

InterWeaver: interaction reports for discovering potential protein interaction partners with online evidence

Zhuo Zhang* and See-Kiong Ng

Knowledge Discovery Department, Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613

Received February 6, 2004; Revised March 18, 2004; Accepted April 15, 2004

ABSTRACT

InterWeaver is a web server for discovering potential protein interactions with online evidence automatically extracted from protein interaction databases, literature abstracts, domain fusion events and domain interactions. Given a new protein sequence, the server identifies potential interaction partners using two approaches. In the homology-based approach, the system performs sequence homology searches to find similar proteins in other species, and then searches the protein interaction databases and the biomedical literature for interaction partners. In the domain-based approach, the system detects the domains in the input protein sequence and searches databases of domain fusion events and putative domain interactions to suggest potential interacting partners. The results are compiled into a personalized and downloadable interaction report to aid biologists in their discovery of protein interactions. InterWeaver is freely available for academic users at <http://interweaver.i2r.a-star.edu.sg/>.

INTRODUCTION

A rapidly increasing number of uncharacterized proteins are being generated by large-scale proteomic studies. Understanding the biological roles of these proteins requires knowledge of their interactions with other proteins. Identification of protein interactions is therefore the subject of many post-genome projects; many of the interaction data are available online. Interactions that were experimentally determined en masse using high-throughput methods such as two-hybrid screening have been curated and deposited in online interaction databases (1,2). A large number of interactions reported in journals and conference papers can also be extracted from online biomedical literature databases (3,4). At the same

time, computational methods have also been developed to predict protein interactions. For example, computationally detected domain fusion events (5) as well as computationally derived domain–domain interactions (6) can be used to infer protein interactions.

METHOD

Given a new uncharacterized protein sequence, a biologist can mine the rich online resources of protein interactions to discover its potential interaction partners. We have created InterWeaver, a server providing interaction reports, to help biologists discover potential protein interaction partner proteins using online evidence. Figure 1 depicts the system framework of InterWeaver for generating customized protein interaction reports.

The InterWeaver server currently employs two different approaches to identify potential interaction partners:

- (i) Homology-based approach. Proteins that are known to interact with the source protein's homologs in various selected species are mined from two different data sources: online protein interaction databases and biomedical literature. InterWeaver first performs sequence homology searches using BLAST (7) to find proteins similar to the source proteins in the other species. InterWeaver then searches the online protein interaction databases DIP (2) and BIND (1), as well as the Protein Data Bank (PDB) (8), a database containing data on protein complexes, for experimentally derived protein interactions and complexes to suggest potential protein interaction partners for the source protein. The system also scans the abstracts in the PubMed database for interactions reported in the biomedical literature using text-mining techniques (4).
- (ii) Domain-based approach. Here, proteins with domains that putatively interact with a domain in the source protein are listed as potential interaction partners. InterWeaver uses computationally derived domain fusion events (5) as well as domain–domain interactions (6) for inference.

*To whom correspondence should be addressed. Tel: +65 68748809; Fax: +65 67748056; Email: zzhang@i2r.a-star.edu.sg

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated.

