ProLoc: Prediction of 24 Subcellular Protein Locations

Novik, A., Hazkani-Covo, A.* and Levanon, E.

Compugen Ltd., Tel Aviv, Israel

*Current address: Department of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University

We have developed *ProLoc*, a program that can accurately predict the sub cellular localization of a protein solely from its amino acids sequence. ProLoc predicts, with high accuracy, the localization of a protein among 24 compartments, the cell organelles themselves, and their membranes. In addition, it divides the membrane proteins into three groups: Type I, Type II, and integral membrane proteins.

To achieve high levels of accuracy, several different approaches were applied concomitantly. Among these were distributions of the protein's length according to compartment, amino acid composition, prediction of trans-membranous regions, recognition of unique patterns that tend to be specific to a certain organelle (such as NLS), signal peptide and anchor modeling and using unique domains from Pfam that are specific to a single compartment.

Testing the program on Swissprot non-redundant, well annotated proteins that were not part of the training set, the sub cellular location of a protein was accurately predicted as the first choice among the 24 compartments in 73% of the cases, and as the second choice in 12% of the instances. When the possibilities are narrowed down to only five compartments (the secretory pathway, transmembrane, nuclear, cytoplasmic and mitochondrial), the predictions are better.

*Current address: Department of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv 69978, Israel