**LARGE SCALE SCREENING FOR TRANSCRIPTION REGULATORY SEQUENCES RECOGNISED BY HOX HOMEODOMAIN PROTEINS.**

JAKT L.M., SHAM M.H.+

Department of Biochemistry, The University of Hong Kong, Sassoon Road, Pokfulam, Hong Kong; e-mail: mhsham@hkucc.hku.hk

+Corresponding author

Keywords: recognition, transcription, regulatory sequences, homeodomain proteins, eukariots, evolutionarily conserved sequence, hoxb3 binding site

We are interested in developing sequence descriptions for sequences which can confer regulation by *Hox* proteins on adjacent promoters, with a view to using the information from proceeding genome projects to predict hox target genes. As a first step in this process we are attempting a large scale screen of at least 2 genomes for fragments which contain specific binding sites for a *Hox* protein (*Hoxb3*), and which are evolutionarily conserved. We are doing this by a an initial selection for binding to *Hoxb*-3 protein, using the in vitro selection scheme of Kinzler and Vogelstein1, followed by a novel application of the whole genome pcr method, to isolate sequences which have been conserved between species. These fragments will be sequenced and analysed using initially software and algorithms freely available, but we are looking to develop or modfy present algorithms to fit the specific nature of the Hox proteins

**1. Background**

Regulation of eukaryotic gene transcription is a complex mechanism involving a hierachy of interactions between *trans*-acting protein factors and *cis*-acting sequence elements, between different protein factors, and modulation of chromatin structure. To understand how genes are regulated scientists have used a wide variety of approaches depending on their interests and objectives, examples include cell transfection and promoter analysis, *in vitro* footprinting and gel-shift analysis, and *in vivo* transgenic mice experiments. The spectrum of laboratory tests are broad, but the experimental outcomes are similar: promoters, enhancers and *cis*-acting regulatory regions of genes and their *trans*-acting factors are identified and characterised. However, compared with the amount of gene coding sequences already available in the databases, information on regulatory regions of genes are still limited. It is expected that strategies for efficient recognition and prediction of genomic regulatory sequences will be in demand.

The *Hox* homeobox genes code for transcription factors which are involved in the specification of regional identity along the anterior-posterior axis of developing vertebrate embryos. The structure and expression of the 39 mammalian *Hox* genes have been well characterized, the *cis*-acting regulatory regions and the biological functions of many of these *Hox* genes have been examined by transgenic mice and gene targeting studies. The clusters of *Hox* genes are highly conserved, and the structures, sequences and regulatory regions of some of the chick, Xenopus and *Fugu* homologues have been described. However, despite the extensive studies of *Hox* genes, it is still unclear what genes are regulated by the Hox proteins. The objective of our work is to experimentally define the DNA sequence environment which can confer regulation by *Hox* proteins on a promoter in *cis* . Having defined the consensus sequence environment, it will be possible to screen the DNA sequences in the various genome databases and identify potential *Hox* target genes. The target sequence information obtained can then be used as the basis for further laboratory testing to verify the regulatory functions of the *Hox* genes.

**2. Methods**

The gene we selected is the mouse *Hoxb3* gene which codes for a polypeptide of 434 amino acids. A partial Hoxb3 cDNA is fused to the GST expression construct to produce a fusion protein containing the homeodomain and N-terminus of *Hoxb3*. We adopted the whole genome PCR approach1 and used mouse genomic DNA cleaved to fragments of 200 - 500 bp as starting material. These fragments were ligated to oligonucleotide linkers before amplification by PCR. The amplified DNA was incubated with the GST-Hoxb3 fusion protein, fragments bound by the fusion protein were immobilised on glutathione Sepharose and amplified by PCR. The amplified DNA fragments were used for another round of binding with the fusion protein. After several rounds of protein-binding enrichment the final PCR products were cloned into plasmid vector to generate a library of clones containing Hoxb3 binding sites.

To isolate hox binding sites present in an evolutionarily conserved sequence context, selected fragments from one species is amplified in the presence of biotin-dUTP, (the bait pool) to incorporate biotin moieties into the DNA. This Biotinilated pool of fragments is then mixed with a pool of fragments selected from a different species (the target pool) and linked to a different linker (such that each pool can be independantly amplified). The mixture of DNA pools are then denatured and annealing is allowed to proceed to a given temperature. DNA duplexes containing biotin are immobilised on streptavidin magnetic beads The DNA bound to the biotin is amplified using primers specific to the target pool, such that fragments from the target pool which are capable of crosshybridising to the bait pool can be selectively amplified. This selection may be repeated as many times as is necessary to give the desired degree of selection.

The sequences from the resulting libraries sites will initially be analysed by existing software available on the internet, and if needed by algorithms designed with the peculiar action of hox proteins in mind.

**3. Results and Discussions**

A library of DNA clones containing potential Hoxb3 binding sites was generated. To assess the quality of the library in terms of specificity and redundancy, the DNA sequences of 41 clones (total of ca. 9000 bp) were determined. Among them there were 37 independent clones, indicating that there was a very low level of repetition even though the enrichment process was PCR based.

The sequences of clones were analysed for the presence of core TAAT motifs, and 2 matrix descriptions generated by MatInd2 from previously published binding sites. The selected sequences contained on average a 5 fold increase in TAAT core motifs over randomly picked nonselected clones. Eight matches to a matrix description based on published Hoxb-3 binding sites in the TTF1 gene3 were found, and 30 matches to a matrix description based on previously detemined Hoxa-3 binding sequences4. No matches (at the same threshold score) to either of these matrixes were found in 6000 bp of sequence of nonselected clones. Furthermore, the sequences of these clones was analysed using the program CoreSearch5, and two separate consensus descriptions which contained core TAAT motifs, typical for Hox proteins were generated.

Though we feel confident that these clones are enriched for Hox protein binding sites, we would only expect a small fraction of them to be physiologically relevant. We are proceeding to select from similarly constructed libraries, sequences which have been conserved between species (as described above). The clones from these experiments will be analysed by functional assays, and the sequences used to attempt a characterisation of the sequence requirements for transregulation by hox proteins.

**References**

1. W.K. Kinzler and B. Vogelstein, “Whole Genome PCR: application to the identification of sequences bound by gene regulatory proteins” Nucleic Acids Res. 17, 3645 (1989)
2. K. Quandt et al, “MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data” Nucleic Acids Res. 23, 4878 (1995)
3. .S. Guazzi et al, “The thyroid transcription factor-1 gene is a candidate target for regulation by Hox proteins” EMBO J. 15, 3339 (1994)
4. K.M. Catron et al, “Nucleotides flanking a conserved TAAT core dictate the DNA binding specificity of three murine homeodomain proteins” Mol. Cell Bio. 14, 4532 (1994)
5. F. Wolfertstetter et al, “Identification of functional elements in unaligned nucleic acid sequences by a novel tuple search algorithm” Comput. Appl. Biosci. 12, 71 (1996)