A method for calculating the pK_a values of small and large molecules

József Szegezdi[+] and Ferenc Csizmadia, ChemAxon Ltd, Maramaros koz 3/a, 1037 Budapest, Hungary. [+]Corresponding author: jszegezdi@chemaxon.com

Introduction

A method was developed for predicting of the aqueous ionization constants (pK_a) of organic molecules. The method is based on empirically calculated physico-chemical parameters that are obtained from ionization site-specific regression equations.

Submit molecule

↓

Take the major resonant form or/and the major tautomeric form

↓

Calculating the micro pk_a from empirical increments

- partial charge increment
- polarizability increment
- structure specific increments

↓

Calculating the microspecies distribution

↓

Assigning calculated pK_a values to the atoms of the submitted molecule

Scheme of pK_a calculation

Calculation of pK_a

pK_a of the monoprotic molecules to be calculated as the sum of the next three increments.

$$pK_a = a*Q + b*P + c*S+d$$

where,

Q is the partial charge increment
P is the polarizability increment
S is sum of the structure specific increments
a,b,c, and d are regression coefficients specific to the ionization site

Structure specific increments may contain two quantities: steric strain or/and H-bond increments.

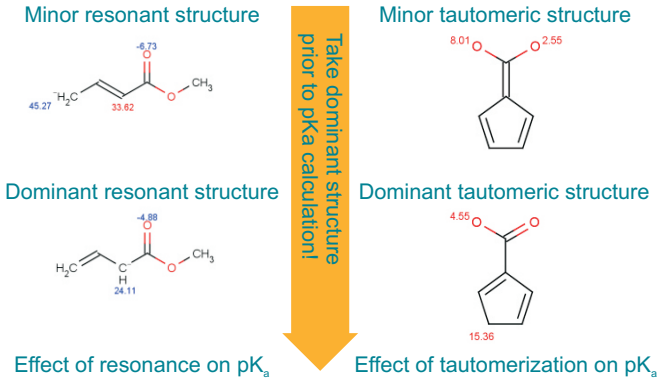
All of the pK_a increments above are calculated from ionization-site specific regression equations.

We should pay special attention to molecules with a large number of ionization sites. In these molecules ionization sites may directly perturb each other through proximity effects like H-bonding or altering electron withdrawing effects by protonation/deprotonation. 'Macro effects' take place when micro pK_a values of sites are close to each other, which shifts the macro pK values measured experimentally. The 'macro effect' is independent of the distance between the ionizable groups.

The pK_a calculation of multiprotic molecules is governed by theoretically derived kinetic equations in our model. One of the outstanding characteristics of the theoretical expressions of the pK_a calculation is that interactions among the ionizable atoms are taken into account also in the case when the ionizable atoms are far from each other. We also take advantage of this behavior in the case of large molecules that may have ionization sites far from each other. Proteins are typical molecules in this category.

One of the most difficult issues is how the effect of tautomerization and resonance can be taken into account in pK prediction. These two isomerization processes may bring about significant difference between predicted and experimentally obtained pK_a's.

This is why we developed a preprocessing algorithm for pK_a calculation which generates the most dominant tautomeric and resonant structures of the submitted molecules prior to the pK_a calculation.



Effect of resonance and tautomerization on pK_a calculation

Types of the acidic and the basic groups in our pK_a model

Acidic groups:

1. X-AH, where X is any atom without lone pair electron which does not take part in delocalization with the surrounding atoms. e.g. ethanol

2. X-AH, where X has lone pair electron or p_z orbital, e.g. acetic acid

3. AH, where A takes part in extended delocalization. e.g. pyrrole

Type of the basic groups

Atom B should have a lone pair

Basic groups:

1. X-B, where X is any atom without a lone pair and it does not take part in a delocalization with the surrounding atoms, e.g. methylamine

2. X-B, it has lone pair or p_z orbital, e.g. amidine

3. B, where B takes part in an extended delocalization, e.g. pyridine

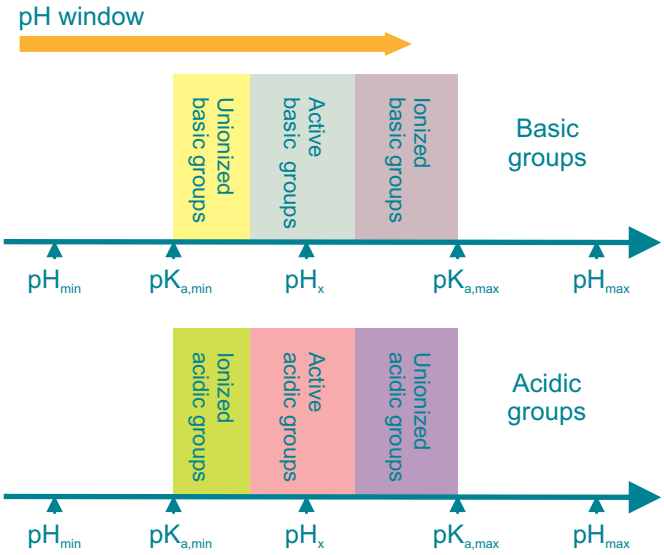
The small and large models of pKa calculations

We use the small pK_a model if the number of the ionizable atoms (N) is smaller or equals to 8.

