Prediction of dissociation constant using microconstants

József Szegezdi* and Ferenc Csizmadia: ChemAxon Ltd.Máramaros köz 3/a, 1037 Budapest, Hungary.
*corresponding author, e-mail: jszegezdi@chemaxon.com

Introduction

A new method for predicting the aqueous ionization constants (pK_a) of organic molecules has been developed. The method is based mainly on empirically calculated partial charges. Hydrogen bonds are also parameterized and taken into account within the calculation.

Predicted and experimental values are in good correlation (r^2=0.95, s=0.72, n=1670) for common organic compounds and pharmaceutical molecules.

- ■ Input molecule
- ↓
- ● Assign ionization sites to the molecule
- ↓
- ● Generating all microspecies
- ↓
- ● Calculation of partial charge distribution of microspecies
- ↓
- ● Setting hydrogen bond interactions
- ↓
- ● Calculation of micro ionization constant pK_a of microspecies
- ↓
- ● Calculate ratio of microspecies
- ↓
- ■ Calculate macro ionization constant pK_a

Fig.1. Scheme of pK_a calculation

Calculation of pK_a

The i-th macro ionization constant $K_{a,i}$ of a multiprotic molecule can be calculated with the next expression.

$$K_{a,i} = \frac{\sum_j [i \text{ protons lost from the fully protonated acidic form}]_j}{\sum_k [(i-1) \text{ protons lost from the fully protonated acidic form}]_k}[H^+] = \frac{\sum_j c_j^{(i)}}{\sum_k c_k^{(i-1)}}[H^+]$$

where,

$i=1...N$

N is the number of ionizable atoms in the molecule.
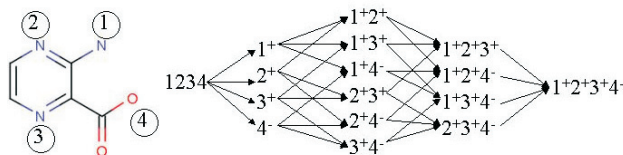
$[H^+]$ is the proton concentration of the aqueous solution

$c_j^{(i)}$ is the concentration of the j-th microspecies released i protons from the fully protonated molecule

$c_k^{(i-1)}$ is the concentration of the k-th microspecies released $(i\text{-}1)$ protons from the fully protonated molecule.

$c_j^{(i)}$ and $c_k^{(i-1)}$ are calculated from the micro ionization constants k_a

Fig.2. Ionization steps of 2-carboxylic-3-amino-pyrazine

Observable macro ionization constants are obtained from the concentration distribution of the microspecies at a given pH. e.g. $K_{a,2}$ is shown.

$$K_{a,2} = \frac{[1^+]+[2^+]+[3^+]+[1^+2^+4^-]+[1^+3^+4^-]+[2^+3^+4^-]}{[1^+2^+]+[1^+3^+]+[2^+3^+]+[1^+2^+3^+4^-]}[H^+]$$

Altogether 2^4 = 16 microspecies can be formed in aqueous solution. Their concentration depends on the micro ionization constants k_a's and the solution's pH. When molecules have a tendency to form tautomers we should consider the most stable isomer during pK_a calculation.
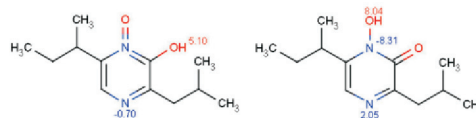

Fig.3. Tautomer forms of aspergillic acid

It is generally true that $2^N\text{-}1$ microspecies can be formed from a molecule containing N ionizable sites. Calculation of the relative concentration of microspecies requires $2^N\text{-}1$ microconstants[1]. Ratios can be calculated if micro ionization constants are available.

The micro ionization constant (pK_a) value is calculated from the partial charge distribution and the atomic polarizability of microspecies[2,3] using empirical linear or non-linear equations.

