still don't know what proportion of schizophrenics in the population at large have this genetic predisposition on chromosome 5." says Elliot Gershon of the National Institute of Mental Health. "To do that we need to study a larger number of families." People should also realize that having the abnormal gene does not necessarily mean that someone will develop schizophrenia, he says.

Gershon thinks the new findings should lead to other advances. They may enable researchers to see if the chromosome 5–linked schizophrenics are a subgroup that might respond better to particular treatments. And, "if the actual gene is found, it will tell us a lot about the biology of this disease," he says. ■ **DEBORAH M. BARNES**