

**1st ANNUAL MEETING OF PORTUGUESE PROTEOMIC NETWORK- PROCURA,
“Functional Genomics and Proteomics”**

ReBIL : Relating Biological Information through Literature

ReBIL aims to improve the efficiency of information extraction systems applied to biological literature, using the correlation between structural and functional classifications of gene products. The dogma of molecular biology stating, that sequences should be correlated with their biological activity, supports our approach. We developed a new method that evaluates extracted information by checking if gene products from a common family match a common set of biological properties. To evaluate the method, we developed an information extraction system that automatically annotates carbohydrate-active enzymes issued from CAZy (which public interface is available at <http://afmb.cnrs-mrs.fr/CAZY>) with functional properties extracted from literature. CAZy attributes each carbohydrate-active enzyme to one or more families of catalytic and carbohydrate-binding modules according to its modular structure. A biological ontology (GO) structures the functional properties as a graph.

So far, we measured a correlation between the modular structures and functional properties of annotations automatically extracted from literature. Based on this correlation, our method yielded a significant increase in precision of the extracted annotations. This result demonstrates that our method is a viable approach for automatic validation of extracted biological information.

To compute the relatedness between functional properties, we implemented a semantic similarity measure in GO. It has also been used to compute the functional similarity between gene-products. The measure is available at <http://xldb.fc.ul.pt/rebil/ssm/>, where it is possible to obtain the similarity between GO terms and SWISS-PROT/TrEMBL proteins.

More information about Rebil project is available at <http://xldb.fc.ul.pt/rebil/>.