

DOMINO: a database of domain–peptide interactions

Arnaud Ceol, Andrew Chatr-aryamontri, Elena Santonico, Roberto Sacco,
Luisa Castagnoli and Gianni Cesareni*

Department of Biology, University of Rome, Tor Vergata, Via della Ricerca Scientifica, 00133 Rome, Italy

Received August 8, 2006; Revised and Accepted October 25, 2006

ABSTRACT

Many protein interactions are mediated by small protein modules binding to short linear peptides. DOMINO (<http://mint.bio.uniroma2.it/domino/>) is an open-access database comprising more than 3900 annotated experiments describing interactions mediated by protein-interaction domains. DOMINO can be searched with a versatile search tool and the interaction networks can be visualized with a convenient graphic display applet that explicitly identifies the domains/sites involved in the interactions.

INTRODUCTION

Cell function is governed by an intricate web of physical and functional links between proteins. Information about the details of this interaction network is dispersed in the scientific literature in a format that is not easily accessible for large scale analysis.

Over the past few years, a number of protein-interaction databases have made an effort to retrieve interaction information from published experiments (1–6). The stored information is freely available and can be downloaded and conveniently represented as graphs where interacting proteins are nodes connected by edges. This mode of representation, however, does not allow the extraction of important information such as the number of partners that any given protein is capable of binding to simultaneously. This question is particularly relevant for proteins (hubs) that have a large number of putative partners and where it is not clear, from a simple protein-interaction graph representation, whether all the partners compete for the same binding site on the hub protein or rather bind in a noncompetitive manner to different domains/sites (7). This limitation can be overcome by taking into account the modular nature of proteins and by mapping each interaction to the binding domains/sites on the partner proteins (8).

A few databases have focused on domain–domain interactions. Although they differ somewhat in scope, InterDom and DIMA aim at integration of multiple data sources and

prediction techniques to assemble a domain interaction graph linking domains that are likely to interact (9,10). iPfam is a resource that describes domain–domain interactions that are observed in protein complexes whose 3D structure is known (11).

None of these resources, however, aim at collecting all experimental observations of interactions mediated by protein-interaction domains.

A fairly large fraction of the links in a protein-interaction network is supported by families of small conserved modular domains binding to relatively short peptides in an extended conformation (12). Although the peptide ligands of most domains within a family (for instance SH3, SH2, PDZ etc...) share specific sequence/structure characteristics, each member of the family displays some degree of specificity (8). For instance SH3 domains bind to peptides that are rich in proline, mostly containing the motif PxxP, but while the SH3 domain of the yeast protein RVS167 has affinity for peptides containing an Arg at position P–3 (RxxPxxP), the SH3 domain of SHO1 prefers a Lys at the same position (13).

Over the past 15 years, the preferred targets of several members of these domain families have been studied and reported in the scientific literature thus allowing one to infer the physiological network mediated by these relatively low-affinity interactions.

In this report, we present DOMINO: A relational database designed to store protein interactions mediated by protein recognition modules (8). PDZBase has a similar scope, although limited to the PDZ domain (14). All the PDZ mediated interactions stored in DOMINO have been freshly curated to meet the Proteomics Standards Initiative Molecular Interactions (PSI-MI) standards (15).

DATABASE STRUCTURE

The data model of DOMINO is based on Intact (1), an open source database, and runs on the PostgreSQL relational database system (<http://www.postgresql.org>). The Intact data model has been extended to provide convenient and faster access to information about interacting domains. Moreover, new tables have been added for storing annotation retrieved

*To whom correspondence should be addressed. Tel: +39 0672594315; Fax: +39 062023500; Email: Cesareni@uniroma2.it

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

© 2006 The Author(s).

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

from Pfam. These are used to display the information about interacting modules in the context of the structure of the protein partners.

The API of Intact was used as a library for the development of DOMINO applications and web tools. The web interface was developed using the Struts framework (<http://struts.apache.org/>). The applications and the web interface were developed with Java 5. To limit compatibility problems, the Viewer applet has been compiled for Java 1.

