

# The complex portal - an encyclopaedia of macromolecular complexes

Birgit H.M. Meldal<sup>1,\*</sup>, Oscar Forner-Martinez<sup>1</sup>, Maria C. Costanzo<sup>2</sup>, Jose Dana<sup>1</sup>, Janos Demeter<sup>2</sup>, Marine Dumousseau<sup>1</sup>, Selina S. Dwight<sup>2</sup>, Anna Gaulton<sup>1</sup>, Luana Licata<sup>3</sup>, Anna N. Melidoni<sup>1</sup>, Sylvie Ricard-Blum<sup>4</sup>, Bernd Roechert<sup>5</sup>, Marek S. Skzyrpek<sup>2</sup>, Manu Tiwari<sup>6</sup>, Sameer Velankar<sup>1</sup>, Edith D. Wong<sup>2</sup>, Henning Hermjakob<sup>1</sup> and Sandra Orchard<sup>1</sup>

<sup>1</sup>European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, United Kingdom, <sup>2</sup>Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305–5477, USA, <sup>3</sup>Department of Biology, University of Rome, Tor Vergata, Rome 00133, Italy, <sup>4</sup>UMR 5086 CNRS, Université Lyon1, Institut de Biologie et Chimie des Protéines, 7 passage du Vercors, 69367 Lyon Cedex 07, France, <sup>5</sup>Swiss-Prot Group, SIB Swiss Institute of Bioinformatics, Centre Medical Universitaire, Geneva, Switzerland and <sup>6</sup>Stammzellbiologie, Institut für Anatomie und Zellbiologie, GZMB Universitätsmedizin Göttingen, Ernst-Caspari-Haus, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany

Received August 15, 2014; Revised September 30, 2014; Accepted October 01, 2014

## ABSTRACT

The IntAct molecular interaction database has created a new, free, open-source, manually curated resource, the Complex Portal ([www.ebi.ac.uk/intact/complex](http://www.ebi.ac.uk/intact/complex)), through which protein complexes from major model organisms are being collated and made available for search, viewing and download. It has been built in close collaboration with other bioinformatics services and populated with data from ChEMBL, MatrixDB, PDB, Reactome and UniProtKB. Each entry contains information about the participating molecules (including small molecules and nucleic acids), their stoichiometry, topology and structural assembly. Complexes are annotated with details about their function, properties and complex-specific Gene Ontology (GO) terms. Consistent nomenclature is used throughout the resource with systematic names, recommended names and a list of synonyms all provided. The use of the Evidence Code Ontology allows us to indicate for which entries direct experimental evidence is available or if the complex has been inferred based on homology or orthology. The data are searchable using standard identifiers, such as UniProt, ChEBI and GO IDs, protein, gene and complex names or synonyms. This reference resource will be maintained and grow to encompass an increasing number of organisms. In-

put from groups and individuals with specific areas of expertise is welcome.

## INTRODUCTION

Biological processes are driven by the interactions of proteins with other molecules in an organism. These interactions may be transient (e.g. signalling receptor–ligand interactions) or lead to the formation of stable biological complexes (e.g. minichromosome maintenance complex, EBI-913604). Many molecules exist only in an obligate complex formation (e.g. collagen type I, EBI-2325312). Many databases deal with protein, small molecule and polysaccharide functions, for example, UniProtKB (1), Gene Ontology (GO) (2), ChEBI (3), ChEMBL (4) and their interactions (e.g. IntAct (5), MINT (Molecular Interaction Database (5,6)), MatrixDB (7) or DIP (Database of Interacting Proteins (8))) but no central, integrated resource has so far been available to hold this type of information for stable, macromolecular complexes. The Complex Portal, hosted by the IntAct team at the European Bioinformatics Institute (EBI) ([www.ebi.ac.uk/intact/complex](http://www.ebi.ac.uk/intact/complex)), provides a framework for the capture of such data, and the existence of web-based editorial tools not only enables collaborating groups to capture standards-compliant experimental data (5) but also to contribute to the curation of the growing collection of manually curated protein complexes. Several databases (such as CORUM (9)) exist that have previously collated information on protein complexes but they provide limited data integration across resources and have tended to concentrate on a more limited species