Example for pK_a - partial charge relation

The next figure shows the pK_a - partial charge distribution of a basic oxygen atom in pyridine N-oxide. (para-NH_2 has the largest pK_a and ortho-NO_2, the smallest).

The pK_a, as a function of the partial charge (q) is obtained through linear regression analysis using the quadratic funtion form.

$$pk_a = p_1(1+q)^2 + p_2(1+q) + p_3$$

where p_1, p_2 and p_3 are regression coefficients
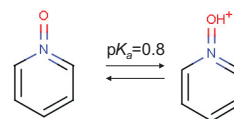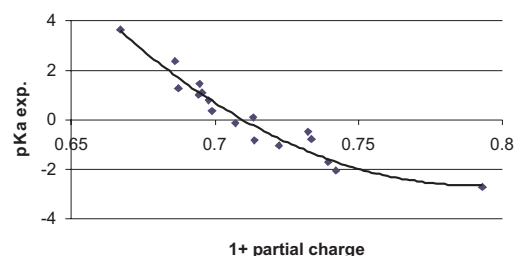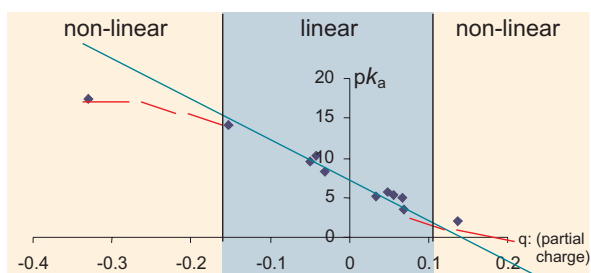

Fig.4. The basic pk_a v. the partial charge of oxygen atom

Sometimes non-linear regression analysis according to the regression equation,

$$pk_a = p_1 \exp(p_1 q) + p_2 q + p_3$$

provides a better fit than a quadratic equation. The example below is the pK_a charge relation of substituted pyrrole derivatives.

It seems to be generally true that the micro ionization constant-partial charge relation has three separable ranges. The majority of data fall into the linear domain range, however very weak or very strong acids or bases are in the non-linear domain.

Explanation of this phenomenon is beyond the scope of this presentation.
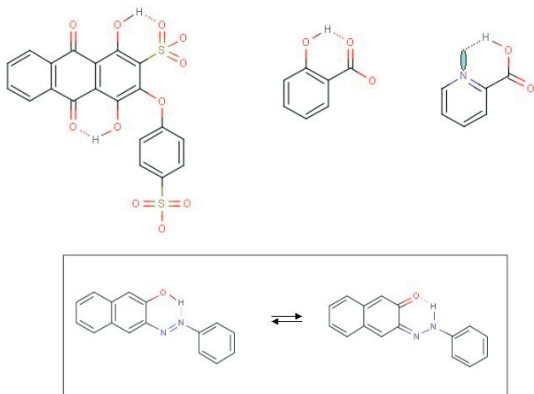
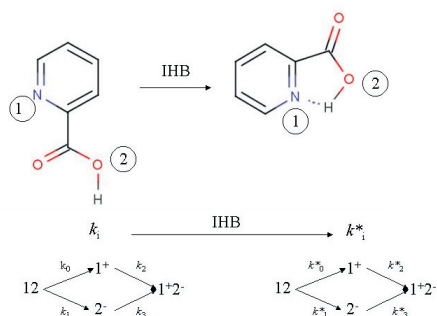## Modeling of the intramolecular hydrogen bond (IHB)

Fig.5. Some typical IHB's

Fig.6. Example for the parameterization of the IHB in picolinic acid.

where,

k_i is the i-th micro ionization constant, predicted with Marvin

$K^*_{a,i}$ is the i-th observed macro ionization constant

k^*_i is the i-th micro ionization constant which is calculated from the equilibrium equations and $K^*_{a,i}$

The i-th macro ionization constant $K^*_{a,i}$ is a function of the four micro ionization constants k^*_i's.

$$K^*_{a,1} = k^*_0 \left(1 + \frac{k^*_1}{k^*_3}\right) \quad or \quad K^*_{a,1} = k^*_0 + k^*_2$$

(and additional two relation for $K^*_{a},2$)

The optimized values of k^*_i's are calculated from the non-linear equation set of $K^*_{a,i}$'s. (Four equations and four variables.)

