



Christoph Steinbeck

PhD Rheinische Friedrich-Wilhelm-Universität, Bonn, 1995. Postdoc at Tufts University, Boston, 1996-1997. Head of Research Group for Structural Chemoinformatics, Max Planck Institute of Chemical Ecology, Jena, 1997-2002. Habilitation in Organic Chemistry, Friedrich-Schiller-Universität, Jena, 2003. Head of Research Group for Molecular Informatics, Cologne University Bioinformatics Centre, Cologne, 2002-2007. Lecturer in Cheminformatics, University of Tübingen, 2007. At EMBL-EBI since 2008.

Cheminformatics and Metabolism

DESCRIPTION OF SERVICES

The Cheminformatics and Metabolism team provides the biomedical community with information on small molecules and their interplay with biological systems. Our database portfolio includes ChEBI, the EBI's database and ontology of chemical entities of biological interest, as well as Rhea and IntEnz, our enzyme-related resources. The group develops methods to decipher, organise and publish the small-molecule metabolic content of organisms. We develop algorithms to predict metabolomes based on genomic and other information, to determine quickly the structure of metabolites by stochastic screening of large candidate spaces and to enable the identification of molecules with desired properties. This requires algorithms for the prediction of spectroscopic and other physicochemical properties of chemical graphs based on machine learning and other statistical methods.

We are further investigating the extraction of chemical knowledge from the scientific literature by text- and graph-mining methods. This, as well as our work on chemical database technology and curation, is supported by research into chemical ontologies. Together with an international group of collaborators we have developed a number of widely known and used open-source cheminformatics software packages. The Chemistry Development Kit (CDK), which originated in our lab, is the leading open-source Java library for structural cheminformatics. Based on this, we have developed the cheminformatics workflow/pipelining system CDK-Taverna, which allows researchers to build executable data-processing workflows in a Lego™-like manner, as well as OrChem, our structure-registration and -searching system for the Oracle™-database. In collaboration with partners in Uppsala we initiated Bioclipse, an award-winning, rich client for chemo- and bioinformatics.

SUMMARY OF PROGRESS

- Released OrChem, our open-source chemical-search cartridge for Oracle™;
- Issued releases 59 to 69 of ChEBI, our ontology and database of chemical entities of biological interest;
- Issued releases 3 to 14 of Rhea and 50 to 61 of IntEnz, our enzyme resources;
- Made ChEBI fully structure-searchable based on OrChem;
- Substantially improved our chemical structure editor JChemPaint with R-groups and reactions;
- Secured a grant to establish the MetaboLights database at the EBI;
- Developed the first prototypes of the Enzyme Portal.

MAJOR ACHIEVEMENTS

The release of OrChem allows users to perform substructure, similarity and exact searching within OrChem. Recently it has included the ability to search for R-groups, enabling wild-card chemical-structure searching. It has also included descriptor calculation and exact chemical-structure searching.

In view of successful use of the ChEBI submission tool throughout the reporting period (576 submissions received from 16 individual external submitters), ChEBI has focused its efforts on improving the submission process. A bulk submission facility was developed to enhance the chemical submission experience of regular submitters by allowing programmatic provision of submitted data. As part of a collaboration with the La Jolla Institute of Allergy and Immunology in the USA, approximately 1500 entities associated with immunology were curated, involving use of a new facility for including citations in compound records. A ChEBI user workshop was organised and run, comprising both training elements and scientific talks and complemented by discussions on the future direction of ChEBI. Major improvements to Rhea/IntEnz have included the transformation of chemical compounds used in Rhea reactions from their neutral forms to their ionised states at pH 7.3 to be closer to common physiological conditions.

In 2009/2010, we were glad to host a considerable number of very talented interns, trainees and visiting scientists. Nikolas Fechner and Georg Hinselmann, University of Tübingen, worked with us in autumn 2009 on a novel method for classification of chemical compounds into ontological hierarchies, as well as on a visualisation technique for chemical spaces. Duan Lian of the East China University of Science and Technology in Shanghai, China optimised fingerprint performance in substructure search pre-screening by selecting fingerprint patterns with a novel analysis method. Leonid L. Chepelev of Carleton University in Ottawa, Canada worked on developing self-organising structure- and function-based chemical-entity hierarchies and tools for automated chemical classification in ChEBI. Laura Daniels, a student from Cranfield University in the UK, worked with our Enzyme Portal Team for her MSc thesis project, a case study about the challenges of data integration in the EBI enzyme portal. Kalai Vanii Jayaseelan of Vellore Institute of Technology (VIT) in India worked with the EBI text-mining infrastructure, developing a method for the association of small molecules with their biological targets. She has now joined our group as a bioinformatician and regular group member.