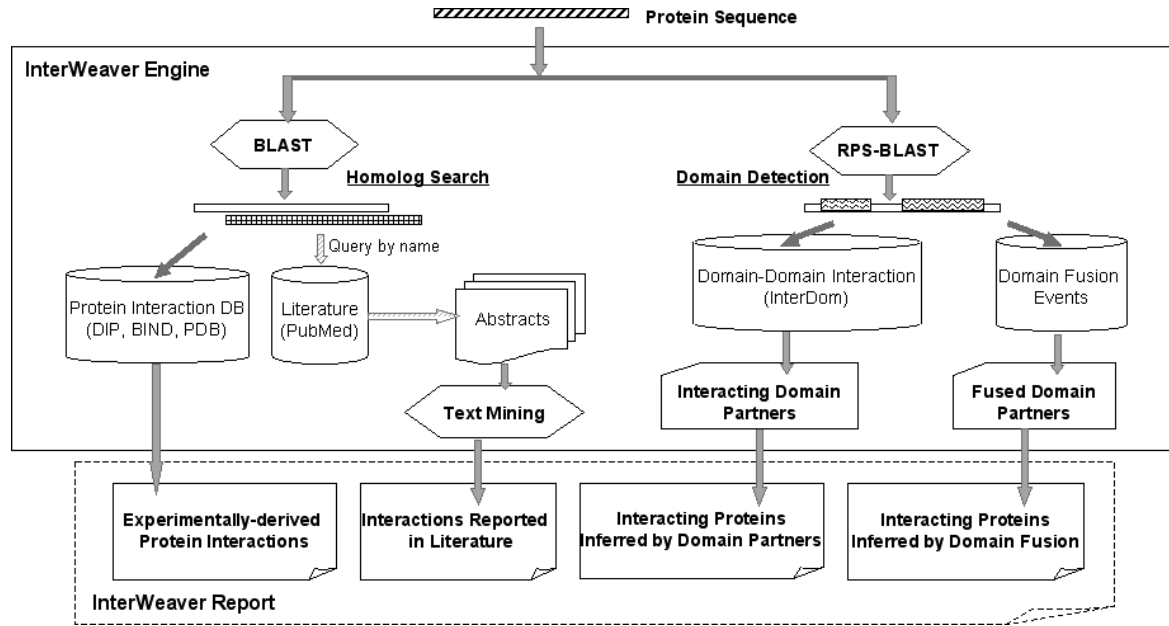


Figure 1. InterWeaver server framework.

The detection of domains in the source protein is done using RPS-Blast (9).

To help biologists in their research, the online evidence for the various potential protein interaction partners is compiled together with cross-reference links to the original databases.

USAGE

InterWeaver provides both online query and offline (batch) query facilities. Offline queries yield interaction reports for

novel proteins. Users submit their protein sequences in FASTA format. They may personalize their reports by specifying the species and *E*-values for BLAST, and by selecting the types of online evidence for inclusion in their reports, namely, protein interaction databases, biomedical literature, domain fusion events and/or domain interactions. When results are available, users receive an email with a password and a link to the compiled InterWeaver report. Users can then browse their reports on the InterWeaver site (each report will be kept on the site for two weeks) or download their

(A) **InterWeaver**

Protein homologs (E-value < 0.01) found for input sequences: 25

Organism	Protein Homologs	Total interacting proteins
Human	10	12
Mouse	3	6
Fruit fly	11	16
C.elegans	1	1

Human			
potential interactors	score	evidences from experiment	homolog value
gi 4507891	2	bind_13089	gi 4504385 (1.2e-21)
		bind_13098	gi 4503577 (8.3e-18)
gi 4502231	1	bind_45545	gi 4504385 (1.2e-21)
gi 1345703	1	dip_10882E	gi 1877215 (2.7e-08)
gi 17738294	1	bind_13115	gi 4504385 (1.2e-21)
gi 105325	1	dip_10884E	gi 1877215 (2.7e-08)
gi 8400738	1	bind_45557	gi 4504385 (1.2e-21)
gi 30795242	1	bind_45549	gi 4504385 (1.2e-21)
gi 72219	1	dip_752E	gi 2134717 (0.0)
gi 544257	1	dip_10881E	gi 1877215 (2.7e-08)
gi 5803046	1	bind_19165	gi 4504385 (1.2e-21)
gi 88553	1	dip_10883E	gi 1877215 (2.7e-08)
gi 12667	1	bind_13093	gi 4504385 (1.2e-21)

Mouse			
potential interactors	score	evidences from experiment	homolog value
gi 423392	1	dip_340E	gi 1363097 (2.8e-05)
gi 544257	1	dip_10823E	gi 1853980 (1.3e-07)
gi 1363097	1	dip_340E	gi 423392 (0.0)
gi 2119731	1	dip_564E	gi 423392 (0.0)
gi 728984	1	dip_10891E	gi 1853980 (1.3e-07)

