Journal of **Chem**informatics

Open Access

In silico pK_a prediction

Robert Körner[1,2*], Iurii Sushko[1], Sergii Novotarskyi[1,2], Igor V Tetko[1,2]

From 7th German Conference on Chemoinformatics: 25 CIC-Workshop
Goslar, Germany. 6-8 November 2011

The biopharmaceutical profile of a compound depends directly on the dissociation constants of its acidic and basic groups, commonly expressed as the negative decadic logarithm pK_a of the acid dissociation constant (K_a). The acid dissociation constant (also protonation or ionization constant) K_a is an equilibrium constant defined as the ratio of the protonated and the deprotonated form of a compound. The pK_a value of a compound strongly influences its pharmacokinetic and biochemical properties. Its accurate estimation is therefore of great interest in areas such as biochemistry, medicinal chemistry, pharmaceutical chemistry, and drug development. Aside from the pharmaceutical industry, it also has relevance in environmental ecotoxicology, as well as the agrochemicals and specialty chemicals industries.

In literature, a vast number of different approaches for pK_a prediction can be found [1]. These approaches can be divided into two different classes. On the one hand there are direct calculations, so called ab initio methods, trying to determine the pK_a value by quantum chemical or mechanical computation. On the other hand, statistical models, trained on chemical or structural descriptors. These descriptors can be, for example, of quantum chemical, semi empirical, graph topological or simple statistical nature. This type of modeling is called QSPR (Quantitative Structure Property Relationship).

In our recent work, we develop such a QSPR model using localized molecular descriptors to train multiple linear regression and artificial neural networks to estimate dissociation constants (pK_a). The performance of our approach is similar to that of a semi-empirical model based on frontier electron theory [2] as well as a prediction model based on Graph Kernels [3].

How such a prediction model can be built, is shown by an example performed with OCHEM, an *online chemical database with an environment for modeling* (http://ochem.eu/). It is a publicly accessible database for chemical compound data and predictive models. Further, users get the facility to develop, apply, and distribute predictive models, so it is unique in its combination of compound data and predictive models.

Author details
[1]eADMET GmbH, Munich, 85764, Germany. [2]Helmholtz Zentrum München, Munich, 85764, Germany.

References
1. Rupp M, Körner R, Tetko IV: **Predicting the pKa of small molecule.** Combinatorial chemistry & high throughput screening. 2011, **14(5)**:307-327.
2. Tehan BG, *et al*: **Estimation of pKa Using Semiempirical Molecular Orbital Methods. Part 1: Application to Phenols and Carboxylic Acids.** *Quant Struct-Act Relat* 2002, **21**:457-472.
3. Rupp M, Körner R, Tetko IV: **Estimation of Acid Dissociation Constants Using Graph Kernels.** *Molecular Informatics* 2010, **29**:731-740.

* Correspondence: r.koerner@eadmet.com
[1]eADMET GmbH, Munich, 85764, Germany
Full list of author information is available at the end of the article