

# Complex Portal 2018: extended content and enhanced visualization tools for macromolecular complexes

Birgit H.M. Meldal<sup>1,\*</sup>, Hema Bye-A-Jee<sup>1</sup>, Lukáš Gajdoš<sup>2</sup>, Zuzana Hammerová<sup>2</sup>, Aneta Horáčková<sup>2</sup>, Filip Melicher<sup>2</sup>, Livia Perfetto<sup>1</sup>, Daniel Pokorný<sup>2</sup>, Milagros Rodriguez Lopez<sup>1</sup>, Alžběta Türková<sup>2</sup>, Edith D. Wong<sup>3</sup>, Zengyan Xie<sup>4</sup>, Elisabeth Barrera Casanova<sup>1</sup>, Noemi del-Toro<sup>1</sup>, Maximilian Koch<sup>1,5</sup>, Pablo Porras<sup>1</sup>, Henning Hermjakob<sup>1,6</sup> and Sandra Orchard<sup>1</sup>

<sup>1</sup>European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK, <sup>2</sup>Department of Biochemistry and National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno-Bohunice, Czech Republic, <sup>3</sup>Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305-5477, USA, <sup>4</sup>Chongqing Key Laboratory on Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, No. 2 Chongwen Road, Nan'an District, Chongqing 400065, China, <sup>5</sup>Novartis Institutes for BioMedical Research (NIBR), Basel, Canton of Basel-Stadt, Switzerland and <sup>6</sup>State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Life Omics, Beijing 102206, China

Received September 13, 2018; Revised October 05, 2018; Editorial Decision October 08, 2018; Accepted October 09, 2018

## ABSTRACT

The Complex Portal ([www.ebi.ac.uk/complexportal](http://www.ebi.ac.uk/complexportal)) is a manually curated, encyclopaedic database that collates and summarizes information on stable, macromolecular complexes of known function. It captures complex composition, topology and function and links out to a large range of domain-specific resources that hold more detailed data, such as PDB or Reactome. We have made several significant improvements since our last update, including improving compliance to the FAIR data principles by providing complex-specific, stable identifiers that include versioning. Protein complexes are now available from 20 species for download in standards-compliant formats such as PSI-XML, MI-JSON and ComplexTAB or can be accessed via an improved REST API. A component-based JS front-end framework has been implemented to drive a new website and this has allowed the use of APIs from linked services to import and visualize information such as the 3D structure of protein complexes, its role in reactions and pathways and the co-expression of complex components in the tissues of multi-cellular organisms. A first draft of the complete complexome of *Saccharomyces cerevisiae* is now available to browse and download.

## INTRODUCTION

All biological processes rely on the interaction of proteins with other cellular components and in many cases these proteins function as part of macromolecular complexes. These complexes may be composed only of proteins or also contain nucleic acids and/or small molecules. Complexes can be stable or transient in nature and vary in their function and composition depending on the tissue type, developmental stage and biological process. Many complexes have been well characterized by experimental methods while the existence or composition of other complexes has been inferred from related evidences.

The Complex Portal ([www.ebi.ac.uk/complexportal](http://www.ebi.ac.uk/complexportal)) is a manually curated, encyclopaedic resource that summarizes all functional aspects of stable macromolecular complexes (1). Since our last update, we have improved the data structure by providing more data download formats, new identifiers and expanded our data content in terms of both, the number of complexes curated and the species represented in the resource. We have significantly improved interoperability with other resources, such as the Gene Ontology (2), and updated the website to provide better data access and more visualization options, all making the Complex Portal more fully compliant with the FAIR principles of Findability, Accessibility, Interoperability and Reusability (3).

The characterization of the complex composition in a model organism can be used to predict the existence of similar complexes in other species (4). Such inferences can contribute to our understanding of evolutionary development

\*To whom correspondence should be addressed. Tel: +44 1223 494107; Email: [bmeldal@ebi.ac.uk](mailto:bmeldal@ebi.ac.uk)

at the molecular level. We are currently developing a tool that can predict orthologous complexes in other species as well as find homologous complexes that have already been curated in the Complex Portal.