\*To whom correspondence should be addressed. Tel: +44 1223 494107; Fax: +44 1223 494468; Email: [bmeldal@ebi.ac.uk](mailto:bmeldal@ebi.ac.uk)

range. UniProtKB/Swiss-Prot contains a wealth of data pertaining to the behaviour of individual proteins within larger complexes in the form of free-text statements. Curators from the UniProt Consortium collaborate with the Complex Portal to help extract and structure these data. The Protein Ontology (PRO) provides an ontological representation of protein-related entities by explicitly defining them and describing their relationship (10). This includes the existence of proteins within larger macromolecular complexes. The PRO consortium, by collaborating with the IntAct group, will contribute to both extend the number of complexes which are available as objects within the PRO hierarchy while making more richly annotated data on these entities available in a broader range of formats through the Complex Portal. Model organism groups, such as *Saccharomyces* Genome Database (11) and WormBase (12), are contributing expertise and curation effort within their organism of interest, while the MINT database is augmenting the resource with additional expertise on the composition of protein complexes. Curators at the EBI ensure that entries relating to conserved complexes remain consistent across multiple species, or that differences that have evolved across a taxonomic range are suitably highlighted. Long-term data maintenance, such as regular updates of controlled vocabulary terms and underlying protein sequence data, are also provided by the IntAct team.

Entries combine biological information about the molecular function, biological processes, cellular location and pathways of a complex with information about its constituents (proteins, nucleic acids, small molecules and polysaccharides), topology, stoichiometry and structural assembly. Structured information about complex ligands, diseases they are involved in and drugs they act as targets for is also added to appropriate entries. The use of the Evidence Code Ontology (13) allows us to indicate for which entries direct experimental evidence is available in a molecular interaction databases, such as IntAct ([www.ebi.ac.uk/intact](http://www.ebi.ac.uk/intact)), or if the complex has been modelled from more limited evidence, or inferred from component sequence homology or orthology.

## CONTENT

### Coverage

The Complex Portal contains complexes from more than 10 model organisms, with a current emphasis on human, mouse, *Saccharomyces cerevisiae* and *Escherichia coli*. The working definition of a stable protein complex is an entity that contains two or more macromolecules that can be co-purified and for which there is evidence (experimental or inferred) that these molecules interact with each other and have a demonstrated biological or MF. The definition currently specifically excludes more transient interactions, such as enzyme-substrate and receptor-ligand complexes, unless the binding of the substrate or ligand is required for complex formation to occur, e.g. PDGF receptors only dimerize when the ligand binds so the tetrameric receptor-ligand complex has been curated (EBI-9080360), in contrast to the dimeric receptor which does not exist *in vivo*. Nucleic acids are only included in the list of participants if they are integral to the formation of the complex (e.g. DnaA-DNA

complex, EBI-6552438). Similarly, small molecules, such as cofactors, are captured when stably bound into complexes. Homology modelling is only allowed across closely related species (e.g. mammals), as one use case for this resource is to enable the study of the evolution of protein complexes and highlight examples where function but not subunit content may have been conserved, or vice versa. As indicated above, modelled or inferred complexes are clearly marked using the evidence ontology terms and may be filtered out from experimentally verified data using a simple query.

### Quality control

As with IntAct and many other manually curated databases, each entry is checked by a second curator before it is released, enabling the provision of highly accurate and up-to-date information. However, no data set can claim to be totally error-free and new information is constantly being published. Therefore, the user is encouraged to feedback, or request curation of a specific complex via the <Feedback> button on the website or by directly emailing [intact-help@ebi.ac.uk](mailto:intact-help@ebi.ac.uk).

### Search

The portal query interface allows users to search the complex information based on the identifier, name or synonyms of the complex or of the individual components (proteins, nucleic acid or small molecules) present in the complex as well as cross-references and species of complexes. Users can search for a single term or a mix of terms. The latter can be of the same or different kind, such as a single UniProt ACs, a list of mixed protein and gene identifiers, a complex name and its species or a GO ID (e.g. GO:0016491). The use of wildcards ('?' and '\*') has been enabled, e.g. for a search of isoforms of 'Q07817' search with 'Q07817\*'. Blanks in search strings are by default interpreted as 'OR'. For more precise searches use the Boolean operators 'AND' and 'NOT', use prepends '+' (= AND) or '-' (= NOT), or use parentheses () or double quotes (") to specifically separate or combine certain terms (e.g. <'nuclear pore'> is more precise than <nuclear AND pore> which is more precise than <nuclear pore> as the latter equals <nuclear OR pore>). More refined searches can be performed using the Complex Query Language, CQL. CQL has been specifically designed for searching complexes but has retained those fields that overlap with the Molecular Interactions Query Language, MIQL, such as ID, alias, species, biological role and feature type. More detail on the search options can be found at <http://www.ebi.ac.uk/intact/complex/help/>. The search results can either be filtered or a particular entry can be selected in order to view its details. Currently, it is possible to filter by species, and by biological role (e.g. enzyme) and interactor type (e.g. protein) as defined using the Proteomics Standards Initiative - Molecular Interactions Controlled Vocabulary (PSI-MI CV) (14).

