CONTRACT REFERENCE NUMBER LRI-Q2-MN-081212

Overcoming current Limitations In Metabolism Prediction of Industrial Chemicals (OLIMPIC).

Aleksey Tarkhov[1], Lothar Terfloth[1], Johann Gasteiger[1], Melanie Bausen[2], Eric Fabian[2], Robert Finking[2], Thorsten Bernshausen[3]

[1]Molecular Networks GmbH, [2]BASF SE, [3]Archeiron

Final summary, December 2011

Background: Initiatives such as Canadian Domestic Substance List and REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) and 7th Amendment for Cosmetic Directive in European Union have put heavy emphasis on non-testing methods for hazard and risk assessment of chemicals. In particular, computational methods have become increasingly important for profiling environmental fate and evaluating toxicity of chemical compounds, which lead to risk assessment. Many of the currently available computational methods do not consider chemical reactivity and fate either in environment and biological compartments. The aim in the OLIMPIC project is to provide improvements in the prediction of metabolism of industrial chemicals in mammals.

Objectives: The holistic investigation in the OLIMPIC project considered *in vivo*, *in vitro* and *in silico* metabolism data to identify differences and to improve, in particular, the *in silico* prediction of industrial chemicals in mammals.

Results: Initially, the literature on rat S9-mix metabolism data was reviewed. A literature database was compiled based on this review. The metabolic reactions were extracted from literature and compiled in a data set. This data set contains 414 unique compounds. 340 different metabolites are formed starting from 77 unique substrates. An independent validation set covering particular classes of chemicals was selected and experimentally studied. More than 50 metabolites were identified for this validation data set. The software TIMES (http://oasis-lmc.org/?section=software&swid=4) was used to predicted the metabolism of the validation set.

Due to the limited size of data set with metabolic reaction from literature on rat S9-mix it turned out not to be feasible to derive a statistical model to estimate the likelihood of the formation of a particular metabolite. Therefore, the scope of the study was extended on rat microsomal metabolism data. A reaction database with 3,390 reactions of 1,044 unique substrates yielding 3,018 unique metabolites was used for modeling. The reaction database was clustered and likelihoods for a set of more than 100 reaction types were derived.

Furthermore, models for the prediction of metabolic reactions were developed which are based on physic-chemical descriptors in order to come up with more generic models instead of a set of expert rules. Classification models for various reaction types were investigated. In particular for aromatic hydroxylation reactions it could be shown that a generic classification model based on physico-chemical descriptors of the atoms involved in the metabolic reactions can make reliable predictions instead of a more or less comprehensive and manually encoded set of rather specific expert rules.

Finally, a selected set of 16 reaction rules was implemented within the software package CRAFT which is publicly available. For these rules a prediction model is also available as plugin for CRAFT implemented as web service hosted on Molecular Networks servers.

An Windows installer package of the extended software CRAFT with the OLIMPIC knowledge base was compiled which is available for download from Molecular Networks GmbH web server (http://www.molecular-networks.com/products/craft)

Conclusions: Within the scope of this project a proof of concept was achieved showing the merit of the use of generic physico-chemical descriptor-based models. Physico-chemical descriptors capture reactivity in a generic way which would have to be encoded in a lot of very specific expert rules in a traditional rule based system for metabolism prediction. This could be shown for example for the reaction type of aromatic hydroxylation.

Further work, in particular more detailed experimental measurements on metabolism is needed as precondition to get improved quality of predictive *in silico* models.