Drosophila development

Homing in on homoeo boxes

from Geoffrey North

SINCE its discovery^{1,2} as a short repeated sequence within the homoeotic gene complexes of the fruitfly *Drosophila melanogaster*, research on the 'homoeo box' has taken three more or less independent directions. One has been the pursuit of evidence that the presence of homoeo boxes in the genomes of other species really does signify that they have equivalents of the fruitfly's homoeotic genes. The state of this particular art was discussed recently in these columns³. A second has been cataloguing of all *Drosophila* genes that have homoeo boxes, the newest member of the club being *caudal*⁴, which is the first example that is 'maternally expressed' and so presumably concerned with the very earliest events of development. And the third has been the attempt to ascribe a function to the homoeo box. This has now borne fruit, with the evidence reported by C. Desplan, J. Theis and P.H. O'Farrell on page 630 of this issue⁵ that the protein domain encoded by the homoeo box of the *engrailed* gene of *Drosophila* has sequence-specific DNA-binding activity.

But first some recent history. Shortly after their discovery^{1,2} as repetitive sequences of both the bithorax and Antennapedia gene complexes of *Drosophila*, homoeo boxes were found to encode roughly 60 highly-conserved amino acids, which would form a domain within the protein products of each of the genes that constitute these complexes. Although most of these genes are strictly homoeotic in that their mutations cause clean transformations of one body-part into another, one of the first homoeo boxes to be sequenced belongs to a rather different kind of gene. This is *fushi tarazu*, the odd-man out in the Antennapedia complex, which falls into the 'pair-rule' class of segmentation genes: in *fushi tarazu* mutant embryos, equivalent regions of every other segment are deleted.

The diversity of genes that contain a homoeo box was increased still further with *engrailed*^{6,7} — the first to be found outside the bithorax and Antennapedia complexes. *Engrailed* has some properties in common with the classical homoeotic genes, but others that are more reminiscent of segmentation genes such as *fushi tarazu*. Thus it seems to specify the posterior compartment of every segment of the developing fruitfly, for if clones of mutant cells lacking wild-type *engrailed* activity are created in a posterior compartment they develop as though they were in the equivalent region of the anterior compartment of the same segment. But in *engrailed* mutant embryos, unlike classical homoeotic mutants, the pattern of seg-

mentation is disrupted. In fact, the effect resembles that of a pair-rule mutation, which may relate to the observation that the developing pattern of *engrailed* expression, which in its final form is a series of stripes corresponding to the eventual pattern or morphological segmentation, goes through an intermediate stage with a clear-cut two-segment periodicity^{8,9} (see figure).

Thus, although all of the *Drosophila* homoeo-box-containing genes, including *caudal*, seem to have interesting roles in

Expression of *engrailed* with a two-segment periodicity (from ref. 8).

development, at the macroscopic level they seem to have a variety of types of function. But the implication of their shared homoeo-box domain is that at the molecular level the products of these genes work in basically similar ways. A clue as to what this common activity might be has come from the homology of the *Drosophila* homoeo boxes to regions of the yeast *MAT a* and *MAT a2* genes^{10,11}, which are known to regulate sets of genes that determine the mating type of haploid yeast cells. The protein product of the *MAT a2* gene has recently been shown to bind specifically to a sequence located upstream of the genes that it regulates¹².

It is against this background that Desplan *et al.* have tried to determine whether, as widely predicted, the *Drosophila* homoeo boxes also encode DNA-binding domains. Following a similar approach to that of Johnson and Herskowitz in their study of the yeast *a2* protein, Desplan *et al.* have constructed fusion proteins in which different regions of the *engrailed* protein product are tagged on to the end of the enzyme β -galactosidase. Expressed from hybrid genes introduced into the bacterium *Escherichia coli*, high enough levels of these fusion proteins can be obtained for biochemical studies.

Compared with Johnson and Herskowitz, however, Desplan *et al.* were at a disadvantage, for they had no prior indication of the sequences to which the *engrailed* protein might bind. So at first they simply tested whether their fusion proteins would bind non-specifically to frag-

ments of the bacteriophage lambda genome. Indeed, they found that those fusion proteins that contained the *engrailed* homoeo-box domain did have DNA-binding activity. But was this simply due to a non-specific interaction between acidic DNA fragments and the basic protein molecules? It seems not, for under more stringent conditions it was revealed that some of the lambda fragments are bound much more avidly than others. This is a clear sign that the DNA-binding activity of the fusion proteins is of the legitimate, sequence-specific kind. Presumably it is just by chance that some lambda fragments fulfill some of the sequence requirements for binding the *engrailed* protein, but it could well be that the fusion proteins are rather less choosy *in vitro* than *in vivo* (where their specificity might well be increased by interactions with other proteins).

Finding the functionally significant sequences of the *Drosophila* genomes to which the *engrailed* protein is presumed to bind is a more difficult problem, which Desplan *et al.* have approached by following the increasing genetic evidence that genes such as *engrailed* are involved in mutual regulatory interactions. They have therefore looked for and found binding sites for the *engrailed* protein upstream of the *engrailed* gene itself — but, as the authors themselves stress, this highly intriguing result must be interpreted with caution, for the affinity of the protein for these sites is not significantly different from that for its favourite fragments of the lambda genome. The binding sites near the *engrailed* gene are conserved between two different *Drosophila* species, however, which implies some functional significance. If it turns out that they are not true binding sites for the *engrailed* protein perhaps they are binding sites for some other protein that contains the homoeo-box domain. An example would be the product of the *fushi tarazu* gene, which recent genetic evidence suggests is involved in establishing the pattern of *engrailed* expression during the early development of *Drosophila* embryos (Phil Ingham and Ken Howard, personal communication). □

1. McGinnis, W. *et al.* *Nature* **308**, 428 (1984).
2. Scott, M.P. & Weiner, A.H. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4115 (1984).
3. Robertson, M. *Nature News and Views* **318**, 12 (1985).
4. Mlodzik, M., Fjose, A. & Gehring, W.J. *EMBO J.* **4**, 2961 (1985).
5. Desplan, C., Theis, J. & O'Farrell, P.H. *Nature* **318**, 630 (1985).
6. Fjose, A., McGinnis, W.J. & Gehring, W.J. *Nature* **313**, 284 (1985).
7. Poole, S.J., Kauver, L.M., Drees, B. & Kornberg, T. *Cell* **40**, 37 (1985).
8. Weir, M.P. & Kornberg, T. *Nature* **318**, 433 (1985).
9. DiNardo, S., Kurer, J.M., Theis, J. & O'Farrell, P. *Cell* **43**, 59 (1985).
10. Shepherd, J.C.W. *et al.* *Nature* **310**, 70 (1984).
11. Laughon, A. & Scott, M.P. *Nature* **310**, 25 (1985).
12. Johnson, A.D. & Herskowitz, I. *Cell* **42**, 237 (1985).

Geoffrey North is Deputy Biology Editor of *Nature*.