

DASMIweb: online integration, analysis and assessment of distributed protein interaction data

Hagen Blankenburg*, Fidel Ramírez, Joachim Büch and Mario Albrecht

Max Planck Institute for Informatics, Campus E1.4, 66123 Saarbrücken, Germany

Received February 28, 2009; Revised April 25, 2009; Accepted May 11, 2009

ABSTRACT

In recent years, we have witnessed a substantial increase of the amount of available protein interaction data. However, most data are currently not readily accessible to the biologist at a single site, but scattered over multiple online repositories. Therefore, we have developed the DASMIweb server that affords the integration, analysis and qualitative assessment of distributed sources of interaction data in a dynamic fashion. Since DASMIweb allows for querying many different resources of protein and domain interactions simultaneously, it serves as an important starting point for interactome studies and assists the user in finding publicly accessible interaction data with minimal effort. The pool of queried resources is fully configurable and supports the inclusion of own interaction data or confidence scores. In particular, DASMIweb integrates confidence measures like functional similarity scores to assess individual interactions. The retrieved results can be exported in different file formats like MITAB or SIF. DASMIweb is freely available at <http://www.dasmiweb.de>.

INTRODUCTION

Protein interactions play an important role in many cellular processes (1). Different small- and large-scale experimental techniques together with the manual curation of the scientific literature as well as numerous computational prediction methods generate ever increasing amounts of publicly accessible protein interaction data (2). However, this rapid accumulation of data renders it difficult for researchers to keep track of all available information because they are scattered over multiple online repositories. As of April 2009, the pathway resource list Pathguide (3) gives the impressive number of 118 databases providing protein interaction data. Some of these projects are highly specialized and focus, for example, on interactions of molecular subcomponents or specific

classes of proteins, on specific diseases or organisms, or on experimentally observed or computationally predicted interactions. Moreover, doubts have been raised about the quality and reliability of protein interaction data and particular detection methods (2,4,5).

Databases that collect and curate experimentally observed protein–protein interactions reported in the literature (6–13) are essential pillars of interactomics, but they cover only a small fraction of the complete set of interactions, and thus proteome-wide predictions are also required (2,4). All these efforts have resulted in a multitude of resources that the user has to query individually. Initiatives like IMEx (14) that promote data exchange between some of the databases are very important, but are still in an early implementation phase. One of the possible solutions to integrate protein interaction data is the creation of data warehouses as composite databases that centrally store and merge the available data from multiple sources (10,11,15–23). However, the static data unification procedure underlying data warehouses has the considerable drawback of providing only a snapshot of a fixed number of data sources at a certain point of time. Once the data have been included into the central repository, curation efforts are required to keep it up to date and in sync with the original data sources. Furthermore, data warehouses are rather inflexible as the inclusion of additional datasets, for example, new experimental or predicted data or improved confidence scores, can normally be accomplished solely by the central authority and not by the user.

In the context of the European BioSapiens network (24), we have developed DASMIweb as a gateway to interactome data from multiple resources. In contrast to composite databases, data are not stored in a local repository, but queries are distributed to the original data sources and the unified results are displayed (25). Due to this novel realization as a distributed and dynamic system, DASMIweb bypasses the inherent rigidity of static databases and addresses their problem of data update cycles. In addition, DASMIweb allows access to distributed servers with confidence scores, which can be used to evaluate the quality of individual interactions with different scoring methods.

*To whom correspondence should be addressed. Tel: +49 681 9325 328; Fax: +49 681 9325 399; Email: hagen.blankenburg@mpi-inf.mpg.de

