Complex Portal: Quick tour

Birgit Meldal [1]

- Systems
- Ontologies
- Proteins
- Structures
- Cross-domain

- Beginner
- 0.5 hour

This quick tour provides a brief introduction to EMBL-EBI's Complex Portal [2]: a manually curated, encyclopedic resource of macromolecular complexes from a number of key model organisms.

The accompanying webinar can be found here [3].

This course was last updated in June 2018.

Learning objectives:

- Outline the scope of the available data in the Complex Portal
- Describe how to search for macromolecular complexes of interest
- List manual and computational methods for downloading Complex data
- Outline how to submit curation requests for Complexes
- Identify sources of more information about the Complex Portal

What is the Complex Portal?

The Complex Portal is a manually curated, encyclopedic resource of macromolecular complexes from a number of
key model organisms. Complexes contain proteins as well as small molecules [4] and nucleic acids as long as they are integral parts of the complex.

Evidence for complexes is derived from one or more of the following methods:

1) Physical molecular interaction and functional evidence extracted from the literature

2) Inferred by curators from information on orthologues in closely-related species or paralogues in the same species where the original complex has experimental evidence

3) Inferred by curators from background biological knowledge based on other evidence (for example, pharmacological studies).


The data in the Complex Portal can be searched and downloaded interactively or by programmatic access.

The details pages for any given complex are structured by topics: Function, Properties, Expression, Cellular Location, Diseases/Pathology, Further Reading. They contain an interactive viewer with details of stoichiometry [7] and binding regions and provide further graphical information for Reactions and Pathways in Reactome [8], structures in the PDBe [9] and expression data from the Expression Atlas [10].

It also provides cross-references to a given complex from many other databases, such as ChEMBL [11], Electron Microscopy Data Base [12] and MatrixDB [13]. The Portal also contains annotations to all three classes of The Gene Ontology [14], EC numbers [15], diseases and literature evidence in EuropePMC [16].

Content is added by expert curators and direct user requests are prioritised. It shares a database with the IntAct [17] molecular interaction database.

Why do we need the Complex Portal?

The Complex Portal provides annotations for the whole complex (i.e. the functional biological entity), along with descriptions of their function and other properties. It differs from UniProt [18] because the latter only provides annotations for the protein subunits.

What is a macromolecular complex?

A macromolecular complex is a stable set of (two or more) interacting protein molecules which:

1. Can be co-purified, and
2. Have been shown to exist as a functional unit in vivo.

Non-protein molecules (e.g. small molecules [4], nucleic acids) may also be present in the complex if they are an integral part of the complex.

The following are not considered to be stable complexes:

- Two or more proteins associated in a pulldown / coimmunoprecipitation with no functional link.
- Genetic interaction evidences.
- Transient interactions such as enzyme-substrate or receptor-ligand complexes that have a 1:1 relationship

There are some exceptions to these rules, as follows:
• An enzyme or receptor is a complex in its own right
• An enzyme or receptor is an obligate complex that requires substrate/ligand binding for its functional assembly

For example:

• Maltose transporter (the enzyme) requires maltose (the substrate) binding for formation of the functional enzyme (CPX-1932 [19])
• PDGF receptors: the PDGF ligands are obligate dimers and receptors require ligand binding for dimerisation of the two receptor chains forming an obligate tetramer (e.g. CPX-2885 [20])

What can I do with the Complex Portal?

With the Complex Portal you can:

• search for complexes by name, synonym [21], identifier [22], species and cross-reference [23] identifier;
• view complex details;
• save favourite complexes in a basket;
• download complexes in PSI-MI standard format [24] or in tabular format [25] from our FTP site [26] or in JSON [27] format from our webservice [28];
• request curation [29] of new complexes.

Searching and visualising data from the Complex Portal

You can search complexes using:

• Any name(s) or identifier [22](s) of participants in the complex (for example, using ChEBI [30], RNACentral or UniProt [18] IDs)
• Any name(s) or identifier(s) of the complex itself, including identifiers from external databases such as ChEMBL [31], The Gene Ontology [32], The PDB [33] or Reactome [34]
• Gene Ontology terms
• The species

A search can be performed via the search box on the homepage (Figure 1), the small box on top of all other pages, or by altering the URL.