The change of the i-th micro ionization constant Δk_i due to the IHB, is obtained from the relation:

$$\Delta k_i = k^*_i - k_i$$

Different picolinic acid derivatives result in different Δk_i values, depending on the partial charge of nitrogen (acceptor) and hydroxyl group (donor). Δk_i as a function of partial charge of donor q_{donor} or acceptor $q_{acceptor}$ is obtained with linear regression analysis.

$$\Delta k_i = a_i q_{donor} + b_i \quad or \quad \Delta k_i = c_i q_{acceptor} + d_i$$

Where a_i, b_i, c_i, d_i are regression parameters

When the IHB is detected in a molecule k'_i is used instead of k_i within Marvin.

$$k'_i = k_i + \Delta k_i$$

The change of $K_{a,i}$ caused by the IHB is approximately proportional to the partial charge difference between the acceptor and donor atoms.

$$\Delta K_{a,i} \approx c_i (q_{acceptor} - q_{donor}) + d_i$$

Where c_i and d_i are regression coefficients

■ Input n molecules and their observed K^*_a's

● Calculate all k^*_i with non-linear equ. solver

● Predict the IHB free k_i with Marvin

● Calculate $\Delta k_i = k^*_i - k_i$ for all i

■ Calculate $\Delta k_i = a_i q + b_i$ with regression analysis for all i

Fig.7. Scheme for the parameterization of the internal hydrogen bond

Test of pK_a calculations

After a preliminary data preparation, altogether 1670 molecules were used for testing the performance of the pK_a calculation model. Observed pK_a data are taken from the PhysProp database.

Certain CH acid derivatives, molecules with ether type oxygen and some tautomeric compounds were excluded from the testing. (We are working on the prediction of these types of molecules.) Several erroneous structures were also omitted.
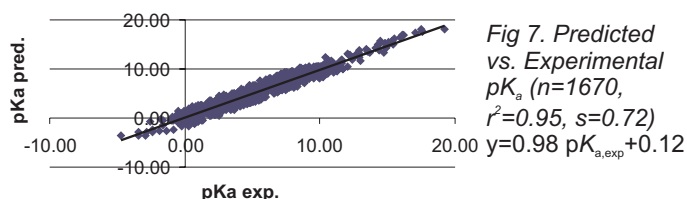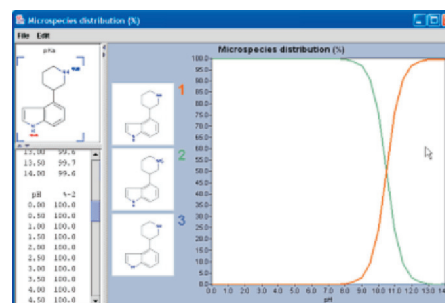

Fig 7. Predicted vs. Experimental pK_a (n=1670, r^2=0.95, s=0.72) y=0.98 p$K_{a,exp}$+0.12

Implementation

All calculations performed using Marvin version 3.4.pre1 Calculation Plugins available through ChemAxon's Marvin and JChem software suites (100% Java)

Hardware and software requirements: any system running Java Runtime Environment 1.1 or above.

Marvin implementation showing predicted pK_a's and microspecies distribution chart

References

1. Csizmadia, F.; Tsantili-Kakoulidou, A.; Panderi, I.; Darvas, F., *J.Pharm. Sci.* **1997**,86, 866-871
2. Gasteiger, J.; Marsili, M.: *Tetrahedron*, **1980**, 36, 3219
3. Miller, K.J.; Savchik, J.A., *J.Am.Chem.Soc.*, **1979**, 101, 7206-7213

ChemAxon