Chemistry Central Journal

Oral presentation

Calculation of molecular lipophilicity: state of the art and comparison of methods on more than 96000 compounds

M Mannhold*[1], GI Poda[2], C Ostermann[3] and IV Tetko[4]

Address: [1]Düsseldorf, Germany, [2]Chesterfield, USA, [3]Konstanz, Germany and [4]Neuherberg, 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):O7 doi:10.1186/1752-153X-3-S1-O7

This abstract is available from: http://www.journal.chemistrycentral.com/content/3/S1/O7

We first review the state-of-the-art in development of log *P* prediction approaches falling in two major categories: substructure-based and property-based methods. Then, we compare the predictive power of representative methods for one public (*N = 266*) and two *in house* datasets from Nycomed (*N = 882*) and Pfizer (*N = 95809*). A total of 30 and 18 methods were tested for public and industrial datasets, respectively. Accuracy of models declined with the number of non-hydrogen atoms. The Arithmetic Average Model (AAM), which predicts the same value (the arithmetic mean) for all compounds, was used as a baseline model for comparison. Methods with *Root Mean Squared Error* (*RMSE*) greater than *RMSE* produced by the AAM were considered as unacceptable. The majority of analyzed methods produced reasonable results for the public dataset but only seven methods were successful on the both *in house* datasets. We proposed a simple equation based on the number of carbon atoms, NC, and the number of hetero-atoms, NHET: log P = 1.46(± 0.02) + 0.11(± 0.001) NC – 0.11(± 0.001) NHET. This equation outperformed a large number of programs benchmarked in this study. Factors influencing the accuracy of log P predictions were elucidated and discussed (the article is in press in *J. Pharm. Sci.*).