# Chapter 3

# The MIntAct Project and Molecular Interaction Databases

Luana Licata and Sandra Orchard

## Abstract

Molecular interaction databases collect, organize, and enable the analysis of the increasing amounts of molecular interaction data being produced and published as we move towards a more complete understanding of the interactomes of key model organisms. The organization of these data in a structured format supports analyses such as the modeling of pairwise relationships between interactors into interaction networks and is a powerful tool for understanding the complex molecular machinery of the cell. This chapter gives an overview of the principal molecular interaction databases, in particular the IMEx databases, and their curation policies, use of standardized data formats and quality control rules. Special attention is given to the MIntAct project, in which IntAct and MINT joined forces to create a single resource to improve curation and software development efforts. This is exemplified as a model for the future of molecular interaction data collation and dissemination.

**Key words** Molecular interactions, Databases, Manual curation, Molecular interaction standards, Controlled vocabulary, Bioinformatics

## 1 Introduction

Each organism, from the simplest to the more complex, is an ensemble of interconnected biological elements, for example, protein–protein, lipid–protein, nucleic acids–protein, and small molecules–protein interactions, which orchestrates the cellular response to its immediate environment. Thus, a system wise understanding of the complexity of biological systems requires a comprehensive description of these interactions and of the molecular machinery that they regulate. For this reason, techniques and methods have been developed and used to generate data on the dynamics and complexity of an interaction network under various physiological and pathological conditions. As a result of these activities, both large-scale datasets of molecular interactions and more detailed analyses of individual interactions or complexes are constantly being published.

In order to archive and subsequently disseminate molecular interaction data, numerous databases have been established to system-

Oliviero Carugo and Frank Eisenhaber (eds.), *Data Mining Techniques for the Life Sciences*, Methods in Molecular Biology, vol. 1415, DOI 10.1007/978-1-4939-3572-7_3, © Springer Science+Business Media New York 2016

atically capture molecular interaction information and to organize it in a structured format enabling users to perform searches and bioinformatics analyses. In the early 2000s, DIP [1] and BIND [2] were the first protein–protein interaction (PPI) repositories to contain freely available, manually curated interaction data. Since then, many others have been established (Table 1). A fuller list of molecular interaction databases is available at: http://www.pathguide.org.

However, due to the increasing amount of interaction data available in the scientific literature, no individual database has sufficient resources to collate all the published data. Moreover, very often these data are not organized in either a user-friendly or structured format and many databases contain redundant information, with the same papers being curated by multiple different resources. In order to allow easier integration of the diverse protein interaction data originating from different databases, the Human Proteome Organisation Proteomics Standards Initiative (HUPO-PSI) [3] developed the PSI-MI XML format [4], a standardized data format for molecular interaction data representation. Following on from this, a number of databases have further cooperated to establish the International Molecular Exchange (IMEx) consortium (http://www.imexconsortium.org/) [5], with the aim of coordinating and synchronizing the curation effort of all the participants and to offer a unified, freely available, consistently annotated and nonredundant molecular interaction dataset. Active members of IMEx consortium are IntAct [6], MINT [7], DIP, MatrixDB [8], MPIDB [9] and InnateDB [10], I2D, Molecular Connections, MBInfo, and the UniProt Consortium [11]. MPIDB was a former member of the IMEx Consortium but no longer exists as an actively curated database. Under the IMEx agreement, however, when MPIDB was retired, the IMEx data it contained was imported into the IntAct data repository and has since been updated and maintained by the IntAct group. In September 2013, MINT and IntAct databases established the MIntAct project [12], merging their separate efforts into a single database to maximize their developer resources and curation work.

## 2 Molecular Interaction Databases

To date, more than 100 molecular interaction database exist (as listed in the PathGuide resource). Many of these resources do not contain experimentally determined interactions but predictions of hypothetical interactions or protein pairs obtained as a result of text-mining or other informatics strategies. Primary repositories of experimentally determined interactions use expert curators to annotate the entries while others import their data from these primary resources. The primary molecular interaction databases can be further divided into archival database, such as IntAct, MINT, and DIP

**Table 1**
**Active molecular interaction databases**

| Database name | Data types | Main Taxonomies | Archival/thematic | Curation depth | IMEx Member | PSICQUIC service | References |
|---|---|---|---|---|---|---|---|
| IntAct | All | Full | Archival | IMEx/MIMIx | Full | Yes | [6] |
| MINT | PPIs | Full | Archival | IMEx/MIMIx | Full | Yes | [7] |
| InnateDB | PPIs | Human and mouse | Proteins involved in innate immunity | IMEx/MIMIx | Full | Yes | [10] |
| MPIDB | PPIs | Bacteria and archaea | Microbial proteins | IMEx/MIMIx | Full | Yes | [9] |
| I2D | PPIs | Model organisms | Cancer related proteins | IMEx/MIMIx | Full | Yes | |
| DIP | PPIs | Full | Archival | IMEx | Full | Yes | [1] |
| MatrixDB | PPIs; PSMIs | Human and mouse | Extracellular matrix | IMEx | Full | Yes | [8] |
| BioGRID | PPIs | Model organisms | Archival | Limited | Observer | Yes | [13] |
| HPRD | PPIs | Human | Human | Limited | No | No | [38] |
| ChEMBL | Drug-target PSMIs | Targets mainly human or pathogens | Drug-target | MIABE [39]/ MIMIx | No | Yes | [16] |
| BindingDB | Drug-target PSMIs | All | Drug-target | MIABE/MIMIx | No | Yes | [40] |
| PubChem BioAssay | Drug-target PSMIs | Targets mainly human or pathogens | Drug-target | MIABE/MIMIx | No | No | [19] |
| PrimesDB | PPIs | Human and mouse | EGFR network | Limited | Observer | No | |
| HPIDB | PPIs | Model organisms and pathogens | Host–pathogen systems | IMEx | Full | Application pending | [34] |

