Journal of **Chem**informatics

POSTER PRESENTATION

Open Access

ChEBI: a chemistry ontology and database

Paula de Matos*, A Dekker, M Ennis, Janna Hastings, K Haug, S Turner, Christoph Steinbeck

From 5th German Conference on Cheminformatics: 23. CIC-Workshop
Goslar, Germany. 8-10 November 2009

The bioinformatics community has developed a policy of open access and open data since its inception. This is contrary to chemoinformatics which has traditionally been a closed-access area. In 2004, two complementary open access databases were initiated by the bioinformatics community, ChEBI [1] and PubChem. PubChem serves as automated repository on the biological activities of small molecules and ChEBI (Chemical Entities of Biological Interest) as a manually annotated database of molecular entities focused on 'small' chemical compounds. Although ChEBI is reasonably compact containing just over 18,000 entities, it provides a wide range of data items such as chemical nomenclature, an ontology and chemical structures. The ChEBI database has a strong focus on quality with exceptional efforts afforded to IUPAC nomenclature rules, classification within the ontology and best IUPAC practices when drawing chemical structures.

ChEBI is currently undergoing a period of restructuring which will allow it to incorporate the small molecule structures from (and link to) EBI's new chemogenomics database ChEMBL [2], increasing its small molecules coverage to over 500,000 entities. We have restructured the chemical structure search facility to use Orchem [3] an Oracle chemistry plug-in using the Chemistry Development Kit [4]. The facility allows a user to draw a chemical structure or load one from a file and then execute either a substructure or similarity search. Furthermore the ChEBI text search will have extensive facilities for querying based on not only names but formula, a range of charges and molecular weight. The ability to query the ChEBI ontology and retrieve all children for a given entity will also be included.

In order to aid the distribution of ChEBI to the chemoinformatics community we have extended our export formats to include an MDL sdf format with a lighter version consisting only of compound structure, name and identifier. A complete version is available with all the ChEBI data properties such as synonyms, cross-references, SMILES and InChI. Furthermore cross-references in ChEBI have been extended to include BRENDA the enzyme database, NMRShiftDB the database for organic structures and their nuclear magnetic resonance (nmr) spectra, Rhea the biochemical reaction database and IntEnz the enzyme nomenclature database.

ChEBI is available at http://www.ebi.ac.uk/chebi.

Published: 4 May 2010

References
1. Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, Alcántara R, Darsow M, Guedj M, Ashburner M: **ChEBI: a database and ontology for chemical entities of biological interest.** *Nucleic Acids Res* 2008, **36**:D344-D350.
2. [http://www.ebi.ac.uk/chembl/].
3. Rijnbeek M, Steinbeck C: **An open source chemistry search engine for Oracle**.
4. Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E, Willighagen EL: **The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics.** *J Chem Inf Comput Sci* 2003, **43**(2):493-500.

Cheminformatics and Metabolism Team, European Bioinformatics Institute, Hinxton Cambridge, CB10 1SD, UK