Cell, Vol. 57, 347–349, May 5, 1989, Copyright © 1989 by Cell Press

Hox and HOM: Homologous Gene Clusters in Insects and Vertebrates

Minireview

Michael Akam
Department of Genetics
University of Cambridge
Cambridge CB2 3EH
England

In Drosophila, homeotic genes of the Antennapedia (ANT-C) and Bithorax (BX-C) complexes specify the distinctions between different segments of the body. Each of these genes contains a conserved DNA sequence, the homeobox, that encodes a DNA binding homeodomain in the protein. When the homeobox was first identified, it was hailed as a motif unique to segmentation genes (Gehring, 1987). It is now clear that homeoboxes are present in a wide range of eukaryotic regulatory genes. Yet results on the detailed structure, organization, and expression of *Antennapedia* (*Antp*)-like genes in vertebrates (Gaunt et al., 1988; Duboule and Dolle, 1989; Graham et al., 1989) suggest that the vertebrate *Hox* clusters containing these genes are true homologs of the insect homeotic gene complexes (HOM-C). It appears that both HOM-C and *Hox* contain, in the same chromosomal order, the descendants of a gene family whose main members were already distinct when insect and vertebrate lineages diverged. Moreover, corresponding genes of the vertebrate and insect complexes show the same relative boundaries of expression along the antero-posterior (A–P) axis of the developing embryo. It is hard to escape the conclusion that insects and vertebrates have inherited from a common ancestor a conserved molecular representation of front, middle, and back.

The *Antp* homeodomain adopts in solution a conformation closely similar to the helix–turn–helix structure of bacterial repressor proteins (Otting et al., 1988). Limited sequence similarity underlies this common structure, but these critical residues are encoded by virtually all homeoboxes, and are also seen in the yeast mating type proteins (Scott et al., 1989), suggesting that all these proteins utilize variants of the same DNA binding domain.

The first homeoboxes to be identified were detected by DNA hybridization with *Antp*-like probes, and so of necessity were relatively similar in sequence to *Antp*. As more divergent homeoboxes are detected by direct comparison of sequence data, it is becoming harder to define precisely what is and what is not a homeobox. At the end of the day (and particularly when higher plants have been adequately screened), the family of homeodomain proteins may merge indistinguishably with other helix–turn–helix proteins that show no more homology to *Antp* than they do to the yeast mating type proteins. In Drosophila the most divergent members of this family have no functional association with segmentation, and in vertebrates they include both general and cell-type-specific transcription factors. While the homeodomain may characterize proteins with a particular structure, it is not the hallmark of genes playing any single role in development.

Antp-like Genes

Within the broad family of homeobox containing genes, the *Antp*-like genes are distinguished both by structure, and, in insects, by function. In Drosophila, six genes, all within the ANT-C or BX-C, are related by the possession of homeoboxes closely resembling that of *Antp*. Five of these genes, *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*), *Antp*, *Ultrabithorax* (*Ubx*), and *abdominal-A* (*abd-A*), are homeotic segment selector genes. Each is expressed only within a particular region along the A–P axis of the developing embryo, and mutations in each alter segment identity within these regions (see Akam, 1987, for a review). These five genes share further structural features—most notably a short conserved protein sequence encoded as part of a separate exon preceding the homeobox. The one exception among this set is the pair-rule segmentation gene *fushi tarazu* (*ftz*). This gene has an *Antp*-like homeobox, but in other respects its structure, regulation, and developmental role render it an anomaly within the *Antp*-like family.

Specific homologs of these *Antp*-like homeobox genes have been characterized in vertebrates. Many of the vertebrate genes encode proteins that are clearly *Antp*-like, on the basis both of homeodomain sequence, and of conserved regions elsewhere in the protein. In several cases the vertebrate homeodomains are virtually identical to that of *Antp* itself (59/60 residues identical); other genes encode proteins that are specifically related to those of the Drosophila genes *Scr* (Graham et al., 1989) or *Dfd* (Regulski et al., 1987). In the latter case there is more than 40% overall identity between the Drosophila and the vertebrate proteins.

Four additional homeobox genes are expressed at specific positions along the A–P axis of the Drosophila embryo. Two of these are clearly involved in the control of segment identity. *Abdominal-B* (*Abd-B*), in the BX-C, is required for the development of posterior abdominal segments; *proboscipedia* (*pb*), in the ANT-C, is active in anterior segments and required for the development of normal mouthparts in the adult (Pulz et al., 1988). The two others, *labial* (*lab*) (Diederich et al., 1989) and *caudal* (*cad*) (MacDonald and Struhl, 1986), are characterized by mutations that disrupt the development of anterior and posterior structures, respectively. But interpretation of these phenotypes is difficult; it is not clear whether their effects are homeotic.

