

Topic: Databases and Bioinformatics Tools
PI: DBT05

PREDIP: AN IN SILICO APPROACH FOR AUTOMATED PROTEIN INTERACTION PREDICTION IN GENOMIC AND TRANSCRIPTOMIC SEQUENCE DATABANKS

Alvarez J C¹, Herai R H^{1,2}, Carazolle M F¹, Pereira G G A¹

¹Laboratorio de Genômica e Expressão, Universidade Estadual de Campinas - Unicamp

²Laboratório de Bioinformática Aplicada, Embrapa Informática Agropecuária - EMBRAPA

The fast-growing amount of protein interaction databanks have made possible to analyze deeply several metabolic or signaling pathways in different organisms. The information available in such databanks can be used to predict new protein interactions in transcriptome or proteome sequences, most of times this task is done by sequence homology or by structure similarity. In this way, several *in silico* systems were already proposed, like Peimap, which makes its predictions based on experimentally validated protein-protein interaction databanks of different organisms. A restricting criteria to use such tools is the input data, that only accepts proteomic sequences. Based on the importance of the protein complex prediction, and in the fact that several genomes still do not have a transcriptome or an annotated proteome, we propose an integrative Web-based system that can perform, automatically, protein-protein interaction predictions considering also genomic sequences.

Our prediction approach was organized in two steps. In the first, a system analyses the input dataset and uses FrameDP and Augustus, two free softwares, to predict proteins. In the second step, the predicted proteins are aligned against available databases (DIP, Mint, BioGrid, Bind, HPRD, StringInts, IntAct) of protein-protein interactions using psi-blast software. Finally, the aligned sequences are analyzed by our pipeline, called PredIP, that creates an interactive chart representing all relationships between each protein and its respective interactor. Our *in silico* strategy to predict pairs of protein complexes can be applied in genomic, transcriptomic or proteomic data, is not necessary knowing previously the transcriptome or proteome of a considered organism. PredIP has already been tested to predict protein interactions between proteins of *Moliniophthora pernicioso* fungus genome. Hundreds of possible protein complex have been predicted, and have already been described in the literature as possible.

The PredIP system structure can be incrementally updated with new discovered protein-protein interactions to perform new predictions. Predicted protein complex in *M. pernicioso* has already been proved to exist, and new predicted complex will be experimentally validated by our group, enabling a better understanding of relationship between the fungus proteins. Furthermore, PredIP can be a useful tool to predict interactions between two distinct organisms. As described in specialized literature, PredIP is the first automated pipeline to predict protein complexes from genomic or transcriptomic sequences.