The small pK_a model generates altogether 2^N microspecies from the submitted molecule.

The micro ionization constant of these microspecies are calculated with the empirical relation given above. Finally, macro pK_a values are obtained from the theoretical relations that hold between macro-micro pK_a values.

The large model is invoked if the number of the ionizable atoms (N) is larger than 8. Since calculation time and memory requiment strongly depends on N, therefore, we developed a so-called pH window method for reducing N. Only pivot atoms are taken into account in the pK_a calculation at a given pH. This technique described in Fig.3. below.



Only those basic or acidic groups are considered to be protonable or deprotonable at a given pH_x which have a micro pK_a value inside the pH window. The width of the pH window is 2 pH units. All atoms that are outside the pH window are considered inactive. Number of ionizable atoms are reduced to N=8 in this way. The reduced microspecies set contains only 2^8 microspecies.

pK_a of the active groups at a given pH can be calculated according to this relation.

$$K_{a,i} \quad \frac{i \text{ protons lost from the fully protonated acidic form}}{(i-1) \text{ protons lost from the fully protonated acidic form}} \quad H \quad \frac{c_j^i}{c_k^{i-1}} H$$

where,

[H+] is the proton concentration of the aqueous solution

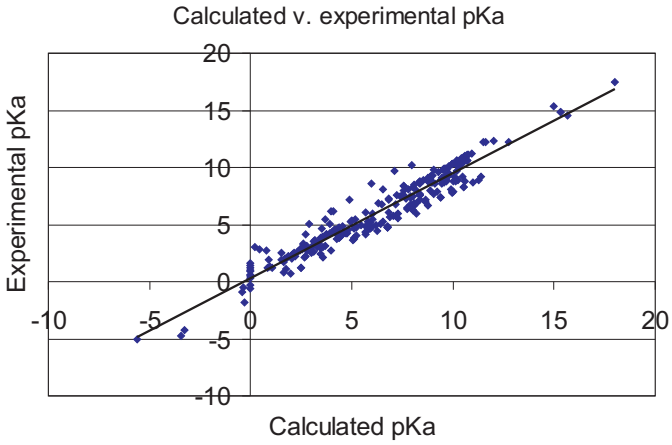
c_j^i is the concentration of the j-th microspecies that released i protons from the fully protonated molecule

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

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

Test of the large and the small models

For proteins we defined H-bond interaction between carboxyl (COOH) and amide groups ($CONH_2$). Conditions of the H-bond interaction calculated from the 3D geometry of the protein.



Test results for small molecules . For a diverse set of small molecules we used the small pK_a prediction model. n=269, r2=0.93 , s=0.75

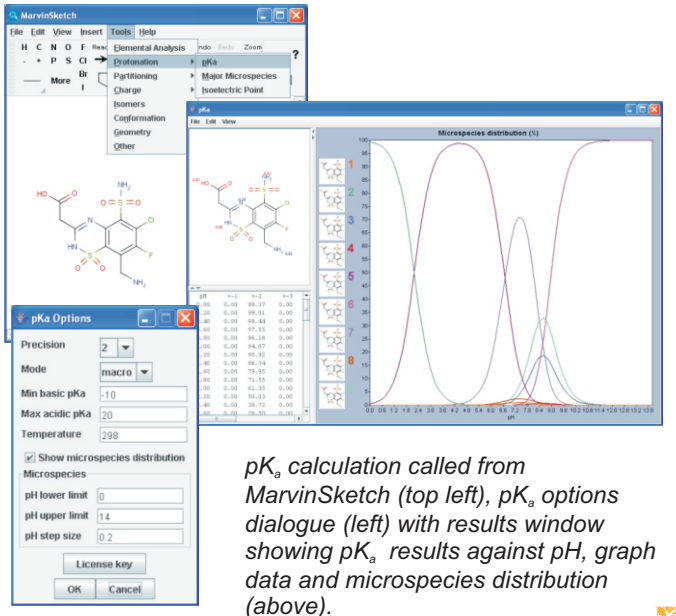
Acidic groups Predicted pK_a	Residue number	Exp. pK_a	Basic groups Predicted pK_a	Residue number	Exp. pK_a
3.55	2	-	12.20	41	
4.16	37	-	11.61	50	
4.86	7	5.60	10.19	34	
5.36	44	5.60	7.71	1	
6.14	61	7.00	-0.26	3	
9.21	73	-	-0.41	41	
9.81	10	-	-0.69	42	
14.12	23	-	-1.04	52	

Table of test results for a protein (1A91.pdb) molecule (SUBUNIT C OF THE F1FO ATP SYNTHASE OF ESCHERICHIA COLI)

Implementation

Calculation plugins are available through Marvin and JChem software suites

Hardware and software requirements: any system running Java Runtime Environment 1.4 or above (The API is also accessible from .NET).



pK_a calculation called from MarvinSketch (top left), pK_a options dialogue (left) with results window showing pK_a results against pH, graph data and microspecies distribution (above).

First presented at American Chemical Society Spring meeting, March 25-29th, 2007

ChemAxon