Journal of
Cheminformatics

POSTER PRESENTATION

Open Access

QSPR designer–a program to design and evaluate QSPR models. Case study on pK$_a$ prediction

O Skřehota[1,2], RS Vařeková[1,2*], S Geidl[1], M Kudera[1], D Sehnal[1], C-M Ionescu[1], J Koča[1]

From 6th German Conference on Chemoinformatics, GCC 2010
Goslar, Germany. 7-9 November 2010

Nowadays, a large amount of experimental and predicted data about the 3D structure of organic molecules and biomolecules is available. Advanced computational methods and high performance computers allow us to obtain large sets of descriptors that can be used to estimate physicochemical properties. It is often of interest to study the correlations between descriptors and properties using multilinear regression and to design, parameterize, and test different QSPR (Quantitative Structure Property Relationship) models.

We developed a modular and easily extensible program, called QSPR Designer, which can read or calculate structural properties of atoms and bonds, employ them as QSPR descriptors, and evaluate correlations between the descriptors and the examined physicochemical property of a molecule. Furthermore, the software allows us to effectively design and parameterize QSPR models, calculate physicochemical properties via the models, test the quality of the models, and provide graphs and tables summarizing the results.

The performance of the software is demonstrated by a case study on the prediction of pK$_a$. The pK$_a$ is of fundamental relevance for chemical, biological and pharmaceutical research, because many important physicochemical properties are pK$_a$ dependent. Unfortunately, pK$_a$ is also one of the most challenging properties to calculate [1]. Atomic charges have proven very successful descriptors for the prediction of pK$_a$[2]. Charges can be calculated using a variety of methods (HF, MP2, functionals, etc.), population analyses (Mulliken, ESP, NPA, etc.) and basis sets. Consequently, the procedure of charge calculation strongly influences their correlation with pKa [3]. Using the QSPR Designer, we have successfully designed,

evaluated, and compared 75 different QSPR models for the prediction of pK$_a$ from charges. Our best model predicted the pK$_a$ for 143 phenols with a correlation coefficient 0.969, RMSE (root mean square error) 0.416 and the average pK$_a$ error 0.329.

Author details
[1]National Centre for Biomolecular Research, Masaryk University, Brno, Czech Republic, 625 00, Czech Republic. [2]ANF DATA, a Siemens company, Prague, Czech Republic, 140 00, Czech Republic.

References
1. Lee AC, Crippen GM: **Predicting pK$_a$.** *J Chem Inf Model* 2009, **49**:2013-2033.
2. Dixon SL, Jurs PC: **Estimation of pK$_a$ for organic oxyacids using calculated atomic charges.** *J Comput Chem* 1993, **12**:1460-1467.
3. Gross KC, Seybold PG, Hadad CM: **Comparison of different atomic charge schemes for predicting pK$_a$ variations in substituted anilines and phenols.** *Int J Quantum Chem* 2001, **85**:569-579.

* Correspondence: svobodova@chemi.muni.cz
[1]National Centre for Biomolecular Research, Masaryk University, Brno, Czech Republic, 625 00, Czech Republic
Full list of author information is available at the end of the article