IMEX/MIMIx—the database contains both IMEx and MIMIx standards data
*PPIs* Protein–protein interactions
*PSMIs* Protein–small molecule interactions

that extract all PPIs described in the scientific literature, and thematic databases that select only the interactions related to a specific topic, often correlated to their research interest. MatrixDB (extracellular matrix protein interactions), InnateDB (innate immunity interactions network), and MPIDB (microbial protein interactions) are typically examples of thematic databases.

Molecular interaction databases can also be classified by the type of data that are captured or by their curation policy. Many resources curate only protein–protein interactions (PPIs), for example MINT and DIP. However, there are others (MatrixDB, IntAct) that also collect interactions between proteins and other molecule types (DNA, RNA, small molecules). Additional resources, such as BioGRID [13], collect genetic interactions in addition to physical protein interactions. Finally, databases can be differentiated accordingly to their curation policies and by the accuracy of their quality control procedures. For example, the IMEx consortium databases have committed to curating all the articles they incorporate to a consistent, detailed curation model. According to this standard, all the protein–protein interaction evidences described in the paper, in enough detail to be captured by the database, must be annotated and the entries thus created are curated to contain a high level of experimental details. All entries are subject to strict quality control measures. Other databases may choose to describe interaction evidences in less detail, which may allow curators to curate a larger number of papers. However, significant increases in curation throughput may come at the expenses of data quality.

## 3 The Manual Curation Process

Irrespective of the curation level adopted by a database, the curators have the task of manually extracting the appropriate data from the published literature. Any interaction is described by a specific experiment, and all the details of that experiment, such as how the interaction was detected, the role each participant played (for example bait, prey), experimental preparation, and features such as binding sites have to be carefully annotated. In this meticulous annotation, the identification and mapping of the molecular identifier is the most critically important piece of information.

In the literature, there are several ways the authors may choose to describe molecules, especially proteins. Commonly, the authors utilize the gene name together with a general or detailed description of the characteristics of the protein. Occasionally, a protein or genomic database identifier is specified. It is also very common that authors of a paper give an inadequate description of protein constructs; in particular, there is frequently a lack of information on the taxonomy of a protein construct. Consequently, curators have to try to trace the species of the construct by going back to the original publication in which the construct had been described or by writing to the author and asking for information about the species of the construct. Both procedures are time consuming and often do not lead to any positive results.

In 2007, in order to highlight this problem, several databases worked together in writing the "The Minimum Information about a Molecular Interaction experiment (MIMIx)" paper [14]. The main purpose of MIMIx was to assist authors by suggesting the information that should be included in a paper to fully describe the methodology by which an interaction has been described, and also to encourage journals to adopt these guidelines in their editorial policy.

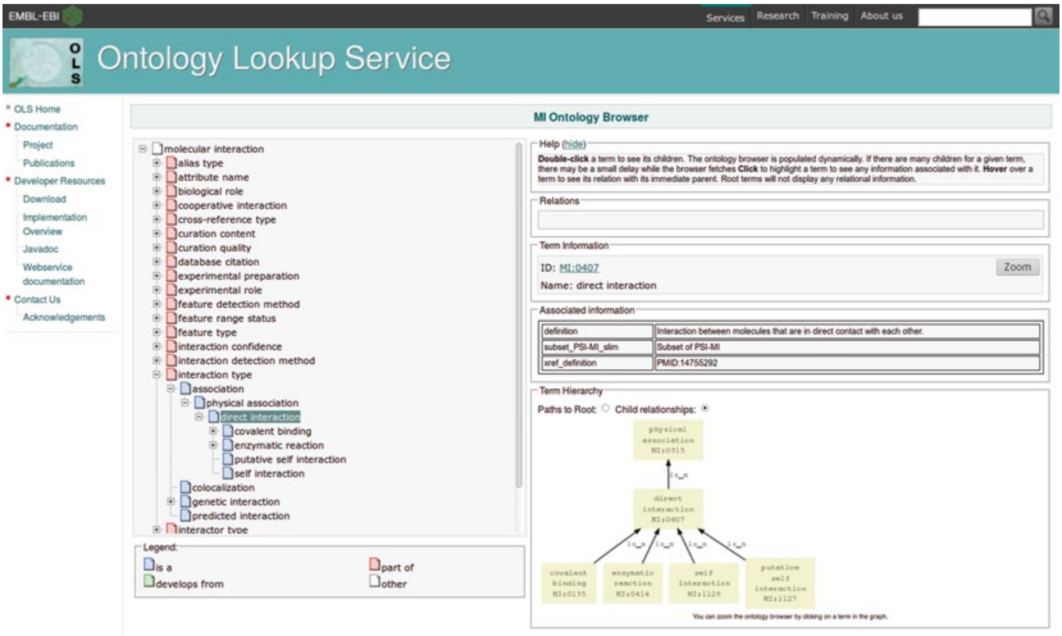
Once a protein has been identified, the curator has to map it onto the reference sequence repository chosen by its database. UniProtKB [15] is the protein sequence reference database chosen by the majority of the interaction databases. Choosing UniProtKB has the advantage of enabling the curator to annotate the specific isoform utilized in an experiment or to describe all isoforms simultaneously, by using the canonical sequence, or to specify a peptide, resulting from a post-translational cleavage. As interaction databases started to collate protein–small molecule data, and drug target databases such as ChEMBL [16] and DrugBank [17] came into existence, a need for reference resources for small molecules was recognized. ChEBI [18] is a dictionary of chemicals of biological interest and serves the community well as regards naturally occurring compounds and metabolites and small molecules approved form commercial sale but larger, less detailed resources such as PubChem [19] and UniChem [20] are required to match the production of potential drugs, herbicides and food additives produced by combinatorial chemistry. The annotation of nucleic acid interactions provides fresh challenges. Genome browsers, such as Ensembl [21], and model organism databases provide gene identifiers for gene–transcription factor binding. RNA is described by in an increasing number of databases, unified by the creation of RNAcentral [22], which enables databases to provide a single identifier for noncoding RNA molecules.

