

Computational Analysis of a New Methyltransferase Family and of the Destruction Box Motif

Reichmann, D.¹ and **Pietrokovski, S.**² ¹ Department of Biological Chemistry, Weizmann Institute of Science ² Department of Molecular Genetics, Weizmann Institute of Science

While protein sequences are relatively easy to determine, identification of their function and structure features by experimental methods is an extremely elaborate and complicated task. Here we demonstrate the use of computational sequence analysis methods, in order to identify functional properties of different protein families and function prediction of uknown proteins. This work is based on the identification and comparison of conserved ungapped sequence motifs ("blocks"). The strength of block analysis was tested in two different ways:

1. Identification of conserved motifs in one protein super-family with very weak sequence similarity,

2. Identification of a conserved motif common to different protein families and apparently responsible for one biological process.

Protein family study: Identification of a new methylatransferase super-family

Methylases (or methyltransferases) form a large and diverse group of enzymes that catalyze the transfer of a methyl group from S-adenosyl-L methionine (AdoMet or SAM) to a wide range of substrates. Methylases are found in organisms from all three domains of life (eukaryotes, bacteria and archaea), as well as in viruses and phages. X-ray studies show that methylases from all classes have an alpha/beta type fold with common core structure typical to them. Global sequence similarity between methylases from different families is very weak to undetectable. Using block-to-block comparison methods, thorough sequence analysis of known methylase protein families, we identified a new super-family of proteins from bacteria, archaea and eukaryotes, that appears to be a previously unknown class of methylases. Sequence motifs present in all members of the super-family are most similar to the catalytic and SAM-binding sites of DNA methylases. The significance of this study is in two areas. First, the highly ancient origin of the proteins, and their presence in prokaryotes to mammals, suggests an important role. Second, we demonstrate the strength of blocks analysis for suggesting the function of protein families by identifying subtle sequence similarity to proteins with known function.

Protein motif study: *Defining of the Destruction box (D-box) motif*

Proteolysis by ubiquitin and 26S proteasome pathway is a fundamental mechanism for protein turnover, cell cycle control and signal transudation. The anaphase promoting complex (APC), or cyclosome, is a multi-subunit ubiquitin ligase protein, responsible for ubiquitination of cell regulators at the metaphase - anaphase and mitosis - G1 transitions. D-box, destruction se-





quence motif, is essential for APC mediated protein degradation. A sequence pattern for the Dbox was previously suggested as RxxLxxx(N/Q). We characterized more accurately the D-box in various known APC subtracts so it can be used for searching for new putative D-box families. D-boxes are typically distinct between different protein families. A D-box block from one family does not always identify the D-boxes from other families at a significant level, making the identification of a new putative D-boxes a difficult task. However, we succeeded in identifying the D-box in the HspII protein that was experimentally verified. We constructed a web server to identify putative D-boxes:

http://bioinfo.weizmann.ac.il/~danag/d-box/ main.html

D-box identification by this server is recommended for known APC substrates or proteins already suspected as targets of D-box dependent degradation. Here we describe a computational approach to identify proteins involved in interactions where one protein, or protein complex, binds to many other proteins (one to many). The strategy of this study can be applied to other studies where one-to-many protein interactions are involved e.g., protein phosphorylation, specific protein cleavage, and cell cycle control.