STORED DATA

DOMINO aims at annotating all the available information about domain-peptide and domain-domain interactions. The core of DOMINO, of July 24, 2006 consists of more than 3900 interactions extracted from peer-reviewed articles and annotated by expert biologists. A total of 717 manuscripts have been processed, thus covering a large fraction of the published information about domain-peptide interactions. The curation effort has focused on the following domains: SH3, SH2, 14-3-3, PDZ, PTB, WW, EVH, VHS, FHA, EH, FF, BRCT, Bromo, Chromo and GYF. However, interactions mediated by as many as 150 different domain families are stored in DOMINO. The pie chart in Figure 1A reports the fraction of interactions mediated by each of the major domain families.

More than 75% of the annotated entries describe interactions between mammalian domains and their target peptides, while most of the remaining entries (22%) involve yeast proteins (see Figure 1C for detailed statistics).

The interactions deposited in DOMINO are annotated according to the PSI-MI 2.5 (15) standard and can be easily analyzed in the context of the global protein-interaction network as downloaded from major interaction databases

like MINT (3), BIND (16), INTACT (1), DIP (5) and Mpack (6).

The curation process follows the PSI-MI 2.5 standard but with special emphasis on the mapping of the interaction to specific protein domains of both participating proteins. This is achieved by paying special attention to the shortest protein fragment that was experimentally verified as sufficient for the interaction. Whenever the authors report only the name of the domain mediating the interaction (i.e. SH3, SH2 . . .), without stating the coordinates of the experimental binding range, the curator may choose to enter the coordinates of the Pfam domain match in the protein sequence. Finally whenever the information is available, any mutation or post-translational modification affecting the interaction affinity is noted in the database.

WEB INTERFACE

DOMINO is accessible through a web interface at <http://mint.bio.uniroma2.it/domino/>. The search page offers the possibility of searching either for any given protein of interest or for all the proteins in the DOMINO database containing a specific domain. The protein search can be carried out by entering identifiers of the main protein databases (Uniprot, SGD, FlyBase and WormBase). However, gene names or synonyms can also be used. A list of all domains included in DOMINO is also provided to facilitate the search. For domain restricted searches, only proteins containing the query domain, and for which the domain has been shown to mediate an interaction stored in DOMINO, will be displayed. If desirable, all types of queries can be restricted to a given organism.

The result of the search is an HTML page containing all the proteins matching the query terms and the list of the corresponding InterPro domains (Figure 2A). By clicking the check boxes corresponding to the specific protein of interest

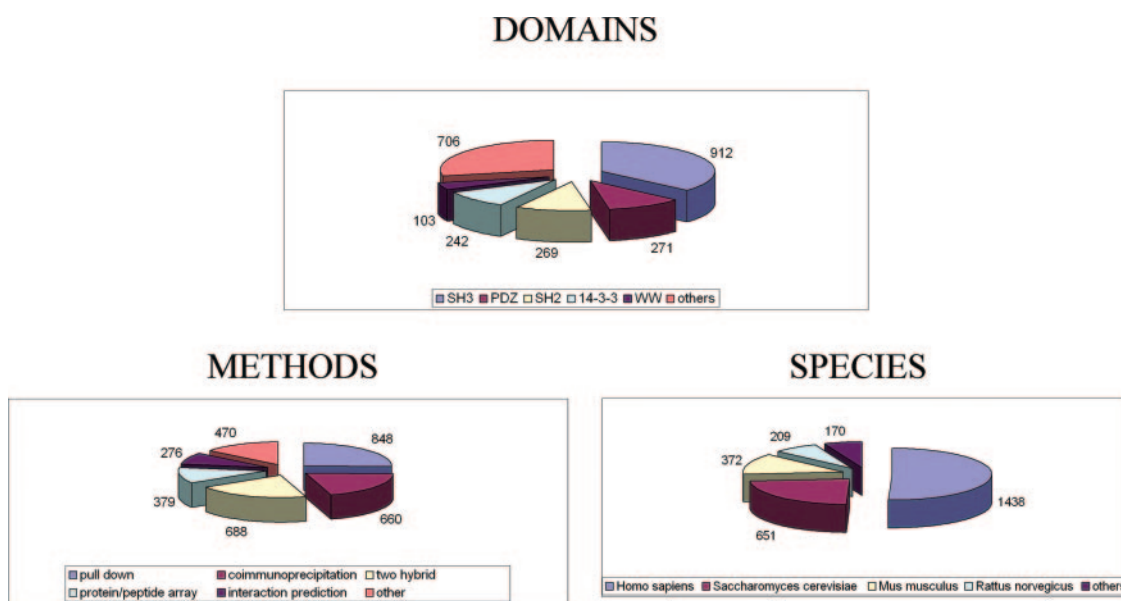


Figure 1. DOMINO statistics. (A) The pie chart represents the number of interactions mediated by each domain family in the DOMINO database. Only the five domains with the largest number of annotated interactions are shown in detail, while the remaining domains are grouped under 'others'. (B) Number of annotated protein interactions supported by each method. (C) Number of protein interactions between proteins in different proteomes.

