An update on the vertebrate homeobox

GREGORY R. DRESSLER
DEPARTMENT OF MOLECULAR CELL BIOLOGY,
MAX PLANCK INSTITUTE OF BIOPHYSICAL CHEMISTRY,
3400 GÖTTINGEN, FRG.

Given the complexity of development and the multitude of differentiated cell types in higher eukaryotic organisms, one might expect that a large variety of proteins would be required to regulate specific developmental processes. However, if the differentiated state of a particular cell is specified not by a single protein but by a combination of factors, significantly fewer regulatory proteins are required to constitute a unique genetic program for a single cell type. Furthermore, if the expression level of a single factor, in combination with others, is a determinative event, an even smaller set of governing factors would suffice.

Thus considering the finite number of controlling factors, it is remarkable that more and more potential developmental and cell-specific factors contain homeoboxes. While the evidence that vertebrate homeobox genes encode transcription factors active during development is mostly circumstantial, several functionally defined tissue-specific transcription factors have also been found to contain homeobox domains.

New classes of mouse homeoboxes

Drosophila homeoboxes fall into various categories depending on the genes in which they are found[1]. The *Antennapedia* (*Antp*) homeobox is the prototype but other types include the *engrailed* (*en*), *bicoid*, *caudal* (*cad*), *paired* (*prd*), and *even-skipped* (*eve*) homeoboxes. Although at least 24 mouse genes contain an *Antp*-type homeobox and at least two genes have the *en*-type homeobox[2,3], murine sequences homologous to other *Drosophila* homeobox classes have not been isolated until recently.

A new mouse homeobox locus on chromosome 5, *Hox 7*, was defined recently in two independent reports. Using the *Drosophila muscle segment homeobox* (*msh*) sequence, a genomic cosmid library was screened and a homologous mouse sequence with 92% amino acid identity was identified[4]. Surprisingly, the same gene was found in a mouse cDNA library by an entirely different screening method. Hill *et al.*[5] isolated cDNA clones that hybridized to *Hox 1.6* at low stringency but did not hybridize to *Drosophila fushi tarazu* (*ftz*), which contains an *Antp*-type homeobox. The resulting clone, *Hox 7.1*, shares only 56% nucleotide identity with *Hox 1.6* and less than 50% with *Antp* or *ftz*. Preliminary evidence indicates that there may be several related genes with the *muscle segment homeobox*[5]; however it is not known if these genes map to the same locus.

Many mouse homeobox genes are expressed in ectoderm- and mesoderm-derived tissues in a region-specific manner during embryogenesis[6]. However, the expression of *Hox 7.1* in neural crest cells, but not neural tube, at 9 days gestation is a unique feature of this gene. In addition, expression of *Hox 7.1* in the first visceral arch and subsequent mandible and maxillary structures may indicate a function during cranial neural crest cell predetermination. Such predetermination is apparent in transplantation experiments in the chick, which indicate that neural crest cells are predisposed to a particular developmental fate before they migrate to their final destination[7]. Thus, a mechanism for predisposition must exist that defines the eventual fate of specific subsets of neural crest cells. Whether homeobox genes are involved in this neural crest predetermination is an intriguing possibility that certainly warrants further investigation.

A second new type of homeobox gene characterized recently is the mouse *Cdx1* gene[8], which contains a homeobox homologous to the *Drosophila caudal* gene. In fact, *Cdx1* shares many sequence similarities with *caudal* besides the homeobox. The expression of *Cdx1* is unusual, relative to other mouse *Hox* genes, because it peaks relatively late in development, continues into adulthood, and is restricted to epithelial cells of the intestinal tract that are derived from embryonic endoderm. These intestinal epithelial cells consist of four major differentiated cell types, all derived from a single stem cell population that continuously renews the cells of the villi; thus it would be interesting to determine if the *Cdx1* gene is restricted to a single differentiated cell type or if it is found only in the progenitor cells.

