Fluid Phase Equilibria 427 (2016) 246–258

Contents lists available at ScienceDirect

Fluid Phase Equilibria

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

Solubility modelling for phytochemicals of *Misai Kucing* in different solvents

Wai Lip Theo [a], Azizul Azri Mustaffa [a, b, *], Jeng Shiun Lim [a, b]

[a] Faculty of Chemical and Energy Engineering, Universiti Teknologi Malaysia (UTM), 81310 UTM Johor Bahru, Johor, Malaysia
[b] Process Systems Engineering Centre (PROSPECT), Research Institute for Sustainable Environment (RISE), Universiti Teknologi Malaysia (UTM), 81310 UTM Johor Bahru, Johor, Malaysia

ARTICLE INFO

Article history:
Received 4 June 2016
Received in revised form
19 July 2016
Accepted 21 July 2016
Available online 22 July 2016

Keywords:
Solid-liquid equilibrium (SLE)
Quantitative Structure-Property Relationship (QSPR) model
KT-NIST-UNIFAC model
Solubility modelling
Flavonoid
Orthosiphon *stamineus*

ABSTRACT

Recently development of computer-aided herbal plant extraction process design has received attention for highly positive future outlook of global herbal industries. Solid-liquid equilibrium (SLE) prediction models comprise a core part of the chemical engineering knowledge base required for the computer-aided process design. In this study, solubility model was developed to predict the solubility of several Orthosiphon *stamineus* (*Misai Kucing*) phytochemicals (i.e. Sinensetin (SEN), Eupatorin (EUP) and Tetramethylscutellarein (TMF)) in seven solvents, in addition to creating physicochemical property database comprising of melting point temperature, fusion enthalpy, and SLE solubility data. Quantitative Structure-Property Relationship (QSPR) model was developed for fusion enthalpy prediction with absolute average relative deviation (AARD) of 9.33%. For improvement of KT-NIST-UNIFAC model, new subgroups (i.e. aC − CO (fused), aC − O (fused) and C=C_{cyclic} − OH) were introduced to better represent the molecular structure of studied flavonoid compounds, and binary interaction parameters were regressed using compiled SLE solubility data. Regressed KT-NIST-UNIFAC model exhibited better performance with AARD of 5.27%. Subsequent simulation study using improved model indicated that acetone was compatible for SEN, EUP, and TMF extraction.

1. Introduction

Herbal industries are anticipated to flourish globally to cater the ever-rising demand worldwide. The market values of herbal healthcare products are speculated to be USD 20 to 70 billion, with average annual growth rate of 15–20% [1]. Their popularity has led to intensive research and development in extraction process design, and solvent design for solubility optimisation is one of the core aspects.

A number of empirical optimisation studies on extraction of phytochemicals from Orthosiphon *stamineus* had been carried out. Specifically, Chew et al. [2] empirically optimised ethanol concentration, extraction time and temperature for maximising phenolic compound yield and antioxidant capacity of Orthosiphon *stamineus* extract. In another study, ethanol-based Orthosiphon *stamineus* oil extraction was optimised using artificial neural network (ANN), via manipulation of extraction temperature, extraction time and number of extraction cycle [3]. Response surface methodology (RSM) was also used to determine the optimal operational parameters for ultrasound assisted extraction process of Orthosiphon *stamineus* using ethanol [4], and water-based solid-liquid extraction (SLE) of Orthosiphon *stamineus* [5]. In addition, Azni [6] explored several extraction options i.e. microwave-assisted extraction, ultrasound-assisted extraction and maceration for aqueous ethanol-based Orthosiphon *stamineus* extraction optimisation.

However, these empirical studies exhibited several weaknesses. Firstly, it involves trial-and-error solvent selection in extraction process optimisation [1]. Cost, time and technical constraints of experimentation have made analyses using extensive array of different solvents impractical and infeasible [7]. As indicated by the review above, solvents examined for Orthosiphon *stamineus* extraction were limited to water, ethanol and aqueous ethanol solution, other better solvents for extraction of phytochemicals in Orthosiphon *stamineus* might be overlooked. Moreover, experimental approach may involve high risk of chemical hazards from solvent application, and restricted feasibility due to pure

* Corresponding author. Faculty of Chemical and Energy Engineering, Universiti Teknologi Malaysia (UTM), 81310 UTM Johor Bahru, Johor, Malaysia.
E-mail address: azizulazri@utm.my (A.A. Mustaffa).

http://dx.doi.org/10.1016/j.fluid.2016.07.019
0378-3812/

phytochemical scarcity at the early stage of extraction technology development [8,9]. Computer aided molecular design (CAMD) could be a better alternative for extensive and hazard-free preliminary solvent screening for phytochemical extraction [10], and a reliable solubility model is its core element.

Several solubility models have been developed for active pharmaceutical ingredient extraction optimisation. In particular, Pharma Modified UNIFAC model was developed with better performance compared to modified UNIFAC (Dortmund) model [11]. Another modified UNIFAC (Dortmund)-based group contribution model tailor-made for polyphenolic compound solubility prediction, named MPP-UNIFAC, was developed to take into account the protic interactions of polyphenolic functional group [12]. In another study, Okuniewski et al. [13] found that modified UNIFAC (Dortmund) model had comparable prediction performance as PC-SAFT equation of state for terpene solubility prediction in organic solvents.