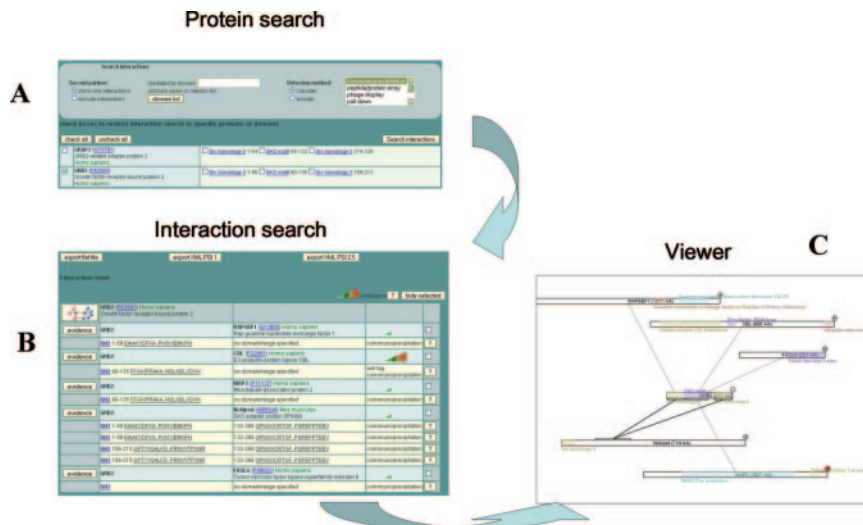


Figure 2. DOMINO WEB interface: (A) a typical result of a protein search. In this case, the search term was GRB2 and the search was restricted to *Homo sapiens*. (B) A partial view of the results of an interaction search for ligands of the GRB2 SH3 and SH2 domains. By clicking the check boxes on the right it is possible to remove irrelevant interactions from the list. (C) A selected number of interactions in the output in B were displayed using the Viewer applet.

or to a specific protein domain, one can direct the search either to the partners of the selected proteins or limit it to the partners binding to the selected domain(s). For instance, in the case of the growth factor receptor-bound protein 2 (GRB2) containing two SH3 and one SH2 domains, it is possible to restrict the search to ligands of the second SH3 domain, or to exclude them. Searches can also be limited to interactions discovered by a specific experimental method. A choice of six main method categories is given (multiple selection is possible), but any of these categories also includes all 'children' techniques, as defined in the PSI controlled vocabulary hierarchy. Among other applications, this filtering tool can be used to exclude results of large scale experiments, if so desired.

Once the appropriate choice is made, after clicking the 'search interaction' button, an HTML page is shown displaying all pairs of relevant interacting proteins and a summary of the interaction details. A full description of the entry, including experimental procedures or biological features such as required post-translation modification or defective mutations, is displayed after pressing the 'evidence' button. The HTML page can be edited by removing interactions that are deemed irrelevant to the specific query.

The edited interaction list can be exported either as a tab delimited file or as a PSI-MI document (PSI-MI version 1 or 2.5). Finally, interactions can be displayed in a graph representation through the Viewer applet (Figure 2C).

In the DOMINO Viewer applet, proteins are represented as rectangles. The protein domain structure is illustrated with a colored background (one color for each domain family). Interactions are represented as edges in the graph. Whereas most protein-interaction display tools only link entire proteins, in DOMINO the viewer utilizes the information stored in the database to link the partner domains involved in the interaction. The extent of the binding site is made clear by drawing a line under the protein fragment involved in the interaction. This representation permits an immediate visualization of the proteins that compete for binding to the same partner (Figure 2C). Whenever the interaction range in one

of the two partners has not been determined experimentally, edges are drawn in grey.

