

Poster I-55

Discrimination of beta-barrel Membrane Proteins based on Statistical Methods and Machine Learning Algorithms



Authors:

M. Michael Gromiha (*CBRC, AIST*)
Makiko Suwa (*CBRC, AIST*)

Short Abstract: We have developed statistical methods and machine learning techniques for discriminating β -barrel membrane proteins from other folding types of globular and membrane proteins using amino acid composition, residue pair preference and motifs. We obtained the accuracy of 95% for correctly identifying the beta-barrel membrane proteins using statistical methods and 99% for excluding globular proteins using machine learning techniques.

Long Abstract:

Discriminating beta-barrel membrane proteins from other folding types of globular and membrane proteins is an important task both for identifying beta-barrel membrane proteins from genomic sequences and for the successful prediction of their secondary and tertiary structures. We have developed statistical methods and machine learning techniques for discriminating beta-barrel membrane proteins using amino acid composition, residue pair preference and motifs. The information about amino acid composition could correctly identify the beta-barrel membrane proteins at an accuracy of 89% and exclude globular and alpha-helical membrane proteins at the accuracy level of 80% (1,2). The residue pair preferences and motifs have more information than amino acid composition and these methods improved the accuracy of more than 95% in detecting beta-barrel membrane proteins (3). On the other hand, we have used support vector machines and neural networks for discriminating beta-barrel membrane proteins. These machine learning techniques improved the overall accuracy to 92%. The sensitivity and specificity are, respectively, 89% and 94%, which indicate that the machine learning techniques excluded the globular and alpha-helical membrane proteins at better accuracy than identifying the beta-barrel membrane proteins (4). From the comparison of statistical methods and machine learning techniques we observed that the statistical methods could identify the beta-barrel membrane proteins at high accuracy while an opposite trend is observed for machine learning techniques, which correctly excluded other folding types of globular and membrane proteins. References: 1. M.M. Gromiha and M. Suwa (2005) *Bioinformatics* 21, 961-968. 2. M.M. Gromiha et al. (2005) *Nucl. Acids Res.* 33, W164-167. 3. M.M. Gromiha et al. (2005) *Comp. Biol. Chem.* 29, 135-142. 4. M.M. Gromiha and M. Suwa (2006) *PROTEINS: Structure Function. Bioinf.* (in press).