Chemical Engineering Science 183 (2018) 95–105

Contents lists available at ScienceDirect

Chemical Engineering Science

journal homepage: www.elsevier.com/locate/ces

Prediction of acid dissociation constants of organic compounds using group contribution methods

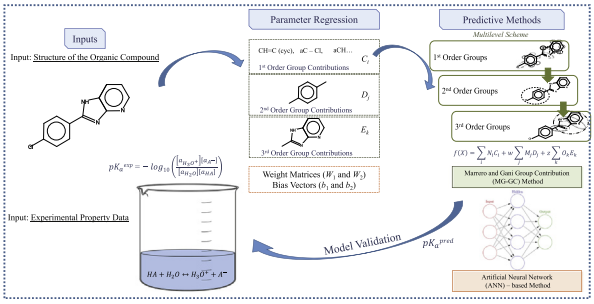
Teng Zhou [a,b], Spardha Jhamb [b], Xiaodong Liang [b], Kai Sundmacher [a], Rafiqul Gani [b,c,*]

[a] Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, D-39106 Magdeburg, Germany
[b] KT Consortium, Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark
[c] PSE for SPEED, Skyttemosen 6, DK-3450 Allerod, Denmark[1]

HIGHLIGHTS

- Prediction of acid dissociation constants (K_a) for a large set of organic compounds.
- The Marrero and Gani–Group Contribution (MG-GC) method to develop the property models.
- Linear and nonlinear GC models for amino acids and other classes of compounds.
- An Artificial Neural Network (ANN) based GC model for organic compounds.
- Modeling details and model parameters provided.
- Accuracy of the models demonstrated through application examples.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 27 October 2017
Received in revised form 6 February 2018
Accepted 2 March 2018
Available online 5 March 2018

Keywords:
Acid dissociation constant
pK_a
Group contribution method
Artificial neural network
Amino acids
Organic compounds

ABSTRACT

In this paper, group contribution (GC) property models for the estimation of acid dissociation constants (K_a) of organic compounds are presented. Three GC models are developed to predict the negative logarithm of the acid dissociation constant pK_a: (a) a linear GC model for amino acids using 180 data-points with average absolute error of 0.23; (b) a non-linear GC model for organic compounds using 1622 data-points with average absolute error of 1.18; (c) an artificial neural network (ANN) based GC model for the organic compounds with average absolute error of 0.17. For each of the developed model, uncertainty estimates for the predicted pK_a values are also provided. The model details, regressed parameters and application examples are highlighted.

1. Introduction

The acid dissociation constant (K_a) of a compound, which expresses the extent to which the compound in its aqueous solution is dissociated into its ionic form, is sought after by many chemists, biochemists and product formulators. Although experimental measurements would yield the most satisfactory results,

* Corresponding author at: KT Consortium, Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark.
E-mail address: rgani2018@gmail.com (R. Gani).
[1] Current address.

https://doi.org/10.1016/j.ces.2018.03.005
0009-2509/

it is not always convenient to setup and conduct experiments for K_a determination. This is because the organic compounds that weakly dissociate lack adequate spectral differences in the dissociated and undissociated forms. Besides, in the cases where a compound is unstable or is insufficiently soluble in water, experimental K_a determination is impossible (Tong and Wen, 2008).

The currently available pK_a (negative logarithm of K_a) compilations provide values for only a small fraction of known or possible acids and bases (Perrin et al., 1981). This motivates the development of advanced pK_a prediction models.

This paper is organized as follows. First, we give a definition on pK_a and highlight its significance in several research areas (Section 1.1). After a brief introduction of the main existing methods for pK_a prediction (Section 1.2), we focus on the powerful group contribution (GC) methods and present more details about these methods in Section 2. Three different GC models are then developed to predict pK_a for amino acids and other classes of organic compounds. The performances of these models are evaluated and compared in Section 3.1. Finally, in Section 3.2, several examples are shown to help the reader in understanding how to apply the developed models for predicting pK_a.