The homeobox sequences of *lab*, *cad*, and *Abd-B* are sufficiently diverged to set them apart from the *Antp*-like inner family (the sequence of the *pb* homeobox has not yet been published). Recently, specific vertebrate homologs for all three of these genes have been reported. In the cases of *lab* (Baron et al., 1987) and *cad* (Duprey et al., 1988), homologies in both the homeobox and in other

regions of the protein are sufficiently strong to make the relationships unequivocal; rather weaker sequence homologies identify the specific homologs of the Drosophila *Abd-B* gene (Graham et al., 1989; Duboule and Dolle, 1989).

Conserved Gene Organization and Embryonic Expression

In both mouse and Drosophila, all the *Antp*-like genes, together with *lab*, *pb*, and *Abd-B*, are present in chromosomal clusters. The mouse has at least four such *Hox* clusters. *Hox-1* (Duboule et al., 1986), and *Hox-2* (Graham et al., 1988) appear to be complete duplications of an ancestral cluster that contained representatives for all members of this gene set. The *Hox-3* (Breier et al., 1988) and *Hox-5* (Duboule and Dolle, 1989) clusters may be incomplete duplications. In Drosophila the *Antp*-like genes are split between two clusters, the ANT-C and the BX-C, but this split probably occurred only relatively recently in the lineage leading to flies. In the beetle Tribolium, mutations analogous to those of both the ANT-C and the BX-C map to a single chromosomal locus, implying the existence of a single homeotic complex, HOM-C (Beeman, 1987). In Drosophila, the chromosomal sequence of genes *lab*–pb–Dfd–Scr–Antp–//–Ubx–abd-A–Abd-B corresponds to the antero-posterior sequence in which these genes are expressed in the embryo (// indicates the separation between ANT-C and BX-C, present in Drosophila but not in Tribolium). Recent papers (Gaunt et al., 1988; Duboule and Dolle, 1989; Graham et al., 1989) show that the mouse homologs of *lab*, *Dfd*, *Antp*, and *Abd-B* are similarly ordered within each *Hox* cluster, and show a similar A-P sequence of expression, most clearly apparent in the nervous system and somitic mesoderm. This is illustrated by Graham et al. (Figure 6, this issue), who show successively more anterior boundaries of expression in the central nervous system for genes of the *Hox-2* cluster.

Although comparable expression data are not available for other vertebrate species, clusters of *Hox* genes have been identified in zebra fish, Xenopus, chicken, and man. The conservation of protein coding sequences allows a direct correspondence to be established between specific *Hox* genes in each of these species. Where data on the chromosomal order of genes are available, they suggest that the colinearity of Drosophila and vertebrate gene types has been preserved in other species, including humans (Boncinelli et al., 1988).

An obvious explanation for the correspondence between the HOM-C of insects and the *Hox* clusters of vertebrates is that they are truly homologous, representing the descendants of an ancestral cluster of homeobox genes already present in the common ancestor. Alternative views require that convergent evolution should establish, in the same chromosomal order, four distinct gene types with corresponding homeobox sequences and spatial regulation.

Evolutionary Considerations

Insects and vertebrates represent the pinnacles of the two major metazoan lineages—the protostomes and the deuterostomes. Any common ancestor must lie among the origins of the bilaterally symmetric metazoans, conventionally thought to be at a grade of organization comparable to that of a flatworm. Clearly such an ancestor must have known its head from its tail. The corresponding deployment of region-specific homeobox genes in insects and vertebrates suggests that flies, mice, and hence, ourselves, have inherited from this common ancestor not only a conserved gene cluster, but also a regulatory network that provides a molecular representation of front, middle, and back.

If this is so, then in the vertebrate and arthropod lineages the evolution of metameric segmentation has most probably been superimposed independently onto a regionalized embryo. In insects, regional identity has been refined into unique segment identities, and the processes of segmentation and regional specification have become very tightly coupled. There is evidence to suggest that the segmentation gene *ftz* may play a key role in this process (Ingham and Martinez-Arias, 1986)—an explanation, perhaps, for the anomalous presence of an *Antp*-like homeobox in this gene. The vertebrates have not utilized their metameric origins in the same way to generate specialized structures. Instead they have tended to suppress overt signs of primary segmentation in the generation of limbs and internal organs. In mammals, the coupling between regional identity and segmentation is not always precise; variation in the number of homologs within each region is not uncommon, and may be found in natural populations (Bateson, 1894). We may therefore expect to see a looser relationship between segmentation and regionalization in mammals than in insects.