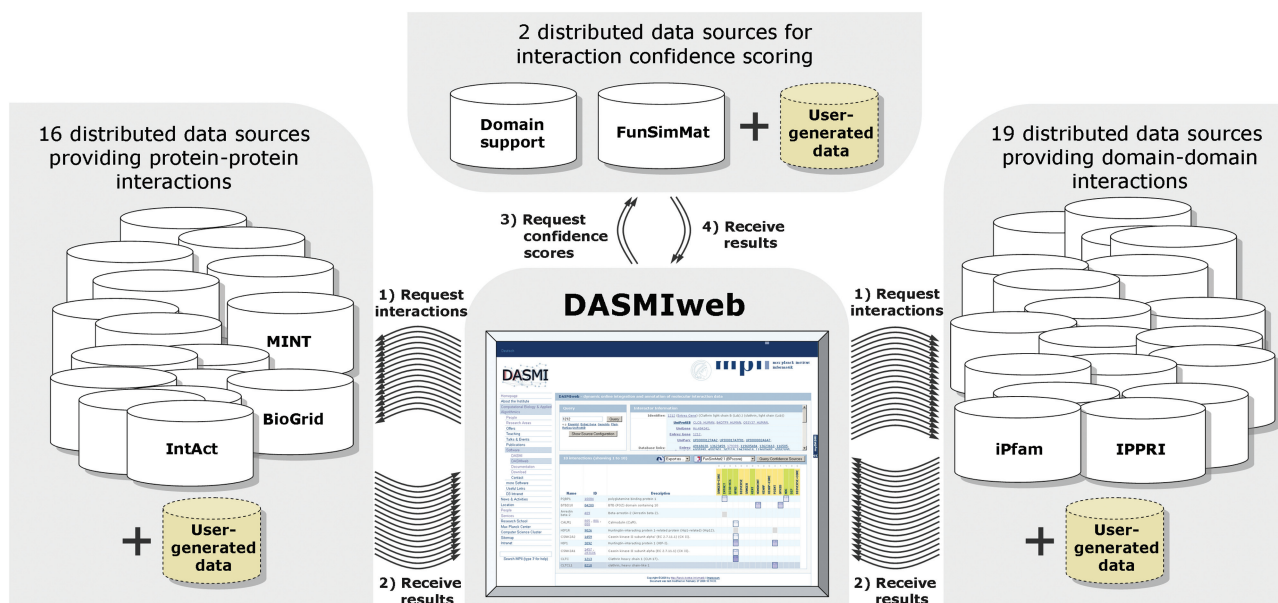


Figure 1. Decentralized architecture of DASMIweb. Data sources for protein and domain interactions as well as for interaction confidence scores are distributed over the Internet and are contacted by DASMIweb upon user request.

MATERIALS AND METHODS

Distributed architecture

The fundamental concept of DASMIweb is decentralization (Figure 1). Here, the interaction data remain distributed with their original providers instead of being periodically aggregated into central data repositories (10,11,15–23). Subsequent to a user request, DASMIweb independently queries each original data provider for interactions, additional annotations, and interaction confidence scores. Then it unifies the retrieved results and presents them to the user.

The technical architecture of DASMIweb, based on an extension of the Distributed Annotation System (DAS) (25,26) and different types of web services (27,28), has the great advantage of being easily extendable with new data sources. In addition, data update cycles every few weeks or months are not necessary because all data is left in its source database and is only retrieved on request. This use of a distributed architecture greatly empowers the end-user who can instantly add data sources, for instance, own private interactions or the results of an improved confidence scoring method. DASMIweb also gives data providers the possibility to easily share their results without the time-consuming development of own web interfaces. Since the distributed architecture supported by DASMIweb is driven by the community that provides the contents, there is no need for central authorities to decide on the available resources. This is exemplarily evident in the case of confidence scoring methods, which can be based on various criteria, for instance, co-expression, co-localization, functional co-annotation, network topology and evolutionary conservation. Apparently, it would be impractical to implement and maintain all these different methods at a single site. Instead, each method

developed by some independent institution can be queried through DASMIweb. It is noteworthy that our decentralization approach has also found the interest of the Proteomics Standards Initiative (PSI) of the Human Proteome Organization (HUPO), which is currently defining standards for distributed interaction data retrieval and interaction confidence scoring (28). We are actively contributing to these projects and all servers developed in this context are accessible via DASMIweb (Table 1).