1.1. Definition and significance of pK_a

In aqueous solution, acids (generically represented by HA) undergo a protolytic reaction with water. This equilibrium reaction is given as:

$$HA + H_2O \leftrightarrow H_3O^+ + A^- \tag{1}$$

The equilibrium constant (in this case, the acid dissociation constant K_a) for the reaction given in Eq. (1) is expressed as Eq. (2), which relates the activity of the dissociated form of the acid (a_{A^-}) to the activity of its undissociated form (a_{HA})

$$K_a = \frac{[a_{H_3O^+}][a_{A^-}]}{[a_{H_2O}][a_{HA}]} \tag{2}$$

As the K_a measurements are generally made in dilute aqueous solutions, the concentration of water remains nearly constant and therefore, its activity can be taken as unity. The general expression of K_a is then derived from Eq. (2), as

$$K_a = \frac{[a_{H_3O^+}][a_{A^-}]}{[a_{HA}]} \tag{3}$$

By taking negative logarithm on both sides of Eq. (3) and rearranging the terms, the relation between the pH of the solution and the pK_a of HA can be obtained, given as Eq. (5).

$$-\log(K_a) = -\log([a_{H_3O^+}]) - \log\left(\frac{[a_{A^-}]}{[a_{HA}]}\right) \tag{4}$$

$$\Rightarrow pK_a = pH + \log\left(\frac{[a_{HA}]}{[a_{A^-}]}\right) \tag{5}$$

In the special case, when the activity of HA equals that of A^-, pK_a is identical to pH.

pK_a is very significant in many different areas. For example, during liquid-liquid extraction, when an organic compound is to be separated from an aqueous solution, the undissociated form of the compound usually is more soluble in the organic phase. Hence, the pH of the aqueous phase can be adjusted to its optimum value if the pK_a of the organic compound is known (Green and Perry, 2008). In preparative chemistry, considering the effects of pH on the properties of reactants as well as the possible intermediates and products, conditions for synthesis are selected by making use of pK_a (Perrin et al., 1981).

1.2. Existing methods for pK_a prediction

Nowadays a large number of experimental pK_a data are available, thus one can predict pK_a of new compounds by extrapolating or interpolating the pK_a of database compounds of the same type. Besides this, theoretical calculations and semi-empirical correlations based on thermodynamics and quantum chemical foundations have also been used for pK_a prediction in various works (e.g., Jensen et al., 2017 use isodemic reactions, where the pK_a is estimated relative to a chemically related reference compound, to make COSMO-based and SMD-based predictions. The pK_a values of 53 amine groups in 48 druglike compounds are computed.)

1.2.1. Linear free energy relationships (LFER)

The Hammett-Taft equation quantifies the electronic effect of organic functional groups (or substituents) on other groups to which they are attached. This equation is a linear free energy relationship (LFER). It is widely used for pK_a prediction (Metzler, 2012) and is as shown in Eq. (6).

$$pK_a = pK_a^0 - \rho \sum \sigma_i \tag{6}$$

where pK_a^0 indicates the pK_a value for unsubstituted reference compounds; σ_i is the substituent constant for the substituent *i;* and ρ is the proportionality constant for the particular equilibrium dissociation reaction i.e. it is the measure of the sensitivity of the reaction to the presence of electron-withdrawing or electron-donating substituents, for example the ρ for phenylacetic acids is 0.49, while that for phenols is 2.23. It should be noted that, currently only a limited number of substituent constants are available, which limits the applicability of the LFER method for pK_a prediction.

1.2.2. Theoretical calculations

There are several first-principle theory based methods for pK_a prediction. The Kirkwood-Westheimer equation (Kirkwood and Westheimer, 1938) quantifies ΔpK_a for a charged or a dipolar substituent as follows,

$$\Delta pK_a = \frac{e\mu\cos\phi}{2.3kTR^2D_{eff}} \tag{7}$$

In Eq. (7), ϕ is the angle between the line joining the centre of the ionizing group to the centre of the dipole and the axis of the dipole, e is the electronic charge, k is the Boltzmann constant, T is the temperature in K, μ is the dipole moment, R is the distance between two charges, D_{eff} is the effective dielectric constant. The largest limitation of the Kirkwood-Westheimer method is that it is applicable only to ellipsoidal molecules with point charges at their foci only.

pK_a can also be estimated based on thermodynamic cycles that relate the gas phase to the solution phase, where state-of-the-art quantum chemical techniques coupled with an appropriate solvation model are used (Shields and Seybold, 2013). Jang et al. (2001) predicted the pK_a values for a series of 5-substituted uracil derivatives using density functional theory (DFT) calculations in combination with the Poisson-Boltzmann continuum-solvation model (Im et al., 1998).

Even though theoretical calculations can yield good results in predicting pK_a, these methods are not very attractive for some applications due to their high computational cost. For instance, in drug formulation design, the pK_a of active ingredients (AIs) is a very important property for selecting AIs because the pK_a value indicates the aqueous solubility of the AI and the ability of the AI to permeate through the gastro-intestinal membrane. In order to perform a fast AI pre-screening, a quick and reliable pK_a prediction method is more preferable than an accurate but very computationally expensive one.