FUTURE PLANS

We have received a grant from the BBSRC to build the missing community resource for metabolomics at the EBI. The database, tentatively named *MetaboLights*, will be cross-species and cross-application and will cover metabolite structures and their reference spectra as well as their biological roles, locations and concentrations. The project will be fully compliant with open standards in metabolomics, including existing minimum reporting standards, or will actively contribute to their creation where these have not been developed. It will further provide the community with a repository for metabolomics experiments reference in scientific publications, matching the functionality of other 'omics' repositories at the EBI, such as the proteomics resource PRIDE. Our team is also leading an effort to integrate all enzyme-related information resources at the EBI into what we call the Enzyme Portal. The Enzyme Portal will provide unified access to resources like IntEnz and Rhea developed in our group as well as to some of the resources of the Thornton group, including their Cofactor Database, EC2PDB, the Catalytic Site Atlas and others.

SELECTED REFERENCES

- Griffin, J.L. and Steinbeck, C. (2010) So what have data standards ever done for us? The View from Metabolomics. *Genome Med.* 2, 38.
- Kuhn, T., et al. (2010) Cdk-Taverna: an open workflow environment for cheminformatics. *BMC Bioinform.* 11, 159.
- Rijnbeek, M., and Steinbeck, C. (2009) Orchem - an open source chemistry search engine for Oracle. *J. Cheminformatics* 1, 17.

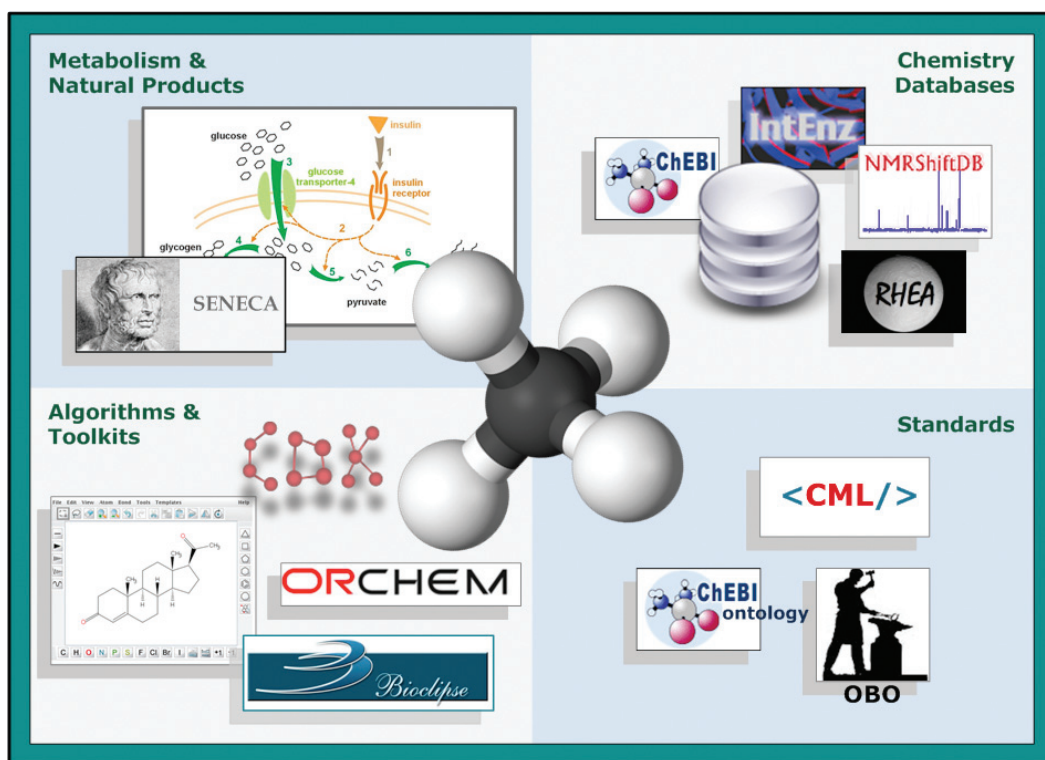


Figure. Scope of work in the cheminformatics and metabolism group.