## CONTENT

### Curation update

The working definition of a protein complex is that it should contain two or more macromolecules for which there is evidence (experimental or inferred) that these molecules stably interact with each other and have a demonstrated molecular function. Each entry (Figure 1) describes the composition, topology and function of a single complex from a defined species and links out to external databases that provide domain-specific information of the complex, such as structural details from PDB (5) or the role of the complex in metabolic reactions or signalling pathways in Reactome (6). A complex is uniquely defined by its protein components using canonical UniProt identifiers, i.e. the protein sequence selected as the reference sequence in a UniProtKB/Swiss-Prot entry. Isoform-specific complex variants are only created when the function of the complex is specific for a given isoform. The Cyclin L-CDK11 complexes provide interesting examples where this is the case (CPX-341, CPX-343, CPX-344, CPX-347). Complex participants are described using the UniProtKB identifiers for post-processed chains when complexes are formed only by a specific cleavage product, for example the amyloid-beta protein 40/42 heterodimer (CPX-1062). Post-transcriptional modifications are described only if they are essential to the formation of the complex as, for example, in the case of the formation of the SMAD1 homotrimer (CPX-144).

The Complex Portal currently focuses on curating stable protein complexes from a limited number of model organisms to provide a reference from which complexes in other species can be inferred and the evolutionary development of a conserved functional unit can be traced. As of release 219 (8 October 2018), 2472 complexes have been curated and released.

Recent curation priorities have included:

- a. Mammalian complexes with available 3D structures. This is a long-term program in collaboration with CEITEC and PDBe. 297 curated mammalian complexes have crystal structures, 248 of them are human.
- b. Human and mouse epigenetic complexes. This set can be accessed by searching on the high-level Gene Ontology term 'chromatin organization' (GO:0006325)
- c. *Saccharomyces cerevisiae*: All known complexes (589) for the budding yeast have been curated, in collaboration with UniProt (7) and the Saccharomyces Genome Database (8). Newly described yeast complexes are being added into SGD and the Complex Portal in direct collaboration.
- d. *Arabidopsis thaliana*: 221 hormone signalling complexes of *Arabidopsis thaliana* are the first plant complexes to be added to the database.
- e. *Escherichia coli*: known *E. coli* complexes are being curated with a view to completion of the complexesome in 2019.

### Quality control

All entries in the Complex Portal are manually annotated and carefully checked by a second curator before publication. Entries are standards compliant, for example using stable identifiers (UniProt for proteins, ChEBI for small molecules (9), RNAcentral for ncRNAs (10) and controlled vocabularies such as the Gene Ontology and PSI-MI ontology (11). Curators work to a detailed curation manual that is constantly updated and reviewed as we discover new data concepts. A summary of our curation practices is available via the <Documentation> tile on the public website; this enables users to understand the annotation process. We have implemented new syntax checks that are run before each release cycle, making the data more robust to human error. Following strict rules and quality control checks means that our entries are consistent and of high quality.


### Complex identifiers and versions

The new complex-specific stable identifiers have a version number that can be incrementally increased to indicate a major change to the underlying record, thus making the resource more fully compliant with the FAIR standard. Identifiers are of the type: CPX-xxxx.y, where xxxx is the running number and y the version. The previously published EBI-xxxxxxx identifiers are retained as secondary identifiers and can still be retrieved as they remain in the download files and are indexed in the webservice. The new identifiers also provide an easy distinction from IntAct experimental evidences with which they previously shared an identifier space. A complex version will be incrementally increased if the composition of the complex has changed, for example additional participants have been added. A change in stoichiometry will currently not trigger reversioning.

If a complex has been obsoleted without replacement, the accession number (AC) will no longer be searchable in the current release files or the website but will remain accessible from the archive on our ftp site (<ftp://ftp.ebi.ac.uk/pub/databases/intact/complex/current>). If a complex is split into several new entries the original AC will be retained as a secondary AC on the new complexes and, similarly, if two complexes are merged, the original AC from one complex becomes a secondary AC on the retained and updated complex. The website only displays the current AC whilst previous entry versions and obsoleted ACs are stored in the archive that can be retrieved from the ftp site.

