PFMutStats: A new method for describing missense mutations by Annotations, Conservation, Coevolution, Interactions and Structural Feature

Marcelo Querino Lima Afonso, Lucas Bleicher

Universidade Federal de Minas Gerais

Abstract

Predicting the functional effect of missense mutations in proteins is essential for understanding the mechanisms of several diseases, and for the rational design of proteins for various applications. In this work we aimed to develop a highly integrative methodology for analysing the possible effects of a given mutation in the functions exerted by a protein. This work is based on analysing Multiple Sequence Alignments of Protein Domains from the Pfam database. Several mutation descriptors are already implemented in a Web Application such as: amino acid frequency before and after mutation, position conservation, UniProt Protein Families annotations, residue interactions established before and after a mutation, centrality metrics related to the residue interaction network of a protein, residue depth and secondary structure assignment, and Ramachandran Distribution analysis before and after the mutations. Various dynamic plots were developed in the D3.js library in order to illustrate and interpret the descriptive results. For the Ramachandran Distribution analysis a Database of reference structures was assembled. All Pfam Protein Family alignments from the UniProt database were downloaded and parsed for calculating and storing the amino acid frequency and column conservation by the Jensen-Shannon entropy. Residue depth was calculated using the latest theoretical model for maximum accessibility as a reference and assignment to the surface or protein interior was based on calculating two vectors of the amino acids frequencies at either site using different assignment thresholds and looking for maximum divergence between these vectors. Residue Interactions Networks were generated by the RING Software and a given chain rank in graph centrality metrics (Betweenness, Closeness and Degree) were calculated and weighted based on the interactions energies. Other features have been planned and should be implemented in the following year.

Funding: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).