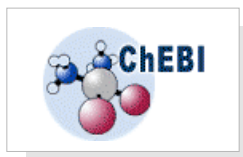


EMBL-EBI



European Bioinformatics Institute  
is an Outstation of the  
European Molecular Biology Laboratory.



# ChEBI: an Open-access Chemistry Resource for the Life Sciences: Facilities for On-line Submission and Curation

M. Ennis, N. Adams, A Dekker, P. de Matos, J. Hastings, K. Haug, D. Hull, Z. Josephs, G. Owen, S. Turner and C. Steinbeck

EMBL Outstation - European Bioinformatics Institute (EBI), Hinxton, Cambridge CB10 1SD, UK

**ChEBI (Chemical Entities of Biological Interest)** [1] is a manually curated database of molecular entities which also provides a chemical ontology. ChEBI contains approximately 22,000 manually curated entities and approximately 620,000 entities in total. It provides a wide range of information including nomenclature, structures and related calculated values, and interrelationships between entities in the chemical ontology. The ChEBI database has a strong focus on quality with exceptional efforts applied to upholding IUPAC nomenclature rules and best IUPAC practices when drawing chemical structures. ChEBI is available at <http://www.ebi.ac.uk/chebi/>. To invite the community to participate more directly in the future growth and development of ChEBI, we have developed a web-based software utility to enable direct user submissions.

## Coping with user requests – ChEBI Submission Tool

Owing to the high demand for good quality curated chemical entities, ChEBI now operates completely on a user-request basis. This means that all our curators work full-time on annotation of requests made by users. In order to harness efficiently the data submitted by users we have introduced the ChEBI submission tool. The ChEBI submission tool is available at: <http://www.ebi.ac.uk/chebi/submissions/>

All main data items can be added directly by the submitter:

- A unique, unambiguous recommended ChEBI Name
- Structural diagram
- Definition
- Synonyms, including an IUPAC-recommended name where appropriate
- Cross-references to other databases
- Links to the ChEBI ontology

## Structure Search

As a cheminformatics resource it is essential to provide a facility for chemical structure searching. ChEBI offers chemical substructure and similarity searching using the Chemistry Development Kit [2,3]. The facility allows a user to draw a chemical structure using the JChemPaint applet [4] developed by the Steinbeck group at the EBI. Both simple and advanced searches are offered, with the advanced option allowing filters (e.g. for formula, mass range, charge range, ontology term, source database) to be applied.

## ChEBI Ontology

We have introduced the term 'role' which forms a sub-ontology branch along with 'molecular structure' and 'subatomic particle' branches. The term 'role' encapsulates 'biological role', 'chemical role' and 'application'. A new relationship 'has role' has been introduced which links chemicals from their 'molecular structure' and 'subatomic particle' sub-ontologies to the 'role' sub-ontology.

## Star rating

All entries are rated according to their level of curation using a star system:

- ★★★★ Manually curated by the ChEBI team
- ★★★ Manually curated by curators from another project or by a ChEBI submitter
- ★ Preliminary Entry loaded automatically from another data source - not manually curated

## Acknowledgements

ChEBI is funded by the European Commission under SLING, contract number 226073 within the Research Infrastructure Action of the FP7 "Capacities Specific" Programme.

## Example of user-submitted entry

## Web Services and Downloads

ChEBI can be accessed programmatically via our SOAP Web Service available at <http://www.ebi.ac.uk/chebi/webServices.do>.

ChEBI is also available for download in various formats including flat file, Oracle binary dumps, generic SQL insert statements and OBO format. You can get further information on downloads at <http://www.ebi.ac.uk/chebi/downloadsForward.do>.

## References

- [1] de Matos, P., Alcántara, R., Dekker, A., Ennis, M., Hastings, J., Haug, K., Spiteri, I., Turner, S. and Steinbeck, C. (2010) Chemical Entities of Biological Interest: an Update. *Nucleic Acids Res.* **38**(Database issue), D249–D254.
- [2] Steinbeck, C., Han, Y.Q., Kuhn, S., Horlacher, O., Luttmann, E., and Willighagen, E.L. (2003) The Chemistry Development Kit (CDK): An open-source Java library for chemo- and bioinformatics. *J. Chem. Inf. Comput. Sci.* **43**, 493–500.
- [3] Steinbeck, C., Hoppe, C., Kuhn, S., Guha, R., Willighagen, E.L. (2006) Recent Developments of The Chemistry Development Kit (CDK) – An Open-Source Java Library for Chemo and Bioinformatics. *Curr. Pharm. Des.* **12**, 2111–2120.
- [4] Kuhn, S. (2006) An Applet Release of JChemPaint. *CDK News* **3**(1), 21–23.

The ChEBI Team  
chebi-help@ebi.ac.uk

EMBL- EBI  
Wellcome Trust Genome  
Campus  
Hinxton  
Cambridge  
CB10 1SD  
UK

T +44 (0) 1223 494 444  
F +44 (0) 1223 494 468  
<http://www.ebi.ac.uk>