

Protein-protein interaction databases

GW School of Medicine & Health Sciences (February 18, 2020)

1.Introduction

The use of new technologies to study molecular biology in the past decades has revolutionized biological and biomedical research, allowing researchers to systematically study the set of proteins, including their structures and functions (Proteomics). There are many protein-protein databases and toolkit for the storage, presentation, and analysis of rich curated molecular data. DIP, IntAct, MINT, STRING, and UniProtKB are the main protein-protein interaction databases.

2. Protein-protein interaction databases

2.1 MINT

MINT: the Molecular INTeraction database. MINT focuses on experimentally verified protein-protein interactions mined from the scientific literature by expert curators. Protein interaction databases represent unique tools to store, in a computer-readable form, the protein interaction information disseminated in the scientific literature.

2.2 STRING

STRING is a database of known and predicted protein-protein interactions. The interactions include direct (physical) and indirect (functional) associations; they stem from computational prediction, from knowledge transfer between organisms, and interactions aggregated from other (primary) databases.

2.3 IntAct

IntAct is an open-source database and toolkit for the storage, presentation, and analysis of rich curated molecular interaction data in the community accepted standard formats. IntAct provides relevant experimental details of protein interactions curated from literature or directly deposited. The IntAct web site provides multiple search functionalities:

- 1) search by anything that might be related to interactions, for example, gene name, identifiers, GO term, publication, and experimental method, etc.;
- 2) search on four ontologies: Gene Ontology, InterPro, PSI-MI, ChEBI ;
- 3) draw all or part of a chemical structure and search for chemical compounds. IntAct release in February 2020 consists of 1043844 binary interactions from 21157 curated publications (Table 2)

2.4 DIP

DIP: is a database that documents experimentally determined protein-protein interactions. It provides the scientific community with an integrated set of tools for browsing and extracting information about protein interaction networks.

2.5 UniProtKB

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent, and rich annotation. UniProtKB interaction

section provides information on the quaternary structure of a protein and interaction(s) with other proteins or protein complexes. The information is filed in different subsections. (Subunit structure, Binary interactions, Protein-protein interaction databases.

3. Comparison of the result of the main databases

A comparison of the main databases indicating the sources of the data. Table 1 shows two different approaches in the collection and presentation of interaction data: (i) primary databases, which include experimentally proven protein interactions coming from either small-scale (Ssc) or large-scale (Lsc) published studies that have been manually curated; DIP, IntAct and MINT are primary databases; (ii) prediction databases, which include mainly predicted PPIs derived using different approaches, combined with experimentally proven PPIs. Computational methods for predicting protein interaction partners were previously reviewed. STRING is the prediction database in Table 1. Table 2 shows the main databases and its number of proteins, publications, interactions, and organism which are involved.

Database	Database Full Name	URL	PPI Sources
Primary databases	include PPIs from large- and small-scale (Lsc & Ssc) experimental data that are usually obtained from the curation of research articles		
DIP	Database of Interacting Protein	http://dip.doe-mbi.ucla.edu/	Ssc & Lsc published studies (literature-curved)
IntAct	IntAct Molecular Interaction Database	http://www.ebi.ac.uk/intact	Ssc & Lsc published studies (literature-curved)
MINT	Molecular INTERaction Database	http://mint.bio.uniroma2.it/mint	Ssc & Lsc published studies (literature-curved)
UniProtKB	UniProt Knowledgebase (UniProtKB)	http://uniprot.org	IntAct, MINT, STRING, DIP, BioGaid,ELM,CORUM,C omplexPortal,
Prediction databases	include PPIs from experimental analyses together with predicted PPIs obtained from the analyses of heterogenous biological data		
STRING	Known and Predicted Protein-Protein Interactions	http://string.embl.de/	BIND, BioGRID, DIP, HPRD, IntAct, MINT, & nonPPI data

Table 1 Description of DIP, IntAct, MINT, STRING and UniProtKB PPI Sources

Database	<i>n</i> Proteins (Feb 2020)	<i>n</i> Interactions (Feb 2020)	Publications	Organism
DIP	28850	81923	8234	834
IntAct	115255	1043844	21157	Homo sapiens 59.2%
MINT	26344	131695	6024	647
STRING	24584628	>20000000		5090
UniProtKB	561568(Reviewed)			188349

Table 2 Description of DIP, IntAct, MINT, STRING, and UniProtKB interactions

Reference

1. The UniProt Consortium, UniProt: a worldwide hub of protein knowledge, *Nucleic Acids Research*, Volume 47, Issue D1, 08 January 2019, Pages D506–D515, <https://doi.org/10.1093/nar/gky1049>
2. Chen, C., Huang, H., & Wu, C. H. (2017). Protein Bioinformatics Databases and Resources. *Methods in molecular biology* (Clifton, N.J.), 1558, 3–39. https://doi.org/10.1007/978-1-4939-6783-4_1
3. Proteome and proteomics: new technologies, new concepts, and new words. N. L. Anderson, N. G. Anderson *Electrophoresis*. 1998 Aug; 19(11): 1853–1861. doi: 10.1002/elps.1150191103