

Databases and ontologies

MPIDB: the microbial protein interaction database

Johannes Goll*, Seesandra V. Rajagopala, Shen C. Shiau, Hank Wu, Brian T. Lamb and Peter Uetz

The J. Craig Venter Institute, Rockville, MD 20850, USA

Received on March 10, 2008; revised on June 6, 2008; accepted on June 8, 2008

Advance Access publication June 13, 2008

Associate Editor: Jonathan Wren

ABSTRACT

Summary: The microbial protein interaction database (MPIDB) aims to collect and provide all known physical microbial interactions. Currently, 22 530 experimentally determined interactions among proteins of 191 bacterial species/strains can be browsed and downloaded. These microbial interactions have been manually curated from the literature or imported from other databases (IntAct, DIP, BIND, MINT) and are linked to 24 060 experimental evidences (PubMed ID, PSI-MI methods). In contrast to these databases, interactions in MPIDB are further supported by 8150 additional evidences based on interaction conservation, co-purification and 3D domain contacts (iPfam, 3did).

Availability: <http://www.jcvi.org/mpidb/>

Contact: jgoll@jcvi.org

1 INTRODUCTION

As of today, about 600 bacteria have been completely sequenced and fragments of thousands of others have been determined as part of metagenomics projects (<http://www.genomesonline.org/>). Relationships among these genes and proteins are now being determined in a systematic manner. One such relationship that can be measured experimentally is a physical interaction between two proteins. High-throughput methods have generated large interaction networks. Such datasets have not only been shown to reveal functional clues about hypothetical proteins (Titz *et al.*, 2008), but also that highly connected proteins are important for survival—a fact that makes them ideal targets for antibiotics (Jeong *et al.*, 2001). Several databases have been established for interaction data (Table 1). In addition to experimental data, databases such as STRING (Von Mering *et al.*, 2007) provide predicted interactions. However, except for the latter, there is a significant bias towards eukaryotes. Furthermore, there is surprisingly little overlap between different databases. Hence users need to download and combine several microbial datasets.

The aforementioned databases have other shortcomings: none of them provides information on whether an interaction is conserved in another species and how similar the homologous proteins are. They do not include domain pairs that are known to form close contacts in 3D structures. Some even do not allow to search for species-specific information or only to a very limited extent, e.g. IntAct currently provides only searches for cryptic species abbreviations or NCBI taxon ids.

*To whom correspondence should be addressed.

Table 1. Overview of imported interaction data

Dataset	No. of interactions	No. of unique interactions	Reference
INTACT	16 322	16 059	(Kerrien <i>et al.</i> , 2007b)
MPI-EXP	3717	3544	(Titz <i>et al.</i> , 2008)
MINT	1720	284	(Zanzoni <i>et al.</i> , 2002)
BIND	1564	180	(Alfarano <i>et al.</i> , 2005)
DIP	1413	117	(Salwinski <i>et al.</i> , 2004)
MPI-LIT	748	656	unpublished data
MPIDB's top species*	No. of interactions	No. of evidences	
<i>Campylobacter jejuni</i>	11 873	12 401	
<i>Treponema pallidum</i>	3644	3818	
<i>Synechocystis sp</i>	3128	3148	
<i>Helicobacter pylori</i>	1629	1647	
<i>Escherichia coli</i>	1292	1811	

Date April 6, 2008. *Numbers may include different strains.

We describe the microbial protein interaction database (MPIDB), a new web resource that addresses such questions and provides unified access to available microbial interaction data.

2 MICROBIAL INTERACTION EVIDENCES

Protein–protein interactions are defined as unique pairs of microbial UniProt accessions (UniProt Consortium, 2008). For each interaction, MPIDB features experimental and additional evidences (Fig. 1). An interaction is reported if there is at least one experimental evidence.

2.1 Experimental evidences

Experimental evidences are based on interaction experiments that identify two physically associated proteins. An evidence is defined by its interaction detection method [PSI-MI vocabulary (Kerrien *et al.*, 2007a)] in combination with its publication. Interactions from pull-down studies were only included if a bait protein purified only one prey protein. Overall, we integrated 24 060 experimental evidences from the scientific literature (MPI-LIT), from in-house two-hybrid experiments (MPI-EXP) and from other databases (Table 1). These experiments describe a unique set of 22 530 protein–protein interactions.