The search is 'exact' but allows the use of Boolean [35] operators [36] ("OR" by default, but also "AND" and "NOT") and the Wildcard character (*). If you want to get results for complexes containing isoforms and post-processed chains append the UniProt AC with the Wildcard, e.g. P21127* to retrieve complexes containing isoform P21127-12 [37].

You can search with single terms, lists of identifiers or a combination of specified fields.

To see how the search results will be presented (and what filters are available), click on one of the examples listed below the search box.
Figure 1 Complex Portal homepage

Search Results

The results of a search (Figure 2):

- Are ordered by relevance
- Are paginated displaying 10 hits per page
- Can be further filtered by species, Biological Role (e.g. enzyme, cofactor [38]) or Interactor Type (e.g. protein, small molecule, RNA)
Complexes can be further explored by clicking on the complex name.

Total number of results: 2420

Golgi SNARE complex SED5-GOS1-YKT6-SFT1

Complex AC: CPX-1855 / Organism: (Saccharomyces cerevisiae (strain ATCC 204508 / S288c); 559292)
Description: When membranes are tethered, their soluble N-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs) bind to each other in trans, that is, with SNAREs anchored in each apposed tethered membrane. The Golgi membrane tangl (1) SNAREs bind.

Serine/threonine-protein phosphatase PP2A variant 1

Complex AC: CPX-1856 / Organism: (Saccharomyces cerevisiae (strain ATCC 204508 / S288c); 559292)
Description: A serine/threonine phosphatase, the activity of the catalytic subunit of which is highly regulated by members of a family of regulatory subunits, which determine the substrate specificity, subcellular localization and catalytic activity of the PP2A.

Serine/threonine-protein phosphatase PP2A variant 2

Complex AC: CPX-1857 / Organism: (Saccharomyces cerevisiae (strain ATCC 204508 / S288c); 559292)
Description: A serine/threonine phosphatase, the activity of the catalytic subunit of which is highly regulated by members of a family of regulatory subunits, which determine the substrate specificity, subcellular localization and catalytic activity of the PP2A.

Serine/threonine-proprotein phosphatase PP2A variant 3

Complex AC: CPX-1858 / Organism: (Saccharomyces cerevisiae (strain ATCC 204508 / S288c); 559292)
Description: A serine/threonine phosphatase, the activity of the catalytic subunit of which is highly regulated by members of a family of regulatory subunits, which determine the substrate specificity, subcellular localization and catalytic activity of the PP2A.

GET4-GET5 transmembrane domain recognition complex

Complex AC: CPX-1861 / Organism: (Saccharomyces cerevisiae (strain ATCC 204508 / S288c); 559292)
Description: Escorts newly-synthesised tail-anchored proteins to GET3 (C12154a) of the GET complex (CPX-196) for insertion into the endoplasmic reticulum membrane. The complex acts both as an active tethering device that bringsGET3 in proximity with its substrate.

PeB0W complex

Complex AC: CPX-1862 / Organism: (Saccharomyces cerevisiae (strain ATCC 204508 / S288c); 559292)
Description: Role in coordinating ribosome biogenesis with cell cycle progression, necessary for the conversion of 27SA3 pre-rRNA to 25SBS pre-rRNA.

TAP42-RDD1-SIT4 phosphatase complex

Complex AC: CPX-1863 / Organism: (Saccharomyces cerevisiae (strain ATCC 204508 / S288c); 559292)
Description: A serine/threonine protein phosphatase complex that plays a major role in TOR1/2 (PFS16P/PS2600)-mediated signaling and gene expression. Rapamycin causes release of the phosphatase-RBD dimer from TAP42.