## 4 Molecular Interaction Standards

The first molecular interaction databases independently established their own dataset formats and curation strategies, resulting in a mass of heterogeneous data, very complicated to use and interpretable only after downstream meticulous work by bioinformaticians. This made the data produced unattractive to the scientific community and it was therefore rarely used. The molecular interaction repositories community recognized that it was therefore necessary to move toward unification and standardization of their data. From 2002 onwards, under the umbrellas of the HUPO-PSI, the molecular interaction group has worked to develop the PSI-MI XML [23] schema to facilitate the description of interactions between diverse molecular types and to allow the capture of information such as the biological role of each molecule participating in an interaction, the mapping of interacting domains, and the capture of any kinetic parameters generated. The PSI-MI XML format is a powerful mechanism for data exchange between multiple sources molecular interaction resources, moreover data can be integrated, analyzed, and visualized by a range of software tools. The Cytoscape open source software platform for visualizing complex networks can input PSI-MI XML files, and then integrate these with any type of 'omics' data, such as the results of transcriptomic or proteomics experiments. A range of applications then enables network analysis of the 'omics' data. A simpler, Excel-compatible, tab-delimited format, MITAB, has been developed for users who require only minimal information but in a more accessible configuration. PSI-MI XML has been incrementally developed and improved upon. Version 1.0 was limited in capacity; PSI-MI XML2.5 was developed as a broader and more flexible format [23], allowing a more detailed representation of the interaction data.

More recently, the format has been further expanded and PSI-MI XML3.0 will be formally released in 2015, making it possible to describe interactions mediated by allosteric effects or existing only in a specific cellular context, and capture interaction dependencies, interaction effects and dynamic interaction networks. Abstracted information, which is taken from multiple publications, can also be described and can be used, for example to interchange reference protein complexes such as are described in the Complex Portal (www.ebi.ac.uk/intact/complex) [24]. The HUPO-PSI MITAB format has also been extended over time to contain more data, with MITAB2.6 version and 2.7 being released [23]. The PSI-MI formats have been broadly adopted and implemented by a large number of databases and are supported by a range of software tools. Having the ability to display molecular interactions as a single, unified PSI-MI format has represented a milestone in the field of molecular interactions.

A common controlled vocabulary (CV) was developed in parallel and has been used throughout the PSI-MI schema to standardize interaction data and to enable the systematic capture of the majority of experimental detail. The controlled vocabularies have a hierarchical structure and each object can be mapped to both parent and child terms (Fig. 1). The adoption of the CV enables users to search the data without having to select the correct synonym for a term (two hybrid or 2-hybrid) or worry about alternative spelling, and allows the curators to uniformly annotate each experimental detail. For example, using the Interaction Type CV, it is possible to specify whether the experimental evidences have shown if the interaction between two molecules is direct (direct interaction, MI:0407)



Help (hide)

**Double-click** a term to see its children. The ontology browser is populated dynamically. If there are many children for a given term, there may be a small delay while the browser fetches **Click** to highlight a term to see any information associated with it. **Hover** over a term to see its relation with its immediate parent. Root terms will not display any relational information.

Relations

Term Information

ID: MI:0407

Name: direct interaction

Associated information

| | |
|---|---|
| definition | Interaction between molecules that are in direct contact with each other. |
| subset_PSI-MI_slim | Subset of PSI-MI |
| xref_definition | PMID:14755292 |

Term Hierarchy

Paths to Root: ○ Child relationships: ●

You can zoom the ontology browser by clicking on a term in the graph.

**Fig. 1** The hierarchical structure of the PSI-MI controlled vocabularies as shown in the Ontology Lookup Service [41], a portal that allows accessing multiple ontologies from a centralized interface

or only that the molecules are part of a large affinity complex (association, MI:0914). Over the years, the number of controlled vocabulary terms has increased dramatically since the original release and have been expanded and improved in order to be in line with the data interchange standard updates. The use of CV terms has also enabled a rapid response to the development of novel technique such as proximity ligation assays (MI:0813), which have been developed over the past few years. New experimental methodologies can be captured by the simple addition of an appropriate CV term, without a change to the data interchange format.