DATA ACCESS

Data stored in DOMINO are released under the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.5/>). According to this license, it is possible to copy, distribute, display and make commercial use of all data if appropriate credit is given. Data can be downloaded at <http://mint.bio.uniroma2.it/domino/download.do>, either as a tab delimited file that can be imported directly into spreadsheet applications, or in PSI-MI 1 and PSI-MI 2.5 XML documents. Users can either download a file containing the full dataset or files containing only the interactions mediated by specialized domains (SH3, SH2, PDZ, 14-3-3 and WW). As stated above, any result of an interaction search can be conveniently downloaded in two file formats.

FUTURE DIRECTIONS

The long-term goal of DOMINO is for it to develop into a stable repository of interactions mediated by protein domains thus offering a unique tool for interpreting protein-interaction networks. We are committed to make the database more comprehensive by entering new data as they become available.

Finally, we plan to use the sequence fragments that have been shown to bind specific domains to automatically identify the consensus ligand peptide for any domain for which sufficient experimental information is available.

ACKNOWLEDGEMENTS

We wish to thank Giuliano Nardelli, Maria Victoria Schneider and Francesca Palmerio for curating some of the DOMINO entries. We also like to thank Lars Kiemer for

critical reading of the manuscript and suggestions. This work is supported by AIRC and by the European Union FP6 Interaction Proteome project and the ENFIN network of excellence. Funding to pay the Open Access publication charges for this article was provided by the FP6 of the EU.

Conflict of interest statement. None declared.

REFERENCES

1. Hermjakob,H., Montecchi-Palazzi,L., Lewington,C., Mudali,S., Kerrien,S., Orchard,S., Vingron,M., Roehert,B., Roepstorff,P., Valencia,A. *et al.* (2004) IntAct: an open source molecular interaction database. *Nucleic Acids Res.*, **32**, D452–D455.
2. Stark,C., Breitkreutz,B.J., Reguly,T., Boucher,L., Breitkreutz,A. and Tyers,M. (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.*, **34**, D535–D539.
3. Zanzoni,A., Montecchi-Palazzi,L., Quondam,M., Ausiello,G., Helmer-Citterich,M. and Cesareni,G. (2002) MINT: a Molecular INTeraction database. *FEBS Lett.*, **513**, 135–140.
4. 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.
5. 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.
6. Guldener,U., Munsterkotter,M., Oesterheld,M., Pagel,P., Ruepp,A., Mewes,H.W. and Stumpflen,V. (2006) MPact: the MIPS protein interaction resource on yeast. *Nucleic Acids Res.*, **34**, D436–D441.
7. Santonico,E., Castagnoli,L. and Cesareni,G. (2005) Methods to reveal domain networks. *Drug Discov. Today*, **10**, 1111–1117.
8. Cesareni,G., Sudol,M. and Yaffe,M. (2004) *Modular Protein Domains*. Wiley-VCH Verlag GmbH and Co. KGaA, Weinheim.
9. 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.
10. Pagel,P., Oesterheld,M., Stumpflen,V. and Frishman,D. (2006) The DIMA web resource—exploring the protein domain network. *Bioinformatics*, **22**, 997–998.
11. 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.
12. Pawson,T. and Nash,P. (2003) Assembly of cell regulatory systems through protein interaction domains. *Science*, **300**, 445–452.
13. Tong,A.H., Drees,B., Nardelli,G., Bader,G.D., Brannetti,B., Castagnoli,L., Evangelista,M., Ferracuti,S., Nelson,B., Paoluzi,S. *et al.* (2002) A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science*, **295**, 321–324.
14. Beuming,T., Skrabanek,L., Niv,M.Y., Mukherjee,P. and Weinstein,H. (2005) PDZBase: a protein-protein interaction database for PDZ-domains. *Bioinformatics*, **21**, 827–828.
15. Hermjakob,H., Montecchi-Palazzi,L., Bader,G., Wojcik,J., Salwinski,L., Ceol,A., Moore,S., Orchard,S., Sarkans,U., von Mering,C. *et al.* (2004) The HUPO PSI's molecular interaction format—a community standard for the representation of protein interaction data. *Nat. Biotechnol.*, **22**, 177–183.
16. Alfarano,C., Andrade,C.E., Anthony,K., Bahroos,N., Bajec,M., Bantoft,K., Betel,D., Bobechko,B., Boutilier,K., Burgess,E. *et al.* (2005) The Biomolecular Interaction Network Database and related tools 2005 update. *Nucleic Acids Res.*, **33**, D418–D424.