The manual curation of the complexes using hierarchical ontologies, such as the GO or the National Center for Biotechnology Information (NCBI) taxonomy (15), also allows queries based on these hierarchies. For example, it is possible to pull out all enzyme complexes by searching on

EMBL-EBI

Complex Portal

GO:0016491

Home About Documentation Search Help Download Contact Us

Complex Portal > Search results

Complex Portal results for **GO:0016491**

Showing 10 results from a total of 40

Species:

- ☐ Escherichia coli (strain K12) (12)
- ☐ Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (10)
- ☐ Homo sapiens (7)
- ☐ Mus musculus (6)
- ☐ Rattus norvegicus (3)
- ☐ Schizosaccharomyces pombe (strain 972 / ATCC 24843) (2)

Interactor type:

- ☐ protein (40)
- ☐ small molecule (28)

Participant's biological role:

- ☐ unspecified role (26)
- ☐ enzyme (22)
- ☐ cofactor (17)
- ☐ electron donor (4)
- ☐ electron acceptor (3)
- ☐ acceptor (1)
- ☐ ancillary (1)
- ☐ electron donor and electron acceptor (1)
- ☐ proton donor (1)

Filter

Page 1 of 4

- Alkanesulfonate monooxygenase complex (Escherichia coli (strain K12); 83333)** EBI-8786862  
An alkanesulfonate monooxygenase that catalyses the conversion of alkanesulfonates into aldehydes and sulfite under conditions of sulfate or ...
- DXP reductoisomerase (Escherichia coli (strain K12); 83333)** EBI-6475852  
Enzyme involved in the MEP pathway of IPP biosynthesis. It catalyzes the NADP-dependent rearrangement and reduction of 1-deoxy-D-xylulos...
- Succinate dehydrogenase complex (Escherichia coli (strain K12); 83333)** EBI-6476056  
Key enzyme linking the Krebs cycle with the respiratory chain in aerobic respiration: Catalyzes the oxidation of succinate to fumarate and the ...
- Ribonucleotide-diphosphate reductase variant 1 (Saccharomyces cerevisiae (strain ATCC 204508 / S288c); 559292)** EBI-1396664  
Catalyzes the reduction of ribonucleotides to the corresponding deoxyribonucleotides, an essential step in the de novo synthesis of monomeric...
- Ribonucleotide-diphosphate reductase variant 2 (Saccharomyces cerevisiae (strain ATCC 204508 / S288c); 559292)** EBI-1396793  
Catalyzes the reduction of ribonucleotides to the corresponding deoxyribonucleotides, an essential step in the de novo synthesis of monomeric...
- mitochondrial respiratory chain complex IV variant 1 (Saccharomyces cerevisiae (strain ATCC 204508 / S288c); 559292)** EBI-2436444  
Key role in aerobic respiration, in which mitochondrial enzymes accept electrons from electron carriers reduced in glycolysis and the tricarboxy...
- mitochondrial electron transport complex IV variant 2 (Saccharomyces cerevisiae (strain ATCC 204508 / S288c); 559292)** EBI-2436468  
Key role in aerobic respiration, in which mitochondrial enzymes accept electrons from electron carriers reduced in glycolysis and the tricarboxy...
- Pyrroline-5-carboxylate reductase 1 complex (Homo sapiens; 9606)** EBI-848646  
Catalyzes the transfer of a reducing equivalent from NAD(P)H to pyrroline-5-carboxylate, yielding products NAD(P)+ and proline.
- Pyrroline-5-carboxylate reductase 1 complex (Mus musculus; 10090)** EBI-848703  
Catalyzes the transfer of a reducing equivalent from NAD(P)H to pyrroline-5-carboxylate, yielding products NAD(P)+ and proline.
- Nitrate reductase A (Escherichia coli (strain K12); 83333)** EBI-6880351  
Involved in electron transport during anaerobic respiration: electrons are passed from the formate dehydrogenase-N (Fdh-N) complex (EBI-68...