Data sources

DASMIweb has been developed to support different levels of molecular interactions, for example, interactions of proteins as well as of protein domains. In the following, we will refer to the distributed data servers that provide interactions, additional annotations or interaction confidence scores, as *data sources* (in contrast to our *server* DASMIweb). As of April 2009, DASMIweb provides access to 35 data sources containing experimentally determined and computationally derived protein and domain interaction datasets (Table 1). In addition, there are two data sources for scoring the confidence of protein–protein interactions.

In our current setup, the majority of data sources are temporarily cached and maintained at our institute, even if this appears to be a contradiction to the actual DASMIweb aim of leaving the interaction data with their original providers. At present, this setup is unavoidable for demonstrating the capabilities of our system; otherwise, many more external data source providers would already be needed from the beginning. Nevertheless, several major protein interaction databases like BioGrid (10), IntAct (12) and MINT (6) are not cached as they already support external web service access to their data. Moreover, since most of the cached

Table 1. Interaction datasets currently available as data sources through DASMIweb and corresponding references

Dataset name	Interaction type	Determination method	Server type	Reference
CCSB-HI1	PPI	Large-scale experiment	DASMI	(37)
MDC	PPI	Large-scale experiment	DASMI	(38)
BioGRID	PPI	Literature curation	PSI WS	(10)
DIP	PPI	Literature curation	DASMI	(8)
HPRD	PPI	Literature curation	DASMI	(13)
IntAct	PPI	Literature curation	PSI WS	(12)
MINT	PPI	Literature curation	PSI WS	(6)
MPIDB	PPI	Literature curation	PSI WS	(11)
Bioverse	PPI	Prediction	DASMI	(39)
HiMAP	PPI	Prediction	DASMI	(40)
HiMAP core	PPI	Prediction	DASMI	(40)
HomoMINT	PPI	Prediction	DASMI	(41)
OPHD	PPI	Prediction	DASMI	(42)
POINT	PPI	Prediction	DASMI	(43)
Sanger	PPI	Prediction	DASMI	(44)
Sanger core	PPI	Prediction	DASMI	(44)
3did	DDI	3D structure analysis	DASMI	(45)
iPfam	DDI	3D structure analysis	DASMI	(46)
PiNS	DDI	3D structure analysis	DASMI	(47)
APMM1	DDI	Prediction	DASMI	(48)
APMM2	DDI	Prediction	DASMI	(48)
DIMA 2.0 dprof	DDI	Prediction	DASMI	(18)
DIMA 2.0 dpea	DDI	Prediction	DASMI	(18)
DIMA 2.0 string	DDI	Prediction	DASMI	(18)
DPEA	DDI	Prediction	DASMI	(49)
InterDom	DDI	Prediction	DASMI	(50)
IPPRI	DDI	Prediction	DASMI	(51)
IPPRI core	DDI	Prediction	DASMI	(51)
LDSC	DDI	Prediction	DASMI	(52)
LDSC core	DDI	Prediction	DASMI	(52)
LLZ	DDI	Prediction	DASMI	(53)
LP	DDI	Prediction	DASMI	(54)
RCDP50	DDI	Prediction	DASMI	(55)
RDFD	DDI	Prediction	DASMI	(56)
TW	DDI	Prediction	DASMI	(57)
FunSimMat	PPI-CM	Gene Ontology-based	DASMI	(58,59)
Domain support	PPI-CM	DDI-based	DASMI	(2)

The interaction type PPI indicates protein–protein interactions, DDI domain–domain interactions, and PPI-CM confidence measures that can be used to assess the quality of protein–protein interactions. The server type DASMI indicates data sources that are available by an extension of the DAS protocol, and PSI WS denotes web services following the standard currently being developed by HUPO-PSI.

datasets are the results of single studies, they are not updated by their authors and do not require any maintenance efforts. The other data sources are updated by us on a monthly basis. Of course, we will replace a temporary cached source as soon as the respective original provider supports web service access to its data.

