IntAct: Molecular interactions at EMBL-EBI

Pablo Porras Millan [1]

- Systems
- Proteins
- Beginner
- 2 hours

This course will show you what kind of information you can find in IntAct (one of the EBI's molecular interactions databases), how to access it, the different tools that can be used to access data and some practical examples on how to search and use the database.

Learning objectives:

- Understanding what the IntAct database is, what type of data it stores and how the data are organised
- Knowing the different ways you can access and query the data that are stored in IntAct
- Understanding the information given in an IntAct entry and the results table pages
- Performing complex queries that can help to refine your search results
- Representing simple networks with the networkView function of IntAct

What is IntAct?

What is IntAct?

IntAct [2] is a central, public repository where molecular interactions data can be stored and accessed (Figure 1). It is hosted in the European Bioinformatics Institute (EMBL-EBI) in Hinxton, UK, where it is maintained by a group of curators and developers.
Figure 1. Graphical display of the IntAct entry page. The homepage may change with new releases, but the above is a correct representation at time of writing (October 2014).

Where do the data come from?

We populate IntAct [3] with interaction data from literature curation [4] or direct user submissions (Figure 2). Most of the data refer to protein–protein interactions, but interactions involving other types of molecules, such as small chemical compounds or nucleic acids, can also be found in IntAct.
Because our aim is to provide molecular interaction datasets to the scientific community, IntAct is freely available and uses an open source database system and analysis tools: it can be locally installed and adapted to the needs of the local organisation. This reduces development time and encourages researchers to build consistent interaction datasets by using the same infrastructure and annotation [5] system.

**Controlled vocabularies**

The database makes extensive use of controlled vocabularies, (lists of defined terms), which allow consistent description of the experimental detail that was used to generate the data. These vocabularies are hierarchically organised and where possible cross-referenced with other controlled vocabularies such as the NCBI taxonomy [6] or Gene Ontology [7]. The main vocabulary that is used by IntAct [3] to define and describe molecular interactions is the PSI-MI controlled vocabulary [8], defined by the IMEx [9] consortium (see next page). This can be navigated using the Ontology Lookup Service hosted by the EBI [10] (look it up under "Molecular Interactions (PSI-MI 2.5)" in the drop-down menu, Figure 3).
Figure 3. A glimpse of how the PSI-MI controlled vocabulary is organised. Here we see how a controlled vocabulary/ontology term is shown in the ontology lookup service. In these three different views you can navigate the ontology following its tree-like structure (left), see the relationships and dependencies that a particular term has in this structure (middle) or check its definition and other information such as accessions, synonyms or cross references (right).

International Molecular Exchange Consortium

IntAct [3] is a member of the International Molecular Exchange (IMEx [11]) Consortium – a group of major public interaction data providers that share curation [4] effort and exchange completed records on molecular interaction data (Figure 4). The PSI-MI controlled vocabulary [8] shown in the previous slide is one of the tools that the consortium has produced with the aim of standardising the way that interactions are represented by consortium members.

When you query data in IntAct, you also access millions of interactions from different data resources via the PSICQUIC [12] (Proteomics Standard Initiative Common QUery InterfaCe) service or a consistently annotated, non-redundant, experimentally determined subset from the IMEx Consortium.
Each data provider in the IMEx consortium contributes to varying depths of curation as exemplified by the previous figure. **Shallow curation** is used if the main goal is to re-publish selected content with minimum effort so that it complies to the MIMIx standards – the minimum information required for reporting a molecular interaction experiment and a lightweight version of the IMEx guidelines [13]. **Deep curation** requires a detailed description of all the features involved in the interaction and interacting partners and complies with the full version of the IMEx guidelines, therefore requiring more time and resources.

### Why do we need IntAct?

We need [IntAct][3] because:

- Molecular interactions provide a valuable resource for the elucidation of cellular function, and protein interaction studies have been the focus of much recent biomolecular research.