No specific vertebrate homologs have yet been found for the Drosophila genes *Ubx* and *abd-A*. There is perhaps good reason for this. These two genes, together with *Antp*, control the difference between the major trunk regions of the insect, the thorax and abdomen. This distinction is believed to have arisen within the arthropod lineage. Early arthropodan ancestors of the insects already possessed specialized anterior and posterior structures, but the trunk segments are generally assumed to have been very similar to each other. We would therefore expect that the genetic mechanisms controlling the distinctions between these trunk segments would have no specific homologs in vertebrates. The origin of the genes *Ubx* and *abd-A* may have been associated with the evolving specialization of segments in the trunk. Even so, it is surprising that these are the only two homeotic genes to have arisen since the divergence of the arthropod and the vertebrate lineages. This limits the extent to which the duplication and divergence of homeotic genes can be used to account for insect evolution; we must look elsewhere for the molecular changes that account for the increasing complexity of segment morphology. A good place to search may be the unusually large regulatory regions associated with many of the homeotic genes.

In Drosophila, and even more obviously in vertebrates, the complex patterns of homeobox gene expression in later development suggest that the HOM-C/*Hox-C* genes are utilized in many developmental processes other than axial specification. In view of the radically different organization of insect and vertebrate organ systems—an organi-

zation that must largely have been evolved after these lineages diverged—it seems likely that these later roles will prove to be analogous rather than homologous, associated with the evolution of new regulatory networks. It is perhaps in relation to this requirement for the elaboration of new control mechanisms that vertebrates now contain multiple copies of very similar protein coding sequences.

References

Akam, M. (1987). Development *101*, 1–22.

Baron, A., Featherstone, M. S., Hill, R. E., Hall, A., Galliot, B., and Duboule, D. (1987). EMBO J. *6*, 2977–2986.

Bateson, W. (1894). Materials for the Study of Variation (London: Macmillan).

Beeman, R. (1987). Nature *327*, 247–249.

Breier, G., Dressler, G. R., and Gruss, P. (1988). EMBO J. *7*, 1329–1336.

Boncinelli, E., Somma, R., Acampora, D., Pannese, M., D'Esposito, M., and Simeone, A. (1988). Hum. Reprod. *3*, 880–886.

Diederich, R. J., Merrill, V. K. L., Pulz, M. A., and Kaufman, T. C. (1989). Genes Dev. *3*, 399–414.

Duboule, D., Baron, A., Mahl, P., and Galliot, B. (1986). EMBO J. *5*, 1973–1980.

Duboule, D., and Dolle, P. (1989). EMBO J. *8*(5), in press.

Duprey, P., Choudhury, K., Dressler, G. R., Balling, R., Simon, D., Guenet, J. L., and Gruss, P. (1988). Genes Dev. *2*, 1647–1654.

Gaunt, S. J., Sharpe, P. T., and Duboule, D. (1988). Development (suppl.) *104*, 169–179.

Gehring, W. (1987). Science *236*, 1245–1252.

Graham, A., Papalopulu, N., Lorimer, J., McVey, J. H., Tudenham, E. G. D., and Krumlauf, R. (1988). Genes Dev. *2*, 1424–1438.

Graham, A., Papalopulu, N., and Krumlauf, R. (1989). Cell *57*, 367–378.

Ingham, P., and Martinez-Arias, A. (1986). Nature *324*, 592–597.

MacDonald, P. M., and Struhl, G. (1986). Nature *324*, 537–545.

Otting, G., Qian, Y., Muller, M., Affolter, M., Gehring, W., and Wuthrich, K. (1988). EMBO J. *7*, 4305–4309.

Pulz, M. A., Diederich, R. J., Cribbs, D. L., and Kaufman, T. C. (1988). Genes Dev. *2*, 901–920.

Regulski, M., McGinnis, N., Chadwick, R., and McGinnis, W. (1987). EMBO J. *6*, 767–777.

Scott, M. P., Tamkun, J. W., and Hartzell, G. W. (1989). BBA Rev. Cancer, in press.