The use of common standards has also allowed the development of new applications to improve the retrieval of PSI-MI standard data. One example has been the development of the PSI Common QUery InterfaCe (PSICQUIC) [25] service that allows users to retrieve data from multiple resources in response to a single query. PSICQUIC data are directly accessible from the implementation view and can be downloaded in the current MITAB format. MIQL, the language for querying PSICQUIC has been extended according to the new MITAB2.7 format. From the PSICQUIC View Web application (http://www.ebi.ac.uk/Tools/webservices/psicquic/view/home.xhtml), it is possible to query all the PSICQUIC Services and to search over 150 million binary interactions. Currently there are 31 PSICQUIC Services and they are all listed in the PSICQUIC Registry (http://www.ebi.ac.uk/Tools/webservices/psicquic/registry/registry?action=STATUS).

Users are assured that the data is continuously updated as each PSICQUIC service is locally maintained.

## 5 IMEx Databases

As stated above, the IMEx consortium is an international collaboration between the principal public interaction repositories that have agreed to share curation powers and to integrate and exchange protein interaction data. The members of the consortium have chosen to use a very detailed curation model, and to capture the full experimental details described in a paper. In particular, every aspect of each experiment is annotated, including full details of protein constructs such as the minimal region required for an interaction, any modifications and mutations and their effects on the interaction, and any tags or labels. A common curation manual (IMEx Curation Rules_01_12.pdf) has been developed and approved by IMEx databases and it contains all the curation rules and the information that has to be captured in an entry.

The IMEx Consortium has adopted the PSI-MI standardized CV for annotation purposes and utilized the PSI-MI standard formats to export Molecular Interaction data. Controlled Vocabulary maintenance is achieved through the introduction of new child or root terms, the improvement description of existing terms, and the upgrading of the hierarchy of terms. Every IMEx member and every database curator contribute to CV maintenance during annual meetings, events or Jamborees or in an independent manner by using the tracker that allows the request of changes to the MI controlled vocabulary. Curation rule updates are also agreed with the consortium and workshops at which quality control procedures are unified are organized periodically.

In order to release high fidelity data, quality control uses a "double-checking" strategy undertaken by expert curators and also the use of the PSI-MI validator. A double-check is made on each new entry annotated in the IMEx databases; any annotation is manually validated by a senior curator before public release. The semantic validator [26] is used to check the XML 2.5 syntax, the correctness in using the controlled vocabularies, the consistency of the database cross references using the PSI-MI ontology. Rules linking dependencies between different branches of the CV, for example the interaction detection method "two hybrid (MI:0018)" will be expected to have participant identification method of either "nucleotide sequence identification (MI:0078)" or "predetermined participant (MI:0396)", have been created by the IMEx curators to enable automated checking of entries. Finally, on release, the authors of a paper are notified that the data is available in the public domain, and they are asked to check for correctness. Although it is not possible to dispense with all possible human error, all these quality control steps and rules ensure that IMEx data is of the highest quality.

## 6 The MIntAct Project

IntAct is a freely available open-source (http://www.ebi.ac.uk/intact) database containing molecular interaction data coming either from manually curated literature or from direct data depositions. The elaborate Web-based curation tool developed by IntAct is able to support both IMEx- and MIMIx-level curation. The IntAct curation interface has been developed as a Web-based platform in order to allow external curation teams to annotate data directly into the IntAct database. IntAct data are released monthly, and all available curated publications are accessible from the IntAct ftp site in PSI-MI XML and MITAB2.5, 2.6, and 2.7 formats. Alternatively, the complete dataset can be downloaded directly from the website in RDF and XGMML formats [6, 27]. Data can also be accessible through PSICQUIC Web service IMEx website. The Molecular Interaction team at the EBI also produces the Complex Portal [24], a manually curated resource that describes reference protein complexes from major model organisms. Each entry contains information about the participating molecules (including small molecules and nucleic acids), their stoichiometry, topology and structural assembly. All data are available for search, viewing, and download.

MINT (the Molecular INTeraction Database, http://mint.bio.uniroma2.it/mint/) is a public database developed at the University of Tor Vergata, in Rome, that stores PPI described in peer-reviewed papers. Users can easily search, visualize, and download interactions data through the MINT Web interface. MINT curators collect data not only from the scientific journals selected by the IMEx consortium but also from papers with specific topics, often correlated to the experimental activity of the group, such as for example, SH3 domain-based interactions [28] or virus–human host interactions. From this interest, in 2006 a MINT sister database was developed, VirusMINT, focusing on virus–virus or virus–host interactions [29]. One of the major MINT activities was the collaboration with the FEBS Letters and FEBS Journal editorial boards, which led to the development of an editorial procedure capable to integrate each manuscript containing PPIs experimental evidences with a Structured Digital Abstract (SDA) [30, 31]. MINT data are freely accessible and downloadable via the PSICQUIC Web service, the IMEx website and from the IntAct ftp site. Currently, the MINT website is under maintenance, and from the MINT download page, it is only possible to download data until August 2013. By the end of 2015, an updated version of MINT website will be available and it will be therefore possible to download all the updated information.

