

Sequence based prediction of protein mutant stability and discrimination of thermophilic proteins

Gromiha, M. M. ; Huang, Liang-Tsung; Lai, Lien-Fu

Abstract

Prediction of protein stability upon amino acid substitution and discrimination of thermophilic proteins from mesophilic ones are important problems in designing stable proteins. We have developed a classification rule generator using the information about wild-type, mutant, three neighboring residues and experimentally observed stability data. Utilizing the rules, we have developed a method based on decision tree for discriminating the stabilizing and destabilizing mutants and predicting protein stability changes upon single point mutations, which showed an accuracy of 82% and a correlation of 0.70, respectively. In addition, we have systematically analyzed the characteristic features of amino acid residues in 3075 mesophilic and 1609 thermophilic proteins belonging to 9 and 15 families, respectively, and developed methods for discriminating them. The method based on neural network could discriminate them at the 5-fold cross-validation accuracy of 89% in a dataset of 4684 proteins and 91% in a test set of 707 proteins.

Key words: Protein stability; Rule generator; Discrimination; Prediction; Thermophilic proteins; Neural network; Machine learning techniques