PFstats: An Open Tool for Evolutionary Protein Analysis

Néli José da Fonseca Júnior[1], Marcelo Querino Lima Afonso[2], Lucas Bleicher[1]

1 UFMG

Abstract

PFstats is a software developed for the extraction of useful information from protein multiple sequence alignments. By analyzing positional conservation and residue coevolution networks, the software allows the identification of structurally and functionally important amino acid groups and the discovery of probable functional subclasses. Furthermore, it contain tools for the identification of the possible biological significance of these findings. The goal of this project is to provide a computational tool with interactive graphical user interface and data visualization tools to predict global and specific functional amino acid residues and also find functional subclasses in protein families. The software was developed under a client-server architecture. The client was developed in C++/QT and in the server side a java webservice is made to enable the communication between the client and repositories databases of UniprotKb, PFAM and PDB. PFstats includes methods for alignment filtering, residue conservation and coevolution analysis, automatic UniprotKb queries for residue-position annotation, amino acid alphabets reduction and many possible data visualizations. We have studied four protein family domains: lysozyme C/Alpha-lactoalbumin, phospholipases A2, nitrogen regulatory protein PII, and the DNA binding domain of the nuclear receptors IV. In all of them communities of residues related to catalytic and binding sites were found, and also communities related to structural importance, as hydrophobic putative channel and secondary structures, and communties reated to taxonomic specificity. PFstats is free and open source, being distributed in the terms of the GPLV3 licence. The software is available in GUI and terminal versions at http://www.biocomp.icb.ufmg.br/biocomp/software-and-databases/pfstats/. We provide binaries for Windows and Linux (debian), but also compilation instructions for other systems, in addition to the source code and a manual.