- Experiments vary from large-scale systems that generate sizeable datasets, for example Yeast two hybrid [14] systems or Tandem affinity purification [15], to in-depth analysis of a single interaction in which the interacting domains, or even the individual amino acid [16] residues, can be identified.

- The IntAct database stores this detailed data in a systematic way and makes it available for search over the web and for local download.

Additional data, curated to the same level of detail, is available from our [IMEx][9] partners via the [PSICQUIC][12] service, along with many other data types, including predictive interactions, interologues and results of text-mining-based approaches for inferring molecular interactions.
What can/can’t I do with IntAct?

What can I do with IntAct [3]?

- Find the molecules that interact with your protein of interest.
- Look more deeply into individual experiments to gain both a degree of confidence in the specific interaction and its functional consequence.
- Graphically display small interaction networks.
- Query across additional resources via PSICQUIC [12].
- Rapidly transfer data into Cytoscape for further analysis.
- Visualise minimal connecting networks for protein sets.
- Download data in PSI-MI XML [17] and tabular formats.

What can’t I do with IntAct?

- IntAct is not a network analysis tool. For in-depth analysis of your network you need to use other software tools, such as Cytoscape.
- The graphical display capabilities of IntAct are limited. You will not be able to represent very large sets of results in IntAct.
- IntAct represents only physical molecular interactions. Other interactions such as genetic interactions can be found in other resources such as BioGRID [18] but not in IntAct.

Where do the data come from?

IntAct [3] mainly gathers interaction data through active, in-depth curation [4] of the available molecular interactions literature (Figure 5). Our curators read published scientific articles containing molecular interactions evidence and
represent the interactions following the IMEx [19] consortium guidelines [20]. This allows for consistent representation between the databases in the IMEx consortium, which aims to improve curation efficiency and to provide interaction-data users with a large, uniform set of interactions. In addition to classical, low-throughput experimental publications, the data can be curated from large datasets produced in high-throughput protein–protein interaction [21] projects. Direct user submissions of both high- and low-throughput datasets are also accepted.

**Figure 5.** Sources of interaction data available for curation.

**IntAct data inputs and cross-referencing**

The interactions represented in IntAct [3] come from experimental data obtained through a plethora of different methodologies [Figure 6A].

IntAct [22] makes heavy use of cross-references [23] in order to enrich the representation of our molecular interactions datasets [Figure 6B]. For a given interaction, other services from inside and outside the EBI such as UniProtKB [24], InterPro [25] or the Gene Ontology [26] can be used to depict the proteins interacting, the domains that actually take part in the interaction and the functional context in which it takes place, whenever this information is available.
Figure 6. IntAct data input pipeline: [A] manual curation [4] based on published literature and direct submission (including high throughput projects). [B] cross reference to several relevant databases allows the users to access a variety of other resources (e.g. ChEBI [27], UniProt).

How do I submit data to IntAct?

We actively encourage the submission of data from external users, particularly if the submitter intends to publish all or part of the information in question. An accession number [28] will be supplied for inclusion in a paper, and the data can be held in confidence until the publication date. Contact the intact [3]-help [at] ebi.ac.uk (IntAct help desk) if you are interested in using this service and to discuss the most appropriate submission format. You can also see the IMEx [9] Consortium submission guidelines [29] to get an idea of the kind of information we would be requiring from you.
How do I access and navigate IntAct?

The easiest way to access IntAct [3] is through its website [30]. In addition to allowing us to search and visualise the interaction data hosted in IntAct, the website contains tools that allow you to:

- produce simple visual representations of the data,
- use the cross-referenced ontologies,
- perform refined complex searches,
- obtain detailed documentation about the database and its services.

![IntAct homepage](image)

The main features in the IntAct homepage are:

[A] The **search bar** allows both simple and complex queries for single or multiple molecules. You can use gene or protein names (e.g. Grx2) or UniProtKB [31] accessions and IDs (e.g. P17695; lck_human) to perform the search.

[B] The **tabs** help you navigate through the different sections of the IntAct website and provide extra tools, specific information for developers, extensive documentation and contact information.

[C] If you want to give IntAct valuable feedback, or if you simply have a question about the service, use the **feedback link**.