### Evidence and data provenance

Many of the curated complexes are unambiguous in content and 63% are supported by good experimental evidence, which is collated and cross-referenced to the IntAct molecular interaction database, PDBe or EMBD (23). In many cases, alternative compositions exist for a complex, often caused by proteins that arose from gene duplications. Each possible variant is curated as long as there is reason to believe these exist in nature. Other complexes, however, have little or no experimental evidence and their component content is currently poorly understood. Curator judgement is often required to make a decision on the essential core components of a complex, and it is accepted that many of these


Complex Portal

🔍

Examples: GO:0016491 Ndc:80 Q05471 PCNA CPX-2158 nuclear pore Q15554, P54274, Q96AP0


Home
About
Feedback
Basket 0

## Hemoglobin HbA complex

ComplexAc: CPX-2158  
*Homo sapiens*; 9606

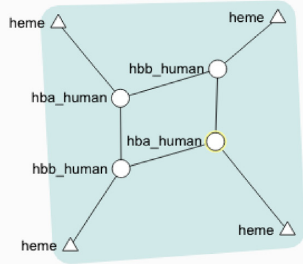
Download
Basket

Evidence by physical interaction  
[evidence](#) [in IntAct](#) [EBI-1029796](#)



## Participants

☰
[Go to](#)



Expand All
Collapse all
Export SVG
Binding Region ▾

Legend	Description	Stoichiometry
○	<b>protein - HBA1 (unspecified role)</b> <a href="#">P69905</a> Hemoglobin subunit alpha	2
○	<b>protein - HBB (unspecified role)</b> <a href="#">P68871</a> Hemoglobin subunit beta	2
△	<b>small molecule - heme (cofactor)</b> <a href="#">CHEBI:30413</a> heme	4

## Function

☰
[Go to](#)

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 CO<sub>2</sub> 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).

### GO - Molecular Function (6)

- [oxygen transporter activity](#)
- [ferrous iron binding](#)
- [ferric iron binding](#)
- [nitric oxide binding](#)
- [oxygen binding](#)

### GO - Biological Process (5)

- [carbon dioxide transport](#)
- [heme oxidation](#)
- [positive regulation of cell death](#)
- [oxygen transport](#)
- [nitric oxide transport](#)

[Show all](#)

**Figure 1.** Screenshot of the details page of Haemoglobin HbA complex (CPX-2158).

entities will have to be revisited and updated as new information on their composition becomes available, with a new version of a complex being issued when appropriate. All entries are annotated with Evidence and Conclusion Ontology (ECO) codes (12) that identify the level of experimental evidence that exists for any given complex. New codes that are specific for complexes and pathways have been added to the ontology:

- *ECO:0000353 physical interaction evidence used in manual assertion* has been retained to indicate cases where the complex is fully experimentally validated.
- *ECO:0005610 biological system reconstruction evidence based on homology evidence used in manual assertion* and appropriate child terms are used to indicate orthologous complexes (inter-species comparison) and paralogous complexes (intra-species comparison where the vari-

ants are related to proteins that arose from gene duplications but retain the same function) if experimental evidence exists for a related complex not only containing orthologous/paralogous participants but that also retains compositional and functional homology.

- *ECO:0005543 biological system reconstruction evidence by experimental evidence from mixed species used in manual assertion* is used if the researcher has used constructs from several species to reconstruct a complex and if the orthologous sequences are identical or very similar.
- *ECO:0005547 biological system reconstruction evidence based on inference from background scientific knowledge used in manual assertion* describes complexes where the expert community believes the complex exists but only limited or circumstantial experimental evidence exists (e.g. pharmacological or ligand-binding assays).

### Variants, molecule sets and complexes as participants

Complex variants can be the result of (a) duplication of one or more genes, (b) the existence of isoforms that result in different functions of the complex or (c) compositional variations related to tissue type or developmental stage of the organism. In these cases, we normally create separate entries, keeping the root recommended name constant and adding an extension that describes the variation (either using terms used by the community or the gene symbols that vary). In a very limited number of cases the possible number of combinations is very large, leading to a combinatorial explosion in the number of possible variants, e.g. the yeast ribosomal sub-complexes (CPX-1599 & CPX-1601) in which most of the subunits have 2 paralogous proteins. In these cases, the protein paralogues are grouped as molecule sets. Each set only contains proteins that have the same evolutionary origin and the same function. The resource now also includes examples of hierarchical builds of complexes assembled from functionally active sub-complexes, for example the yeast CMG-Pol epsilon complex (CPX-1556).

