

## Poster H-87

### CaPSuLo, an integrative web tool for predicting the subcellular location of proteins



#### Authors:

Perret (*Swiss Institute of Bioinformatics*)

Bairoch (*Swiss Institute of Bioinformatics*)

Veuthey (*Swiss Institute of Bioinformatics*)

**Short Abstract:** CaPSuLo provides a common access to a set of selected subcellular location prediction programs in an integrative tool, preventing an easy interpretation of inconsistent results. It displays results together with information on sequence homology and topology to help users take the most reliable prediction (<http://biosapiens.isb-sib.ch/capsulo>).

#### Long Abstract:

Knowing the subcellular location of a protein is a crucial step in the discovery of its function. The huge increase in available biological data requires the development of prediction tools to better functionally characterize genomes and proteomes. For instance, knowing how a protein is targeted to a subcellular compartment can lead to further investigations regarding its function. As a consequence, scientists have taken advantage of the implementation of experimental data in sequence databases and have developed many subcellular location prediction programs. These programs can be categorized into two groups according to the protein features they use to perform the prediction, namely (i) the composition and physicochemical properties of amino acids from part of or the entire protein sequence, or (ii) the detection of a targeting peptide. They also diverge depending on the methods used to perform the prediction. These differences along with the increasing number of tools can be confusing for users; although it is generally admitted that relying on several predictors improves confidence, inconsistent results may make the decision on the correct prediction difficult. These observations lead us to build CaPSuLo, a web tool, which includes several subcellular location prediction programs, as well as additional information such as protein topology prediction and BLAST homology search (<http://biosapiens.isb-sib.ch/capsulo>).

CaPSuLo consists of three different parts:

(i) The query interface allows the user to enter either a protein sequence, or a UniProtKB identifier or accession number. Additional information on the taxonomic group to which the sequence belongs is also required, as a number of programs have only been trained on sequences from a specific range of organism. The user can select the predictors to run on the sequence from a table, with optional parameters for some programs, a link to the original website of each predictor, and the corresponding PubMed citation. The range of subcellular locations that each program is able to predict coupled with the taxonomic specificity of these predictions, thanks to a documented color code, completes the table. In addition, the protein topology can also be predicted using TMHMM.

(ii) A set of CGI scripts calls the user-selected programs, via the network, with the parameters required. In addition, we run a BLAST automatically on the query sequence against UniProtKB/Swiss-Prot.

(iii) CaPSuLo parses the retrieved results of each selected program and displays them

together with their respective confidence measurements. When the user calls the programs with a UniProtKB identifier, CaPSuLo displays the annotation corresponding to the protein's subcellular location if it is available in the UniProtKB entry. Moreover, a link to the complete entry annotation is provided. The best-scored BLAST results (above 70% identity) can be screened, together with their full UniProtKB description and the corresponding subcellular location annotation, when available. Additional predicted features, e.g. target peptide cleavage sites and transmembrane regions, can be seen directly on the sequence.

Predicting the subcellular location of a protein is a step towards its functional characterization, from where further research work can be carried out. We hope CaPSuLo will help experimentalists in such a task by providing an easy way to compare and analyze the results of up to ten different subcellular location prediction programs. We also plan to improve the UniProtKB annotation by providing the database curators with a decision support system for protein subcellular location prediction.