Homeobox

A **homeobox** is a DNA sequence, around 180 base pairs long, found within genes that are involved in the regulation of patterns of anatomical development (morphogenesis) in animals, fungi, plants, and numerous single cell eukaryotes. Homeobox genes encode **homeodomain** protein products that are transcription factors sharing a characteristic protein fold structure that binds DNA to regulate expression of target genes. Homeodomain proteins regulate gene expression and cell differentiation during early embryonic development, thus mutations in homeobox genes can cause developmental disorders.

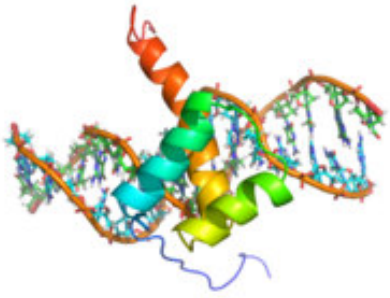
Homeosis is a term coined by William Bateson to describe the outright replacement of a discrete body part with another body part, e.g. antennapedia—replacement of the antenna on the head of a fruit fly with legs. The "homeo-" prefix in the words "homeobox" and "homeodomain" stems from this mutational phenotype, which is frequently observed when these genes are mutated in animals. The homeobox domain was first identified in a number of *Drosophila* homeotic and segmentation proteins, but is now known to be well-conserved in many other animals, including vertebrates.

Homeodomain structure

The characteristic homeodomain protein fold consists of a 60-amino acid long domain composed of three alpha helixes. The following shows the consensus homeodomain (~60 amino acid chain):

```
Helix 1          Helix 2         Helix 3/4

         _______________    __________    ___________________
RRRKRTAYTRYQLLELEKEFHFNRYLTRRRRIELAHSLNLTERHIKIWFQNRRMKWKKEN
....|....|....|....|....|....|....|....|....|....|....|....|
          10        20        30        40        50        60
```



The vnd/NK-2 homeodomain-DNA complex. Helix 3 of the homeodomain binds in the major groove of the DNA and the N-terminal arm binds in the minor groove, in analogy with other homeodomain-DNA complexes.

Helix 2 and helix 3 form a so-called helix-turn-helix (HTH) structure, where the two alpha helices are connected by a short loop region. The N-terminal two helices of the homeodomain are antiparallel and the longer C-terminal helix is roughly perpendicular to the axes established by the first two. It is this third helix that interacts directly with DNA via a number of hydrogen bonds and hydrophobic interactions, as well as indirect interactions via water molecules, which occur between specific side chains and the exposed bases within the major groove of the DNA.

Homeodomain proteins are found in eukaryotes. Through the HTH motif, they share limited sequence similarity and structural similarity to prokaryotic transcription factors, such as

lambda phage proteins that alter the expression of genes in prokaryotes. The HTH motif shows some sequence similarity but a similar structure in a wide range of DNA-binding proteins (e.g., cro and repressor proteins, homeodomain proteins, etc.). One of the principal differences between HTH motifs in these different proteins arises from the stereochemical requirement for glycine in the turn which is needed to avoid steric interference of the beta-carbon with the main chain: for cro and repressor proteins the glycine appears to be mandatory, whereas for many of the homeotic and other DNA-binding proteins the requirement is relaxed.

Sequence specificity

Homeodomains can bind both specifically and nonspecifically to B-DNA with the C-terminal recognition helix aligning in the DNA's major groove and the unstructured peptide "tail" at the N-terminus aligning in the minor groove. The recognition helix and the inter-helix loops are rich in arginine and lysine residues, which form hydrogen bonds to the DNA backbone. Conserved hydrophobic residues in the center of the recognition helix aid in stabilizing the helix packing. Homeodomain proteins show a preference for the DNA sequence 5'-TAAT-3'; sequence-independent binding occurs with significantly lower affinity. The specificity of a single homeodomain protein is usually not enough to recognize specific target gene promoters, making cofactor binding an important mechanism for controlling binding sequence specificity and target gene expression. To achieve higher target specificity, homeodomain proteins form complexes with other transcription factors to recognize the promoter region of a specific target gene.

Biological function

Homeodomain proteins function as transcription factors due to the DNA binding properties of the conserved HTH motif. Homeodomain proteins are considered to be master control genes, meaning that a single protein can regulate expression of many target genes. Homeodomain proteins direct the formation of the body axes and body structures during early embryonic development. Many homeodomain proteins induce cellular differentiation by initiating the cascades of coregulated genes required to produce individual tissues and organs. Other proteins in the family, such as NANOG are involved in maintaining pluripotency and preventing cell differentiation.

Regulation

Hox genes and their associated microRNAs are highly conserved developmental master regulators with tight tissue-specific, spatiotemporal control. These genes are known to be dysregulated in several cancers and are often controlled by DNA methylation. The regulation of Hox genes is highly complex and involves reciprocal interactions, mostly inhibitory. Drosophila is known to use the polycomb and trithorax complexes to maintain the expression of Hox genes after the down-regulation of the pair-rule and gap genes that occurs during larval development. Polycomb-group proteins can silence the HOX genes by modulation of chromatin structure.

Mutations

Mutations to homeobox genes can produce easily visible phenotypic changes in body segment identity, such as the Antennapedia and Bithorax mutant phenotypes in *Drosophila*. Duplication of homeobox genes can produce new body segments, and such duplications are likely to have been important in the evolution of segmented animals.

Types of homeobox genes

1. **Hox genes:** Hox genes are the most commonly known subset of homeobox genes. They are essential metazoan genes that determine the identity of embryonic regions along the anterior-posterior axis. The first vertebrate Hox gene was isolated in *Xenopus* by Edward De Robertis and colleagues in 1984. The main interest in this set of genes stems from their unique behavior and arrangement in the genome. Hox genes are typically found in an organized cluster. The linear order of Hox genes within a cluster is directly correlated to the order they are expressed in both time and space during development. This phenomenon is called colinearity.

2. **LIM genes:** LIM genes encode two 60 amino acid cysteine and histidine-rich LIM domains and a homeodomain. The LIM domains function in protein-protein interactions and can bind zinc molecules. LIM domain proteins are found in both the cytosol and the nucleus. They function in cytoskeletal remodeling, at focal adhesion sites, as scaffolds for protein complexes, and as transcription factors.

3. **Pax genes:** Most Pax genes contain a homeobox and a paired domain that also binds DNA to increase binding specificity, though some Pax genes have lost all or part of the homeobox sequence. Pax genes function in embryo segmentation, nervous system development, generation of the frontal eye fields, skeletal development, and formation of face structures. Pax 6 is a master regulator of eye development, such that the gene is necessary for development of the optic vesicle and subsequent eye structures.

4. **POU genes:** Proteins containing a POU region consist of a homeodomain and a separate, structurally homologous POU domain that contains two helix-turn-helix motifs and also binds DNA. The two domains are linked by a flexible loop that is long enough to stretch around the DNA helix, allowing the two domains to bind on opposite sides of the target DNA, collectively covering an eight-base segment with consensus sequence 5'-ATGCAAAT-3'. The individual domains of POU proteins bind DNA only weakly, but have strong sequence-specific affinity when linked. The POU domain itself has significant structural similarity with repressors expressed in bacteriophages, particularly lambda phage.

Dr. Asheesh Shivam Mishra
Head, Department of Zoology
Nehru Gram Bharati (Deemed to be University), Prayagraj, U.P.