We will show you detailed examples that explain how to search for data in IntAct and about its different services in following sections; but first we will explain how the data is stored in IntAct.

How is data stored in IntAct?

To learn how to use IntAct [3] effectively, you need to understand how the resource is structured (Figure 8).
Each entry in the database corresponds to a publication— a scientific article that is curated by the IntAct team.

The information in the paper is analysed and those experiments that contain molecular interactions are represented in the database (experiment level). Each experiment corresponds to one or more interactions identified in the same organism using the same techniques.

The interactions in each experiment are then listed in the interaction level, adding information such as kinetic binding parameters if they were given by the authors.

Each participant in an interaction is identified using cross-references to existing resources such as UniProt [32] or ChEBI [27], if possible, and represented at the participant level.

Finally, features of each participant, such as binding sites, tags, mutations affecting the interaction or post-translational modifications, can be depicted in detail at the feature level.

Next, we will explain how the interactions are represented in IntAct.

Representing molecular interactions data
Representing molecular interaction data

Representation of molecular interactions can be challenging. There is a great level of detail that can be inferred from experimental datasets, such as the interacting surfaces in a protein–protein interaction [21] (Figure 9). IntAct [3] supports the representation of interacting domains down to the residue level, including required post-translational modifications (PTMs) and sequence mutations that have an impact in the interaction. The domains are represented as specific ranges of the underlying amino acid sequence. These ranges are re-mapped whenever the relevant UniProtKB [31] protein sequence is updated.

![Interacting domains](image)

**Figure 9.** Representation of interacting domains in a protein–protein interaction.

How to deal with complexes

Some experimental protocols, such as tandem affinity purification (TAP) [33], generate complex data sets, depicting interactions in which more than two proteins are involved at the same time (n-ary interactions).

However, interaction data are often stored in tabular formats that aim to be amenable to quick, comprehensive searches. It may be desirable to convert these complexes into sets of binary interactions to simplify and speed up searches. There are two algorithms that will perform such conversion: the matrix model and the spoke model [13], as depicted in Figure 10. In this hypothetical example, take the bottom right protein complex (marked “reality”). A tandem affinity experiment (far left) might tell you that each of the other five proteins interact with the red bait protein in the middle. In reality, the red protein has only one interactor, which is the yellow protein.

![Overlay of Ranges on sequence](image)

As you can see, both algorithms are somewhat misleading, but as the spoke model generates up to 3 times fewer false positives [13], IntAct uses the spoke model when data are exported in tabular format. Many people do not find spoke-expanded data is useful and prefer to exclude them from their analyses, so IntAct gives you the option to filter your results to remove spoke-expanded interactions.
Figure 10. The matrix expansion algorithm and the spoke expansion algorithm converting complex interactions into binary ones. In this hypothetical example, take the bottom right protein complex (marked “reality”). A tandem affinity experiment (far left) might tell you that each of the other five proteins interact with the red bait protein in the middle. In reality, the red protein has only one interactor, which is the yellow protein.

Next, we will show how the data is displayed in the IntAct website.

The tabular view in IntAct

Visualising interactions in IntAct

Now that we have discussed the structure of IntAct [3] and the challenges that come with representing molecular interactions data, let us have a look at how these data are displayed in the IntAct website.

The tabular view allows the visualisation of several interactions together in IntAct.

Figure 11 shows the view that you get once you perform a search for interactions in the IntAct website and it is based in the PSI-MITAB [34] format. This view has the following characteristics:

- It describes a binary interaction.
- A binary interaction collapses multiple sets of experimental evidence into a single line.
- The order of interactor "A" and "B" is non-deterministic.
- Whenever the data originates from an n-ary interaction, it is expanded using the Spoke model.
Figure 11. The view of a search for interactions in the IntAct website.

[A] Table setting tools: These links allow us to change to different preset table displays or to build a personalised one, in which we can choose which columns are shown using the “Show/hide columns” button, as we will see in the next slide.