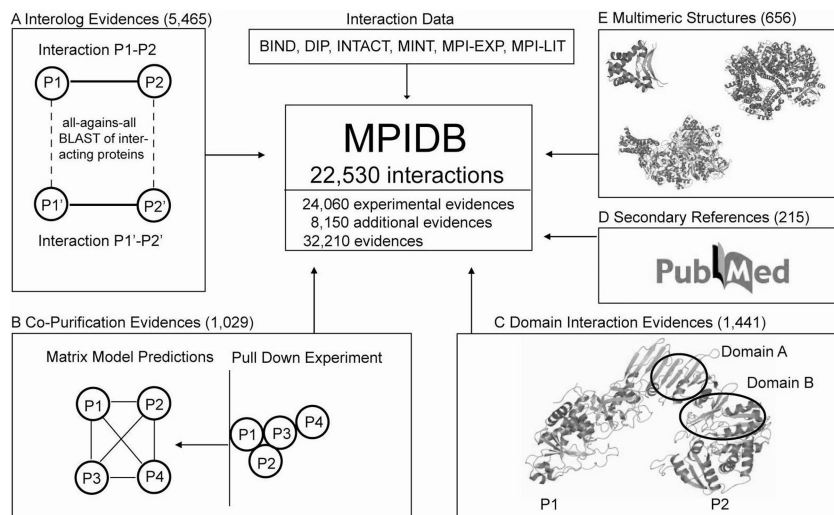


Fig. 1. MPIDB features 22 530 interactions that are supported by experimental and additional (A–D) interaction evidences.

2.2 Additional evidences

Interactions were used to identify homologous interactions (interologs) in the same or different species. A protein sequence comparison of interacting proteins revealed 5465 interolog evidences among 110 species/strains (BLAST E -value $\leq 10^{-5}$, Fig. 1A). Interacting proteins were co-purified by 1029 pull-down experiments (bait–prey or prey–prey, Fig. 1B) and 1441 Pfam domain combinations among interacting proteins have been found to form close contacts in 3D structures [iPfam (Finn *et al.*, 2005), 3did (Stein *et al.*, 2005), Fig. 1C]. Whenever an interaction from another study is mentioned in an article without experimental evidence in the same paper, we add the referenced study as a ‘secondary reference’ (Fig. 1D). Finally, multimeric 3D structures were integrated whenever both interacting proteins were part of a protein complex (homodimers have been excluded, Fig. 1E).

3 WEB INTERFACE

MPIDB can be searched by species/strain and/or by protein names (common gene name, ordered locus name and UniProt accession). More specific requests may be made using the advanced search option. Found interactions are displayed as protein pairs ordered by supporting interaction evidences. Lists of interactions may be downloaded as PSI-MI tab delimited files and can be imported by visualization tools, such as Cytoscape (<http://www.cytoscape.org/>). Clicking on the number of evidences shows experimental and additional evidences along with links to the source databases.

4 DISCUSSION

Experimental and/or predicted protein interactions are provided by several publicly available databases. Although some of them feature microbial interactions, the majority emphasise eukaryotic interactions while none of them provide access to all microbial

data: each database stores interactions that are not reported by another (see No. of Unique interactions in Table 1). While features of these databases have influenced MPIDB, its manual literature curation efforts [meanwhile as an IMEx observer member (<http://imex.sourceforge.net/>)] and supporting evidences such as interologs and domain–domain interactions are unique to our database. In contrast to other repositories, we do not include binary interactions inferred from pull-down experiments. While STRING predicts interactions *de novo*, we treat predictions with caution and only add them as additional evidences to experimentally verified ones. Finally, users are requested to suggest additional features of interest.

Conflict of Interest: none declared.

REFERENCES

- Alfarano *et al.* (2005) The biomolecular interaction network database and related tools 2005 update. *Nucleic Acids Res.*, **33** (Database issue), D418–D424.
- Finn, R.D. *et al.* (2005) iPfam: visualization of protein–protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics*, **21**, 410–412.
- Jeong, H. *et al.* (2001) Lethality and centrality in protein networks. *Nature*, **411**, 41–42.
- Kerrien, S. *et al.* (2007a) Broadening the horizon—level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol.*, **5**, 44.
- Kerrien, S. *et al.* (2007b) IntAct—open source resource for molecular interaction data. *Nucleic Acids Res.*, **35** (Database issue), D561–D565.
- Salwinski, L. *et al.* (2004) The database of interacting proteins: 2004 update. *Nucleic Acids Res.*, **32** (Database issue), D449–D451.
- Stein, A. *et al.* (2005) 3did: interacting protein domains of known three-dimensional structure. *Nucleic Acids Res.*, **33** (Database issue), D413–D417.
- Titz, B. *et al.* (2008) The binary protein interactome of *Treponema pallidum*—the syphilis spirochete. *PLoS ONE*, **3**, e2292.
- UniProt Consortium (2008) The universal protein resource (UniProt). *Nucleic Acids Res.*, **36** (Database issue), D190–D195.
- von Mering, C. *et al.* (2007) STRING 7—recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res.*, **35** (Database issue), D358–D362.
- Zanzoni, A. *et al.* (2002) MINT: the molecular INteraction database. *FEBS Lett.*, **513**, 135–140.