Page 1 of 4

EMBL-EBI

Services

By topic  
By name (A-Z)  
Help & Support

Research

Overview  
Publications  
Research groups  
Postdocs & PhDs

Training

Overview  
Train at EBI  
Train outside EBI  
Train online  
Contact organisers

Industry

Overview  
Members Area  
Workshops  
SME Forum  
Contact Industry programme

About us

Overview  
Leadership  
Funding  
Background  
Collaboration  
Jobs  
People & groups  
News  
Events  
Visit us  
Contact us

EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK +44 (0)1223 49 44 44  
Copyright © EMBL-EBI 2014 | EBI is an outstation of the European Molecular Biology Laboratory | Privacy | Cookies | Terms of use

**Figure 1.** Screenshot of the results for the GO search term 'GO:0003824' (catalytic activity). Search results are displayed as 10 per page and filters for refining the search are shown in the left-hand side panel. The number of hits for each filter option is given in parenthesis.

the parent term 'catalytic activity' (GO:0003824, Figure 1) which will find all examples within the current collection, although each has been more specifically annotated to an appropriate child term. Similarly, it is possible to find all mammalian complexes by using the search term 'species:40674' (mammalia).

### Complex details

Each complex has as a minimum a unique accession number, a recommended name, a list of participants, a functional description and a list of cross-references (Figure 2).

Complexes are curated based either on direct evidence in the literature, or inferred from closely related species (e.g. mouse PCNA, EBI-1208675) or homologous complexes in the same species, for which experimental evidence exists (rat hemoglobin variants, EBI-9105417 and EBI-9105480), or based on background scientific knowledge of an expert submitter (e.g. existence based on pharmacological evidence, e.g. GABA receptors, EBI-9008426). The type of evidence that exists for each complex entry is identified by using Ev-

idence Code Ontology (ECO) codes. For ECO codes currently in use please refer to the online documentation (<http://www.ebi.ac.uk/intact/complex/documentation/>). In cases where the existence of a complex has been inferred by sequence homology of the components, rather than from direct experimental evidence, the details of the experimentally verified complex, on which it has been modelled, are given. These entries will be updated when experimental evidence becomes available.

Complex members are identified via unique accession numbers with direct links to the respective databases (UniProtKB for proteins, ChEBI for small molecules), name and description of the molecule (imported from the above reference databases), stoichiometry (when known), interactor type and the biological role of that molecule within the complex. Known protein binding regions are identified by their start and end residue numbers and cross-referenced to InterPro (16) if they match a functional domain (EBI-9519429), whereas unknown regions are identified by range '?-' (EBI-706546). The linking of binding regions enables the systematic capture of the topology of



EMBL-EBI

ServicesResearchTrainingAbout us

Complex Portal

Search

Home

About

Documentation

Search

Help

Download

Contact Us

Complex Portal > Details

Hemoglobin HbA complex

Species: Homo sapiens; 9606

Accession number: EBI-9008420

Summary

Systematic Name:

2xHBA1:2xHBB

Synonyms:

HBA-HBB complex	HBA1-HBB complex	HBA2-HBB complex
HBA1-HBB heterotetramer	HBA2-HBB heterotetramer	adult hemoglobin HbA complex
HBA-HBB heterotetramer	Haemoglobin HbA complex	

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 (EBI-9108045 & EBI-9108218).

Properties:

Two alpha chains and two beta chains. Each chain has a heme b group attached to it containing either an Fe2+ or Fe3+ ion. Oxygen only binds to Fe2+ ions, not Fe3+ ions. CO2 binds directly to the protein chains and therefore does not compete with oxygen binding. MW = 64 kD

Disease:

Mutations in hemoglobin alpha chain can result in a range of diseases, including alpha-thalassemia, Heinz body anemias and hemoglobin H (HBH) disease. Mutations in hemoglobin beta chain can result in a range of diseases, including sickle-cell anemia, beta-thalassemia and Heinz body anemias.

Complex Assembly:

Heterotetramer.