[B] Download: Using the drop-down menu we can choose different formats in which to download our data. There is a size constraint and large results will only be available in the tab-delimited format PSI-MITAB.

[C] Column headers: In this basic view a subset of basic fields, including the names of the interacting partners and the method used to find the interaction, is shown. Using the tools described in [A] all the information recorded in an IntAct entry can be displayed as column fields.

[D] Detailed interaction view: following this link (or clicking on the “Interaction AC” as shown in the last column in this view) opens the interaction detail view (see corresponding section).

[E] Links field: If you click on the UniProtKB [31] accession [35] shown for each interacting protein (participant) you will be sent to the relevant UniProt [32] entry. If the participant is not a protein, but a small molecule, for example, you will find a ChEBI [27] identifier [36] in this place. Each participant is also assigned a unique EBI identifier that is shown in the links field. If you click on this identifier, you will be re-directed to a Dasty2 visualisation page. We will learn more about Dasty [37] visualisation later.

The next slide will show other display columns including a method of expansion, allowing you to filter out columns:

Changing the view

Here we have a more detailed view of the display options (Figure 12).
As you can see in the magnified section [A], with the “Show/hide columns” menu you can choose precisely which columns to show, accessing information such as interacting partners alias, interaction detection methods or interaction confidence scores. Either from this menu or from the “Change table preset...” one [B], you can also use pre-determined displays that range from the minimal view, showing only the names of the interactors and the accession [35] to the interaction, to the expanded view, which shows 26 different fields comprising most of the information referred to the interaction.

The interaction detail view in IntAct

Accessing the interaction detail view

To open the interaction detail view click on the magnifying glass link [A] or on the appropriate “Interaction AC” [B] displayed in the last column of the table described in Tabular View [38], see Figure 13. Either of these actions will direct you to the interaction view.
Figure 13. The view of a search for interactions in the IntAct website.

Let's have a look at the interaction detail view.

**Detailed view of the interaction tables - Overview**

The interaction detail view follows the general IntAct [3] schema described in “How is data stored in IntAct? [39]”. Each level of the schema has a different section in the interaction view. Let's look through these levels in more detail (Figure 14).

Figure 14. The first two levels of data in the interaction detail view: [A] Experiment, [B] Publication.

[A] **Experiment level**: you can see how many interactions are found in the experiment where the interaction of interest was found. Here you can learn which organism the experiment was performed in and the method used to find the interaction, as well as the method used to identify the interaction participants.

[B] **Publication level**: this level provides a reference to the article in which the experiment was published. The publication is the highest level of an IntAct database entry, so you can also find other information on the database entry here. These may include comments on the origin of the data if was not present in the publication but was...
registered by the submitter or added through deep curation [4]. There is also a link to all the interactions found in the same publication.

We will continue the description of further levels on the next page.

**Detailed view of the interaction tables - Overview**

Continuing view of the interaction table (Figure 15):

![Interaction table and graphical representation](image)

**Figure 15.** Further levels of Interaction Details: [C] Interaction, [D] Participants, [E] Graphical representation of experimental features.

[C] **Interaction level**: here the name and accession [35] code of the interaction are listed. There is also information about what type of interaction we are dealing with, whether it is a direct interaction, an enzymatic reaction, some type of physical association, etc… The figure or table in the article in which the interaction data is given is also shown here.

[D] **Participant level**: this provides a detailed representation of the participants in the interaction. We will describe this in more detail in the next slide.

[E] **Graphical representation of experimental features**: this provides a simplified graphical view of the interaction, which we will describe in more detail in the next slide. In this particular example the second interactor has a mutation.
Detailed view of the interaction tables - Interaction Representation

Here we describe how the interaction is represented:

The contents of the interaction, participant and feature levels give us detailed information about the interacting molecules and their relationship (Figure 16).


[A] Image shows the actual participants of the interaction and their identifiers.

We can also see information about:

- their organism of origin [B],
- their experimental role in the interaction (e.g. bait or prey) [C],
- their biological role (e.g. enzyme or enzyme substrate) [D],
- their type (e.g. protein or small molecule) [E].