### WEBSITE SEARCH AND DATA VISUALIZATION

The original Complex Portal website (1) consisted of a simple, mainly text-based presentation and data access was provided only by ftp. This has now been updated to provide a dynamic, interactive, responsive user interface, created using a component-based JS front-end framework served by an enhanced REST API (<https://www.ebi.ac.uk/intact/complex-ws/>). The layout and functionality of the site has been developed using iterative user-centric design principles. 2D visualization of the topology of each complex is supplied by the ComplexViewer (13) which provides a scalable vector graphic (SVG) element within a web page. This viewer enables the visualization of the topology of complexes containing multiple participants, sequence features such as binding domains relevant to complex formation, a range of biomolecules as interactor types (proteins, small molecules, nucleic acids, sub-complexes), discontinuous sequence features and stoichiometry information.

The existing simple search option has been supplemented by an organism-centric view which enables researchers to download all the complexes from a particular species in a

choice of formats. This can be accessed from one of a series of ‘tiles’ visible on the front-page, designed as easy-to-use entry points. Other tiles lead users to training materials, documentation, a ‘basket’ (a user-defined list of stored complexes that can be populated from the individual complex details pages), and a webform for curation requests or general feedback about the service.

Use of the component-based JS front-end technology enables the easy integration of external APIs, enabling Complex Portal entries to be enriched with data from collaborating resources (Figure 2). The Reactome viewer (14) allows browsing through pathways related to the complex of interest. Clicking on any element changes the focus of the diagram to this element and clicking on the Reactome symbol opens it in the Reactome website providing their full functionality. PDBe’s LiteMol viewer (15) allows detailed investigation of 3D structures where these are available. Users can further investigate the original structural data entry via the link to the PDBe website. Finally, the Body Map tool from the Gene Expression Atlas displays tissue-specific gene expression data for all members of the complex for most multi-cellular species (16).

### DOWNLOAD OPTIONS

All complexes are available for download in the molecular interaction standard formats PSI-MI XML2.5 (17) and PSI-MI XML3.0 (18). PSI-MI XML2.5 is the older, more widely adopted format, but has limitations in the representation of complexes. Its schema was designed to represent experimental interactions, therefore an experiment description is required for each interaction. This is not appropriate for abstracted interactions such as curated complexes, where data from multiple publications may be combined, therefore the Complex Portal became an early adopter of the updated PSI-MI XML3.0 format, as being more appropriate for this data type.

Many visual applications now preferentially use JSON as an input format including the ComplexViewer which inputs newly-developed MI-JSON (13). This is written by the Complex Portal webservice (1) and *export* and *complex* methods have been added to the existing *search* and *details* methods (<https://www.ebi.ac.uk/intact/complex-ws/>).

Additionally, on request from our users, a tab-delimited format, ComplexTab, has been developed, modelled on the HUPO-PSI MITAB format (17). ComplexTab exports data in 18 columns (Table 1) listing details for each complex in a single, tab-delimited line with pipes separating elements in the same column. Fields including names, lists of synonyms, participants and cross-references, description, properties and assembly. It does not include any participant feature descriptions including binding regions or required PTMs as they cannot be simply assigned to the right complex member. If more detailed data is required the user is advised to use the MI-JSON or PSI-MI XML3.0 format.

Access to the data is available in all formats described above via the <Programmatic Access> tile on the homepage and access to the PSI-MI XML and MI-JSON files from the header of each individual complex page. On the ftp site (<ftp.ebi.ac.uk/pub/databases/intact/complex/current/>), each complex is stored in a unique XML file,

## A Pathways

Complex Identifier	Complex Name	Pathway Identifier
<a href="#">R-HSA-1237312</a>	Protonated Carbamino DeoxyHbA [cytosol]	<ul style="list-style-type: none"> <li><a href="#">R-HSA-1237044</a></li> <li><a href="#">R-HSA-1247673</a></li> </ul>
<a href="#">R-HSA-1237320</a>	OxyHbA [cytosol]	<ul style="list-style-type: none"> <li><a href="#">R-HSA-1237044</a></li> <li><a href="#">R-HSA-1247673</a></li> </ul>
<a href="#">R-HSA-2168856</a>	ApoHemoglobin [extracellular region]	<ul style="list-style-type: none"> <li><a href="#">R-HSA-2168880</a></li> </ul>
<a href="#">R-HSA-2168866</a>	Methemoglobin [extracellular region]	<ul style="list-style-type: none"> <li><a href="#">R-HSA-2168880</a></li> </ul>

Selected Complex:

