

A simple recipe for the non-expert bioinformaticist for building experimentally-testable hypotheses for proteins with no known homologs

Alexander Zawaira · Youtaro Shibayama

Abstract

The study of the protein-protein interactions (PPIs) of unique ORFs is a strategy for deciphering the biological roles of unique ORFs of interest. For uniform reference, we define unique ORFs as those for which no matching protein is found after PDB-BLAST search with default parameters. The uniqueness of the ORFs generally precludes the straightforward use of structure-based approaches in the design of experiments to explore PPIs. Many open-source bioinformatics tools, from the commonly-used to the relatively esoteric, have been built and validated to perform analyses and/or predictions of sorts on proteins. How can these available tools be combined into a protocol that helps the non-expert bioinformaticist researcher to design experiments to explore the PPIs of their unique ORF? Here we define a pragmatic protocol based on accessibility of software to achieve this and we make it concrete by applying it on two proteins—the ImuB and ImuA' proteins from *Mycobacterium tuberculosis*. The protocol is pragmatic in that decisions are made largely based on the availability of easy-to-use freeware. We define the following basic and user-friendly software pathway to build testable PPI hypotheses for a query protein sequence: PSI-PRED MUSTER metaPPISP ASAVIEW and ConSurf. Where possible, other analytical and/or predictive tools may be included. Our protocol combines the software predictions and analyses with general bioinformatics principles to arrive at consensus, prioritised and testable PPI hypotheses.