The final column [F], labelled "more", allows access to extra information on: annotations (indicated with an A), experimental parameters (indicated with an P), such as binding affinities, stochiometric values (indicated with an S) and features (indicated with an F). The latter can be particularly interesting as shown in Figure 17 below see: [G].
Figure 17. The features can comprise labelling tags, binding regions or any modification that the interacting molecule might have. The graphical display allows simple representation of binding regions and other experimental features that the molecules might have, as explained in the help box magnified here.

Next, we will explain how to search IntAct [3].

How to search IntAct

Basic IntAct search

Searching for interactions in IntAct [3] is very simple (Figure 18): Open the Search tab [A] and just paste the name or accession [35] of your molecule(s) of interest in the search bar and press "Search". In this example, we use a list of proteins and we give their UniProtKB [31] accessions.
Different ways of searching in IntAct shown in the figure above:

[A] The Search bar can be accessed using the search tab.

[B] The Search tab view allows the use of the normal quick search.

[C] Ontology Search: can be used to search using more advanced features such as terms from the different controlled vocabularies that IntAct uses, such as the Gene Ontology [26] or the PSI-MI ontology [41].

[D] Other Searches: can be used to do searches like a chemical search, in which we can look for interacting partners of a particular chemical substance or structure.

You can also use filters to refine your search, which we will describe next.

Advanced IntAct Search

Advanced searches in IntAct [3] can be performed using the filtering [42] options (Figure 19).
In the search bar, look for the “Fields” link [A]. Using this link we can use specific fields to filter our *search for a particular organism or method*. Once the search is performed, *additional filters* can be added to refine our search results.

Here we will show you how to use advanced search to filter your search with a particular organism and biological role.

![IntAct search interface]

**Figure 19.** Advanced filtering search for a particular organism or method.

1. On the homepage open the advanced search option as shown.
2. In the ‘Field’ box select ‘organism’ and type in ‘human’ in the next box. You can now search for all human protein interactions in IntAct and other databases by clicking the ‘Add and Search’ button.
3. To add more options in the advanced search, type 'species: human' in the main 'Search' box, open the ‘Show Advanced Fields’ and select the ‘And’ button.
4. Now you can choose ‘biological role’ from the Field box and type 'enzyme' in the next box.

Next, we will describe how to search using the Molecular Interaction Query Language (MIQL).
Searching with MIQL

You can also build complex queries using the Molecular Interaction Query Language (MIQL). Press the "MIQL syntax reference" link to obtain a detailed description of the field aliases that you can use to build your MIQL search (Figure 20). The operators "NOT", "AND" and "OR" can be used to filter for more than one field and different fields can be grouped using parentheses.

For example, imagine that you need to look for all the interactions involving the human protein HTT (Huntingtin) and any other human protein, but you want only those interactions detected using two-hybrid methodologies. You could use the following query:

\[(\text{alias:(htt)} \text{ AND (taxidA:9606 AND taxidB:9606)}) \text{ AND detmethod:} "\text{two hybrid}" \]

Figure 20. Example of a MIQL query (boxed text) and the MIQL syntax summary that you can use to construct a MIQL query.

Next, we will show how to download your results.

How do I get data from IntAct?

Download data from IntAct
Once you have your results in the IntAct [3] website, you can download them in a variety of formats (Figure 21).

![Figure 21. Interaction Tab. Drop-down menu [A] and the Download button [B].]

[A] The format can be chosen from the drop-down menu.

[B] It can be downloaded by clicking the “Download” button.

You can now download the results in a format that can be read by Cytoscape [45], a popular open source [46] bioinformatics [47] software platform for visualising [48] molecular interaction networks [49]. This is important for the next step in the analysis.

For large datasets, only the tab-delimited format PSI-MI-TAB [50] is available.

**Download data from IntAct**

You can get a complete copy of all the data in IntAct [3] using the Downloads section (Figure 22).

![Figure 22. The downloads section of IntAct. [A] Access to the Downloads section on the side menu. [B] Download formats.]

