Universal Journal of Green Chemistry
http://ojs.wiserpub.com/index.php/UJGC/

Research Article

Use of Group Contribution Methods, Hansen's Solubility Theory and Microsoft Excel in Solvents Selection for Extraction of Natural Products

Manuel Díaz de los Ríos*, Evelyn Faife Pérez

Cuban Research Institute for Sugarcane Derivatives, Vía Blanca 804, Esquina C. Central, La Habana, Cuba
E-mail: mddelosrios@yahoo.com

Received: 16 August 2024; **Revised:** 13 November 2024; **Accepted:** 13 November 2024

Abstract: Usually, Hansen's theory and the COSMO continuous solvation models have been used for the selection of solvents related to a solute, using the HSPiP and COSMO-RS computational tools, respectively. Those tools are not always available to many researchers; for this reason, Microsoft Excel (ME) has been used for the estimation of good solvents in the extraction of natural products of high value and interesting for the food, pharmaceutical and cosmetic industries. This method is applicable to cases where there is little, or practically no information about the compound of interest, for which functional group contribution methods are used, among them, the Joback method for the estimation of properties necessary for the determination of solubility models, and Van Krevelen method for the initial estimation of the Hansen solubility parameters for solubility estimates, the Flory-Hugging model is used, from which and its classification, according to the North American Pharmacopeia, the good and bad solvents are defined in the selected database for the final estimation of Hansen's parameters and the solubility region. The methodology was validated with ten study cases reported in the literature. The results of this study showed that other green solvents, not reported in the literature consulted, are suggested as good solvents for polyphenols, such as diacetone alcohol, diethylene glycol monobutyl ether and tetrahydrofurfuryl alcohol. In the consulted literature there are no references on the use of a structured methodology in Microsoft Excel (ME) for the estimation of good solvents for the extraction of natural products from the estimation of solubility with existing models. All the consulted papers refer the use of various professional computer programs for this purpose, specifically HSPiP, and COSMO-RS, which are not accessible to all researchers. With these estimates and an initial classification based on pharmacopoeia criteria, a first classification of good and bad solvents is made, which is refined by means of optimization methods in ME.

***Keywords*:** hansen solubility parameters, group contribution methods, solubility, Microsoft Excel, hansen's sphere

1. Introduction

The extraction of natural products for pharmaceutical purposes or for their use in cosmetics using solvents requires arduous experimental work, where sometimes the selection of the appropriate solvent leads to a trial and error process, particularly when no previous references are available. The final destination of natural active pharmacological ingredients (API) currently imposes limits on possible solvents, beyond their technical and economic feasibility, due to growing environmental restrictions and current trends to use products of organic origin, not only because their nature, but also considering the character of the process used. For this reason, many works focus on the use of so-called green solvents

DOI: https://doi.org/10.37256/ujgc.2220245505

for the extraction of API from plant sources [1–3]. The first two articles presented reviews on the subject, which focus on biomass-derived solvents, particularly terpenes, and biodiesel as substitutes of volatile organic solvents [1]. The work of Chemat, F. et al. [2] delves into the characteristics of these solvents and the extraction methods in which they have been used, as well as other solvents substitutes of green solvents, such as ionic liquids, deep eutectic liquids and natural deep eutectic liquids, which are characterized by their non-flammability, thermal stability and low vapor pressure. Ramdas R. et al. [3] focused on the techniques that use those solvents to extract active ingredients from plant sources and agricultural production residues to increase their benefit. However, once the green solvent database has been defined, it is necessary to face the challenge of selecting those are recommendable from the technical and economic point of view. Various strategies and methods have been reported in the literature for the selection of optimal solvents for a given process, for which the estimation of the thermodynamic properties that allow predicting the solubility of a drug is of great importance [4]. The solubility of the API is a function of the fusion enthalpy (ΔH_{fus}), the melting temperature (T_m), the difference in caloric capacity (ΔC_p); usually neglected, the temperature of the medium (T) and the activity coefficient between the solid and the solvent used (γ); being this parameter the one whose estimation is more complex (Equation (1)).

