BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 22 2012, pages 2998–2999
doi:10.1093/bioinformatics/bts539

Structural bioinformatics

Advance Access publication September 6, 2012

SAPIN: A framework for the structural analysis of protein interaction networks

Jae-Seong Yang[1], Anne Campagna[1], Javier Delgado[1], Peter Vanhee[1], Luis Serrano[1,2,*] and Christina Kiel[1,*]

[1]EMBL/CRG, Design of Biological Systems, Systems Biology Research Unit, Centre for Genomic Regulation (CRG), UPF, 08003 Barcelona, Spain and [2]ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain

Associate Editor: Burkhard Rost

ABSTRACT

Summary: Protein interaction networks are widely used to depict the relationships between proteins. These networks often lack the information on physical binary interactions, and they do not inform whether there is incompatibility of structure between binding partners. Here, we introduce SAPIN, a framework dedicated to the structural analysis of protein interaction networks. SAPIN first identifies the protein parts that could be involved in the interaction and provides template structures. Next, SAPIN performs structural superimpositions to identify compatible and mutually exclusive interactions. Finally, the results are displayed using Cytoscape Web.

Availability: The SAPIN server is available at http://sapin.crg.es.

Contact: jae-seong.yang@crg.eu or christina.kiel@crg.eu

Supplementary information: Supplementary data are available at *Bioinformatics* Online.

Received on December 8, 2011; revised on August 8, 2012; accepted on August 27, 2012

1 INTRODUCTION

Protein–protein interaction (PPI) networks are used to describe cellular functions in normal and disease perturbed states (Vidal *et al*., 2011). However, they often lack the information on physical binary interactions, e.g. if non-binary interaction methods, such as co-immunoprecipitation (co-IP) or tandem affinity purification (TAP)-tag, were used (reviewed in Drewes and Bouwmeester, 2003). At the moment, only statistical approximations, such as the socio-affinity index, are used to decide which of the putative interactions detected in a co-IP or TAP-tag experiments are direct (Gavin *et al*., 2006). Further, PPI networks do not always inform whether several proteins binding to the same protein 'hub' are compatible ('AND') or mutually exclusive ('XOR') structurally (Kim *et al*., 2006). Integration of structural information with PPI networks can partially solve these limitations: first, information from 3D structures of protein/domain complexes can be used to distinguish 'AND' from 'XOR' binding sites. Adding 'AND'/'XOR' structural information into protein interaction networks is important to identify competing interactions where changes in concentration of one component could result in different pathways being active (Kiel *et al*., 2011). Also, it can help understanding different functional effects of mutations taking place in the same protein but at different interaction surfaces (Kar *et al*., 2009; Wang *et al*., 2012). Limitations in the availability of structures of protein complexes to decide whether interactions are of 'AND' or 'XOR' type can be partly circumvented by using structures of homologous proteins, as homologous pairs of binding proteins tend to have a similar interaction topology (Aloy *et al*., 2003).

In part, the structural knowledge in the aforementioned structural interaction network (SIN) was increased by compiling the information of homologous sequences, as homologous pairs of binding proteins tend to have a similar interaction topology (Aloy *et al*., 2003). Second, we propose that 3D information of homologous structures, when combined with experimental information, can be extremely valuable for predicting physical binary interactions among proteins that were found together in a complex in a co-immunoprecipitation experiment.

Here, we introduce SAPIN, a framework dedicated to the Structural Analysis of Protein Interaction Networks. It encompasses features allowing (i) a full analysis of the protein sequence for the identification of the parts potentially involved in an interaction; (ii) a mapping of the available structural data involving the previously identified parts; and (iii) the identification of compatible and mutually exclusive interactions at the network level.

2 METHODS AND IMPLEMENTATION

SAPIN has been implemented using Python programming language and takes as input an interaction file and the related protein sequences provided as a FASTA file (Fig. 1).

2.1 Sequence analysis

SAPIN uses various applications to predict the protein parts that could mediate the interaction with its binding partners: (i) the domain composition using HMM collection from PFAM (Finn *et al*., 2010); (ii) the possible phosphorylation sites and motifs derived from experimental data (Phospho.ELM; Dinkel *et al*., 2011) or from prediction methods [Scansite (Obenauer *et al*., 2003) and NetPhorest (Miller *et al*., 2008)]; (iii) the disordered regions (Disopred; Ward *et al*., 2004); and (iv) secondary structure element composition (Jones, 1999).

2.2 Search for structural templates of domain–domain interactions

SAPIN searches for potential structural templates in the 3DID database (Stein *et al*., 2009) (filtered for crystal packing artifacts, see Supplementary Material) that could model the interaction between two proteins. The resulting matches are evaluated

*To whom correspondence should be addressed.

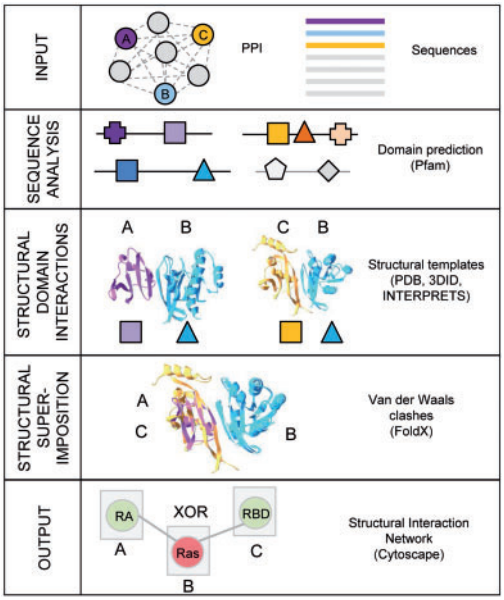

Fig. 1. The input of SAPIN is a txt-file of interactions and protein sequences. Sequences are used to predict domains, and possible structural templates are searched for. If a protein has at least two interacting partners, the domains mediating the interaction are superimposed on the reference domain, and the interacting domains are analysed for compatibility, and the results are displayed in the final structural interaction network

using InterPreTS (Aloy and Russell, 2003) to select the most relevant domain interaction.

