MotiFinder: Improvement of the Search for Protein Functional Sites using Phylogenetic and Physicochemical Information

Wollman, R. and Ben-Tal, N.

Department of Biochemistry, George S. Wise Faculty of Life Sciences, Tel Aviv University

Functionality assignment to proteins is one of the main goals in molecular biology. The classical way to accomplish this involves expansive and time-consuming mutagenesis studies to determine the residues comprising the functional site(s). Accumulative experimental data have been documented in databases such as PROSITE, which are commonly used to suggest putative functional sites in proteins of unknown function. To this end, stretches of residues comprising a functional (e.g., amidation) site in related proteins were aligned and a signature, corresponding to the functional residues, was derived for the site. Signature derivation is error prone. For example, the signature of a particular functional site reflects the currently documented proteins having this site, and a search using the signature might miss a true functional site even if it is only marginally different from the documented signature.

We describe here a novel method for the identification of signature-like putative functional sites in proteins. The new method, implemented in the MotiFinder program, uses current signature definitions but performs a more permissive search. Each putative signature is assigned a score that reflects its physicochemical similarity to the original signature using an amino acid replacement matrix (1). Another score is assigned to the putative signature based on its evolutionary conservation within homologous proteins (2). These two scores are used to estimate the statistical significance of the putative signature. Currently the program only uses signatures that are defined as PROSITE patterns but its extension to other signatures, e.g., Hidden Markov Model-based signatures, is straightforward.

The method is demonstrated on the Apo-Dethiobiotin Synthetase from *E. coli*; MotiFinder identified a known ATP/GTP binding site (3), which was overlooked by PROSITE, BLOCKS and PRINTS scans.

References

(1) Miyata, T., S. Miyazawa, and T. Yashunaga (1979). Two types of amino acid substitutions in protein evolution. *J. Mol. Evol.* **12**, 219-236.

(2) Pupko, T., R.E. Bell, I. Mayrose, F. Glaser and N. Ben-Tal (2002). Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics* (in press).

(3) Yang G, Sandalova T, Lohman K, Lindqvist Y, Rendina AR (1997). Active site mutants of E. coli dethiobiotin synthetase: effect of mutations on enzyme catalytic and structural properties. *Biochemistry* **36**, 4751-4760.