生物物理学报(2001) 17(4), 703-712

PDF文档

用离散量预测蛋白质的结构型

吕志清、李前忠
内蒙古大学物理学系理论生物物理研究室

基于蛋白质的结构类型决定于它的二级结构序列的概念,用二级结构序列参数N_{α}, N_{β}, $N_{\beta\alpha\beta}$, $N_{(\beta\alpha\beta)}$构成离散源,并计算离散量$D(X_{\alpha})$,$D(X_{\beta})$,$D(X_{\alpha/\beta})$,$D(X_{\alpha+\beta})$,利用离散增量预测蛋白质的结构类型,它是由这个蛋白质的离散量$D(X_{n})$与四个标准离散$D(X_{\alpha})$,$D(X_{\beta})$,$D(X_{\alpha\beta})$,$D(X_{\alpha+\beta})$之间离散增量的最小值所决定的。预测结果表明,准确率分别达到84.8%(标准集)和83 8%(检验集)。

THE PREDICTION OF THE STRUCTURAL CLASS OF PROTEIN: APPLICATION OF THE MEASURE OF DIVERSITY

According to the concept that structural class of protein is mainly determined by its secondary structure sequence, the structural class of a protein can be predicted by using the increment of diversity between the protein and a set of standard set of proteins. The standard sources of diversity are respectively determined by the secondary structure parameters N_{α}, $N_{\beta\beta\alpha\beta}$, $N_{(\beta\alpha\beta)}$. The four increments of diversity between the measure of diversity $D(X_{\alpha})$, $D(X_{\beta})$, $D(X_{\alpha/\beta})$, $D(X_{\alpha+\beta})$ and a measure of diversity $D(X_{n})$ of a new protein (n) (in the 1998 test set and the 1995 standard set) are respectively calculated. The structural class of a protein is determined by the lowest increments of diversity. The average rates of correct prediction are 84.8% for standard set and 83.8% for test set.

关键词

结构型(Structural class); 框架结构(Framework structure); 二级结构序列(Secondary structure sequence); 离散量(Measure of diversity); 离散增量(Increment of diversity)