Participants

ID	Name	Description	Stoichiometry	Biological Role	Interactor Type	Linked Features	Other Features
P69905 EBI-714680	HBA1	Hemoglobin subunit alpha	2	unspecified role	protein	binding-associated region CHEBI:30413 [7-7] binding-associated region P68871 [7-7]	
P68871 EBI-715554	HBB	Hemoglobin subunit beta	2	unspecified role	protein	binding-associated region CHEBI:30413 [7-7] binding-associated region P69905 [7-7]	
CHEBI:30413 EBI-6880399	heme	heme	4	cofactor	small molecule	binding-associated region P69905 [7-7] binding-associated region P68871 [7-7]	

Cross References

Type	Database	Identifier	Description
Identical object in an external resource	chembl	CHEMBL2095168	
	evidence ontology	ECO:000353	
biological process	gene ontology	GO:0006788	heme oxidation
cellular component	gene ontology	GO:0005833	hemoglobin complex
molecular function	gene ontology	GO:0005344	oxygen transporter activity
Experimental evidence that interaction occurs	intact	EBI-1029796	
Identical object in an external resource	omim	140700	
additional information	pubmed	16765986	
Identical object in an external resource	reactome	REACT_123853.1	
Identical object in an external resource	wwpdb	2dn3	

**Figure 2.** Screenshot of the details page of the human haemoglobin HbA complex (EBI-9008420). The Cross References table has been shortened to fit in the figure. Clicking on the hyperlinks will open the details directly in the external database.

a complex in a format which will subsequently enable the automated derivation of a graphical representation of this topology. Other features, such as post-translational modifications and disulphide bridges, which are required for binding, are also listed (e.g. EBI-8869931).

Each complex has up to three name categories: firstly, a mandatory recommended name that is either taken from the GO or is the most commonly used name in the literature. In

parallel, missing terms are added to the GO and the protein complex hierarchy within the GO is being revised and augmented. Secondly, additional names for the complex are added to the list of *synonyms*. Thirdly, in collaboration with Reactome (17,18), we give each complex a *systematic name*, which is essentially a string of gene names of the complex participants, separated with a colon (:) and stoichiometry

indicated, e.g., 2xHBA1:2xHBB for haemoglobin HbA. For further details of the Reactome nomenclature rules, see (19).

Further information on the complex is added by the curator providing a detailed description of its *function*, and any processes it is involved in, as well as information about *properties*, such as molecular weight, complex size or internal topology of the complex, where available. Complexes with complete stoichiometry are also annotated with the relevant *assembly* terms, such as homodimer or heterotrimer. Where applicable, important *ligands* are listed and short descriptions are given about *diseases* with which the complex is associated.

To aid data integration and to provide maximum additional information to the user, related classes or instances of the complex in other databases are provided as cross-referenced links. These may include links to ChEMBL, EMDatabank (20), the GO, MatrixDB, PDBe (21) and Reactome. Links to experimental evidence proving the existence and content of a particular complex are provided as IMEx (International Molecular Exchange Consortium (22)), IntAct, MINT, MatrixDB or DIP cross-references and, as indicated, ECO codes are used to further indicate the level of evidence available for the complex. For enzymes we provide a link to the Integrated relational Enzyme database (IntEnz (23)). When a complex is involved in a disease we provide cross-references to Online Mendelian Inheritance in Man (OMIM, <http://omim.org/>), Experimental Factor Ontology (EFO) and/or Orphanet (<http://www.ebi.ac.uk/efo/about.html> and [http://www.orphadata.org/cgi-bin/inc/ordo\\_orphanet.inc.php](http://www.orphadata.org/cgi-bin/inc/ordo_orphanet.inc.php)). As EFO is in the process of compiling a new disease ontology we will progressively be cross-referencing to EFO/Orphanet/ORDO. Further information in the literature on the complex, such as an appropriate review or a functional study, is cross-referenced to Europe PubMed Central (24).

All complexes are annotated to GO terms. Every complex is annotated to at least one cellular component term and to relevant molecular function and biological process terms for the complex as a whole, rather than annotation of individual components as is the focus of other annotation efforts. New terms are created as appropriate.

## WEB APPLICATION, WEB SERVICE AND DOWNLOAD OPTIONS

A new index has been built and a public RESTful web service has been provided (<http://www.ebi.ac.uk/intact/complex-ws/>) for programmatic retrieval of complexes. The current webservice provides only two methods but will be expanded as new features are added. A method `/search` will take as parameter for any CQL query and return a custom JSON which only gives a summary of the existing complexes hitting the user query (complex name, species, description and internal identifier). The user can then get the full details of the complex of interest using the `/details` method which returns another custom JSON describing the complex cross-references, participants, etc.