Various methods for activity coefficient estimating are reported in the literature, including Wilson, Functional-group Activity Coefficients (UNIFAC), Universal Quasichemical Activity Coefficients (UNIQUAC), NonRandom TwoLiquid (NRTL), and various modifications of those, which have been widely used to estimate or model the solubility of drugs in organic solvents [4,5].

$$\ln x_1\gamma_1 = \frac{\Delta H_{fus}}{R}\left(\frac{1}{T_m}-\frac{1}{T}\right) - \frac{1}{RT}\int_{T_m}^{T}\Delta C_p dT + \frac{1}{R}\int_{T_m}^{T}\frac{\Delta C_p}{T}dT \quad (1)$$

Usually, these parameters are obtained through regression adjustments of experimental data, although they can also be estimated through theoretical methods based on functional groups contribution [6].

Jouyban. A. et al. [7] presented a review on solubility estimation methods for drugs, which summarizes various models to describe the solubility of API in aqueous media, and includes reports in periodicals, patents and software developments; in this last field. Zhuyifan Ye and Defang Ouyang [8] have developed a computational learning model (machine learning algorithm) that predicts the solubility of compounds in organic solvents, for which they used a database of more than 5000 solubility values at various temperatures, 266 compounds and 166 solvents. Although the Activity Coefficient is the parameter that represents the greatest challenge in the process of estimating the solubility of an API, knowledge of the fusion temperature and fusion enthalpies is essential, which can be determined experimentally by means of a Differential Scanning Calorimetric test (DSC). However, sometimes experimental data of these variables are not available, in which case the estimation methods constitute a practical tool for a first screening of feasible solvents. Farhad Gharagheizi, F and Salehi, G.R. [9] have developed a procedure based on neural networks implemented in the software Mat lab for the estimation of the fusion enthalpy, although perhaps, the most widespread method for this purpose, is the Joback and Reid method [10] based on the contribution of functional groups.

Several authors have formulated strategies for the identification and selection of solvents related to organic compounds of interest pharmacological, nutritional or for use as cosmetics [11–13]. Most of the proposed strategies are based on two fundamental theories: the COSMO continuous solvation model and its extension beyond the dielectric approximation COSMO-RS (Conductor-like Screening Model for Real Solvents) developed by Klamt, A. et al. [14] and the Hansen's solubility theory [15]. Both studies have become computational tools: HSPiP and COSMOS-RS respectively, widely extended in the research, development and innovation of new products obtained from natural sources and wastes.

Laboukhi-Khorsi, S et al. [11] use Hansen's methodology with HSPiP software to select suitable green solvents substitutes of n-hexane for the extraction of Artemisinin, an antimalarial drug from the plant *Artemesia annua* L. This methodology was ratified by Sánchez-Camargo, A.P et al. [13], who formulated a first screening stage with the use of the HSPiP software, which are compared with the estimates obtained using COSMO-RS and conclude that both tools provide good estimates, although COSMO-RS can be superior to HSPiP whenever accurate solubility calculations is required. They remark that both HSPiP and COSMO-RS, require further experimentation in order to validate the predicted solubility. In the case of Hansen's theory, the solute/solvent affinity is established by the spatial proximity between both compounds in a three-dimensional space,

defined by expression (2); however, the definition of a good or bad solvent is established by determining a hypothetical region, called the Hansen sphere, whose radius (R_0) should be determined experimentally. For this reason, although the value of R_a, is an important indicator of affinity, is not conclusive if information on the solubility region is not available.

$$R_a = \left[4*(\delta_{D1}-\delta_{D2})^2 + (\delta_{P1}-\delta_{P2})^2 + (\delta_{H1}-\delta_{H2})^2\right]^{1/2} \quad (2)$$

$$RED = \frac{R_a}{R_0} \quad (3)$$

Another study developed a prediction method for solids solubility's using excess Gibbs energy models (GE models) together with analysis of data, model parameter estimation, and calculations of solids solubility [16]. From this work a computer-aided model-based framework for solids solubility's calculations and solvent selection and design, called SolventPro was developed.