Transcription factors with homeodomains

The *Drosophila en*[9], *eve*[10] and *Antp*[11] homeodomains bind DNA in a sequence-specific manner. Furthermore, evidence supporting the helix–turn–helix tertiary structure model was recently obtained with nuclear magnetic resonance[12]. Whether these data can be extrapolated to vertebrate *Hox* genes remains to be shown. However, while many investigators are eager to prove that *Hox* genes encode transcription factors, several groups studying known transcription activators have cloned and sequenced the respective genes and have found homeodomains.

A protein that binds the immunoglobulin promoter octamer sequence, the human Oct-2 protein, contains a homeodomain that is only 33% homologous to the consensus homeodomain but, more importantly, it retains the helix–turn–helix structural motif[13,14]. A three amino acid substitution in the recognition helix disrupts the binding of Oct-2 to the octamer sequence[13]. Two other

homeodomain-containing vertebrate transcription factors have recently been identified: Oct-1 (Ref. 15) is another immunoglobulin octamer-binding protein, but unlike Oct-2, which is B cell specific, it is expressed ubiquitously; the second protein, Pit-1 (Ref. 16) or GHF-1 (Ref. 17), has been isolated from the rat pituitary gland, and specifically activates the growth hormone and prolactin genes in cells of the somatotropic and lactotropic lineages. The protein contains a homeodomain with 35% amino acid identity to the *Drosophila prd* homeobox.

These studies clearly demonstrate that mammalian proteins with homeodomains recognize specific sequences and activate transcription. However, target promoter sequences that bind the vertebrate *Hox* genes containing the *Antp*-type homeodomains have not been identified to date.

Functional assays for mammalian *Hox* gene function

Unlike *Drosophila*, mouse mutants for particular *Hox* genes have not been described and thus assigning a developmental function to these genes has proved difficult. Not deterred by lack of naturally occurring or chemically induced mutants, modern molecular biologists have begun to circumvent this problem with innovative transgenic approaches.

The first report of modulated *Hox* gene expression in transgenic mice was recently published by Wolgemuth *et al.*[18]. Through an apparent gene dosage effect, the *Hox 1.4* gene was overexpressed in certain embryonic structures. In particular, the mesenchyme of the developing intestinal tract, which normally expresses low levels of *Hox 1.4*, showed a dramatic increase in mRNA levels. The resultant adult phenotype resembled congenital megacolon, as characterized by an enlarged bowel and the inability to extrude feces. It was postulated that a deficiency of neural crest cells, which innervate the colon and generate the myenteric ganglia, would explain this developmental defect.

That neural crest cells are potential targets of *Hox* gene function was also demonstrated by Balling *et al.*[19] in transgenic mice expressing *Hox 1.1*. By coupling the *Hox 1.1* coding sequence to a ubiquitous promoter, ectopic expression during development was observed. Craniofacial malformations were apparent in transgenic animals expressing *Hox 1.1* in anterior structures that normally do not express the gene. In fact, the resulting phenotype, which included cleft secondary palate, nonfused pinnae of the outer ear, and open eyes at birth, was strikingly similar to retinoic acid embryopathy, a syndrome induced by the administration of retinoic acid during pregnancy. The implications of this study are twofold: (1) that aspects of neural crest cell maturation during craniofacial morphogenesis can be disturbed by the ectopic expression of a homeobox gene, and (2) that retinoic acid may induce *Hox* gene expression *in vivo*, as it does *in vitro*.

In the amphibian *Xenopus laevis*, perturbation of dorsal root ganglia and dorsal fins has been reported by Cho *et al.*[20] after microinjection of antibodies against the *XlHbox1* homeodomain protein. Interestingly, both dorsal fin mesenchyme and dorsal root ganglia are derived form neural crest cells. Furthermore, the defect appeared to be specific to anterior ganglia only. The data again suggest that neural crest cells are at least one potential target for *Hox* gene function.

