用支持向量机识别 B-发夹模体

胡秀珍[1,2]、李前忠[1]
1　内蒙古大学
2　内蒙古工业大学理学院

　　基于蛋白质序列，提出了一种新的超二级结构模体β-发夹的预测方法。利用离散增量构成的向量来表示序列信息，并将6个离散增量输入支持向量机，在六维向量空间中寻找最优超平面，将β-发夹和非β-发夹进行分类。计算结果表明，利用所设计的算法预测β-发夹，有较高的预测能力。对于训练集，5-交叉检验的预测总精度为81.24%，相关系数为0.57，β-发夹敏感性为83.06%；对于独立的检验集，预测总精度为78.34%，相关系数0.56，β-发夹敏感性为77.24%。将此预测模型应用于CASP6的63个蛋白质进行检验，得到较好结果。

The B-hairpin Motifs Prediction Using Support Vector Machine

　　Based on the protein sequence, a new method for predicting supersecondary structure motif, β-hairpins, is proposed. By using of the composite vector with increment of diversity to express the information of sequence, and input the increment of diversity to support vector machine(SVM), SVM can find the optimization hyper plane in six dimension space to classify the β-hairpins and the non-β-hairpins. The result indicates that the higher predicting accuracy of β-hairpin motifs is obtained by using of our method. For training set 5-fold cross validation, the overall accuracy of prediction, Matthew's correlation coefficient (MCC) and sensitivity for β-hairpins are 81.24%, 0.57 and 83.06%, respectively. For independent testing set, the overall accuracy of prediction, MCC and sensitivity for β-hairpins are 78.34%, 0.56 and 77.24%, respectively. In addition, the performance of the method was also evaluated by predicting the 63 proteins in the CAPS6 dataset. And the better results are obtained by using our method.

关键词