Chemical Engineering Science 175 (2018) 148–161

Contents lists available at ScienceDirect

Chemical Engineering Science

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

Estimation of physical properties of amino acids by group-contribution method

Spardha Jhamb [a], Xiaodong Liang [a], Rafiqul Gani [a,*], Amol Shivajirao Hukkerikar [b]

[a] Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark
[b] Systems Engineering, Food Systems Business Unit, Alfa Laval Copenhagen A/S, Maskinvej 5, DK-2860 Søborg, Denmark

HIGHLIGHTS

- Prediction of normal melting point, water solubility and octanol-water partition coefficient of amino acids.
- Group contribution approach to develop the property models.
- Provides uncertainty estimates for the modeled properties.
- Provides modelling details together with model parameters.

ARTICLE INFO

Article history:
Received 24 May 2017
Received in revised form 2 September 2017
Accepted 9 September 2017
Available online 12 September 2017

Keywords:
Normal melting point
Aqueous solubility
Octanol/water partition coefficient
Group contribution method
Amino acids

ABSTRACT

In this paper, we present group-contribution (GC) based property models for the estimation of physical properties of amino acids using their molecular structural information. The physical properties modelled in this work are normal melting point (T_m), aqueous solubility (W_s), and octanol/water partition coefficient (K_{ow}) of amino acids. The developed GC-models are based on the published GC-method by Marrero and Gani (2001) with inclusion of new structural parameters (groups and molecular weight of compounds). The main objective of introducing these new structural parameters in the GC-model is to provide additional structural information for amino acids having large and complex structures and thereby improve the predictions of physical properties of amino acids. The group-contribution values were calculated by regression analysis using a data-set of 239 values for T_m, 211 values for W_s, and 335 values for K_{ow}. Compared to other currently used GC-models, the developed models make significant improvements in accuracy with average absolute error of 10.8 K for T_m and logarithm-unit average absolute errors of 0.16 for K_{ow} and 0.19 for W_s.

1. Introduction

Amino acids, both natural as well as modified types, are widely used in the food and pharmaceutical industry on account of their nutritive and therapeutic properties. They are also extensively used in the cosmetics and personal care products owing to their high solubility and compatibility with the skin. Further, use of amino acids in consumer-oriented chemical products does not cause any harmful effects on the environment, a vital requirement from REACH® regulation (Echa.europa.eu, 2017) which compels European chemical companies to stop the use of hazardous chemical substances. Recently, driven by the demand for biodegradable chemicals, amino acids are also being studied for the development of biodegradable plastics. Table 1 lists selected amino acids and their applications in food, cosmetics and personal-care, therapeutic, and polymer industry.

Abbreviations: AAE, average absolute error; ARE, average relative error [%]; $A_{Log\ ws}$, adjustable parameter of Eq. (4); $A_{Log\ kow}$, adjustable parameter of Eq. (5); α, adjustable parameter of Eq. (3); $B_{Log\ ws}$, adjustable parameter of Eq. (4); C_i, contribution of first-order group of type-i; $COV(\boldsymbol{P}^*)$, covariance matrix; D_j, contribution of second-order group of type-j; E_k, contribution of third-order group of type-k; $f(X)$, function of property X; GC, group-contribution; MG, Marrero and Gani; M_j, occurrence of second-order group of type-j; N, number of experimental data-points used in the regression; N_i, occurrence of first-order group of type-i; O_k, occurrence of third-order group of type-k; $\boldsymbol{P}$, model parameters; $\boldsymbol{P}^*$, estimated values of model parameters; R^2, coefficient of determination; $S(\boldsymbol{P})$, cost function; SD, standard deviation; SSE, minimum sum of squared errors; $t(\nu, \alpha_t/2)$, t-distribution value corresponding to the $\alpha_t/2$ percentile; ν, degrees of freedom; X^{exp}, experimental property value; X^{pred}, predicted property value.

* Corresponding author.
E-mail address: rag@kt.dtu.dk (R. Gani).

https://doi.org/10.1016/j.ces.2017.09.019
0009-2509/

Table 1
Selected applications of amino acids in food, cosmetics, personal-care, therapeutic, and polymer industry.