SIT4-SAP155 phosphatase complex

Complex AC: CPX-1864 / Organism: (Saccharomyces cerevisiae (strain ATCC 204508 / S288c); 559292)
Description: Forms in response to nutrient deprivation, mediates G1 to S cell cycle progression and a number of signaling events controlled by the target of rapamycin TOR signaling cascade including growth, budding, HCR gene expression and G2/M-regulated transil...
Exploring Complexes

Each complex is extensively annotated:

- The title contains the recommended name, accession number, species, link to experimental evidence and options to save the complex in a basket or download the complex file as JSON (Figure 3).
- The recommended name is the name most commonly used in the literature – all alternatives are captured as synonyms at the bottom of the page.
- Each complex has a unique accession number (ComplexAc), in the format: CPX-xxxx (where x = {1-9}). The ComplexAc's are also versioned but only the latest version is displayed on the website. Previous versions can be retrieved from our ftp site.

Figure 3 Hemoglobin complex (CP-2158): page header, ComplexViewer and participant table

- The ComplexViewer automatically displays topology, stoichiometry and internal interaction details as far as they are known. The viewer is interactive allowing detailed explorations of complexes (Figure 3).
- The Participant Table next to the ComplexViewer lists all participants including their stoichiometry (if known), links to the reference database for each molecule and a legend for the viewer (Figure 3).
- Further down the page you will find information on the Function, Properties, Expression and Cellular Location, Diseases and Pathologies, and Additional Information associated with the complex.
Using the ComplexViewer

The ComplexViewer [1] [41] allows you to explore complex topology in greater detail.

Four different features can be selected from the dropdown menu (Figure 4):

1. Binding regions.
4. Display unique interactors in the same colour (up to 20 colours).

---

**Figure 4** ComplexViewer: Drop-down menu to select different complex features.

- All nodes can be dragged around with the mouse to manipulate the layout.
- Known internal interactions are represented by edges between participants (Figure 5a). If the internal topology for any participants is unknown no edges are displayed (Figure 5b).
Clicking on the circles expands proteins into a bar (Figure 6a and b) depicting various attributes (see below).

- The <Expand all> button (Figure 4) below the viewer allows you to expand all proteins simultaneously.
- Holding <shift> while repeatedly clicking on the bar zooms into the protein sequence and eventually shows the amino acid residues (Figure 6c).
- The <Collapse all> button (Figure 4) resets the viewer to the default settings.

Annotated regions (binding evidences, Superfamilies, UniProt features) are coloured in and binding regions are connected with broad edges in the expanded view. Feature names and ranges are displayed with a tooltip by hovering the mouse over it. If there is evidence for binary interactions but not for the regions that interact with each other the edges are connected at position 0 of the protein(s) (Figure 6b-f).
**Function**

All complexes are manually annotated with a **function statement** followed by [Gene Ontology][32] annotations to the Molecular Function (where applicable) and Biological Process class. If applicable, links are provided to [IntEnz][44] for enzymes and **ligands, antagonists** and **agonists** are listed with links to their reference databases (Figure 7).

---

**Figure 6** ComplexViewer: Topology and binding features.
**Function**

Adult hemoglobin A (HbA) is expressed in erythrocytes in the bone marrow. Binds oxygen in the lungs and transports it to the various peripheral tissues. Transports CO2 from cells back to the lungs. It appears in late pregnancy and becomes the dominant hemoglobin type in adults, replacing fetal hemoglobin (CPX-2932 & CPX-2933).

<table>
<thead>
<tr>
<th>GO - Molecular Function (6)</th>
<th>GO - Biological Process (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>oxygen transporter activity of</td>
<td>carbon dioxide transport of</td>
</tr>
<tr>
<td>ferrous iron binding of</td>
<td>heme oxidation of</td>
</tr>
<tr>
<td>ferric iron binding of</td>
<td>positive regulation of cell death of</td>
</tr>
<tr>
<td>nitric oxide binding of</td>
<td>oxygen transport of</td>
</tr>
<tr>
<td>oxygen binding of</td>
<td>nitric oxide transport of</td>
</tr>
<tr>
<td>carbon dioxide binding of</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 7** Hemoglobin complex: Functions

A table of cross-references to equivalent complex in [Reactome](https://www.reactome.org) [34] together with the Reactome Pathway widget are dispayed (human complexes only) (Figure 8).
### Properties

Where applicable, further **physical and biological properties** and cross-references to [PDBe](https://www.ebi.ac.uk/pdbe) [45] and [EMBD](https://www.ebi.ac.uk/EMBL-EBI) are listed and 3D structures displayed using the LiteMol viewer developed by PDBe (Figure 9).
Figure 9 Hemoglobin complex: Properties and 3D structure viewer.