2.3 Identification of compatible and mutually exclusive complexes

For two given complexes AB and AC, the related domain–domain interactions structural models are obtained by structural alignment of A over the templates using the CE program (Shindyalov and Bourne, 2001). The interacting domains A and C are then analysed for backbone van der Waals clashes using the empirical force field FoldX (Schymkowitz *et al.*, 2005). A threshold of van der Waals clashes for a compatible ('AND') or mutually exclusive interaction ('XOR') was defined (see Supplementary Material). Benchmarking of structural superimpositions using a set of 1118 solved structures (7288 heterotrimeric domain interactions) showed that 95% of the complexes are correctly evaluated as compatible.

2.4 The SAPIN webserver

SAPIN is accessible through a web portal at http://sapin.crg.es. The portal is built on the open-source Drupal content management system for full flexibility. The pipeline is designed to deal preferably with UNIPROT accession numbers (The UniProt Consortium, 2011). For each pipeline step, a user-friendly interface is provided to visualize the results. SAPIN provides a user-friendly interface to visualize the results. The sequence features for each protein can be browsed by simply clicking on a link to display the detailed results from the different predictions. The structural information (from 3DID, InterPreTS scores and the identified domain–motif interactions) can be downloaded as a table. The structural representation of compatible and mutually exclusive interactions is available in a table and in an interactive network browser (Lopes *et al.*, 2010). The structural information is represented using Cytoscape Web (Lopes *et al.*, 2010): in addition to the nodes representing proteins, compatible ('AND') and mutually exclusive ('XOR') interactions of domains are displayed as green and red dots, respectively.

3 DISCUSSION AND CONCLUSION

SAPIN is a framework that brings together protein interaction networks and structural data, with the objective of reaching a better understanding of how proteins interact with each other. It predicts structural interactions based on sequence data from proteins that have been experimentally co-purified within a complex. It then uses the knowledge of interaction interfaces to identify compatible and exclusive interactions. However, this approach is based on single domain–domain interactions, without taking into account the fold of the full-length proteins.

Our method highlights the principle of competition, which could be important in signal transduction pathways. Further, it could be combined with statistical approaches (e.g. socio-affinity index) to describe the organization of protein complexes more accurately.

Funding: EU (PROSPECTS, grant agreement number HEALTH-F4-2008-201648).

Conflict of Interest: none declared.

REFERENCES

Aloy,P. and Russell,R.B. (2003) InterPreTS: protein interaction prediction through tertiary structure. *Bioinformatics*, **19**, 161–162.
Aloy,P. *et al.* (2003) The relationship between sequence and interaction divergence in proteins. *J. Mol. Biol.*, **332**, 989–998.
Dinkel,H. *et al.* (2011) Phospho.ELM: a database of phosphorylation sites—update 2011. *Nucleic Acids Res.*, **39**, D261–D267.
Drewes,G. and Bouwmeester,T. (2003) Global approaches to protein—protein interactions. *Curr. Opin. Cell Biol.*, **15**, 199–205.
Finn,R.D. *et al.* (2010) The Pfam protein families database. *Nucleic Acids Res.*, **38**, D211–D222.
Gavin,A.-C. *et al.* (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature*, **440**, 631–636.
Jones,D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.
Kar,G. *et al.* (2009) Human cancer protein-protein interaction network: a structural perspective. *PLoS Comput. Biol.*, **5**, e1000601.
Kiel,C. *et al.* (2011) Structural and functional protein network analyses predict novel signaling functions for rhodopsin. *Mol. Syst. Biol.*, **7**, 551.
Kim,P.M. *et al.* (2006) Relating three-dimensional structures to protein networks provides evolutionary insights. *Science*, **314**, 1938–1941.
Lopes,C.T. *et al.* (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.
Miller,M.L. *et al.* (2008) Linear motif atlas for phosphorylation-dependent signaling. *Sci. Signal.*, **1**, ra2.
Obenauer,J.C. *et al.* (2003) Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.*, **31**, 3635–3641.
Schymkowitz,J.W.H. *et al.* (2005) Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proc. Natl. Acad. Sci. USA*, **102**, 10147–10152.
Shindyalov,I.N. and Bourne,P.E. (2001) CE: a resource to compute and review 3-D protein structure alignments. *Nucleic Acid Res.*, **29**, 228–229.
Stein,A. *et al.* (2009) 3did update: domain-domain and peptide-mediated interactions of known 3D structure. *Nucleic Acids Res.*, **37**, D300–D304.
The UniProt Consortium. (2011) Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res.*, **39**, D214–D219.
Vidal,M. *et al.* (2011) Interactome networks and human disease. *Cell*, **144**, 986–998.
Wang,X. *et al.* (2012) Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nat. Biotechnol.*, **30**, 159–164.
Ward,J.J. *et al.* (2004) The DISOPRED server for the prediction of protein disorder. *Bioinformatics*, **20**, 2138–2139.