

Poster H-25

ProtComp: Hybrid approach of predicting sub-cellular localization of animal, plants and bacterial proteins.



Authors:

Victor Solovyev (*Royal Holloway, University of London, Egham, Surrey, TW20 0EX, UK*)

Alexander Bachinsky (*Softberry Inc., 116 Radio Circle, Suite 400, Mount Kisco, NY, 10549, USA*)

Short Abstract: We developed a group of programs, PortComp, for predicting protein sub-cellular localization. Three taxon-specific programs: for plant, animal, and bacterial proteins, use common general architecture. The test results show that ProtComp significantly outperforms other programs for predicting sub-cellular localization in both accuracy and number of compartments covered.

Long Abstract:

We developed a group of programs, named PortComp, for predicting protein sub-cellular localization. Three taxon-specific programs, one each for plant, animal/fungal, and bacterial proteins, use common general architecture that combines neural networks and/or linear discriminant analysis to integrate different sequence characteristics, such as amino acid composition, predicted functional peptide sequences (signal peptides, signal-anchors, transit peptides of mitochondria and chloroplasts, transmembrane segments, GPI-anchors), certain ProSite motifs, and direct comparison with proteins of known localization from SWISS-PROT+TrEMBL database into a single prediction framework. ProtComp can identify 9 sub-cellular locations for animals/fungi and 8 locations for plants proteins as well as 5 locations for bacterial proteins. For performance tests ~200 sequences, which were absent in any training sets, are extracted from SWISS-PROT+TrEMBL updates. The test results show (Table 1) that ProtComp significantly outperforms other programs for predicting protein sub-cellular localization in both accuracy and number of compartments covered. Table 1. below presents accuracy of sub-cellular localization prediction of animal/fungi proteins for several well known programs.

ProtComp Psort Cello ESLPred SubLoc

Nucleus 70 60 70 60 70

Plasma Membrane 80 70 80 - -

Extracellular 70 50 50 20 50

Cytoplasm 80 70 90 70 70

Mitochondria 70 70 70 40 40

E. reticulum 80 40 10 - -

Peroxisome 80 - 30 - -

Lysosome 70 - 50 - -

Golgi 80 - 20 - -