IntAct is released weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.

Contents of the IntAct FTP directory:
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
- IntAct releases weekly. All IntAct data and software is freely available to all users, academic or commercial, under the terms of the Apache License, Version 2.0, from ftp://ftp.ebi.ac.uk/pub/databases/IntAct/current.
Go to the “Downloads” section using the side menu.

Among all the information found in this section, you can find the option to download the entirety of IntAct in PSI-MI format (an XML-based format complying with the PSI initiative guidelines) or in PSI-MI-TAB.

Next, we will explain how to graphically visualise the results.

Visualising the output

If your results table is too big, the web tool will not be able to represent them and you will need to download your data as explained before in a Cytoscape-compatible format (XGMML would be the best choice) in order to represent it there.

If your results table is not too long (roughly below 500 interactions), you can have a quick look at your data by producing a simple graphical representation using the “Graph” tab (Figure 23). This opens the Network visualisation window and shows a network representation of your interactions using CytoscapeWeb (infobox below). Each interacting molecule is represented as a node (the little balls in the graph) that is linked to other node(s) by an edge, the connecting line that represents the interaction.


Open in Cytoscape: If you want to produce a more complex visualisation, making use of Cytoscape advanced representation options, you can export your representation to Cytoscape Webstart. You can also download your dataset in XGMML format [51] (an XML-based format specifically designed for Cytoscape) as shown in 'How do I get data from IntAct [52]' and then import it into your favourite version of Cytoscape.

CytoscapeWeb controls: CytoscapeWeb has limited visualisation options but you can can use the Cytoscape Web controls box to choose between three different layouts and to represent each piece of evidence for an interaction as a single edge by deactivating the “merge edges” feature. Each edge represents one interaction detected in one publication using one specific methodology, but they are merged by default when more than one...
instance of an interaction is found. By deactivating this feature, you can easily see which interactions have been found more than once.

- CytoscapeWeb is a web-based network visualisation tool.
- It is modelled on Cytoscape: it is open-source, interactive, customisable and easily integrated into web sites.
- CytoscapeWeb contains none of the plugin functionality of Cytoscape, nor its advanced representation options.

Summary

- The data stored in IntAct (mostly, but not exclusively, protein–protein interactions) are mapped to UniProtKB [31] accessions or any other suitable identifier [36] for the molecule involved in the interaction (e.g., ChEBI [27] accessions are used for small molecules [53]). Controlled vocabularies are used to describe the interactions, and other resources, such as the Gene Ontology [26], are cross-referenced to enrich the information given for the interaction.
- The interaction data comply with the IMEx [9] consortium guidelines. The interactions are described using a consistent level of detail and in a format that can be easily merged with data from other members of the consortium.
- The IntAct webpage [30] is the main interface for accessing the IntAct database. Here you can:
  - Search for interactions using molecule identifiers, ontology terms, gene names and different synonyms.
  - Filter your results using featured information, such as the interaction detection method or the organism where the interaction takes place.
  - Visualise all the information stored for one or more interactions using customisable tables and detailed representations.
  - Download the data in a variety of formats, ranging from the standard PSI-MI XML [17] and tabular formats to the Cytoscape-specific XGMML.
  - Represent your results as simple graphs of interaction networks.

When to use IntAct: guided example

The following guided example is intended to help you reinforce what you have learned so far on this course by providing a real example of how IntAct [3] can be used.

The human signalosome complex

Initial simple search
Imagine that you are interested in the human signalosome complex, a multifunctional protein complex essential for development and possibly involved in the regulation of protein degradation. You want to generate a network that will show the proteins that are known to physically interact to form the complex or regulate its function.

Since IntAct [3] supports searching with terms coming from Gene Ontology [26], we can use “signalosome” (a GO term) to perform a simple search using the search bar [A]. 993 binary interactions are found, as you can see in the close-up (Figure 24).

However you want to refine the search because you want to study the signalosome in humans, not in other organisms.

![Figure 24](image_url) Performing a simple search on IntAct.

