International Molecular Exchange Consortium - IMEx

Sandra Orchard
EMBL-EBI

EBI is an Outstation of the European Molecular Biology Laboratory.
Why do we need interaction databases

• Issues with all interaction data – true picture can only be built up by combining data derived using multiple techniques, multiple laboratories

• Problematic for any bench researcher to do – issues with data formats, molecular identifiers, sheer volume of data

• Molecular interaction databases publicly funded to collect this data and annotate in a format most useful to researchers
Interaction Databases

Deep Curation
IntAct – active curation, broad species coverage, all molecule types
MINT – active curation, broad species coverage, PPIs
DIP – active curation, broad species coverage, PPIs
MatrixDB – active curation, extracellular matrix molecules only
MPACT - no curation, limited species coverage, PPIs
BIND – ceased curating 2006/7, broad species coverage, all molecule types
– information becoming dated

Shallow curation
BioGRID – active curation, limited number of model organisms
HPRD – no curation, human-centric, modelled interactions
* MPIDB – active curation, microbial interactions
*InnateDB - active curation – interactions involved in innate immunity
*I2D – active curation – PPIs involved in cancer
• Community standard for Molecular Interactions

• XML schema and detailed controlled vocabularies

• Jointly developed by major data providers:

• Version 1.0 published in February 2004

• Version 2.5 published in October 2007 – tab-delimited format (MITAB) also released
  Broadening the Horizon – Level 2.5 of the HUPO-PSI Format for Molecular Interactions;
PSI-MI XML/MITAB benefits

- Collecting and combining data from different sources has become easier
- Standardized annotation through PSI-MI ontologies
- Tools from different organizations can be chained, e.g. analysis of IntAct data in Cytoscape.

Home page
http://www.psidev.info/MI
Controlled vocabularies

www.ebi.ac.uk/ols
How can I access PSICQUIC?

As PSICQUIC is a Web Service, you can access the data:

• Via SOAP
  o A WSDL file exists, and it is the same for all the data resources.
  o IntAct has developed a Java client, but any other languages can be used.
  o You can use it to get interactions in two standard formats: PSI-MI XML and PSI-MI TAB.

• Via REST
  o Retrieving data directly by using a URL
  o Easy to access and data can be obtained just using an internet browser.
  o Effective for scripting.
The PSICQUIC Registry

• It contains a **list of the PSICQUIC services** from different providers.

• It is a **web service** itself, and it can be accessed remotely using REST.

• Information can be found about the services, such as the URLs to use, number of interactions provided, versioning, tags, etc.

• The Registry can be found at:

  http://www.ebi.ac.uk/Tools/webservices/psicquic/registry/registry?action=STATUS

PSICQUIC Services Tagging

Content
- protein-protein
- small molecule-protein
- nucleic acid-protein

Interaction representation
- evidence
- clustered

Curation standards
- mimix curation
- imex curation
- rapid curation

Source
- internally curated
- text mining
- predicted
- imported

Complex expansion
- spoke
- matrix
- bipartite
<table>
<thead>
<tr>
<th>Name</th>
<th>Active</th>
<th>Interactions</th>
<th>Version</th>
<th>SOAP URL</th>
<th>REST URL</th>
<th>REST Example</th>
<th>Restricted</th>
<th>Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>APID</td>
<td>YES</td>
<td>416,124</td>
<td>1.1.5</td>
<td><a href="http://cicblade.dep.us">http://cicblade.dep.us</a></td>
<td><a href="http://cicblade.dep.us">http://cicblade.dep.us</a></td>
<td>Example</td>
<td>NO</td>
<td>protein-protein imported spoke clustered</td>
</tr>
<tr>
<td>CHEMBL</td>
<td>YES</td>
<td>606,404</td>
<td>1.1.6</td>
<td><a href="http://www.ebi.ac.uk/">http://www.ebi.ac.uk/</a></td>
<td><a href="http://www.ebi.ac.uk/">http://www.ebi.ac.uk/</a></td>
<td>Example</td>
<td>NO</td>
<td>smallmolecule-protein internally curated mimix curation spoke evidence</td>
</tr>
<tr>
<td>IntAct</td>
<td>YES</td>
<td>232,793</td>
<td>1.1.6-SNAPSHOT</td>
<td><a href="http://www.ebi.ac.uk/">http://www.ebi.ac.uk/</a></td>
<td><a href="http://www.ebi.ac.uk/">http://www.ebi.ac.uk/</a></td>
<td>Example</td>
<td>NO</td>
<td>protein-protein smallmolecule-protein nucleicacid-protein internally curated imported mimix curation spoke evidence</td>
</tr>
</tbody>
</table>
PSICQUIC Services
PSICQUIC