Within the panorama of molecular interaction databases, IntAct and MINT were individually two of the largest databases, as determined both by the number of manuscripts curated and the number of nonredundant interactions. Both have made it their mission to adopt the highest possible data quality standards. Originally both databases were separately created and were independent in funding and organization. The two databases worked closely together on the data formats and standards, together with other partners of the Molecular Interaction work group of the HUPO-PSI, and were founder members of the IMEx Consortium. MINT used a local copy of the IntAct database to store their curated data but, despite their common infrastructure, the two databases remained two physically separate entities. In September 2013, in order to optimize limiting developer resources and improve the curation output, MINT and IntAct agreed to merge their efforts. All previously existing MINT manually curated data has been uploaded into the IntAct database at EMBL-EBI and combined with the IntAct original dataset and all the new entries captured by MINT are curated directly into the IntAct database using the IntAct editorial tool. Data maintenance, and the PSICQUIC and IMEx Web services are the responsibility of the IntAct team, while the curation effort is undertaken by both IntAct and MINT curators. This represents a significant cost saving in the development and maintenance of the informatics infrastructure. In addition, it ensures a complete consistency of the interaction data curated by the MINT and IntAct curation teams. The MINT Web interface continues to be separately maintained and is built on

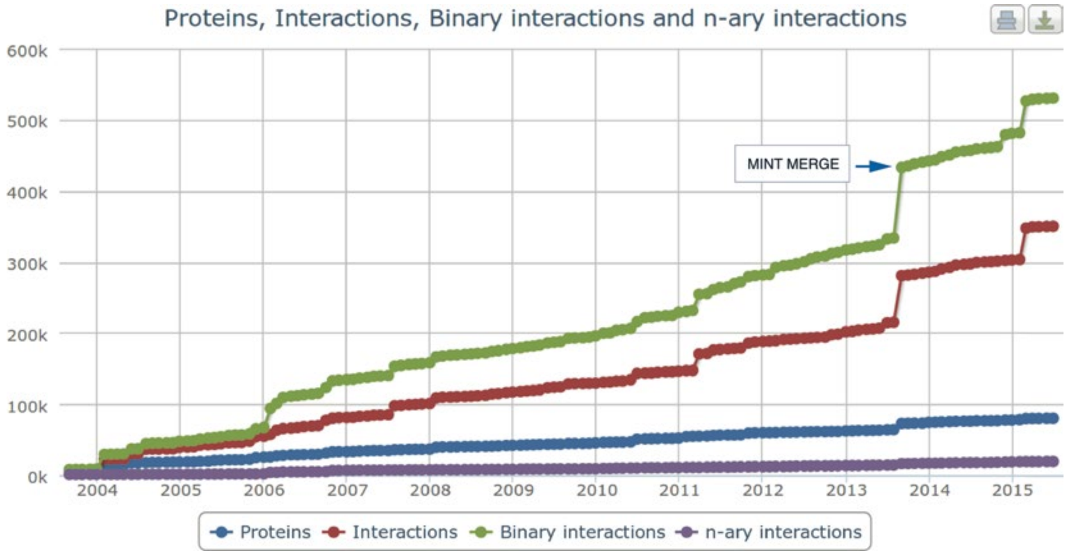


**Fig. 2** IntAct data growing and the effect of the MINT merge on data growth

an IntAct-independent database structure. All the manually curated papers from VirusMINT were tagged under a new tagged data subset called Virus, and increased by additional IntAct papers containing virus–virus or virus–host interactions. The first merged dataset was released in August 2013 and increased the number of publications in IntAct from 6600 to almost 12,000. To date, IntAct stores 529,495 binary interactions and 13,684 publications (*see* Fig. 2). The mentha [32] and virusmentha [33] interactome browser, two resources developed in the MINT group, continue to utilize the PSICQUIC Web services of the IMEx databases and BioGRID to merge all the interaction data in a single resource, as it was before the merge.

The merger of the two databases required intense work by both curators and developers. However, despite the size of the original MINT dataset, the procedure took approximately only 1 month, because of the use of community standard data representation and common curation strategies. The unification of MINT and IntAct dataset, curation activities and optimization of the developer resources provide users with a complete, up-to-date dataset of high quality interactions.

### 6.1 The IntAct Web-Based Curation Tool

The IntAct editorial tool has been designed in such a way as to allow external curators from different institutes to contribute to the dataset but at the same time giving full credit to their work. Institute Manager enables the linking of each individual curator to their parent institute or to a particular grant funding body. Any external database that uses the IntAct website as curation platform, can therefore specifically import its own data back into its own database. Moreover, each group can choose to embed its own dedicated PSICQUIC Web service within a Web page or tool.

The IntAct Web-based editorial tool allows the systematic capture of any molecular interaction experiment details to either IMEx or MIMIx-level. A number of data resources now curate directly into IntAct and utilize the existing IntAct data maintenance pipeline. For example, some UniProtKB/Swiss-Prot and Gene Ontology curators annotate molecular interactions directly into IntAct. Among the various databases, there are I2D (Interologous Interaction Database), which curates PPIs data relevant to cancer development, InnateDB, capturing both protein and gene interactions connected to innate immunity process and MatrixDB a database focusing on extracellular proteins and polysaccharides interactions. The contract curation company, Molecular Connections (www.molecularconnections.com/), carries out pro bono public domain data curation through IntAct. AgBase, a curated resource of animal and plant gene products, captures data subsequently imported into their host–pathogen database, HPIDB [34]. The Cardiovascular Gene Ontology Annotation Initiative at University College London is collecting cardiovascular associated protein interactions (http://www.ucl.ac.uk/cardiovasculargeneontology/) [35].

In order to annotate molecular interactions other than PPIs, the IntAct editorial tool has been extended to enable access to both small molecule data from ChEBI and gene derived information from Ensembl. The ability to access noncoding RNA sequence data from the RNAcentral database will be added soon.