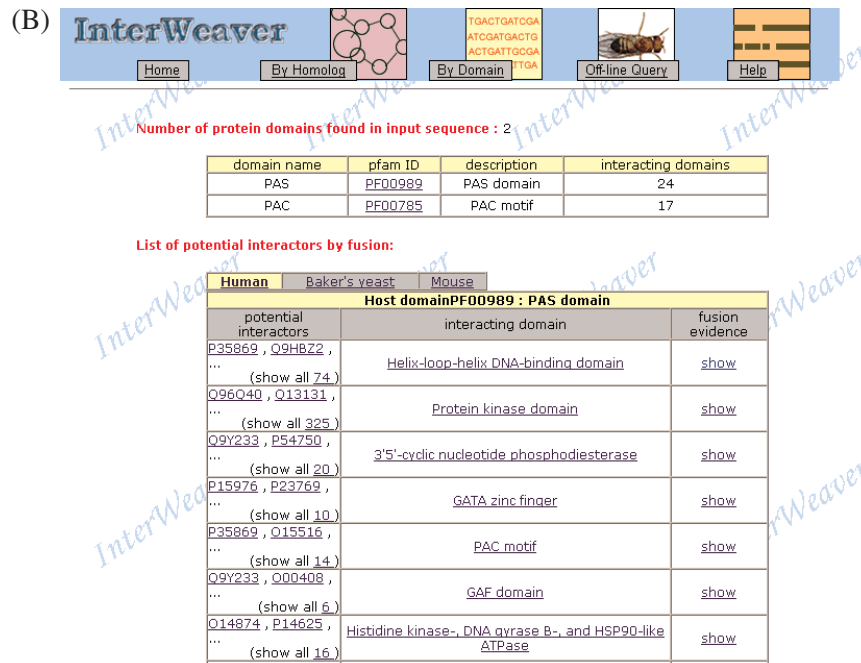


Figure 2. Result pages of the InterWeaver server (A) Searching for interaction partners by homolog, showing evidence from the P–P interaction database. (B) Searching for interaction partners by domain fusion events.

InterWeaver reports in zipped folders for offline browsing. Users can also perform online queries for prompt results. Figure 2 shows the result pages after searching by interaction databases (Figure 2A) and searching by domain fusion events (Figure 2B) respectively.

DISCUSSION

Online databases such as Predictome (10) and STRING (11) contain putative protein–protein interactions pre-computed using various computational methods. Our InterWeaver system is designed as a web server to predict potential protein interacting partners for novel sequences using both homology- and domain-based predictive approaches. The server uses a variety of online resources ranging from experimentally derived protein interaction databases to computationally derived domain interaction databases, and mines data sources from structured databases to unstructured text databases. In fact, InterWeaver is designed to be easily extensible to include other online protein interaction resources and different computational approaches. To help biologists manage the wealth of information at their own pace, InterWeaver generates comprehensive downloadable web reports for offline analysis.

The variety of evidence compiled about the potential interaction partners can be useful in helping biologists validate and annotate the experimental results for their proteins. However, as with other predictive tools, it is important to bear in mind that the potential interaction partners may be predicted by the system based on assumptions that are yet to be conclusively validated. For example, the accuracy of homology-based inference of protein interactions has not yet been proved with conclusive evidence based on significant datasets. However, with suitable prudence and combining evidence

from different approaches and data sources, InterWeaver can serve as a useful hypothesis engine for dissecting the vast interactomes.

REFERENCES

- Bader,G.D., Betel,D. and Hogue,C.W. (2003) BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.*, **31**, 248–250.
- Xenarios,I., Salwinski,L., Duan,X.J., Higney,P., Kim,S.M. and Eisenberg,D. (2002) DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.*, **30**, 303–305.
- Mack,R. and Hehenberger,M. (2002) Text-based knowledge discovery: search and mining of life-sciences documents. *Drug Discov. Today*, **7**, S89–S98.
- Ng,S.K. and Wong,M. (1999) Toward routine automatic pathway discovery from on-line scientific text abstracts. *Genome Inform. Ser. Workshop Genome Inform.*, **10**, 104–112.
- Marcotte,E.M., Pellegrini,M., Ng,H.L., Rice,D.W., Yeates,T.O. and Eisenberg,D. (1999) Detecting protein function and protein–protein interactions from genome sequences. *Science*, **285**, 751–753.
- Ng,S.K., Zhang,Z., Tan,S.H. and Lin,K. (2003) InterDom: a database of putative interacting protein domains for validating predicted protein interactions and complexes. *Nucleic Acids Res.*, **31**, 251–254.
- Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
- Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
- Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Mellor,J.C., Yanai,I., Clodfelter,K.H., Mintseris,J. and DeLisi,C. (2002) Predictome: a database of putative functional links between proteins. *Nucleic Acids Res.*, **30**, 306–309.
- von Mering,C., Huynen,M., Jaeggi,D., Schmidt,S., Bork,P. and Snel,B. (2003) STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.*, **31**, 258–261.