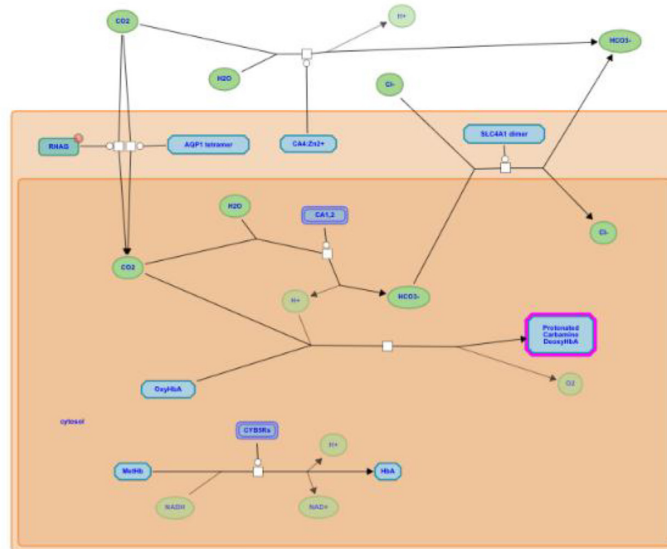
**Stable Identifier:** [R-HSA-1237312](#)

**Name:** Protonated Carbamino DeoxyHbA [cytosol]

Selected Pathway:

**Stable Identifier:** [R-HSA-1237044](#)

**Name:** Erythrocytes take up carbon dioxide and release oxygen



R-HSA-1237312 - 1 entity flagged

B

## Properties

[Go to](#)

Two alpha chains and two beta chains. Each chain has a heme b group attached to it containing either an Fe<sup>2+</sup> or Fe<sup>3+</sup> ion. Oxygen only binds to Fe<sup>2+</sup> ions, not Fe<sup>3+</sup> ions. CO<sub>2</sub> binds directly to the protein chains and therefore does not compete with oxygen binding. MW = 64 kD

## Assembly

Heterotetramer.

## 3D Structure



## Curated structures (3):

[2dn1](#)



[2dn2](#)



[2dn3](#)



C

## Expression and Cellular Location

[Go to](#)

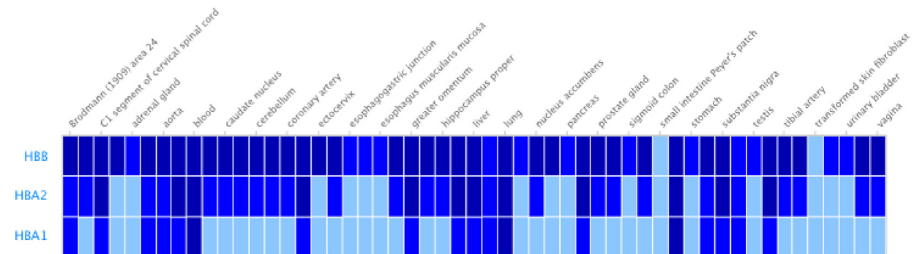
### Gene Expression Map

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

Showing 3 genes:



Expression level in TPM  
0 to 315,499



Select genome browser to view tracks | Download

This page is a summary of the data held in [Expression Atlas](#) for this gene; click here for the full record.  
Please send any queries or feedback via [the EBI Support & feedback form](#).

## GO - Cellular Component (1)

[hemoglobin complex](#)

Figure 2. (A–C) Reactome, PDBe and Gene Expression Atlas tools as integrated in Complex Portal for Hemoglobin HbA complex.

