Czech FOBIA Free & Open BioInformatic Association

presents

Simone Scalabrin

PhD student at
University of Udine
Italy

in collaboration with

Jan Pačes and Zbyněk Kozmik

Institute of Molecular Genetics
Academy of Sciences of the Czech Republic

MODULAR SEARCH OF TRANSCRIPTION FACTOR BINDING SITES IN EUKARYOTES

Thursday, 10th March 2005
3:00 PM
in the small lecture of the
Institute of Molecular Genetics
Flemingovo námestí 2, Praha 6

keywords: phylogenetic footprinting, module/cluster, enhancer, promoter, TFBS

Phylogenetic footprinting is an in vogue technique to discover Transcription Factor Binding Sites, TFBSs, in eukaryotes.

Comparative studies have been very effective at identifying conserved non coding sequences (CNS) that might have regulatory functions; however CNS regions may regulate a broad variety of biological functions, not necessarily confined to transcriptional regulation. For example, CNSs may be involved in the process of DNA replication or mRNA splicing. So one of the main reasons of high false positive rate of present TFBS discovery software is that they are mainly exploiting only the Kimura rule: ''functionally less important molecules or parts of molecules evolve (in term of mutant substitutions) faster than more important ones''. They take into account only the alignment of sequences and the percentage of conservation among them while there is extra information which could be used as the modular structure of TFBSs.

The current view is that once an enhancer or a promoter is accessible it binds to combinations of activators/transcription factors. Binding of proteins is generally cooperative: while one protein binds weakly, multiple activators/transcription factors involved in protein-protein interactions increase their affinities to the regulatory region.

Although a well defined structure for enhancers has not yet been described in detail we can summarize it as follow: 500 to 800 bp long, composed of up to ten different binding sites each of them 5 to 20 bp long, possibly with many mutations. The distance between adjacent binding sites in an enhancer is at least 10 bp (to leave enough space for binding proteins) and at most 200 bp (to allow protein-protein interactions).

Moreover most of current software is based on global alignment while the modular structure of binding sites suggests local alignment to be used. Even in tools which carry out local alignment to discover regulatory regions there is no direct use of the modular structure of TFBSs.