Chemistry Central

Poster presentation

Chemical complexity mapping in QSAR models

T Thalheim*, R-U Ebert, R Kühne and G Schüürmann

Address: Department of Ecological Chemistry, Helmholtz-Centre for Environmental Research, Permoserstraße 15, D-04318 Leipzig, Germany
* Corresponding author

This abstract is available from: http://www.journal.chemistrycentral.com/content/3/S1/P40

QSAR models relate features of chemical structures to target properties or effects. Quality models are supposed to apply validated data sets. Typically, the target data are validated in terms of accuracy and reliability. To each data item, a chemical structure is assigned, and in case of 3D geometry models some more or less sophisticated geometry optimisation is performed. However, usually less attention is drawn to the proper representation of chemical identities themselves before entering the model training set. Reported chemical names or even registry numbers often relate to ambiguous chemical structures. There are chemical aspects such as isomerism, mesomerism, and tautomerism, and measured data may relate to generic compound specifications, or to mixtures of defined or even undefined compositions.

Within the framework of the EU projects OSIRIS and 2-FUN, a database concept is introduced to reflect these aspects of chemical complexity. One of the goals of this development is to provide a tool for obtaining representative data sets for QSAR developments, taking into account the chemical complexity in an appropriate manner.

The importance of this approach is demonstrated by example calculations to show the effect of uncertainties due to ambiguous chemical structures on the output of QSAR models. This study is supported by the EU projects OSIRIS (contract No. 037017) and 2-FUN (contract No. 036976).