Gene expression and subcellular location

Gene expression [46] data (Figure 10) is displayed by means of the widget developed by the Expression Atlas [47] team.

Subcellular [48] locations are annotated using the Gene Ontology [32] Cellular Compartments class.
Expression and Cellular Location

Gene Expression Map

RNA-seq from 53 human tissue samples from the Genotype-Tissue Expression (GTEx) Project

Figure 10 Hemoglobin complex: Expression data and subcellular location.

Diseases and pathologies

Human complexes, where applicable, are annotated with their related diseases and cross-referenced to the Experimental Factor Ontology [49] (EFO) (Figure 11).

Complex that are drug targets are cross-referenced to the ChEMBL [31] database.
Diseases and Pathologies

Hemoglobin complex: Disease annotations and drug target.

Additional information

Additional information, including cross-reference [23] to MatrixDB [13] and the Protein Ontology [50] may be added (Figure 12). Literature references to other relevant publications such as reviews and functional papers are linked to EuropePMC, and all complexes are given a systematic name which is the concatenation of their gene names and it includes stoichiometry [7] – these are based on the Reactome [34] naming rules. [2] [41]
Complex Portal: Quick tour
Published on EMBL-EBI Train online (https://www.ebi.ac.uk/training/online)

Additional Information

Getting data from the Complex Portal

To retrieve data from the Complex Portal you can download complexes in PSI-MI XML 2.5 and XML3.0 [51] and in tabulated formats from the Complex Portal FTP site [52] and in JSON [27] format from the Complex Portal webservice [53] (Figure 13).

https://www.ebi.ac.uk/complexportal/download:

- PSI-MI XML 2.5
- PSI-MI XML 3.0
- ComplexTab
- PSI-MI JSON

1 Folder per species
Please use PSI-MI XML 3.0

REST-API:
http://www.ebi.ac.uk/intact/complex-ws/

ftp://ftp.ebi.ac.uk/pub/databases/intact/complex/current/

Figure 12 Hemoglobin complex: Literature references, synonyms and systematic name.

Figure 13 Programmatic access options.
Complexes are saved as one file per complex and organised in folders for each species.

The webservice has four methods:

- /search/
- /details/
- /export/
- /complex/

## Contributing to the Complex Portal

To make a request for us to curate a particular complex or send correction comments, please fill out our contact form (which you can access via the yellow tile on the homepage) and supply as much detail as possible, including:

- PubMed Identifier (PMID) that provides evidence for the existence of the complex and its function;
- list of complex participants, ideally with UniProt ACs;
- suggested name and synonyms;
- suggested definition/functional annotation/properties.

## Stay in touch

You can sign up to our newsletter via the link on the homepage or follow us on Twitter.

## Get help and support on the Complex Portal

### Support

- We have extensive documentation and help pages.
- For further queries and requests please contact our helpdesk.

## Your feedback

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

## References and Funding
Papers mentioned in this Quick Tour


How to cite Complex Portal


Funding

- European Molecular Biology Laboratories Core Funding.
- Biotechnology and Biological Sciences Research Council [BB/L024179/1].
- National Institutes of Health [268201000035C].
- Wellcome Trust [WT101477MA]

Contributors

Birgit Meldal [1]

EMBL-EBI
Scientific Database Curator (Molecular Interaction & Complexes) - Orchard Team: Molecular Interactions

Birgit Meldal joined EMBL-EBI in 2012 as a Scientific Curator for the molecular interactions database IntAct. Her work focuses on curating macromolecular complexes into the Complex Portal, and she also contributes to the Gene Ontology. Birgit is a trainer for the IntAct and Reactome databases. She gained her PhD from the University of Southampton (in collaboration with the Natural History Museum, London) in 2004, having built the first comprehensive molecular phylogeny of the phylum Nematoda. She followed this with postdoctoral positions at the University of Cambridge working on the function of the breast cancer gene EMSY and on molecular epidemiology and host-virus interactions of transfusion-transmissible, blood-borne viruses (Hepatitis B and Hepatitis E virus).