Search refinement (1)

To refine the search and limit it to those interactions that involve human proteins*, we make use of the advanced fields of the IntAct [3] search box (Figure 25).

*Sometimes you will find interactions in IntAct involve proteins from two different organisms. This may be the case if, for example, a researcher does not have a purified human protein but wants to find out which other human proteins it interacts with. In such studies, homologues from other species such as mouse may be used as a substitute for the human protein.
Figure 25. Refining your search.

1. Click the “Show Advanced Fields” link [A]. An extension of the search box appears below it.
2. From the extension box use a drop-down menu [B] to select “organism”.
3. Type “human” (or "Homo sapiens") into the "Add & Search" box to limit your results to interactions involving human proteins.

Search refinement (2)

In the close-up we can see how the number of interactions has gone down to 339. Our results are now limited to those interactions that involve human proteins. However, if we have another look at the close-up we see that a significant number of these interactions come from spoke expanded co-complexes (see Figure 10). As some of these may be false positives, we can filter them out to get a list of binary interactions. For that, we click on the “filter” link as underlined in red (Figure 26).
Figure 26. Adding a filter to your refined search results.

Filtering spoke expanded expansions

Notice that the whole process can be summarized by a direct MIQL query as depicted in the search bar (Figure 27) [A].

After we have finished filtering our results, we end up with 168 interactions [B] involving human proteins annotated to be related to the signalosome.
Figure 27. Your filtered results.

We can now download the results using the download drop-down menu and button [C].

Alternatively we can represent the network using the “Graph” tab [D] (shown in the next slide).

Graphical representation of the signalosome network

Finally, here we have the representation of the signalosome-related interactions as given by IntAct [3] (Figure 28). This can be used as a starting point to help you unveil the relationships between the different proteins involved in the complex. The CytoscapeWeb Controls box [A] can be used to try different layouts and see if you can represent the networks in a more informative fashion.

Figure 28. A graphical representation of the signalosome-related interactions.

Exercises

Exercises allow you to apply what you have learned so far on this course by providing examples of how IntAct [3]
IntAct: Molecular interactions at EMBL-EBI
Published on EMBL-EBI Train online (https://www.ebi.ac.uk/training/online)

can be used and asking you to perform the task provided. You can start by clicking on one of the exercise titles provided.

If you need help to complete this section you can look in the ‘Need some help?’ and ‘Want to know how we did it?’ sections.

The human Valosin-containing protein or VCP

Scenario

Imagine you are a researcher interested in the human Valosin-containing protein or VCP (UniProtKB [31] AC: P55072), a multimeric [54] protein involved in several cellular processes, including the endoplasmic reticulum unfolded protein response and spindle disassembly at the end of mitosis. As this protein has been inferred in a number of different processes, you would like to know which proteins have been found to interact with VCP and which ones have been detected more than once (in more than one publication or using more than one different interaction detection method).

Exercise

Find out which molecules have been found to interact with VCP in more than one publication, and which ones have been found using more than one different interaction detection method.

Need some help?

Try using the graphical visualisation tool in IntAct [3] to represent each interaction detection method as a single edge linking two different molecules.

Want to know how we did it?

2. Once you have the table with the results (104 interactions as of March 2012), you could try to work out which proteins have been found to interact with VCP more than once in the table. However, it will be much easier to detect these using a graphical representation.

3. Click on the Graph tab.

4. Once you get the network visualisation, look at the molecules that interact with VCP. The proteins are represented by balls and the small molecules [53] are represented by triangles. Check different layouts until you are happy with one.

5. To visualise those partners that have been found to interact with VCP more than once, go to CytoscapeWeb Controls and deactivate the "Merge edges" feature.

6. Proteins such as UBXN-1, -6 and -7 and AMFR have been found to interact with VCP in multiple instances, as shown by the multiple edges that connect them with VCP. You can go back to the tabular view to learn more about each interaction if you want.

