MoiraiSP: a novel mitochondrial localization signal predictor

Yoshinori Fukasawa[1,2], Kenichiro Imai[2,3], Szu-Chin Fu[3], Junko Tsuji[1] and Paul Horton[1,3]

[1] Dept. of Comp. Biol., the University of Tokyo, Japan
[2] JSPS Research Fellow, Japan
[3] CBRC, Institute of Advanced Industrial Science and technology, Japan

Background and Motivations

1000-1500 different proteins are estimated to localize in mitochondria, however numerous mitochondrial proteins remain undiscovered. Prediction of mitochondrial targeting signal is an efficient approach when identifying undiscovered mitochondrial proteins. A cleavable N-terminal presequence is the best characterized mitochondrial targeting signal: it is said that about half of known mitochondrial proteins possess the presequence. Mitochondrial proteins with presequence are imported into mitochondria via the translocase and then the presequence is cleaved off by mitochondrial processing protease (MPP) in the matrix. However, the detail mechanisms remain unclear. Moreover the data of experimentally identified presequences was limited. Thus, current predictors cannot produce sufficient performances in presequence and cleavage site prediction. Fortunately, large scale proteomic analyses of presequence were recently reported in yeast and plant. This proteomic data gave us an opportunity to improve the signal prediction.

Proposed Approaches

In this work, we therefore developed a predictor for presequences and their cleavage sites trained on recent proteomic data as well as annotated sequences extracted from SwissProt. We trained an SVM classifier for this task and applied several features to predict presequence such as amino acid composition, physico-chemical properties and import receptor recognition motif. We furthermore performed novel motif search and generated profiles for cleavage sites of MPP. These novel characteristic features were integrated into our predictor as features of presequence.

Results and Conclusions

Our predictor attains better performances than the present predictors: 0.70 of Predotar, 0.65 of TargetP and 0.77 of our predictor in Matthew's correlation coefficient. In addition, we achieved a significant performance improvement in cleavage site prediction. Prediction of cleavage site shows better performance with comparing TargetP: our predictor predicts 71% of canonical cleavage sites and TargetP does about 54% within predicted as presequence containing proteins. The results indicate that, having the advantage of a large training dataset for cleavage site, our predictor makes more accurate predictions than previous methods. Thus, our method is valuable for finding candidates of undiscovered mitochondria proteins and their signal regions.