**Table 1.** Descriptions for each of the 18 fields in the new ComplexTab format

Field name	Field description
<b>Unique identifier for complex</b>	The unique primary identifier of the complex in the Complex Portal database.
<b>Recommended name</b>	The most commonly used description for the complex.
<b>Aliases for complex</b>	All regularly used synonyms. Multiple annotations are separated by '
<b>Taxonomy identifier</b>	NCBI taxonomy identifier for complex.
<b>Identifiers (and stoichiometry) of molecules in complex</b>	The list of molecules and their stoichiometry (in parenthesis) that make up the complex. For proteins, the accession should be UniProtKB, for chemical entities, ChEBI, and for RNA, RNACentral. When the stoichiometry of a molecule is unknown, it is represented by (0). Multiple identifiers are separated by '
<b>Confidence</b>	Evidence and Conclusion Ontology AC and term name.
<b>Experimental evidence</b>	Identifier of the experimental evidence for this complex in an IMEx member molecular interaction database, PDB or EMDB.
<b>GO annotations</b>	List of Gene Ontology terms the complex is annotated to. Multiple annotations are separated by '
<b>Cross references</b>	List of accession codes for the complex in related databases, such as PDB, Reactome or MatrixDB, as well as links to diseases caused by changes in the complex. Multiple annotations are separated by '
<b>Description</b>	A free-text description of the function(s) and role(s) of the complex.
<b>Complex properties</b>	A free-text description of the physical properties of the complex such as assembly and internal topology details, size, molecular weight or binding to co-factors.
<b>Complex assembly</b>	Description of the assembly type, such as Homodimer, or Heterohexamers.
<b>Ligand</b>	List of natural ligands of the complex. Multiple annotations are separated by '
<b>Disease</b>	List of description of diseases caused by changes in the complex. Multiple annotations are separated by '
<b>Agonist</b>	List of natural agonists of the complex. Multiple annotations are separated by '
<b>Antagonist</b>	List of natural antagonists of the complex. Multiple annotations are separated by '
<b>Comment</b>	List of comments that do not fit any of the other defined fields. Multiple annotations are separated by '
<b>Source</b>	Group the original curator of the complex is associated with.

grouped into folders by species. ComplexTab data is available in a single file per species. The webservice provides a MI-JSON file per complex on demand. Download options for Search Results and the Basket are under development.

## COLLABORATION AND COMMUNITY INVOLVEMENT

The Complex Portal entries are designed to integrate data from multiple resources which hold information on these biologically important entities. Cross-references to UniProt provide information on the component proteins, whilst reciprocal cross-references in UniProt inform users as to the complex membership of individual proteins. Links to Reactome provide information on the role protein complexes play in pathways and cellular processes. All Reactome reactions have now been integrated into WikiData and use the Complex Portal ACs as primary complex identifiers where available (see <https://www.wikidata.org/wiki/Q50250937>). Links to ChEMBL enable the researcher to access a description of drugs which bind to and modulate the activity of a protein complex. The Complex Portal is a member of the Gene Ontology (GO) Consortium: Every complex is annotated to the relevant Gene Ontology terms for the complex as a whole, enabling hierarchical searching of the data set. The GO annotation guidelines have been updated to allow use of Complex Portal identifiers as GO annotation objects in GO editorial tools, such as Protein2GO. Annotations to Complex Portal identifiers are exported to and available through GO annotation files (GPAD, GPI) and can be searched and viewed via the web application QuickGO (<https://www.ebi.ac.uk/QuickGO/search/CPX-591>). The InterMine open source data warehouse system (19) has developed a pipeline to im-

port protein complexes into the InterMine environment using the recently published JAMI library (20) to read and translate the data into interaction objects and export a MI-JSON file to the ComplexViewer. YeastMine and HumanMine are already making relevant data sets publicly available. The IMEx Consortium of interaction databases (21) now describe molecules that interact with complexes using Complex Portal entries as annotation objects.

As described above, the prioritized curation of complexes with a known complete, or partial, 3D structure has been undertaken as a collaboration with students from CEITEC (Masaryk University) and PDBe, and the yeast complexome has been assembled in partnership with the Saccharomyces Genome Database and curators from the UniProt Consortium. UniProt curators have also contributed to the annotation of nematode complexes. We are interested in collaborating with other interest groups to continue to expand and improve this key reference resource and encourage potential collaborators to contact us via the yellow <Request> tile on the homepage or directly by emailing us at [complexportal@ebi.ac.uk](mailto:complexportal@ebi.ac.uk).

## RELATED RESOURCES

The Complex Portal focuses on stable, functional biological units which are the molecular machineries that drive biological processes. Compared to other macromolecular complex repositories, the Complex Portal uniquely adds details of nucleic acid and small molecule components of protein complexes, and describes the roles each contributes to complex function. This level of detail requires the collation of information abstracted from multiple publications. Evidence for the component content of a complex is linked, when

available, to the detailed experimental curation of molecular interactions in the IMEx Consortium databases (21), a consortium of resources which systematically capture experimental interaction data from peer-reviewed journals. A complex may be described within a single interaction in an IMEx record, or may need an expert biocurator to assemble evidence from multiple records.