The web application has been built on top of the RESTful service for searching and viewing of individual complexes.

Files can be downloaded in PSI-MI XML2.5 format from our dedicated ftp site: <http://www.ebi.ac.uk/intact/complex/download/>.

## COLLABORATIONS AND COMMUNITY INVOLVEMENT

The Complex Portal has been designed to provide a framework in which groups are able to contribute their expert knowledge by annotating complexes within a particular defined area. The IntAct molecular interaction database is responsible for training, data quality, data consistency and long-term data maintenance. While coverage is limited at this early stage of the Portal development, the involvement of multiple contributing groups ensures that this will grow rapidly. We hope that further groups will add to the community effort in the near future, in particular, participation of the plant community would assist in filling this gap in the taxonomic coverage.

For suggestions for new complexes to be added to the database and direct submissions please contact us via the `<Feedback>` link on the homepage or by directly emailing at [intact-help@ebi.ac.uk](mailto:intact-help@ebi.ac.uk).

## FUTURE PLANS

Work is continually ongoing to improve the website, search capabilities and download options. The PSI-MI are currently actively working to provide an improved XML format, PSI-MI XML3.0, for the capture of abstracted data, such as protein complexes. Individual complex download files (in both XML2.5 and XML3.0) will be available from both the list of initial search results and the details page. The filter options will be expanded, for example, to make it easier to extract the experimentally verified complexes from those which have been modelled or inferred by homology.

Visualization of complex topology is critical and a complex viewer is currently under development that will display the complex participants and its binding features in a flexible, interactive mode from the information supplied in the underlying XML file. The complex viewer will be wrapped in a BioJS (25) component which will be available in the BioJS registry so it can be re-used by other developers wishing to import and display some or all of these protein complexes, or indeed other data presented in the PSI-MI XML format. Other BioJS components, such as a structural viewer, for those complexes which have been fully, or partially, crystallized and patterns of co-expression data for complex components, will also be incorporated. Ongoing activities include improvements to associated controlled vocabularies and further cross-referencing in addition to increasing the content of the Complex Portal.

We constantly try to improve our databases and services in terms of accuracy and representation and actively encourage user feedback. Please contact the IntAct group if you have any questions via the `<Feedback>` link on the homepage or email us directly at [intact-help@ebi.ac.uk](mailto:intact-help@ebi.ac.uk). Information about curation is provided at <http://www.ebi.ac.uk/intact/complex/documentation/>. Extensive documentation and training material on how to best use our resource is available at <http://www.ebi.ac.uk/training/networks>. Cu-

ration groups interested in capturing macromolecular complex data or experimental interaction data who would like access to the editorial tool are encouraged to contact IntAct to discuss this further (intact-help@ebi.ac.uk).

## FUNDING

National Institutes of Health [268201000035C]; Biotechnology and Biological Sciences Research Council [BB/L024179/1]; European Molecular Biology Laboratories. Funding for open access charge: European Molecular Biology Laboratories Core Funding.  
*Conflict of interest statement.* None declared.