Unfortunately, these computing tools, necessary and useful for the execution of first estimation for API extraction, are not always available. Therefore, the objective of this work is to develop an estimation methodology, based on Hansen's solubility theory, functional group contribution methods for the estimation of key properties and their implementation in Microsoft Excel, as a first estimate of good solvents in extractive processes. For its validation, various case studies reported in the literature have been evaluated.

2. Methodology used

The procedure for the estimation of possible solvents to be used in an extractive process is illustrated in Figure 1. In some occasions, it is desired to extract a product whose chemical structure is not completely known or simply our target is a mixed product, where it is recommended to obtain samples and apply Hansen's experimental methodology to determine the Hansen solubility parameters (HSP) and the solubility region, known as the Hansen sphere with radius R_0. Of course, if the structures of the components that make up the mixture and their composition are known, it is possible to obtain estimates of the HSP of the mixture compound. Once the HSP and R_0 have been determined, the problem reduce to select possible solvents in the database, verifying their solubility experimentally, and evaluating their technical and economic viability.

However, if the chemical structure of the substance is known, but not all the information is available on the properties that allow estimating its solubility, it is possible to resort to functional groups contribution methods to estimate them. In the present work, the Joback and Reid method [10] is used to estimate properties such as molecular mass (MW), melting temperature (T_m) and fusion enthalpy (ΔH_{fus}), among others, implemented in Microsoft Excel by https://chesheets.com/joback_method.html. Solid solubility is directly related with the Fusion heat and activity coefficient as it is showed in Equation (4). The activity coefficient is related with the Flory-Hugging parameter (Equation (5)) and that value with HSP, as it is showed in Equations (6) and (7).

Once the properties have been determined, a first estimate of the HSP is obtained using the Van Krevelen method [17], which has been implemented in Excel by the authors. HSP estimates allow obtaining a preliminary value of the solubility of the compound of interest in a group of solvents through the Flory-Hugging model [18,19] represented by Equations (4)–(6). For the estimation of the Hansen solubility sphere, 40 green solvents recommended by the HSPiP software have been selected.

This model is more appropriate to describe the behavior of polymers and macromolecules, so it has also been used to describe the behavior of solids with high molar volume [20]; which is common in many natural extracts, such as the starch obtained from various plants [21]. In addition, its application is simple, since it allows estimating the activity coefficient for the various solute/solvent pairs and predicting the API solubility.

$$\ln x_1 = \frac{\Delta H_{fus}}{R}\left(\frac{1}{T_m} - \frac{1}{T}\right) - \ln \gamma_1 \tag{4}$$

$$\ln \gamma_1 = \ln \frac{\Phi_1}{x_1} + \frac{\Phi_1}{x_1} + \chi_{12} \cdot \Phi_2^2 \tag{5}$$

$$\chi_{12} = \frac{v_1}{RT}(\delta_1 - \delta_2)^2 \tag{6}$$

where Φ_1, Φ_2 are the volumetric concentrations of the solute and solvent respectively, v_1 is the molar volume of the solute, and χ_{12} is the solute/solvent interaction coefficient.