## 7 Future Plans

One of the principal aims of the IntAct molecular interaction database has always been to be able to increase the literature coverage of database with a view to eventually being able to complete the interactomes of key model organisms. Whilst this remains an ambitious long-term goal, the merge with MINT has significantly increased the amount of molecular interaction data currently stored in IntAct. To date, more than half a million experimentally determined protein interactions are freely available via the IntAct website, PSICQUIC services and ftp site. This number could foreseeably grow to 750,000 binary interaction evidences in the next 5 years. As data become more sophisticated, new ways of visualizing data need to be developed or implemented, with a particular attention to the new generation of dynamic interaction data. IntAct has already developed an extension of the CytoscapeWeb viewer [36] that allows the user to visualize simple dynamic changes but this will to be extended as more parameters, such as molecule concentrations needs to be added to the equation. In the near future, the next challenge for the molecular interaction curation community will be to collect and collate the increasing amount of RNA-based interaction data, and the further development of reference resources such as RNAcentral will became essential.

Finally, as the experience of MIntAct has taught us, the future of the molecular interaction databases requires a move towards the consolidation of yet more disparate resources into a single, central database, where data, curation effort, software and infrastructure development will be harmonized and optimized for the benefit of the end users, thus maximizing return for investment to grant funders and making the most of limited resources. Adopting the wwPDB model [37] of a single dataset, which member databases may then present to the user via their own customized website, will give the benefit of multiple ways of searching and displaying the data whilst removing the confusion engendered by have many separate resources producing overlapping datasets.

## References

1. Xenarios I, Salwínski L, Duan XJ, Higney P, Kim S-M, 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
2. Alfarano C, Andrade CE, Anthony K, Bahroos N, Bajec M, Bantoft K, Betel D, Bobechko B, Boutilier K, Burgess E, Buzadzija K, Cavero R, D'Abreo C, Donaldson I, Dorairajoo D, Dumontier MJ, Dumontier MR, Earles V, Farrall R, Feldman H, Garderman E, Gong Y, Gonzaga R, Grytsan V, Gryz E, Gu V, Haldorsen E, Halupa A, Haw R, Hrvojic A, Hurrell L, Isserlin R, Jack F, Juma F, Khan A, Kon T, Konopinsky S, Le V, Lee E, Ling S, Magidin M, Moniakis J, Montojo J, Moore S, Muskat B, Ng I, Paraiso JP, Parker B, Pintilie G, Pirone R, Salama JJ, Sgro S, Shan T, Shu Y, Siew J, Skinner D, Snyder K, Stasiuk R, Strumpf D, Tuekam B, Tao S, Wang Z, White M, Willis R, Wolting C, Wong S, Wrong A, Xin C, Yao R, Yates B, Zhang S, Zheng K, Pawson T, Ouellette BFF, Hogue CWV (2005) The Biomolecular Interaction Network Database and related tools 2005 update. Nucleic Acids Res 33:D418–D424. doi:10.1093/nar/gki051
3. Taylor CF, Hermjakob H, Julian RK, Garavelli JS, Aebersold R, Apweiler R (2006) The work of the Human Proteome Organisation's Proteomics Standards Initiative (HUPO PSI). OMICS 10:145–151. doi:10.1089/omi.2006.10.145
4. Hermjakob H, Montecchi-Palazzi L, Bader G, Wojcik J, Salwinski L, Ceol A, Moore S, Orchard S, Sarkans U, von Mering C, Roechert B, Poux S, Jung E, Mersch H, Kersey P, Lappe M, Li Y, Zeng R, Rana D, Nikolski M, Husi H, Brun C, Shanker K, Grant SGN, Sander C, Bork P, Zhu W, Pandey A, Brazma A, Jacq B, Vidal M, Sherman D, Legrain P, Cesareni G, Xenarios I, Eisenberg D, Steipe B, Hogue C, Apweiler R (2004) The HUPO PSI's molecular interaction format--a community standard for the representation of protein interaction data. Nat Biotechnol 22:177–183. doi:10.1038/nbt926
5. Orchard S, Kerrien S, Abbani S, Aranda B, Bhate J, Bidwell S, Bridge A, Briganti L, Brinkman FSL, Brinkman F, Cesareni G, Chatr-aryamontri A, Chautard E, Chen C, Dumousseau M, Goll J, Hancock REW, Hancock R, Hannick LI, Jurisica I, Khadake J, Lynn DJ, Mahadevan U, Perfetto L, Raghunath A, Ricard-Blum S, Roechert B, Salwinski L, Stümpflen V, Tyers M, Uetz P, Xenarios I, Hermjakob H (2012) Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. Nat Methods 9:345–350. doi:10.1038/nmeth.1931
6. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, Jandrasits C, Jimenez RC, Khadake J, Mahadevan U, Masson P, Pedruzzi I, Pfeiffenberger E, Porras P, Raghunath A, Roechert B, Orchard S, Hermjakob H (2012) The IntAct molecular interaction database in 2012. Nucleic Acids Res 40:D841–D846. doi:10.1093/nar/gkr1088
7. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, Sacco F, Palma A, Nardozza AP, Santonico E, Castagnoli L, Cesareni G (2012) MINT, the molecular interaction database: 2012 update. Nucleic Acids Res 40:D857–D861. doi:10.1093/nar/gkr930
8. Launay G, Salza R, Multedo D, Thierry-Mieg N, Ricard-Blum S (2015) MatrixDB, the extracellular matrix interaction database: updated content, a new navigator and expanded functionalities. Nucleic Acids Res 43:D321–D327. doi:10.1093/nar/gku1091
9. Goll J, Rajagopala SV, Shiau SC, Wu H, Lamb BT, Uetz P (2008) MPIDB: the microbial protein interaction database. Bioinformatics 24:1743–1744. doi:10.1093/bioinformatics/btn285
10. Lynn DJ, Winsor GL, Chan C, Richard N, Laird MR, Barsky A, Gardy JL, Roche FM, Chan THW, Shah N, Lo R, Naseer M, Que J, Yau M, Acab M, Tulpan D, Whiteside MD, Chikatamarla A, Mah B, Munzner T, Hokamp K, Hancock REW, Brinkman FSL (2008) InnateDB: facilitating systems-level analyses of the mammalian innate immune response. Mol Syst Biol 4:218. doi:10.1038/msb.2008.55
11. UniProt Consortium (2009) The Universal Protein Resource (UniProt) 2009. Nucleic Acids Res 37:D169–D174. doi:10.1093/nar/gkn664
12. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N, Duesbury M, Dumousseau M, Galeota E, Hinz U, Iannuccelli M, Jagannathan S, Jimenez R, Khadake J, Lagreid A, Licata L, Lovering RC, Meldal B, Melidoni AN, Milagros M, Peluso D, Perfetto L, Porras P, Raghunath A, Ricard-Blum S, Roechert B, Stutz A, Tognolli M, van Roey K, Cesareni G, Hermjakob H (2014) The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res 42:D358–D363. doi:10.1093/nar/gkt1115
13. Chatr-Aryamontri A, Breitkreutz B-J, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Breitkreutz A, Kolas N, O'Donnell L, Reguly T, Nixon J, Ramage L, Winter A, Sellam A, Chang C, Hirschman J, Theesfeld C, Rust J, Livstone