The current selection of data sources listed in Table 1 is only a snapshot, additional resources for interactions and confidence scores are currently being prepared at other institutions for public access in the near future. At the moment, there are two alternative ways for providing new data sources. The first option is the download of a server library from our website <http://www.dasmi.de>. This software, available as Java and Perl implementations, parses interaction data from several standard file formats and serves them in a format supported by

DASMIweb (25). An online tutorial on setting up own data sources is available on our website. Data sources that provide confidence scores are handled like sources that contain interaction data and can be set up using the same software library. The second option is the implementation of a web service that follows the standard currently defined by HUPO-PSI (28). However, as this standard is not yet published, it might still evolve.

Identifier mapping

Proteomics research uses a substantial diversity of object identifiers for describing genes, proteins or protein domains. Accordingly, interaction datasets use a variety of identifier systems for their data (2). In order to unify them, DASMIweb maintains internal mapping tables derived from iProClass (29) and Pfam (30) to convert identifiers between the different systems. For protein interactions, DASMIweb currently supports the identifier systems Ensembl (31), Entrez Gene (32), Entrez Geneinfo (32), RefSeq (32) and UniProtKB (33). In the following, we will refer to these identifier systems as compatible systems because mappings exist between them. The mappings enable DASMIweb to merge results from data sources that employ different but compatible identifier system. For example, if the user requests all huntingtin interactions known for the UniProtKB protein ‘HD_HUMAN’, DASMIweb will automatically convert this identifier to the Entrez Gene identifier ‘3064’, the RefSeq identifier ‘NP_002102.4’, and all other compatible identifier systems described above. Subsequently, data sources providing interactions in the Entrez Gene or RefSeq identifier system will be queried in addition to data sources that use the requested UniProtKB identifier system. The final unification of interactions from different data sources is performed by converting all identifiers to Entrez Gene identifiers. Therefore, the DASMIweb results are independent of the particular identifier system used for querying; in the prior example, a query for ‘HD_HUMAN’ will return the same results as the one for ‘NP_002102.4’ or ‘3064’.

It should be noted that the identifier mapping procedure can result in considerable, but unavoidable, computational overhead. While there is usually a one-to-one mapping from UniProtKB to Entrez Gene identifiers, mapping in the opposite direction may produce multiple results as one gene can be responsible for several protein variants or fragments. Therefore, in the exemplary case when the user queries DASMIweb with the Entrez Gene identifier ‘3064’, it will be converted to two UniProtKB entries ‘Q59FF4’ and ‘P42858’. Consequently, all interactions reported for the two protein variants or fragments will be included. Fortunately, the identifier diversity for domain interaction datasets is less problematic as stable Pfam identifiers (30) are predominantly used.

USER INTERFACE

Our primary goal while designing DASMIweb was user-friendliness, which we tried to achieve by an intuitive user interface and a clear representation of the results.

We currently support three approaches for including new data sources in DASMIweb. First, all data sources registered at the central DAS registry (<http://www.dasregistry.org>) will be available automatically to the DASMIweb users. The second option is the local registration of a data source by providing information like its name, URL, and identifier system. The third option is uploading a PSI-MI XML2.5 file (34) to DASMIweb and temporarily creating a data source from its content. The first two options require that the data source to be added is already set up and accessible over the Internet. In contrast, the third option allows for comparing own interactions with existing datasets or for assessing them by different confidence scoring servers. Another distinction can be made with respect to data privacy: data sources added with the first approach are accessible to all DASMIweb users, while the second and third approaches affect only the respective user session.

Interaction confidence scoring

Despite improved methods for generating protein interaction data (35), current interactomes are still incomplete to a large extent and doubts about the reliability of detection methods remain (2,4,5). Quality assessment is crucial not only for interactions determined by large-scale experimental assays, but also for those curated from scientific literature or obtained by computational prediction methods. Therefore, DASMIweb provides access to specialized data sources (Table 1) and, at the same time, supports the convenient evaluation of the quality of individual protein interactions.

