Chapter 4

Identifying HOX paralog groups by the PBX-binding region

Morgan, R., In der Rieden, P. M. J., Hooiveld, M. H. W., and Durston, A. J.

Hubrecht Laboratorium, Nederlands Instituut voor Ontwikkelingsbiologie, Uppsalalaan 8, 3584 CT, Utrecht, The Netherlands.

The Hox genes are a family of transcription factors that define specific anteroposterior identities, both in vertebrate and in invertebrate embryos, and that are characterized by a very highly conserved DNA-binding motif known as the homeodomain. In vitro, most Hox proteins recognize the same four-base-pair consensus sequence that is actually repeated many times in the genome. Far greater binding specificity is achieved when Hox proteins bind as heterodimers with PBX proteins (vertebrate homologs of Drosophila extradenticle). PBX and Hox proteins interact at a specific and highly conserved hexapeptide on the surface of the Hox protein. This short sequence of amino acids is necessary for PBX binding and, apart from the homeodomain itself, is the most characteristic feature of Hox genes.

During the evolution of vertebrates, the ancestral cluster of Hox genes was duplicated at least twice; hence, most vertebrates have at least four independent Hox clusters, referred to as a, b, c, etc. Despite some Hox genes in each cluster having become non-functional or even entirely deleted subsequent to the duplication step, the overall genomic structure of each cluster has been conserved in evolution. In general, the descendants of each of the genes in the ancestral Hox cluster have similar expression patterns and some conserved functions. They are described as paralogs (e.g. hoxa1, hoxb1 and hoxd1). Outside of the homeodomain region, the overall sequence identity between members of each paralog group is very low. As a result, the paralog identity of each gene has often been ambiguous.

Two previous studies have addressed this problem by aligning Hox genes based on their hexapeptide sequences rather than their homeodomain (which forms the usual basis for Hox gene alignment comparisons). Their findings revealed that there were, indeed, some amino acids adjacent to the hexapeptide that are conserved only within individual paralog groups. Here we have extended these studies to include all hexapeptide-containing paralog groups from a wide range of species. Interestingly, this reveals that there are several very highly conserved amino acids clustered around the hexapeptide sequence. These amino acids consistently identify Hox genes as belonging to a particular paralog group (Fig. 1).

Why are the amino acids around the hexapeptide sequence so highly conserved between paralogs but not clusters? Members of one paralog bind to a distinct DNA sequence only when bound to PBX at the hexapeptide site;
however, this hexapeptide is highly conserved. Thus, paralog-specific amino acids surrounding this motif might ‘fine tune’ the PBX–HOX binding interaction, resulting in a unique DNA-binding specificity for each paralog group. Indeed, in the case of the Hox4 paralog group at least, there is already some evidence that this might be the case\textsuperscript{15}. In a recent report\textsuperscript{16}, a Drosophila PBX homolog is shown to determine DNA-binding specificity directly via a 21 bp element. This is distinct from the sequence that is bound by the HOX homeodomain. Hence, the function of PBX might not be limited to simply modifying the binding specificity of HOX proteins. We hope that future studies on the requirement for these amino acids close to the hexapeptide might provide clues as to how this ‘fine-tuning’ works.

\begin{thebibliography}{9}
\item Chang, C.-F. et al. (1998) Flex evolutions of the homeodomain amino-terminal arm establishes different DNA-binding specificities across the fly loci. Mol. Cell. Biol. 18, 1158–1166
\end{thebibliography}