用离散量的方法识别蛋白质的超二级结构

胡秀珍、李前忠
内蒙古大学

　　用离散量的方法，对2208个分辨率在2.5?魡以上的高精度的蛋白质结构中四类超二级结构进行了识别。从蛋白质一级序列出发，以氨基酸（20种氨基酸加一个空位）和其紧邻关联共同为参数，当序列模式固定长取8个氨基酸残基时，对"822"序列模式3交叉检验的平均预测精度达到78.1%，jack-knife检验的平均预测精度达到76.7%；当序列模式固定长取10个氨基酸残基时，对"1041"序列模式3交叉检验的平均预测精度达到83.1%，jack-knife检验的平均预测精度达到79.8%。

The Protein Super-secondary Structure Recognition with the Method of Diversity Measure

　　Four types of super-secondary structures of the 2208 proteins with higher resolution (>2.5?魡) were predicted by using of the diversity increment algorithm. The compositions of amino acids and twin amino acids were chosen as the information parameters of the amino acids sequences. The prediction results show that average prediction accuracies of the "822type" for fixed-length pattern with 8 amino acids were 78.1% and 76.7% by the 3-cross validation test and jack-knife test, respectively. If the fixed-length pattern was taken as 10 amino acids, the average prediction accuracies of the "1041type" were 83.1% and 79.8% by the 3-cross validation test and jack-knife test, respectively.

关键词

　　离散量(Measure of diversity)；离散增量(Increment of diversity)；蛋白质超二级结构(Super-secondary structure of proteins)