Sr. No.	Amino acid	Amino acid derivative	Use	Reference
Applications in food industry				
1	Lysine	L-Lysine hydrochloride	Dietary supplement	Leuchtenberger et al. (2005)
2	Glutamic acid	Monosodium Glutamate	Flavor enhancer	Ault (2004)
3	Tryptophan	L-Tryptophan	Antioxidant for preservation of milk powder	O'Connor and O'Brien (2006)
4	Cysteine	L-Cysteine	Antioxidant for preservation of fruit juices	Lee (1995)
5	Aspartic acid	Aspartyl-phenylalanine methyl ester	Low calorie artificial sweetener	Tandel (2011)
Applications in cosmetics and personal-care products				
1	Arginine	L-Arginine	Neutralizer in soap (alternative base)	Sakamoto (2016)
2	Glutamine	L-Glutamine	Neutralizer in soap (alternative acid)	Sakamoto (2016)
3	Proline	1-carbamimidoyl-L-proline	Improved hydration of stratum corneum (SC)	Sakamoto (2016)
4	Cystine	N-acetyl cysteine	Generation of melanin in UV-protecting creams	Sakamoto (2016)
5	Cysteine	L-Cysteine hydrochloride	Hair waving / Hair straightening agents	Sakamoto (2016)
6	Glycine	N-Acylglycinate	Mild, functional anionic surfactant	Sakamoto (2016)
7	Alanine	N-Acylalanate	Mild, functional anionic surfactant	Sakamoto (2016)
8	Serine	N-(4-pyridoxylmethylene)-L-serine	Anti-oxidative effects against UV Radiation	Sakamoto (2016)
Therapeutic uses				
1	Arginine	L-Arginine hydrochloride	Treat heart and circulatory diseases, combat fatigue	Gad (2010)
2	Phenylalanine	L-dihydroxyphenylalanine	Treatment of Parkinson's disease	Hornykiewicz (1974)
3	Tryptophan	5-hydroxytryptophan	Treatment of depression, anxiety and sleeping disorders	Turner et al. (2006)
4	Glutamine	L-Glutamine	Treatment of mucositis, Gastrointestinal health	Meletis and Barker (2005)
5	L-Carnitine	N-Acetyl Carnitine	Treatment of Alzheimer's disease	Meletis and Barker (2005)
As a solvent				
1	Lysine	Potassium salt solution of L-Lysine	Solvent for CO_2 capture in flue gas	Lerche et al. (2012)
As a bio-degradable polymer (under investigation)				
1	Alanine	DL-Alanine	Degradable polymeric drugs	Domb (1990)
2	Aspartic acid	Polyaspartate	Disposable diapers	Low et al. (1996)

The normal melting point (T_m) of amino acids is a fundamental physical property of amino acids that allows estimation of solubility of amino acids (when used together with heat of fusion data as shown by Eq. (1) (Kontogeorgis and Folas, 2010)) and provides useful information on whether a substance is solid or liquid at normal conditions.

$$\ln x_i\gamma_i = -\frac{\Delta H^{\circ}_{fus}}{RT}\left(1-\frac{T}{T_{m,i}}\right)+\frac{\Delta C_{p,i}}{R}\left(\frac{T_{m,i}-T}{T}\right)-\frac{\Delta C_{p,i}}{R}\ln\frac{T_{m,i}}{T} \quad (1)$$

where x_i is the solubility of component i at temperature T, γ_i is the activity coefficient, ΔH°_{fus} is the standard heat of fusion, R is the universal gas constant, $T_{m,i}$ is the melting point and $\Delta C_{p,i,}$ is the difference between the heat capacity of the liquid and the solid.

The octanol/water partition coefficient (K_{ow}) is the ratio of the equilibrium concentrations of organic compound in the two phases. By convention, the ratio of concentrations is expressed as octanol over water. K_{ow} is a measure of tendency of an organic compound to partition out of water into other environmental compartments (these could be sediments, micro-organisms, etc.). Organic compounds with K_{ow} less than 10^2 will not partition into any environmental compartments. Whereas, compounds with K_{ow} greater than 10^6 will tend to partition and accumulate in the environmental compartments.

Water solubility (W_s), on the other hand, is defined as the maximum amount of an organic compound that is dissolved in pure water at a specific temperature, and it is also an important parameter in environmental studies. The relevance of T_m, W_s, and K_{ow} of amino acids in chemical process-product design and environmental studies is therefore very high. There are three ways in which a property user can obtain these physical properties: (i) using available databases/open literature; (ii) performing laboratory measurements; and/or (iii) using property prediction methods. A key limitation associated with the use of databases is the limited number of compounds and limited amount of property-data stored in them. While use of experimentally measured physical properties is highly desirable, laboratory measurements may be time consuming, expensive, and sometimes may not even be feasible due to unavailability of samples or due to thermal decomposition issue. Therefore, it is more practical and convenient to employ property prediction methods to obtain the physical properties of amino acids, at least in the early stages of chemical product-process design (Gani and Constantinou, 1996).

The normal melting point of organic compounds is a difficult property to correlate because it is dependent upon the arrangement of the molecules in the crystal lattice as well as upon the strength of the pairwise group interactions (Katritzky et al., 2001).

The modelling of melting point of amino acids is even more challenging and difficult due to the fact that these compounds are amphoteric and hence they exist as zwitterions at their isoelectric point (pI). Their ability to form zwitterions leads to stronger electrostatic attraction which in turn leads to larger energy requirement to break the bonds. For the estimation of normal melting point of organic compounds, several property prediction models have been published in the literature. These models can be broadly classified into two types: (i) GC based models; and (ii) Quantitative Structural-Property Relationship (QSPR) based models. In a GC-method, the property of a pure compound is a function of structurally dependent parameters, which are determined as a function of the frequency of the groups representing the pure compound and their contributions. In principle, it is an additive method, where the contributions of each group towards a property are summed to obtain the property value. Simamora and Yalkowsky (1994) proposed a simple GC method to estimate the normal melting point of aromatic compounds using 1690 compounds in the regression. Krzyzaniak et al. (1995) developed a simple GC method for estimation of normal melting of aliphatic, non-hydrogen bonding compounds. The GC methods developed by Joback and Reid (1987), Constantinou and Gani (1994), Marrero and Gani (2001), and Hukkerikar et al. (2012) also allow quick estimations of normal melting points of organic compounds without