

Predicting Catalytic Residues from Sequence and Structural Information

Elisa Cilia⁽¹⁾, Andrea Passerini⁽²⁾, Mauro Brunato⁽¹⁾

(1) Dipartimento di Ingegneria e Scienza dell'Informazione, University of Trento
via Sommarive 14 - I38100 - Trento, Italy, {cilia,brunato}@disi.unitn.it

(2) Dipartimento di Sistemi e Informatica, Università degli Studi di Firenze
via di Santa Marta 3 - I50139 - Firenze, Italy, passerini@dsi.unifi.it

Keywords: catalytic residue prediction, conservation profiles, support vector machine, protein active site, protein structure.

Abstract. One of the aims of modern bioinformatics is to discover the molecular mechanisms which rule the protein operation. This is a fundamental step in understanding the complex processes involved in living systems and would possibly allow us to correct dysfunctions. A protein may have different functions. Such functions are determined by the protein *primary structure*, i.e. the sequence of amino acids that constitute it, and by their spatial disposition (*tertiary structure*). Protein function identification is a challenging problem as it involves the combination of a large number of variables, most of which still unknown. Therefore approaches for the automatic detection of protein functional sites are needed. In this work we concentrate on the prediction of functional residues, i.e. residues which directly interact with the substrate. In its simpler formulation, the problem can be cast into a binary classification task at the residue level. Preliminary experiments showed that evolutionary enriched sequence based information alone achieve performance which are statistically indistinguishable from carefully crafted features extracted from 3D coordinates [1]. While allowing to apply functional residue prediction to the much wider range of sequenced proteins with possibly unknown 3D structure, such results indicate that using structural information in the automatic prediction of protein functional aspects is a non-trivial task. We show that by modeling physico-chemical properties of the residue structural neighbourhood we obtain significant improvements, but further research is needed in order to fully exploit the information provided by protein 3D structure.

References

- [1] N.V. Petrova and C.H. Wu, "Prediction of catalytic residues using support vector machine with selected protein sequence and structural properties". *BMC Bioinformatics*, vol. 7, pp. 312-324, 2006.