As the distributed retrieval of confidence scores for a large number of interactions can be computationally demanding, it has to be explicitly requested by pressing a button in the header of the interaction table. After retrieval, the different scoring methods can be selected in a drop-down menu next to the same button. This menu also lists all original confidence scores provided by the authors of a source dataset. A brief description of the selected scoring method can be found in the bottom right corner of the screen. Confidence scores are printed atop the interaction squares and are also available as new interaction details. If the scores of a method can be normalized to a range between zero and one, they will additionally be color-coded by a white-to-blue gradient into the respective interaction squares, white for the value zero, dark blue for the value one (Figure 2).

Exporting results

Interaction results can be exported in different file formats, enabling the user to analyze the retrieved data further in other applications. Currently, we support the Simple Interaction Format (SIF), defined by the network analysis and visualization program Cytoscape (36), and the tabular MITAB2.5 format as specified by HUPO-PSI (34).

CONCLUSIONS

We presented our new web server DASMIweb that supports the online integration, analysis and assessment of distributed sets of molecular interaction data in a dynamic and user-configurable fashion. DASMIweb provides access to over thirty different interaction and confidence scoring resources, which constitutes one of the largest amounts of protein and domain interaction data available through one web interface. In particular, DASMIweb can be used to assess the quality of arbitrary user-defined sets of protein interactions with different confidence scoring methods. Due to the decentralized architecture, users can easily extend DASMIweb by adding further data sources. Additional data sources for providing protein interactions and confidence scoring methods are already expected to be made available within DASMIweb by different external sites in the near future. Furthermore, additional DASMIweb features are currently under development, ranging from support for full-text searches and batch queries for multiple interactors to additional data import and export formats.

ACKNOWLEDGEMENTS

We are grateful to Dorothea Emig and Sven-Eric Schelhorn for providing several domain interaction datasets and to Andreas Schlicker for his help with the functional similarity scoring servers. We also thank Robert Finn, Andrew Jenkinson, Andreas Prlić and Jonathan Warren for maintaining parts of the DAS infrastructure. In particular, we greatly appreciate the time and work of external providers of interaction data sources because a distributed system can only be successful with broad community support.

FUNDING

German National Genome Research Network (NGFN) and the German Research Foundation (DFG), contract number KFO 129/1-2. The work was conducted in the context of the DFG-funded Cluster of Excellence for Multimodal Computing and Interaction as well as of the BioSapiens Network of Excellence funded by the European Commission under grant number LSHG-CT-2003-503265. Funding for open access charge: Max Planck Society.

Conflict of interest statement. None declared.

REFERENCES

1. Frishman,D., Albrecht,M., Blankenburg,H., Bork,P., Harrington,E.D., Hermjakob,H., Jensen,L.J., Juan,D.A., Lengauer,T., Pagel,P *et al.* (2009) Protein-protein interactions: analysis and prediction. In Frishman,D. and Valencia,A. (eds), *Modern Genome Annotation: The Biosapiens Network*. Springer, Wien, Austria, pp. 353–410.
2. Ramírez,F., Schlicker,A., Assenov,Y., Lengauer,T. and Albrecht,M. (2007) Computational analysis of human protein interaction networks. *Proteomics*, **7**, 2541–2552.
3. Bader,G.D., Cary,M.P. and Sander,C. (2006) Pathguide: a pathway resource list. *Nucleic Acids Res.*, **34**, D504–D506.