However, to date, there has been no UNIFAC-based solubility model specifically designed to facilitate extensive organic solvent screening for Orthosiphon *stamineus* phytochemical extraction application, with major phytochemicals i.e. Sinensetin (SEN), Eupatorin (EUP) and Tetramethylscutellarein (TMF) as biomarkers. Moreover, experimental data of fusion enthalpies, necessary for SLE prediction, for these rare biomarkers are not available, since the current phytochemical and ethnobotanical databases only emphasise on pharmaceutical and biological properties (i.e. activity, lethal dosage, etc.). In this scenario, the regression of flavonoid fusion enthalpy prediction model is necessary, due to scarcity and prohibitive costs of purified phytochemical resources. Therefore, this study aims to overcome two major research gaps. Firstly, prediction models are developed to estimate the (currently missing) normal fusion enthalpies of SEN, EUP and TMF. Secondly, UNIFAC-based solubility prediction model is regressed for examining and comparing more organic solvents for their solvation potentials for these Orthosiphon *stamineus* biomarkers.

Brief review on both pure component property (fusion enthalpy) and mixture property (solubility) prediction models to be developed in this study is presented in Section 2. In Section 3, detailed methodology for developing phytochemical physicochemical property database for model training purpose, development of fusion enthalpy prediction model, and UNIFAC-based solubility model development is elaborated. UNIFAC model regression outcomes are then discussed in Section 4 with preliminary case study of solvent selection using regressed KT-NIST-UNIFAC model, followed by conclusion in Section 5.

2. Property prediction models

In SLE solubility prediction, SLE thermodynamics model shall be the fundamental element [14]. It predicts the solubility of solid compound i in liquid solvent as in Eq. (1), given the system temperature, melting point temperature, fusion enthalpy, and activity coefficient.

$$\ln(x_i\gamma_i) = \left(H_{fus}(T_m)/RT_m\right)[1-(T_m/T)] \tag{1}$$

where x_i refers to molar solubility of compound i, γ_i stands for activity coefficient of compound i in the mixture, T_m is normal melting point temperature of compound i (in K), $H_{fus}(T_m)$ is normal fusion enthalpy of compound i at its melting point temperature (in J/mol), R is ideal gas constant (i.e. 8.314 J/mol K) and T is system temperature (in K).

For pure component property (fusion enthalpy) prediction, Marrero-Gani model and Quantitative Structure-Property Relationship (QSPR) model were the basis for model development in this study. In Marrero-Gani model, pure component property is correlated to the summation of contributions from all independent structural groups making up the molecular structure of the compound [15]. This model involves three levels of property estimation, whereby the second-order groups and third-order groups are assigned for isomeric compound differentiation and better description of complex, heterocyclic and poly-functional acyclic compounds. Marrero-Gani group contribution model for fusion enthalpy prediction are shown in Eq. (2).

$$H_{fus} - H_{fus,0} = \sum N_i H_{fus1,i} + \sum M_j H_{fus2,j} + \sum O_k H_{fus3,k} \tag{2}$$

where $H_{fus,0}$ is universal constant (adjustable) of fusion enthalpy (i.e. −2806 J/mol), N_i is number of first-order group type-i, M_j is number of second-order group type-j, O_k is number of third-order group type-k, $H_{fus1,i}$ is group contribution parameter of first-order group type-i for fusion enthalpy, $H_{fus2,j}$ is group contribution parameter of second-order group type-j for fusion enthalpy, and $H_{fus3,k}$ is group contribution parameter of third-order group type-k for fusion enthalpy.

QSPR model is an empirical model that correlates the molecular structure (i.e. encoded in the form of molecular descriptors) to the compound's properties [16]. Common molecular descriptors could be of electrostatic, topological, constitutional, quantum chemical, and geometrical types [17–26]. Specifically, molecular descriptors associated to hydrogen bonding, molecular interaction, molecular size, shape and geometry are highly correlated to fusion enthalpy. QSPR model could be constructed using multi-linear regression [27], principal component regression [28], partial least square regression [29], artificial neural network [28,30], and support vector machine [31].

UNIFAC (UNIQUAC Functional Group Activity Coefficient) model is a semi-empirical group contribution model that predicts the activity coefficient based on the independent contribution terms of subgroups making up the molecular structure of constituents in a particular mixture. Galanakis et al. [32] had proven that UNIFAC model could appropriately predict the phytochemical solubility trends in organic solvents. However, the existing UNIFAC models has been inadequate for precise *Orthosiphon stamineus* solubility prediction since the property database dedicated for UNIFAC model development does not have ample coverage of its bioactive compounds. KT-NIST-UNIFAC model is one of the most compatible UNIFAC variants with high prediction performance. Kang et al. [33] had introduced new subgroups associated to cyclic compound structure (i.e. cyclic carbonyl, cyclic hydroxyl, and cyclic ether), and second-order group contribution terms to overcome the proximity effect in complex compound with multiple functional groups. This model involves Eqs. (3)–(16), as follows.

$$ln(\gamma_i) = ln\left(\gamma_i^C\right) + ln\left(\gamma_i^R\right) + w_{R_2} ln\left(\gamma_i^{R_2}\right) \tag{3}$$

$$ln\left(\gamma_i^C\right) = 1 - J_i + ln(J_i) - (z/2)q_i[1-(J_i/L_i) + ln(J_i/L_i)] \tag{4}$$

$$ln\left(\gamma_i^R\right) = q_i(1-L_i) - \sum_k^{NMG}(s_{ki}/\eta_k - G_{ki} ln(s_{ki}/\eta_k)) \tag{5}$$