The CORUM database of mammalian complexes (22) defines complexes based on experimental evidences found in a single publication, akin to classic protein-protein interaction databases. Its coverage for mammalian complexes is currently larger than Complex Portal while Complex Portal provides a broader species range. The overlap of complexes between the 2 resources cannot be directly compared as their definition of complexes is not the same: several experimentally-evidenced complexes from CORUM may be combined into one Complex Portal entry while Complex Portal provides separate entries for variants (see Section "Variants, molecule sets and complexes as participants" above).

## SUMMARY AND FUTURE PLANS

The Complex Portal has been designed with 3 major use cases in mind. Firstly, the description of individual protein complexes to allow information on these biological entities to be integrated across multiple biomedical resources. We will continue to add cross-references and further improve our API in order to enable this. Secondly, the completion of the entire set of known protein complexes for key model organisms will enable these to be identified in large-scale data sets. ComplexTAB has been designed as an easy-to-parse option to provide data for this purpose. *Saccharomyces cerevisiae* is the first organism for which we believe we have achieved a complete representation of current knowledge but we anticipate more complexes will be identified in this organism as experimental techniques become more sensitive and we will continue to enhance this set as information becomes available. We look to complete the first bacterial complexome, for *Escherichia coli*, in 2019 and will continue to work towards a fuller representation of the human and murine complex sets. Thirdly, we look to contribute to a broader understanding of complex evolution and development across a range of taxonomic species by identifying how complex content and function alter with the complexity of the host organism. We are currently developing a mechanism by which orthologous and paralogous complexes can be identified and a tool to subsequently visualize these relationships and will be making this publicly available on the Complex Portal website in the near future.

We constantly try to improve our databases and services in terms of accuracy and representation and actively encourage user feedback. If you have any questions or suggestions please contact the Molecular Interaction Team via the yellow <Request> tile on the homepage or email us directly at [complexportal@ebi.ac.uk](mailto:complexportal@ebi.ac.uk). Information about curation is provided at <https://www.ebi.ac.uk/complexportal/documentation>. Extensive training material on how to best use our resource is available at <https://www.ebi.ac.uk/training?ebi-topic%5Bsystems%5D=systems>.

## Computational Resources

The Complex Portal is a community project. Developers can contribute to the code at <https://github.com/Complex-Portal/complex-portal-view>.

Data can be accessed either via our ftp site (<ftp.ebi.ac.uk/pub/databases/intact/complex/current/>) or our REST API (<https://www.ebi.ac.uk/intact/complex-ws/>).

## FUNDING

European Molecular Biology Laboratory Core Funding; Elixir-CZ; OpenTargets and the China Scholarship Council (201708500036). Funding for open access charge: European Molecular Biology Laboratories Core Funding.

*Conflict of interest statement.* None declared.