4. Venkatesan,K., Rual,J.F., Vazquez,A., Stelzl,U., Lemmens,I., Hirozane-Kishikawa,T., Hao,T., Zenkner,M., Xin,X., Goh,K.I. *et al.* (2009) An empirical framework for binary interactome mapping. *Nat. Methods*, **6**, 83–90.
5. Anonymous (2009) Maturing interactions. *Nat Methods*, **6**, 2.
6. Chatr-aryamontri,A., Ceol,A., Montecchi-Palazzi,L., Nardelli,G., Schneider,M.V., Castagnoli,L. and Cesareni,G. (2007) MINT: the Molecular INTeraction database. *Nucleic Acids Res.*, **35**, D572–D574.
7. Güldener,U., Martin,M., Oesterheld,M., Pagel,P., Ruepp,A., Mewes,H.-W. and Volker,S. (2006) MPact: the MIPS protein interaction resource on yeast. *Nucleic Acids Res.*, **34**, D436–D441.
8. Salwinski,L., Miller,C.S., Smith,A.J., Pettit,F.K., Bowie,J.U. and Eisenberg,D. (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.*, **32**, D449–D451.
9. Bader,G.D., Betel,D. and Hogue,C.W.V. (2003) BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.*, **31**, 248–250.
10. Breitkreutz,B.J., Stark,C., Reguly,T., Boucher,L., Breitkreutz,A., Livstone,M., Oughtred,R., Lackner,D.H., Bahler,J., Wood,V. *et al.* (2008) The BioGRID interaction database: 2008 update. *Nucleic Acids Res.*, **36**, D637–D640.
11. Goll,J., Rajagopala,S.V., Shiau,S.C., Wu,H., Lamb,B.T. and Uetz,P. (2008) MPIDB: the microbial protein interaction database. *Bioinformatics*, **24**, 1743–1744.
12. Kerrien,S., Alam-Farouque,Y., Aranda,B., Bancarz,I., Bridge,A., Derow,C., Dimmer,E., Feuermann,M., Friedrichsen,A., Huntley,R. *et al.* (2007) IntAct—open source resource for molecular interaction data. *Nucleic Acids Res.*, **35**, D561–D565.
13. Keshava Prasad,T.S., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al.* (2009) Human Protein Reference Database—2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
14. Orchard,S., Kerrien,S., Jones,P., Ceol,A., Chatr-aryamontri,A., Salwinski,L., Nerothin,J. and Hermjakob,H. (2007) Submit your interaction data the IMEx way. *Proteomics*, **7**, 28–34.
15. Wu,J., Vallenius,T., Ovaska,K., Westermarck,J., Makela,T.P. and Hautaniemi,S. (2009) Integrated network analysis platform for protein–protein interactions. *Nat. Methods*, **6**, 75–77.
16. Chaurasia,G., Malhotra,S., Russ,J., Schnoegl,S., Hanig,C., Wanker,E.E. and Futschik,M.E. (2009) UniHI 4: new tools for query, analysis and visualization of the human protein–protein interactome. *Nucleic Acids Res.*, **37**, D657–D660.
17. Jensen,L.J., Kuhn,M., Stark,M., Chaffron,S., Creevey,C., Muller,J., Doerks,T., Julien,P., Roth,A., Simonovic,M. *et al.* (2009) STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.*, **37**, D412–D416.
18. Pagel,P., Oesterheld,M., Tovstukhina,O., Strack,N., Stümpflen,V. and Frishman,D. (2008) DIMA 2.0: predicted and known domain interactions. *Nucleic Acids Res.*, **36**, D651–D655.
19. Prieto,C. and Rivas,J.D.L. (2006) APID: Agile Protein Interaction Data Analyzer. *Nucleic Acids Res.*, **34**, W298–W302.
20. Raghavachari,B., Tasneem,A., Przytycka,T.M. and Jothi,R. (2008) DOMINE: a database of protein domain interactions. *Nucleic Acids Res.*, **36**, D656–D661.
21. Tarcea,V.G., Weymouth,T., Ade,A., Bookvich,A., Gao,J., Mahavisno,V., Wright,Z., Chapman,A., Jayapandian,M., Ozgur,A. *et al.* (2009) Michigan molecular interactions r2: from interacting proteins to pathways. *Nucleic Acids Res.*, **37**, D642–D646.
22. Razick,S., Magklaras,G. and Donaldson,I.M. (2008) iRefIndex: a consolidated protein interaction database with provenance. *BMC Bioinformatics*, **9**, 405.
23. Li,D., Liu,W., Liu,Z., Wang,J., Liu,Q., Zhu,Y. and He,F. (2008) PRINCESS, a protein interaction confidence evaluation system with multiple data sources. *Mol. Cell Proteomics*, **7**, 1043–1052.
24. Thornton,J. (2009) Annotations for all by all—the BioSapiens network. *Genome Biol.*, **10**, 401.
25. Jenkinson,A.M., Albrecht,M., Birney,E., Blankenburg,H., Down,T., Finn,R.D., Hermjakob,H., Hubbard,T.J.P., Jiménez,R.C., Jones,P. *et al.* (2008) Integrating biological data – the Distributed Annotation System. *BMC Bioinformatics*, **9**(Suppl. 8), S3.
