应用连续小波变换预测蛋白质的二级结构

邱建丁,梁汝萍,邹小勇,莫金垣

中山大学化学与化学工程学院

摘要 将代码为lgca蛋白质的氨基酸序列映射为疏水值序列，在合适的尺度下，通过连续小波变换法分别对其α螺旋,α螺旋和β折叠之间的连接多肽(即部分规则和无 规则二级结构)进行预测，准确率分别为76．5%和85．7%．从PDBsum数据库中随 机抽取100个蛋白质作为测试对象，其中全α螺旋、全β折叠、α／β以及α+β蛋 白质各25个．在100个蛋白质中共有1618个连接多肽和747个α螺旋.本法预测到的连接多肽共有1536个，其中1308个与实际结构一致，平均预测准确率为85.2％；预 测到的α螺旋有770个，其中581个与实际结构一致，平均预测准确率为75．5％． 结果表明：该法可较好地预测蛋白质的α螺旋、连接多肽，具有极大的发展前景.

关键词 蛋白质 多肽 疏水性 氨基酸 序列分析 $小玻变换

分类号 Q5

Prediction of Protein Secondary Structure by Continuous Wavelet Transform

Qiu Jianding,Liang Ruping,Zou Xiaoyong,Mo Jinyuan

School of Chemistry and Chemical Engineering, Sun Yat-sen (Zhongshan) University

Abstract α-Helices and short peptides connecting α-helices and β-strands can be predicted by using continuous wavelet transform (CWT) under the appropriate dilation after the amino acids of lgca protein are transformed into sequences of hydrophobic values per residue, the prediction accuracy is 76.5% and 85.1% , respectively. We randomly choose 100 proteins, which consist of 25 all-α-helices, 25 β, 25 α+ β and 25α/βproteins from PDBsum database as the test objects, there are 1618 connecting peptides and 747 α-helices. It was found that 1536 connecting peptides can be predicted by CWT and 1308 among them are consistent with the actual structure, the average predicted accuracy is 85.2%. Comparing with the 747 a-helices contained in the 100 proteins, 770 of a-helices can be predicted by this method and 581 of them are accurate, the average predicted accuracy is 75.5% . The result indicates that CWT is an efficient tool to predict the secondary structures of proteins, and has a tremendous development foreground.

Key words PROTEIN POLYPEPTIDE HYDROPHOBILITY AMINO ACID SEQUENCE ANALYSIS wavelet transform

DOI:

通讯作者