In addition to gain-of-function mutants, gene targeting in embryonic stem cells can potentially generate loss-of-function mutant animals. In a recent report by Zimmer and Gruss[21], the *Hox 1.1* gene was mutated in embryonic stem cells by homologous recombination. Although these stem cells could contribute to chimeric mice, it is not yet known if germ-line transmission will occur and, subsequently, if heterozygous mice are viable. The viability of mutant lines is essential in addressing gene function and may be problematic considering the variety of tissues that express a single *Hox* gene and the potential for pleiotropism. In fact, the problem of viability has already surfaced in the study of Balling *et al.*, where it was reported that animals expressing the transgene, with one exception, died shortly after birth.

In conclusion, it appears that the numerical limit of homeobox-containing genes is slowly being approached, particularly for the *Antp*-type homeobox genes, as the rate of new gene isolation decreases. New genes cloned recently are more and more divergent from the prototype *Antp* homeodomain. It is likely that mouse homologs to other *Drosophila* homeobox sequences, such as *eve* and *bicoid*, do exist; however, they probably do not represent a family as large as the murine *Hox* genes. Although new genes will continue to be characterized, the current emphasis clearly is on assigning functions to the many *Hox* and *Hox*-related genes already described.

Acknowledgements

I thank Rudi Balling and Michael Kessel for sharing results prior to publication and for stimulating discussions.

References

1 Gehring, W.J. (1987) *Science* 236, 1245–1252
2 Dressler, G.R. and Gruss, P. (1988) *Trends Genet.* 4, 214–219
3 Wright, C.V.E., Cho, K.W.Y., Oliver, G. and De Robertis, E.M.(1989) *Trends Biochem. Sci.* 14, 52–56
4 Robert, B. *et al.* (1989) *EMBO J.* 8, 91–100
5 Hill, R.E. *et al.* (1989) *Genes Dev.* 3, 26–37
6 Holland, P.W.H. and Hogan, B.L.M. (1988) *Genes Dev.* 2, 773–782
7 Noden, D.M. (1988) *Development* 103 (Suppl.), 121–140
8 Duprey, P. *et al.* (1988) *Genes Dev.* 2, 1647–1654
9 Desplan, C., Theis, J. and O'Farrell, P.H. (1988) *Cell* 54, 1081–1090
10 Hoey, T. and Levine, M. (1988) *Nature* 332, 858–861
11 Müller, M. *et al.* (1988) *EMBO J.* 7, 4299–4304
12 Otting, G. *et al.* (1988) *EMBO J.* 7, 4305–4309
13 Ko, H-S., Fast, P., McBride, W. and Staudt, L.M. (1988) *Cell* 55, 135–144
14 Clerc, R. *et al.* (1988) *Genes Dev.* 2, 1570–1581
15 Sturm, R.A., Das, G. and Herr, W. (1988) *Genes Dev.* 2, 1582–1599

16 Ingraham, H.A. *et al.* (1988) *Cell* 55, 519–529
17 Bodner, M. *et al.* (1988) *Cell* 55, 505–518
18 Wolgemuth, D.J. *et al.* (1989) *Nature* 337, 464–467
19 Balling, R., Mutter, G., Gruss, P. and Kessel, M. *Cell* (in press)
20 Cho, K.W.Y. *et al.* (1988) *EMBO J.* 7, 2139–2149
21 Zimmer, A. and Gruss, P. (1989) *Nature* 338, 150–153

Autoregulation - a common property of eukaryotic transcription factors?

EDGAR SERFLING

INSTITUT FÜR VIROLOGIE UND IMMUNBIOLOGIE,
UNIVERSITÄT WÜRZBURG,
VERSBACHER STR. 7, D-8700 WÜRZBURG, FRG

The identification of homeobox-like sequences in mammalian genes coding for tissue-specific transcription factors (see Ref. 1) suggests that many of these factors function in the same way as the proteins of homeotic genes in *Drosophila*, namely by establishing and maintaining the differentiation status of a given cell type. Thus, for example, one may assume that Oct-2, the B-cell-specific octamer factor controlling the expression of immunoglobulin (Ig) genes, plays an important role in B-lymphocyte development.

