A method for in-silico validation of protein interactions

Barbarini N[(1)], Simonelli L[(2)], Azzalin A[(3)], Comincini S[(3)], Bellazzi R[(1)]

[(1)] Department of Computer Science and Systems, University of Pavia, Pavia, Italy
[(2)] Institute for Research in Biomedicine, Bellinzona, Switzerland
[(3)] Department of Genetics and Microbiology, University of Pavia, Pavia, Italy

Motivation

Protein interactions are crucial in most biological processes, therefore several in-silico methods have been recently developed to predict them. These methods follow two different directions to search for interactors: the former is based on sequence similarity and the latter considers the properties of the three dimensional structures of proteins. The aim of this work is to design a knowledge-based tool that can integrate both these approaches. In particular, given a target protein (TP), we select the interactors which really interact, among the ones already observed by experimental techniques but not validated in-vivo, using the knowledge about the other interactions already confirmed in-vivo. The proposed method can be conveniently exploited to select within a list of putative interactors, found by experimental analysis, the ones that have to be further investigated through wet-lab analysis.

Methods

The methodology can be applied to validate interactors only for a TP that satisfies the following requirements: the three-dimensional structure is known; one or more interactors, discovered both in-vivo and in-vitro, are known; one or more structures of complexes including the target protein are known. By a preliminary search of all the known interactors of TP in a database of protein interactions (HPRD, etc…) we get two lists of interactors: in-vivo validated (called confirmed interactors, CINT) and in-vitro discovered that we consider potential interactors (PINT). We consider as CINT also the sequences of the interacting chains of the TP extracted from the list of the structures of complexes including TP, searched in the available databases of protein structures (PDB and PQS) . Every sequence of PINT is globally aligned with each sequence of CINT and the score of the best alignment is saved as Score1 for every PINT. For every CINT with known structure, we evaluate the level of conservation of every aminoacid using the tool ConSurf. Every PINT is globally aligned with the conserved regions of the CINT and the score of the best alignment is Score2. From the known protein complexes, we extract for each pair of interacting proteins the binding sites, i.e. the aminoacids at the interface and their neighbours by looking at the intermolecular geometry and calculating all kinds of possible interactions: disulfur bridges, hydrogen bonds and salt bridges. Then we build from these binding sites a set of interacting motifs; such interacting motif takes into account both the sequence (considering the hydropathy and charge of the aminoacids) and the secondary structure of each proteins. To every motifs is assigned a score dependent on its composition and length. We search for both the sequence motifs and the structural motifs in all the PINT list members. Secondary structure prediction for PINT is performed by three of the most used tools (PREDATOR, NNPREDICT and NPS). Therefore we may compute two other scores for every PINT: the first one depends on the motif with best score found in that PINT (Score3); the second one looks at the subset of the motifs extracted from a CINT belonging to an interaction site which is entirely contained in a possible interactor (Score4), assuming that a single motif is not enough to create an interaction. The four scores are normalized and their sum is considered as final score for every PINT, expressing the likelihood that a possible interactor is a real interactor of the TP.

Results

We applied our method GRB2 protein. By means of an in-silico validation procedure based on known GBR2 interactors, we showed that our method has a precision of 85.71%. We also defined a suitable score-threshold to be applied to select unconfirmed interactors. We then retrieved 46 not already confirmed possible interactors and we computed the probability of each possible interactor to be a real interactor; by applying the previously defined score-threshold we identified 21 of the 46 interactors as the most probable ones. Between them, the molecule with the highest probability of interaction with GRB2 is MAPK14. In particular the global alignments between MAPK14 and both the entire sequence and the conserved aminoacids of ERK2 (an interactor confirmed in-vivo) have the best scores. While the best motif found for MAPK14 is PPP[IVL]; this motif is also an interacting site, because in the complex from which it was in-silico extracted (1GBQ), it allows the SH3 domain to bind the SOS-1 molecule.

Contact : nicola.barbarini@unipv.it