26. Dowell,R.D., Jokerst,R.M., Day,A., Eddy,S.R. and Stein,L. (2001) The Distributed Annotation System. *BMC Bioinformatics*, **2**, 7.
27. Stein,L.D. (2008) Towards a cyberinfrastructure for the biological sciences: progress, visions and challenges. *Nat. Rev. Genet.*, **9**, 678–688.
28. Orchard,S., Albar,J.P., Deutsch,E.W., Binz,P.A., Jones,A.R., Creasy,D. and Hermjakob,H. (2008) Annual Spring Meeting of the Proteomics Standards Initiative, 23–25 April 2008, Toledo, Spain. *Proteomics*, **8**, 4168–4172.
29. Huang,H., Barker,W.C., Chen,Y. and Wu,C.H. (2003) iProClass: an integrated database of protein family, function and structure information. *Nucleic Acids Res.*, **31**, 390–392.
30. Finn,R.D., Tate,J., Mistry,J., Coghill,P.C., Sammut,S.J., Hotz,H.-R., Ceric,G., Forslund,K., Eddy,S.R., Sonnhammer,E.L.L. *et al.* (2008) The Pfam protein families database. *Nucleic Acids Res.*, **36**, D281–D288.
31. Hubbard,T.J., Aken,B.L., Ayling,S., Ballester,B., Beal,K., Bragin,E., Brent,S., Chen,Y., Clapham,P., Clarke,L. *et al.* (2009) Ensembl 2009. *Nucleic Acids Res.*, **37**, D690–D697.
32. Sayers,E.W., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., DiCuccio,M., Edgar,R., Federhen,S. *et al.* (2009) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **37**, D5–D15.
33. (2009) The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.*, **37**, D169–D174.
34. Kerrien,S., Orchard,S., Montecchi-Palazzi,L., Aranda,B., Quinn,A., Vinod,N., Bader,G., Xenarios,I., Wojcik,J., Sherman,D. *et al.* (2007) Broadening the horizon—level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol.*, **5**, 44.
35. Braun,P., Tasan,M., Dreze,M., Barrios-Rodiles,M., Lemmens,I., Yu,H., Sahalie,J.M., Murray,R.R., Roncari,L., de Smet,A.S. *et al.* (2009) An experimentally derived confidence score for binary protein–protein interactions. *Nat. Methods*, **6**, 91–97.
36. Cline,M.S., Smoot,M., Cerami,E., Kuchinsky,A., Landys,N., Workman,C., Christmas,R., Avila-Campillo,I., Creech,M., Gross,B. *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.*, **2**, 2366–2382.
37. Rual,J.-F., Venkatesan,K., Hao,T., Hirozane-Kishikawa,T., Dricot,A., Li,N., Berriz,G.F., Gibbons,F.D., Dreze,M., Ayivi-Guedehoussou,N. *et al.* (2005) Towards a proteome-scale map of the human protein–protein interaction network. *Nature*, **437**, 1173–1178.
38. Stelzl,U., Worm,U., Lalowski,M., Haenig,C., Brembeck,F.H., Goehler,H., Stroedicke,M., Zenkner,M., Schoenherr,A., Koeppen,S. *et al.* (2005) A human protein–protein interaction network: a resource for annotating the proteome. *Cell*, **122**, 957–968.
39. McDermott,J., Bumgarner,R. and Samudrala,R. (2005) Functional annotation from predicted protein interaction networks. *Bioinformatics*, **21**, 3217–3226.
40. Rhodes,D.R., Tomlins,S.A., Varambally,S., Mahavisno,V., Barrette,T., Kalyana-Sundaram,S., Ghosh,D., Pandey,A. and Chinnaiyan,A.M. (2005) Probabilistic model of the human protein–protein interaction network. *Nat. Biotechnol.*, **23**, 951–959.
41. Persico,M., Ceol,A., Gavrila,C., Hoffmann,R., Florio,A. and Cesareni,G. (2005) HomoMINT: an inferred human network based on orthology mapping of protein interactions discovered in model organisms. *BMC Bioinformatics*, **6**(Suppl. 4), S21.
42. Brown,K.R. and Jurisica,I. (2005) Online Predicted Human Interaction Database. *Bioinformatics*, **21**, 2076–2082.
43. Huang,T.-W., Tien,A.-C., Huang,W.-S., Lee,Y.-C.G., Peng,C.-L., Tseng,H.-H., Kao,C.-Y. and Huang,C.-Y.F. (2004) POINT: a database for the prediction of protein–protein interactions based on the orthologous interactome. *Bioinformatics*, **20**, 3273–3276.
44. Lehner,B. and Fraser,A.G. (2004) A first-draft human protein–interaction map. *Genome Biol.*, **5**, R63.
45. Stein,A., Panjkovich,A. and Aloy,P. (2009) 3did update: domain–domain and peptide-mediated interactions of known 3D structure. *Nucleic Acids Res.*, **37**, D300–D304.
46. Finn,R.D., Marshall,M. and Bateman,A. (2005) iPfam: visualization of protein–protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics*, **21**, 410–412.

