ChEBI for systems biology and metabolic modelling

Janna Hastings[1,*] Neil Swainston[2], Venkatesh Muthukrishnan[1], Namrata Kale[1], Adriano Dekker[1], Gareth Owen[1], Steve Turner[1], Pedro Mendes[2] and Christoph Steinbeck[1]

[1] Cheminformatics and Metabolism, EMBL -- European Bioinformatics Institute (EMBL-EBI), Hinxton, UK
[2] Manchester Institute of Biotechnology, School of Computer Science, University of Manchester, UK

1 INTRODUCTION

ChEBI (http://www.ebi.ac.uk/chebi) is a curated database and ontology of biologically relevant small molecules. It is widely used as a reference for chemicals in the context of biological data such as protein interactions, pathways, and models (Hastings et al., 2013). As of the last release (May 2015), ChEBI contains 44,263 fully curated entries, each of which is classified within one of the sub-ontologies: chemical entities (classified according to structural features), roles (classified according to biological or chemical mode of action or use in application), and subatomic particles.

Systems biology brings together a wide range of information about cells, genes and proteins, as well as the small molecules that act on and within these biological structures. It gives a holistic perspective aiming to track and eventually simulate the entire functioning of biological systems. One aspect of systems biology is metabolic modelling, which aims to develop metabolic reconstructions. At the whole-genome scale, these are all-encompassing interlinked maps of all known metabolic reaction pathways for a given organism (Thiele et al., 2013). Chemical data from ChEBI, such as molecular formula, chemical structure and ontology relationships, can fruitfully be used in the model building and refining process to improve model accuracy and enhance the representation of metabolism (Swainston et al., 2011).

Within this context, efforts are currently underway to improve ChEBI for systems biology and metabolic modelling. The enhancements include the addition of a library, libChEBI, for comprehensive programmatic access to ChEBI data, which will be widely applicable but with a particular focus on metabolic modelling. It will include the facility to determine relationships between molecules, such as stereochemistry, tautometrism and redox pairings, to calculate important physicochemical properties, such as pKa and the Gibbs free energy of formation, and to harness these facilities in support of developing, merging and expanding metabolic models. The library will be open source, available in several programming languages including Java and Python.

ChEBI will be providing a facility for bulk submission of novel compounds which will be automatically classified within the ontology. We will also be undertaking curation of the known metabolomes (i.e. all the known metabolites) across four major species (human, mouse, E. coli and yeast).

Finally, we will be introducing into the ChEBI public website novel visualisations of relevance to the systems biology community, such as chemicals in the context of pathways (powered by Reactome, Croft et al., 2011, and MetaboLights, Haug et al., 2013) and models (powered by BioModels, Li et al., 2010).

ACKNOWLEDGEMENTS

ChEBI is funded by the BBSRC, grant agreement number BB/K019783/1 within the "Bioinformatics and biological resources" fund.

REFERENCES

Croft D, O'Kelly G, Wu G, Haw R, et al. (2011) Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. 39:D691-7.

Hastings J., de Matos P., Dekker A., Ennis M., Harsha B., Kale N., Muthukrishnan V., Owen G., Turner S., Williams M. and Steinbeck C. (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Res. 41:D456-63.

Haug, K., Salek, R. M., Conesa, P., Hastings, J. et al. (2013) MetaboLights—an open-access general-purpose repository for metabolomics studies and associated metadata. Nucl. Acids Res. 41: D781-D786.

Li, C., Donizelli, M., Rodriguez, N., Dharuri H. et al. (2010) BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. BMC Systems Biology 4:92.

Swainston N, Mendes P, et al. (2011) The SuBliMinaL Toolbox: automating steps in the reconstruction of metabolic networks. Journal of Integrative Bioinformatics, 8(2):186

Thiele, I., Swainston, N., Fleming, R.M.T, et al. (2013) A community-driven global reconstruction of human metabolism. Nature Biotechnology 31, 419–425.