MS, Dolinski K, Tyers M (2015) The BioGRID interaction database: 2015 update. Nucleic Acids Res 43:D470–D478. doi:10.1093/nar/gku1204

14. Orchard S, Salwinski L, Kerrien S, Montecchi-Palazzi L, Oesterheld M, Stümpflen V, Ceol A, Chatr-aryamontri A, Armstrong J, Woollard P, Salama JJ, Moore S, Wojcik J, Bader GD, Vidal M, Cusick ME, Gerstein M, Gavin A-C, Superti-Furga G, Greenblatt J, Bader J, Uetz P, Tyers M, Legrain P, Fields S, Mulder N, Gilson M, Niepmann M, Burgoon L, De Las RJ, Prieto C, Perreau VM, Hogue C, Mewes H-W, Apweiler R, Xenarios I, Eisenberg D, Cesareni G, Hermjakob H (2007) The minimum information required for reporting a molecular interaction experiment (MIMIx). Nat Biotechnol 25:894–898. doi:10.1038/nbt1324
15. Magrane M, Consortium U (2011) UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford) 2011:bar009. doi:10.1093/database/bar009
16. Davies M, Nowotka M, Papadatos G, Dedman N, Gaulton A, Atkinson F, Bellis L, Overington JP (2015) ChEMBL web services: streamlining access to drug discovery data and utilities. Nucleic Acids Res. doi:10.1093/nar/gkv352
17. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, Djoumbou Y, Eisner R, Guo AC, Wishart DS (2011) DrugBank 3.0: a comprehensive resource for "omics" research on drugs. Nucleic Acids Res 39:D1035–D1041. doi:10.1093/nar/gkq1126
18. Hastings J, de Matos P, Dekker A, Ennis M, Harsha B, Kale N, Muthukrishnan V, Owen G, Turner S, Williams M, Steinbeck C (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Res 41:D456–D463. doi:10.1093/nar/gks1146
19. Wang Y, Suzek T, Zhang J, Wang J, He S, Cheng T, Shoemaker BA, Gindulyte A, Bryant SH (2014) PubChem BioAssay: 2014 update. Nucleic Acids Res 42:D1075–D1082. doi:10.1093/nar/gkt978
20. Chambers J, Davies M, Gaulton A, Hersey A, Velankar S, Petryszak R, Hastings J, Bellis L, McGlinchey S, Overington JP (2013) UniChem: a unified chemical structure cross-referencing and identifier tracking system. J Cheminform 5:3. doi:10.1186/1758-2946-5-3
21. Fernández-Suárez XM, Schuster MK (2010) Using the ensembl genome server to browse genomic sequence data. Curr Protoc Bioinformatics Chapter 1: Unit1.15. doi: 10.1002/0471250953.bi0115s30
22. Bateman A, Agrawal S, Birney E, Bruford EA, Bujnicki JM, Cochrane G, Cole JR, Dinger ME, Enright AJ, Gardner PP, Gautheret D, Griffiths-Jones S, Harrow J, Herrero J, Holmes IH, Huang H-D, Kelly KA, Kersey P, Kozomara A, Lowe TM, Marz M, Moxon S, Pruitt KD, Samuelsson T, Stadler PF, Vilella AJ, Vogel J-H, Williams KP, Wright MW, Zwieb C (2011) RNAcentral: a vision for an international database of RNA sequences. RNA 17:1941–1946. doi:10.1261/rna.2750811
23. Kerrien S, Orchard S, Montecchi-Palazzi L, Aranda B, Quinn AF, Vinod N, Bader GD, Xenarios I, Wojcik J, Sherman D, Tyers M, Salama JJ, Moore S, Ceol A, Chatr-Aryamontri A, Oesterheld M, Stümpflen V, Salwinski L, Nerothin J, Cerami E, Cusick ME, Vidal M, Gilson M, Armstrong J, Woollard P, Hogue C, Eisenberg D, Cesareni G, Apweiler R, Hermjakob H (2007) Broadening the horizon--level 2.5 of the HUPO-PSI format for molecular interactions. BMC Biol 5:44. doi:10.1186/1741-7007-5-44
24. Meldal BHM, Forner-Martinez O, Costanzo MC, Dana J, Demeter J, Dumousseau M, Dwight SS, Gaulton A, Licata L, Melidoni AN, Ricard-Blum S, Roechert B, Skyzypek MS, Tiwari M, Velankar S, Wong ED, Hermjakob H, Orchard S (2015) The complex portal--an encyclopaedia of macromolecular complexes. Nucleic Acids Res 43:D479–D484. doi:10.1093/nar/gku975
25. Aranda B, Blankenburg H, Kerrien S, Brinkman FSL, Ceol A, Chautard E, Dana JM, De Las RJ, Dumousseau M, Galeota E, Gaulton A, Goll J, Hancock REW, Isserlin R, Jimenez RC, Kerssemakers J, Khadake J, Lynn DJ, Michaut M, O'Kelly G, Ono K, Orchard S, Prieto C, Razick S, Rigina O, Salwinski L, Simonovic M, Velankar S, Winter A, Wu G, Bader GD, Cesareni G, Donaldson IM, Eisenberg D, Kleywegt GJ, Overington J, Ricard-Blum S, Tyers M, Albrecht M, Hermjakob H (2011) PSICQUIC and PSISCORE: accessing and scoring molecular interactions. Nat Methods 8:528–529. doi:10.1038/nmeth.1637
26. Montecchi-Palazzi L, Kerrien S, Reisinger F, Aranda B, Jones AR, Martens L, Hermjakob H (2009) The PSI semantic validator: a framework to check MIAPE compliance of proteomics data. Proteomics 9:5112–5119. doi:10.1002/pmic.200900189
27. del-Toro N, Dumousseau M, Orchard S, Jimenez RC, Galeota E, Launay G, Goll J, Breuer K, Ono K, Salwinski L, Hermjakob H (2013) A new reference implementation of the PSICQUIC web service. Nucleic Acids Res 41:W601–W606. doi:10.1093/nar/gkt392
28. Carducci M, Perfetto L, Briganti L, Paoluzi S, Costa S, Zerweck J, Schutkowski M, Castagnoli L, Cesareni G (2012) The protein interaction