• PSICQUIC gives access to multiple databases and large amounts of data but…

• PSICQUIC is redundant – the same data is ‘recycled’ many times

• PSICQUIC contains many different data types – experimental (physical/genetic), predicted, interlogs, gene associations, text-mining. Currently these can only be separated at the database level.
IMEx

- Curation databases formed a consortium to provide users with a single, non-redundant dataset - IMEx

- **Independent** molecular interaction resources all separately funded and with their own curation priorities

- Spent several years developing **Common curation standards** for detailed curation and a joint curation manual

- **Common data formats** – all data downloadable in PSI formats (PSI-MI MITAB/XML)

- IMEx is an instance of PSICQUIC, specific records are tagged as part of the IMEx set and only these records are searchable and downloadable on the website.
Coordinated & non-redundant curation – databases ensure that each paper is curated once, and once only by a single member database.

Each paper is registered with a central database, IMEx Central, which ensures curation is not repeated by a second database.
IMEx

- **Common accession number** space – all submitted data gets an IMEx ID and is searchable on the IMEx site, the site of submission and multiple member database sites.

> Interaction datasets are required to distinguish between conserved and non-conserved (but biologically relevant) interactions and separate them from false positives and false negatives. Such a classification will make it much easier to evaluate the biological significance of individual interactions, either by suggesting additional experiments or by facilitating computational analysis such as protein docking.

**MATERIALS AND METHODS**

Description of datasets and a more extensive description of the applied methods can be found as supporting information ([Discussion S1](#)). The interactions of this study have been submitted to the IntAct database ([http://www.ebi.ac.uk/intact/](http://www.ebi.ac.uk/intact/), accession number EBI-1501350) and to the IMEx consortium ([http://imex.sourceforge.net](http://imex.sourceforge.net)) through the MPIDB database ([http://www.jcvi.org/mpidb](http://www.jcvi.org/mpidb), identifier IM-9152).

**Cloning of baits and preys, Y2H screening**

The ORF clones from McKevitt et al. ([37](#)) were transferred into compatible bait and prey vectors pAS1-loxP, pLP-G8KT7Amp, and pLP-GADT7 [Clontech], by Cre-mediated homologous recombination. After transformation into yeast, all preys were arrayed and screened as described in ([14](#)).
IMEx partners

IntAct – Active
DIP – Active
MINT – Active
MatrixDB – Active
MPIDB - Active
I2D - Active
Innate DB – Active
Molecular Connections – Active
UniProtKB - Active

BioGRID - Observer
www.imexconsortium.org
<table>
<thead>
<tr>
<th>Molecule A</th>
<th>Name A</th>
<th>Links molecule A</th>
<th>Molecule B</th>
<th>Links molecule B</th>
<th>Molecule A Identifier</th>
<th>Molecule B Identifier</th>
<th>Aliases molecule A</th>
<th>Molecule A Species</th>
<th>Molecule B Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>P04637</td>
<td></td>
<td></td>
<td>P04637</td>
<td></td>
<td>refindexyCFlunmihw3ujeB8kawoCWiwh0666</td>
<td>refindexyCFlunmihw3ujeB8kawoCWiwh0666</td>
<td>P53; Tumor suppressor p53; Phosphoprotein p53; Antigen NY-CC-13; p53_human; TF53; DIP:358N; EBI:366083</td>
<td>Homo sapiens</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>000987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>refindexyCFlunmihw3ujeB8kawoCWiwh0666</td>
<td>refindexyCFlunmihw3ujeB8kawoCWiwh0666</td>
<td>P53; Tumor suppressor p53; Phosphoprotein p53; Antigen NY-CC-13; p53_human; TF53; DIP:358N; EBI:366083</td>
<td>Homo sapiens</td>
<td>Homo sapiens</td>
</tr>
</tbody>
</table>
IMEx statistics

Sept 2012 – 237,593 binary interactions from 5668 publications
IMEx

- In production mode since February 2010

- Since 3/2009 supported by the European Commission under PSIMEx, contract number FP7-HEALTH-2007-223411, with additional partners Vital-IT, Nature, Wiley, BiaCore (GE), U. Maryland, CSIC, TU Munich, MIPS, SCBIT (Shanghai)