Numerous homeobox-containing genes in *Drosophila* control their own expression by positive autogenous regulation. One example among several (see Fig. 1 and Table 1) is *fushi tarazu* (*ftz*), a pair rule gene that is expressed in two phases during development: at the blastoderm stage and later during neurogenesis. The expression of *ftz* in blastoid cells is regulated by two different *cis*-acting DNA elements, the so-called zebra-element located next to the transcription start site and an enhancer element situated a few kb further upstream[2]. While the activity of the zebra (= promoter) element is controlled by the products of other pair rule and gap genes, the activity of the enhancer element is positively regulated by the *ftz* gene product[2] (Fig. 1).

At the molecular level, the *ftz* product recognizes the sequence TCAATTAAAT located within its enhancer and autoregulates its own expression through this motif. Two other homeobox proteins, engrailed (en) and even-skipped (eve), can also bind *in vitro* to TCAATTAAAT sequences, but they cannot activate *ftz* expression. Moreover, they strongly antagonize autogenous induction by *ftz*[3,4].

Repeats of the consensus sequences TCAATAAAT and TCAGCACCG are part of putative control regions of *en* and *eve*[5]. When expressed in bacteria, the proteins of both genes bind to these sequences, and it is likely that *en* and *eve* autoregulate their expression by interaction with these motifs. The proteins of the paired (*prd*) and *zerknüllt* (*zen*) genes have similar binding specificities (Table 1).

Homeotic genes also seem to *trans*-activate their own expression and that of other homeotic genes through sequence motifs composed of multiples of the trinucleotide TAA [in particular $(TAA)_5$], the related sequence TAATCG, and ANNNNCATTA[5,6]. $(TAA)_5$ sequences, often intermingled with repeated

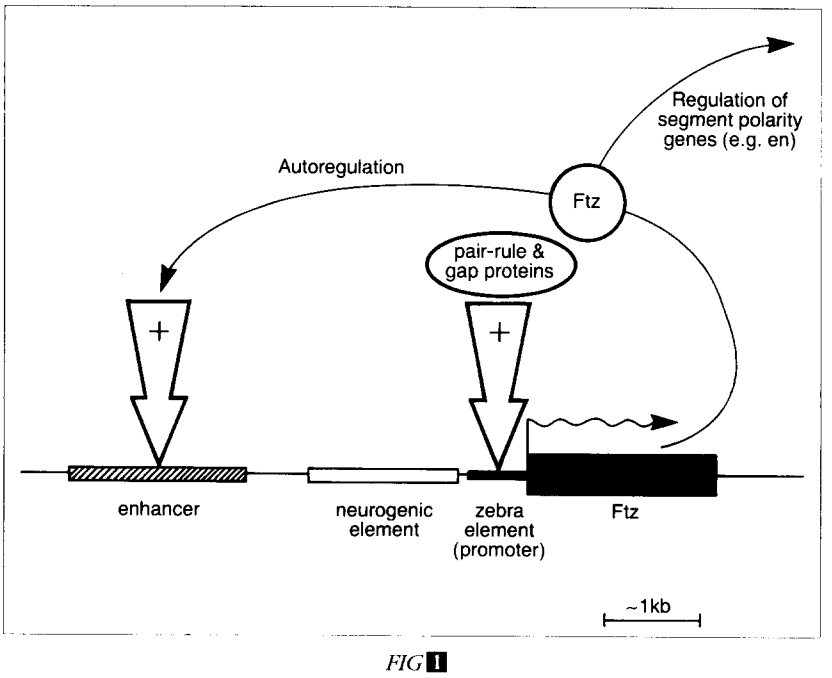

FIG **1**

Scheme for autoregulation of pair rule gene *fushi tarazu* in blastoid cells during early *Drosophila* development. In blastoid cells, the expression of *ftz* is controlled both by the proximal zebra (= promoter) element and the distal enhancer. While the activity of the zebra element is regulated by the products of other pair rule and gap genes, the enhancer is the target of autogenous regulation by the *ftz* gene product[2]. However, *ftz* can also activate other genes, such as *engrailed*, by direct interaction with their regulatory sequences[25].