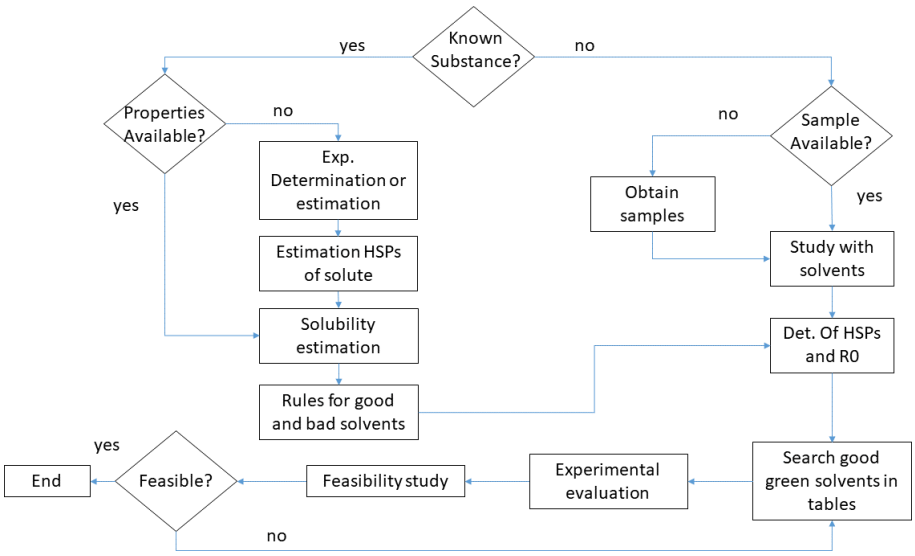


Figure 1. General methodology for good solvents selection for natural products extraction

Equation (6) has a lower fit for compounds with hydrogen bonds and high polarity, and Equation (7) is recommended to estimate the Flory-Hugging parameter, where Hansen [15] recommends the use of alpha equal to 1.

$$\chi_{12} = \alpha \frac{v_1}{RT}\left((\delta_{1,d} - \delta_{2,d})^2 + 0.25\left((\delta_{1,p} - \delta_{2,p})^2 + (\delta_{1,h} - \delta_{2,h})^2\right)\right) \tag{7}$$

It is highly probable that the activity coefficient estimated by this model does not provide a good fit in many of the solvents chosen to determine the Hansen solubility sphere, but the purpose of the described procedure is not to obtain a reliable solubility model, but only to classify each solvent as good or bad by this estimate.

The definition of a good or bad solvent is established from the estimate of the solubility of the compound of interest in each solvent, according to the classification established by the United States Pharmacopoeia (Table 1)[22]. Once the solvents have been classified, the HSP of the solute are calculated in Microsoft Excel, as if they were experimental data, according to the procedure described by Díaz and Hernández[23] with the equations used.

As trial solubility, 1×10^{-4} was used for all solvents, which is practically the limit between soluble and insoluble, so that the final estimated value can be defined on one side or the other of that limit.

A description of the general procedure for each compound is provided in the Microsoft Excel file provided as supplementary information, where Coumarin has been selected as example. The "Data and Instruction" sheet reports the

fundamental properties of the compound used in the prediction and a sequence on how to proceed in the workbook. The "Calculation" and "Van Krevelen" sheets allow obtaining an estimate of the fusion enthalpy and an initial estimate of the Hansen solubility parameters from the information of the functional groups of the component respectively. Subsequently, if information on the solubility of the compound in various solvents is available, the "Solubility Literature" sheet is used to discern between the use of Equations (6) or (7) and where an estimate of $\alpha = 1$ is assumed for Equation (7) in all solvents; γ is determined for each compound according to Equation (5) and the solubilities are estimated by expression (4). Finally, the model that shows the lowest value of the sum of squares of the residuals (SCR) for the group of solvents evaluated is selected.

Table 1. Classification of drug solubility (practically insoluble considered as bad solvents, other class as good solvents)

Class	Part of solvent required for dissolving 1 part of solute
Very soluble	<1
Freely soluble	1–10
Soluble	10–30
Sparingly soluble	30–100
Slightly soluble	100–1000
Very slightly soluble	1000–10,000
Practically insoluble	>10,000

The selected model is evaluated in the sheet whose name describes the compound to be evaluated in the database of solvents used, and the good and bad solvents are defined.