## REFERENCES

- Meldal, B.H.M., Forner-Martinez, O., Costanzo, M.C., Dana, J., Demeter, J., Dumousseau, M., Dwight, S.S., Gaulton, A., Licata, L., Melidoni, A.N. *et al.* (2015) The complex portal—an encyclopaedia of macromolecular complexes. *Nucleic Acids Res.*, **43**, D479–D484.
- Gene Ontology Consortium (2019) The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.*, doi:10.1093/nar/gky1055.
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E. *et al.* (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, **3**, 160018.
- Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G.R., Wu, G.R., Matthews, L. *et al.* (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.*, **33**, D428–D432.
- Mir, S., Alhroub, Y., Anyango, S., Armstrong, D.R., Berrisford, J.M., Clark, A.R., Conroy, M.J., Dana, J.M., Deshpande, M., Gupta, D. *et al.* (2018) PDBE: towards reusable data delivery infrastructure at protein data bank in Europe. *Nucleic Acids Res.*, **46**, D486–D492.
- Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B. *et al.* (2018) The reactome pathway knowledgebase. *Nucleic Acids Res.*, **46**, D649–D655.
- The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **45**, D158–D169.
- Skrzypek, M.S. and Nash, R.S. (2015) Biocuration at the *Saccharomyces* genome database. *Genesis*, **53**, 450–457.
- Hastings, J., Owen, G., Dekker, A., Ennis, M., Kale, N., Muthukrishnan, V., Turner, S., Swainston, N., Mendes, P. and Steinbeck, C. (2016) ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.*, **44**, D1214–D1219.
- The RNAcentral Consortium (2017) RNAcentral: a comprehensive database of non-coding RNA sequences. *Nucleic Acids Res.*, **45**, D128–D134.
- Hermjakob, H., Montecchi-Palazzi, L., Bader, G., Wojcik, J., Salwinski, L., Ceol, A., Moore, S., Orchard, S., Sarkans, U., von Mering, C. *et al.* (2004) The HUPO PSI's molecular interaction format—a community standard for the representation of protein interaction data. *Nat. Biotechnol.*, **22**, 177–183.
- Chibucos, M.C., Siegel, D.A., Hu, J.C. and Giglio, M. (2017) The evidence and conclusion ontology (ECO): Supporting GO annotations. *Methods Mol. Biol.*, **1446**, 245–259.
- Combe, C.W., Sivade, M.D., Hermjakob, H., Heimbach, J., Meldal, B.H.M., Mickle, G., Orchard, S. and Rappsilber, J. (2017) ComplexViewer: visualization of curated macromolecular complexes. *Bioinformatics*, **33**, 3673–3675.
- Fabregat, A., Sidiropoulos, K., Viteri, G., Marin-Garcia, P., Ping, P., Stein, L., D'Eustachio, P. and Hermjakob, H. (2018) Reactome diagram viewer: data structures and strategies to boost performance. *Bioinformatics*, **34**, 1208–1214.



15. Sehnal,D., Deshpande,M., Vařeková,R.S., Mir,S., Berka,K., Midlik,A., Pravda,L., Velankar,S. and Koča,J. (2017) LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. *Nat. Methods*, **14**, 1121–1122.
16. Papatheodorou,I., Fonseca,N.A., Keays,M., Tang,Y.A., Barrera,E., Bazant,W., Burke,M., Füllgrabe,A., Fuentes,A.M.-P., George,N. *et al.* (2018) Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res.*, **46**, D246–D251.
17. Kerrien,S., Orchard,S., Montecchi-Palazzi,L., Aranda,B., Quinn,A.F., Vinod,N., Bader,G.D., Xenarios,I., Wojcik,J., Sherman,D. *et al.* (2007) Broadening the horizon—level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol.*, **5**, 44.
18. Sivade Dumousseau,M., Alonso-López,D., Ammari,M., Bradley,G., Campbell,N.H., Ceol,A., Cesareni,G., Combe,C., De Las Rivas,J., Del-Toro,N. *et al.* (2018) Encompassing new use cases - level 3.0 of the HUPO-PSI format for molecular interactions. *BMC Bioinformatics*, **19**, 134.
19. Kalderimis,A., Lyne,R., Butano,D., Contrino,S., Lyne,M., Heimbach,J., Hu,F., Smith,R., Stěpán,R., Sullivan,J. *et al.* (2014) InterMine: extensive web services for modern biology. *Nucleic Acids Res.*, **42**, W468–W472.
20. Sivade Dumousseau,M., Koch,M., Shrivastava,A., Alonso-López,D., De Las Rivas,J., Del-Toro,N., Combe,C.W., Meldal,B.H.M., Heimbach,J., Rappsilber,J. *et al.* (2018) JAMI: a Java library for molecular interactions and data interoperability. *BMC Bioinformatics*, **19**, 133.
21. Orchard,S., Kerrien,S., Abbani,S., Aranda,B., Bhate,J., Bidwell,S., Bridge,A., Briganti,L., Brinkman,F.S.L., Brinkman,F. *et al.* (2012) Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat. Methods*, **9**, 345–350.
22. Ruepp,A., Waegele,B., Lechner,M., Brauner,B., Dunger-Kaltenbach,I., Fobo,G., Frishman,G., Montrone,C. and Mewes,H.-W. (2010) CORUM: the comprehensive resource of mammalian protein complexes—2009. *Nucleic Acids Res.*, **38**, D497–D501.
23. Lawson,C.L., Patwardhan,A., Baker,M.L., Hryc,C., Garcia,E.S., Hudson,B.P., Lagerstedt,I., Ludtke,S.J., Pintilie,G., Sala,R. *et al.* (2016) EMDatabank unified data resource for 3DEM. *Nucleic Acids Res.*, **44**, D396–D403.