47. Bordner,A.J. and Gorin,A.A. (2008) Comprehensive inventory of protein complexes in the Protein Data Bank from consistent classification of interfaces. *BMC Bioinformatics*, **9**, 234.
48. Wang,R.-S., Wang,Y., Wu,L.-Y., Zhang,X.-S. and Chen,L. (2007) Analysis on multi-domain cooperation for predicting protein–protein interactions. *BMC Bioinformatics*, **8**, 391.
49. Riley,R., Lee,C., Sabatti,C. and Eisenberg,D. (2005) Inferring protein domain interactions from databases of interacting proteins. *Genome Biol.*, **6**, R89.
50. Ng,S.-K., Zhang,Z. and Tan,S.-H. (2003) Integrative approach for computationally inferring protein domain interactions. *Bioinformatics*, **19**, 923–929.
51. Schelhorn,S.E., Lengauer,T. and Albrecht,M. (2008) An integrative approach for predicting interactions of protein regions. *Bioinformatics*, **24**, i35–i41.
52. Lee,H., Deng,M., Sun,F. and Chen,T. (2006) An integrated approach to the prediction of domain–domain interactions. *BMC Bioinformatics*, **7**, 269.
53. Liu,Y., Liu,N. and Zhao,H. (2005) Inferring protein–protein interactions through high-throughput interaction data from diverse organisms. *Bioinformatics*, **21**, 3279–3285.
54. Guimarães,K.S., Jothi,R., Zotenko,E. and Przytycka,T.M. (2006) Predicting domain–domain interactions using a parsimony approach. *Genome Biol.*, **7**, R104.
55. Jothi,R., Cherukuri,P.F., Tasneem,A. and Przytycka,T.M. (2006) Co-evolutionary analysis of domains in interacting proteins reveals insights into domain–domain interactions mediating protein–protein interactions. *J. Mol. Biol.*, **362**, 861–875.
56. Chen,X.-W. and Liu,M. (2005) Prediction of protein–protein interactions using random decision forest framework. *Bioinformatics*, **21**, 4394–4400.
57. Wuchty,S. (2006) Topology and weights in a protein domain interaction network—a novel way to predict protein interactions. *BMC Genomics*, **7**, 122.
58. Schlicker,A. and Albrecht,M. (2008) FunSimMat: a comprehensive functional similarity database. *Nucleic Acids Res.*, **36**, D434–D439.
59. Schlicker,A., Domingues,F.S., Rahnenführer,J. and Lengauer,T. (2006) A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics*, **7**, 302.