生物物理学报(2002) 18(4), 467-473

PDF文档

# 用BP神经网络基于氨基酸特性预测非同源蛋白质二级结构含量

秦红珊[1]、杨新岐[2]
1 天津大学理学院物理系
2 天津大学材料科学与工程学院

依据蛋白质氨基酸特性，以氨基酸组成和有偏自协方差函数为特征矢量，用BP神经网络提出了一种预测非同源蛋白质中α螺旋和β折叠二级结构含量的计算方法。采用相互独立的非同源蛋白质数据库对该方法进行了检验。用Ponnuswamy值时，对二级结构α螺旋和β折叠含量的预测结果是：自检验平均绝对误差分别为0.069和0.065，相应标准偏差分别为0.044和0.047；他检验平均绝对误差分别为0.077和0.070，相应标准偏差分别为0.051和0.049。与仅以氨基酸组成为特征矢量的BP神经网络方法比较，相应的他检验平均绝对误差分别减小了0.024和0.016，标准偏差分别减小了0.031和0.018；与改进的多元线性回归方法比较，相应的他检验平均绝对误差分别减小了0.018和0.011，准偏差分别减小了0.020和0.012。表明：基于氨基酸组成和有偏自协方差函数为特征矢量的BP神经网络预测蛋白质二级结构含量的方法可有效提高预测精度。

# PREDICTION OF THE CONTENT OF α-HELIX AND β-STRAND OF NON-HOMOLOGOUS GLOBULAR PROTEINS BASED ON BP NEURAL NETWORK

The amino acid composition and the biased auto-covariance function are considered as features and multiple layer propagation artificial neural network of BP algorithm is used to synthesize these features. A new method to predict the content of a-helix and b-strand of globular proteins is presented. The prediction accuracy of this method is verified by using the independent non-homologous protein database. We have found that the amino-acid index proposed by Ponnuswamy leads to the optimal predictive result in the case for the database sets studied in this paper. It is shown that the average absolute errors for resubstitution test are 0.069 and 0.065 with the standard deviations 0.044 and 0.047 for the prediction of the content of α-helix and β-strand respectively. The average absolute errors for cross-validation test are 0.077 and 0.070 with the standard deviations 0.051 and 0.049 for the prediction of the content of α-helix and β-strand respectively. Compared with the BP neural network method only using the amino acid composition as features, the average absolute errors for cross-validation test are relatively reduced by 0.024 and 0.016 with the standard deviations reduced by 0.031 and 0.018, respectively. Compared with the multiple linear regression method proposed by Zhang et al (Protein Eng., 11, 971- 979, 1998), the average absolute errors for cross-validation test are relatively reduced by 0.018 and 0.011 with the standard deviations reduced by 0.020 and 0.012. It is shown that the BP neural network method combined with the amino-acid composition and the biased auto-covariance function features could effectively improve the prediction accuracy.

## 关键词

BP神经网络(BP neural network); 氨基酸组成(Amino-acid composition); 有偏自协方差函数(Biased auto-covariance function); 氨基酸特性(Amino-acid index); 二级结构α和β含量(Content of α-helix and β-strand)