AFP-PseAAC User Guide

AFP-PseAAC is a novel support vector machines based predictor that uses Chou's pseudo amino acid composition for identifying antifreeze proteins (AFP). It is rapid and modestly efficient over the diverse group of proteins that AFPs are. Individuals in research and industry, interested in AFP related studies and applications are most likely to find this useful.

To use predCF, the user needs the following:

$ *svm_classify <example_file> <model_file> <output_file>*

where,

svm_classify: The support vector machine classifier, SVM^{light}, is available at http://svmlight.joachims.org/.

example_file: Extracted features for query protein(s) in the following format, as per required by SVM^{light}:

> *0 1:0.53 2:0.12 .... #Protein1*
> *0 1:0.13 2:0.10 ..... #Protein2*
> *0 1:0.87 2:0.12 ......#Protein3*

Features may be extracted as mentioned in the following Table 1 using PseAAC web server available at www.csbio.sjtu.edu.cn/bioinf/PseAAC/ or PseAAC-Builder available at http://pseb.sourceforge.net/.

Table 1. Parameters for pseudo amino acid compositions

Type of protein identified	Pseudo amino acid composition features	
	Mode	Lambda (λ), Weight (ω)
AFP	Amphiphilic	10, 0.05

model_file: The *model_file* with optimal pseudo amino acid composition ($\lambda = 10$ and $\omega = 0.05$) as well as support vector machines parameters (C = 25 and $\gamma = 0.0005$) is provided herewith.

output_file: This file has information on the explored prediction. It may be analyzed as follows:

> sign(+): Indicating a positive finding, implies the protein is most likely an antifreeze protein.

> sign(-): Indicating a negative finding, implies the protein is not likely to be an antifreeze protein.

Thus, the prediction may be deemed complete for further use as an aid to experimental biology or industrial applications. Happy computing!

***