The human proteasome

Scenario

You are a researcher working on the structure of the human proteasome, the main protein degradation machinery of the cell. You are specially interested in knowing if the subunits B5 (UniProtKB [31] id PSB5_HUMAN) and A7 (UniProtKB id PSA7_HUMAN) of the proteasome have been reported to be directly linked or if they are linked through a third protein. You also want to know if the interactions found have been proved to be binary interactions between the proteins involved.

Exercise

Use IntAct [3] to look for interacting partners of the proteasome subunits B5 and A7. Find out if the two proteins interact or if there are other proteins acting as linkers between them. Filter out any interactions from your results that are not binary.

Need some help?
The best way to see complex interaction relationships between two different molecules is to use a network representation. N-ary interactions can be easily filtered out from a table of results in IntAct [3].

Want to know how we did it?

1. Perform a search by putting the UniProtKB [31] ids “PSB5_HUMAN” and “PSA7_HUMAN” in the search box.

2. Once you have your table of results (88 interactions as of March 2012), you can remove the 26 N-ary interactions, where you are not sure whether the proteins are interacting as a binary pair or not, by clicking the "filter" link.

3. Now you end up with just 62 interactions. you can search the table to see whether the two proteasome subunits are interacting or if they share common interacting partners. However, it will be easier if you represent the data as a graphical network.

4. Click on the "Graph" tab.

5. In the graphical representation you can see that PSMA7 and PSMB5 are not directly interacting, but that there are three links between them through PSMA3, EPB41 and RIPK2.

Quiz: IntAct

<table>
<thead>
<tr>
<th>Questions:</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attempts allowed:</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Available:</td>
<td>Always</td>
</tr>
<tr>
<td>Pass rate:</td>
<td>75 %</td>
</tr>
<tr>
<td>Backwards navigation:</td>
<td>Allowed</td>
</tr>
</tbody>
</table>
Your feedback

Please tell us what you thought about this course. Your feedback is invaluable and helps us to improve our courses and thus enhance your learning experience.

Learn more

For additional documentation on IntAct resources follow the links below:

IntAct annotation guidelines
Curated datasets available in IntAct
Resources for developers
IntAct statistics
IntAct FAQs

Further courses:

http://www.ebi.ac.uk/training/
http://www.imexconsortium.org/about/#training
http://www.youtube.com/embed/sSSZUcm8PRM

Get help and support on IntAct

For support, submission and related enquiries, email the intact-help@ebi.ac.uk (IntAct help desk)

References

1. Orchard, S. et al. (2007) The minimum information required for reporting a molecular interaction experiment


Contributors


Pablo Porras got his PhD in 2006 in the University of Córdoba, Spain, having done research about transmembrane protein translocation and redox homeostasis. After that, he moved to Berlin to work in the Neuroproteomics group of the Max Delbrueck Center, getting involved in projects dealing with interactomics, neurodegenerative diseases and the ubiquitin-proteasome system. During this postdoc, he faced the problem of how to represent and analyze molecular interactions data. This experience proved to be of great value once he joined the EBI to work as a scientific curator in the molecular interactions database IntAct in 2011.

Source URL: https://www.ebi.ac.uk/training/online/course/intact-molecular-interactions-ebi

Links
[1] https://www.ebi.ac.uk/training/online/trainers/pporras
[3] https://www.ebi.ac.uk/training/online/glossary/intact
[4] https://www.ebi.ac.uk/training/online/glossary/curation
[5] https://www.ebi.ac.uk/training/online/glossary/annotation
[8] https://www.ebi.ac.uk/training/online/glossary/controlled-vocabulary
[9] https://www.ebi.ac.uk/training/online/glossary/imex
[12] https://www.ebi.ac.uk/training/online/glossary/psicquic
[13] https://www.ebi.ac.uk/training/online/course/intact/references
[14] https://www.ebi.ac.uk/training/online/glossary/yeast-two-hybrid
[15] https://www.ebi.ac.uk/training/online/glossary/tandem-affinity-purification
[16] https://www.ebi.ac.uk/training/online/glossary/amino-acid
[17] https://www.ebi.ac.uk/training/online/glossary/xml