network mediated by human SH3 domains. Biotechnol Adv 30:4–15. doi:10.1016/j.biotechadv.2011.06.012

29. Chatr-aryamontri A, Ceol A, Peluso D, Nardozza A, Panni S, Sacco F, Tinti M, Smolyar A, Castagnoli L, Vidal M, Cusick ME, Cesareni G (2009) VirusMINT: a viral protein interaction database. Nucleic Acids Res 37:D669–D673. doi:10.1093/nar/gkn739

30. Ceol A, Chatr-Aryamontri A, Licata L, Cesareni G (2008) Linking entries in protein interaction database to structured text: the FEBS Letters experiment. FEBS Lett 582:1171–1177. doi:10.1016/j.febslet.2008.02.071

31. Leitner F, Chatr-aryamontri A, Mardis SA, Ceol A, Krallinger M, Licata L, Hirschman L, Cesareni G, Valencia A (2010) The FEBS Letters/BioCreative II.5 experiment: making biological information accessible. Nat Biotechnol 28:897–899. doi:10.1038/nbt0910-897

32. Calderone A, Castagnoli L, Cesareni G (2013) mentha: a resource for browsing integrated protein-interaction networks. Nat Methods 10:690–691. doi:10.1038/nmeth.2561

33. Calderone A, Licata L, Cesareni G (2015) VirusMentha: a new resource for virus-host protein interactions. Nucleic Acids Res 43:D588–D592. doi:10.1093/nar/gku830

34. Kumar R, Nanduri B (2010) HPIDB a unified resource for host-pathogen interactions. BMC Bioinformatics 11(Suppl 6):S16. doi:10.1186/1471-2105-11-S6-S16

35. Lovering RC, Dimmer EC, Talmud PJ (2009) Improvements to cardiovascular gene ontology. Atherosclerosis 205:9–14. doi:10.1016/j.atherosclerosis.2008.10.014

36. Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27:431–432. doi:10.1093/bioinformatics/btq675

37. Berman HM, Kleywegt GJ, Nakamura H, Markley JL (2013) The future of the protein data bank. Biopolymers 99:218–222. doi:10.1002/bip.22132

38. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadran S, Chaerkady R, Pandey A (2009) Human Protein Reference Database--2009 update. Nucleic Acids Res 37:D767–D772. doi:10.1093/nar/gkn892

39. Orchard S, Al-Lazikani B, Bryant S, Clark D, Calder E, Dix I, Engkvist O, Forster M, Gaulton A, Gilson M, Glen R, Grigorov M, Hammond-Kosack K, Harland L, Hopkins A, Larminie C, Lynch N, Mann RK, Murray-Rust P, Lo Piparo E, Southan C, Steinbeck C, Wishart D, Hermjakob H, Overington J, Thornton J (2011) Minimum information about a bioactive entity (MIABE). Nat Rev Drug Discov 10:661–669. doi:10.1038/nrd3503

40. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK (2007) BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res 35:D198–D201. doi:10.1093/nar/gkl999

41. Côté RG, Jones P, Apweiler R, Hermjakob H (2006) The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. BMC Bioinformatics 7:97. doi:10.1186/1471-2105-7-97