

Poster presentation

## Implication of heteroatom tautomer in QSAR models

T Thalheim\*, A Vollmer, 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

from 4th German Conference on Chemoinformatics  
Goslar, Germany. 9–11 November 2008

Published: 5 June 2009

*Chemistry Central Journal* 2009, **3**(Suppl 1):P39 doi:10.1186/1752-153X-3-S1-P39

This abstract is available from: <http://www.journal.chemistrycentral.com/content/3/S1/P39>

© 2009 Thalheim et al; licensee BioMed Central Ltd.

For molecules with mobile H atoms, the result of quantitative structure-activity relationships (QSAR) depends on the actual position of the respective hydrogen atoms. Thus, to obtain reliable results, tautomerism needs to be taken into account. A new algorithm to create all tautomer forms of a given molecule based on the mobile H-layer information in the InChI (The IUPAC International Chemical Identifier) code is presented. Unlike published tautomer generation models, this requires no particular rule set. The algorithm eliminates atoms not participating in tautomerism and finally generates all possible combinations. The algorithm was applied to ca. 70,000 structures of the EINECS (European Inventory of Existing Commercial Chemical Substances) database. Ca. 7,500 structures with mobile H atoms were detected. In total, around 200,000 tautomers were generated. Typically, the number of tautomers for single compounds is below 25, but for some molecules this number is extremely large.

Two estimation models, one for the soil sorption coefficient and one for the water solubility, were applied to the subset of compounds with tautomerism. For each substance, the variability of the results due to the different tautomer forms has been inspected. Calculation results exceeding the range of reasonable values have been excluded from the study. The average variation of the soil sorption coefficient within tautomers of individual compounds was almost 0.5 in logarithmic units, and differences up to 3 orders of magnitude were obtained for particular chemicals. For the water solubility, the average variation was between 1 and 2 orders of magnitude, with maximum differences of more than ten logarithmic units.