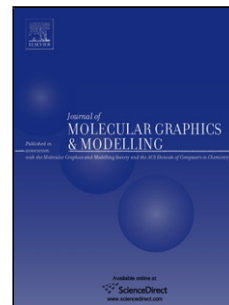


Accepted Manuscript

Title: Predict protein structural class by incorporating two different modes of evolutionary information into Chou's general pseudo amino acid composition

Author: Yunyun Liang Shengli Zhang



PII: S1093-3263(17)30677-0
DOI: <https://doi.org/doi:10.1016/j.jmgs.2017.10.003>
Reference: JMG 7042

To appear in: *Journal of Molecular Graphics and Modelling*

Received date: 30-8-2017
Revised date: 3-10-2017
Accepted date: 3-10-2017

Please cite this article as: Yunyun Liang, Shengli Zhang, Predict protein structural class by incorporating two different modes of evolutionary information into Chou's general pseudo amino acid composition, *Journal of Molecular Graphics and Modelling* (2017), <https://doi.org/10.1016/j.jmgs.2017.10.003>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Predict protein structural class by incorporating two different modes of evolutionary information into Chou's general pseudo amino acid composition

Yunyun Liang^{a,*}; Shengli Zhang^{b,†}

^aSchool of Science, Xi'an Polytechnic University, Xi'an 710048, P. R. China

^bSchool of Mathematics and Statistics, Xidian University, Xi'an 710071, P. R. China

Abstract. Structural class is an important attribute of protein, which plays an important role in both the experiment and theoretical research of protein science. Prediction of protein structural classes has been a challenging task, specifically for low-similarity datasets. In this paper, we develop a feature extraction method PSSS-SOMA-PSSM by incorporating two different modes of evolutionary information into Chou's general pseudo amino acid composition (PseAAC) based on predicted secondary structure sequence (PSSS) and position-specific scoring matrix (PSSM). We construct a 170-dimensional (170D) feature vector for each protein sequence sample, which contains 10D PSSS features that reflect content, alternating word frequency and novel position information, and contains 160D PSSM features that are calculated by second-order moving average (SOMA) algorithm. The SVM classifier with RBF kernel function and the jackknife test are used to predict and evaluate on 1189, 25PDB and 640 datasets with sequence similarity lower than 40%, 25%, and 25%, respectively. Comparison of our results with other methods shows that the proposed method provides the state-of-the-art performance and a cost-effective alternative to structural classes prediction for low-similarity datasets.

Key words: Protein structural classes; Predicted secondary structure sequence; Position information; Position-specific scoring matrix; Second-order moving average; Support vector machine

*Corresponding author. Tel./Fax: +86-29-83116360. E-mail address: yunyunliang88@163.com

†Corresponding author. Tel./Fax: +86-29-88202860. E-mail address: shengli0201@163.com

Download English Version:

<https://daneshyari.com/en/article/4953000>

Download Persian Version:

<https://daneshyari.com/article/4953000>

[Daneshyari.com](https://daneshyari.com)