Identifying specificity determinant residues through decomposition of protein families affiliation network

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

1 UFMG

Abstract

Affiliation networks are widely used in the context of social and ecological systems. In the present work, we embrace the state of art in this field in order to apply it in the mapping of amino acid coevolution patterns. The goal of this project consists in, given a multiple sequence alignment, predict patterns of local residue conservation that may be related to some specificity (functional, structural or taxonomic). A bipartite network is modeled from a multiple sequence alignment, in a way that each protein is connected with their respective residues. This network is then projected to a residue monopartite representation and its backbone is extracted in order to remove statistically insignificant edges. Finally, the resulting network is decomposed into communities of residues that are more likely to co-occur. We evaluated seven methods for network sparsification with simulated data. These virtual alignments were randomly generated with functional and secondary evolutionary constraints. Experiments with real data were also performed using the HIUase/Transthyretin family and the G protein-coupled receptor, rhodopsin-like family. The results showed that most of the sparsification methods evaluated could in fact rise coevolution patterns in this type of networks. We detected specificity determinant residues for both subclass of the HIUase/Transthyretin family using either filters or weighting to treat the alignment bias. Several functional subclasses were also identified in the GPCR analysis. The methodology presented here is fast and useful to analyze specificity determinant sites, functional subclasses and local conservation residues. This pipeline can be used with large multiple sequence alignments as the obtained from Pfam. Depending on the method used to extract the backbone, anti-correlations could also be observed. Stereochemical correlations can also be identified by generating multiple networks with different amino acid alphabets.

Funding: Capes