

Support Vector Machine Approach for Protein Subcellular Localization Prediction

Sujun Hua & Zhirong Sun*

Institute of Bioinformatics,
State Key Laboratory of Biomembrane and Membrane Biotechnology,
Department of Biological Sciences and Biotechnology,
Tsinghua University,
Beijing 100084, China

Abstract

Motivation: *Subcellular localization is a key functional characteristic of proteins. A fully automatic and reliable prediction system for protein subcellular localization is needed, especially for large-scale genome sequences analysis.*

Results: *In this paper, Support Vector Machine has been introduced to predict the subcellular localization of proteins from their amino acid compositions. The total prediction accuracies reach 91.4% for three subcellular locations in prokaryotic organisms and 79.4% for four locations in eukaryotic organisms. Predictions by our approach are robust to errors in the protein N-terminal sequences. This new approach provides superior prediction performance compared with existing algorithms based on amino acid composition and can be a complementary method to other existing methods based on sorting signals.*

Availability: *A web server implementing the prediction method is available at <http://www.bioinfo.tsinghua.edu.cn/SubLoc/>.*

Contact: *sunzhr@mail.tsinghua.edu.cn, huasj00@mails.tsinghua.edu.cn*

Supplementary information: *Supplementary material is available at <http://www.bioinfo.tsinghua.edu.cn/SubLoc/>.*

*To whom correspondence should be addressed.