## REFERENCES

1. UniProt Consortium (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **42**, D191–D198.
2. Gene Ontology Consortium, Blake, J.A., Dolan, M., Drabkin, H., Hill, D.P., Li, N., Sitnikov, D., Bridges, S., Burgess, S., Buza, T. *et al.* (2013) Gene Ontology annotations and resources. *Nucleic Acids Res.*, **41**, D530–D535.
3. Hastings, J., de Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., Muthukrishnan, V., Owen, G., Turner, S., Williams, M. *et al.* (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res.*, **41**, D456–D463.
4. Bento, A.P., Gaulton, A., Hersey, A., Bellis, L.J., Chambers, J., Davies, M., Krüger, F.A., Light, Y., Mak, L., McGlinchey, S. *et al.* (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Res.*, **42**, D1083–D1090.
5. Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., Chavali, G., Chen, C., del-Toro, N. *et al.* (2014) The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.*, **42**, D358–D363.
6. Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., Sacco, F., Palma, A., Nardoza, A.P., Santonico, E. *et al.* (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.*, **40**, D857–D861.
7. Chautard, E., Fatoux-Ardore, M., Ballut, L., Thierry-Mieg, N. and Ricard-Blum, S. (2011) MatrixDB, the extracellular matrix interaction database. *Nucleic Acids Res.*, **39**, D235–D240.
8. Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U. and Eisenberg, D. (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.*, **32**, D449–D451.
9. 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.
10. Natale, D.A., Arighi, C.N., Blake, J.A., Bult, C.J., Christie, K.R., Cowart, J., D'Eustachio, P., Diehl, A.D., Drabkin, H.J., Helfer, O. *et al.* (2014) Protein Ontology: a controlled structured network of protein entities. *Nucleic Acids Res.*, **42**, D415–D421.
11. Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R. *et al.* (2012) Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.*, **40**, D700–D705.
12. Yook, K., Harris, T.W., Bieri, T., Cabunoc, A., Chan, J., Chen, W.J., Davis, P., de la Cruz, N., Duong, A., Fang, R. *et al.* (2012) WormBase 2012: more genomes, more data, new website. *Nucleic Acids Res.*, **40**, D735–D741.
13. Chibucos, M.C., Mungall, C.J., Balakrishnan, R., Christie, K.R., Huntley, R.P., White, O., Blake, J.A., Lewis, S.E. and Giglio, M. (2014) Standardized description of scientific evidence using the Evidence Ontology (ECO). *Database J. Biol. Databases Curation*, doi:10.1093/database/bau075.
14. Orchard, S., Montecchi-Palazzi, L., Hermjakob, H. and Apweiler, R. (2005) The use of common ontologies and controlled vocabularies to enable data exchange and deposition for complex proteomic experiments. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.*, 186–196.
15. Federhen, S. (2012) The NCBI Taxonomy database. *Nucleic Acids Res.*, **40**, D136–D143.
16. Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T.K., Bateman, A., Bernard, T., Binns, D., Bork, P., Burge, S. *et al.* (2012) InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.*, **40**, D306–D312.
17. Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M.R. *et al.* (2014) The Reactome pathway knowledgebase. *Nucleic Acids Res.*, **42**, D472–D477.
18. Milacic, M., Haw, R., Rothfels, K., Wu, G., Croft, D., Hermjakob, H., D'Eustachio, P. and Stein, L. (2012) Annotating cancer variants and anti-cancer therapeutics in reactome. *Cancers*, **4**, 1180–1211.
19. Jupe, S., Jassal, B., Williams, M. and Wu, G. (2014) A controlled vocabulary for pathway entities and events. *Database J. Biol. Databases Curation*, doi:10.1093/database/bau060.
20. Lawson, C.L., Baker, M.L., Best, C., Bi, C., Dougherty, M., Feng, P., van Ginkel, G., Devkota, B., Lagerstedt, I., Ludtke, S.J. *et al.* (2011) EMDatabank.org: unified data resource for CryoEM. *Nucleic Acids Res.*, **39**, D456–D464.
21. Gutmanas, A., Alhroub, Y., Battle, G.M., Berrisford, J.M., Bochet, E., Conroy, M.J., Dana, J.M., Fernandez Montecelo, M.A., van Ginkel, G., Gore, S.P. *et al.* (2014) PDB: Protein Data Bank in Europe. *Nucleic Acids Res.*, **42**, D285–D291.
22. 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.
23. Fleischmann, A., Darsow, M., Degtyarenko, K., Fleischmann, W., Boyce, S., Axelsen, K.B., Bairoch, A., Schomburg, D., Tipton, K.F. and Apweiler, R. (2004) IntEnz, the integrated relational enzyme database. *Nucleic Acids Res.*, **32**, D434–D437.
24. McEntyre, J.R., Ananiadou, S., Andrews, S., Black, W.J., Boulderstone, R., Buttery, P., Chaplin, D., Chevuru, S., Copley, N., Coleman, L.-A. *et al.* (2011) UKPMC: a full text article resource for the life sciences. *Nucleic Acids Res.*, **39**, D58–D65.
25. Corpas, M., Jimenez, R., Carbon, S.J., Garcia, A., Garcia, L., Goldberg, T., Gomez, J., Kalderimis, A., Lewis, S.E., Mulvany, I. *et al.* (2014) BioJS: an open source standard for biological visualisation - its status in 2014. *F1000 Res.*, **3**, 55.