Once the good and bad solvents in the database reported by HSPiP have been estimated, the HSP values are optimized in the "HSP Sphere" sheet, according to the procedure previously reported[23].

3. Results methodology validation

The methods for estimating the chemical properties of various compounds, as well as the estimation of operational conditions, are of great importance for the design of technologies. In the case of extraction of natural products with high added value for the pharmaceutical, food and cosmetic industries, the appropriate selection of solvents is crucial, where the solute-solvent affinity will determine the solubility of the solute in it, which, together with economic, environmental and health considerations will guarantee the sustainability of the technology.

The proposed validation procedure consists on determining if the solvents reported as good for the extraction of a given API in the literature are included in the Hansen solubility sphere determined by the suggested estimation procedure.

To validate this procedure, 10 examples of literature on natural products extracted from various sources were used, whose basic information is summarized in Tables 2a and 2b. Table 3 summarizes the experimental fusion enthalpy values determined and those estimated by Joback and Reid, used in the study, as well as the absolute error of the estimation respect to the experimental value.

Absolute error is determined by the following expression:

$$Abs\ error = 100 * \frac{ABS(\Delta H_{fus\ exp} - \Delta H_{fus\ est})}{\Delta H_{fus\ exp}}$$

It can be seen that the ΔH_{fus} estimates are acceptable for the first six compounds in the table and ferulic acid. However, in the case of rosmarinic acid, contradictory values are reported in the literature, since Aydi, A. et al.[25] report a value of 16.39 kJ/mol, while Santos Veras, K et al.[26] report 142.29 J/g (51.27 kJ/mol). Considering the value estimated by Joback and that estimated by the method of Yamamoto, H.[27] of 59.5 kJ/mol are similar to the value reported by Santos Veras, K et al, this is adopted as a valid experimental value. A similar situation occurs with the experimental values reported for ferulic acid, where Faiyaz Shakeel et al.[44] obtained a value of 15.01 kJ/mol, while Mota Fátima, L et al.[46] reported a value of

33.34 kJ/mol, very similar to that calculated by the Joback and Reid method. The precision in the estimation of the fusion enthalpy is vital for the adjustment of the solubility models and, therefore, in the estimation of the good solvents.

Table 2a. Basic information of API compounds used for procedure validation (left)

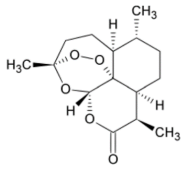
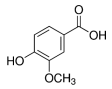
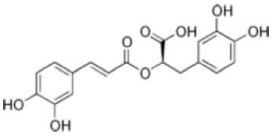
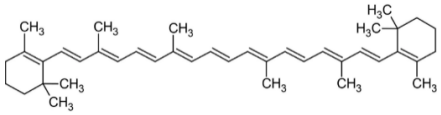
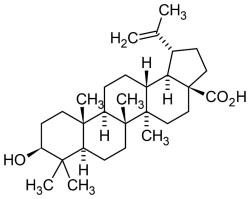
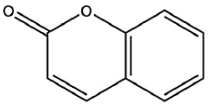
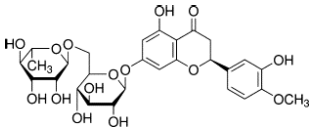
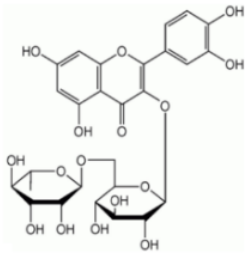
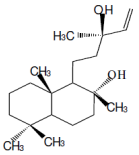
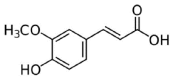
Name	Formulae	MW (g/mol)	Structure
artemisinin [11,24]	$C_{15}H_{22}O_5$	282.34	
vanillic acid [32]	$C_8H_8O_4$	168.14	
rosmarinic acid [25,26,28,33]	$C_{18}H_{16}O_8$	360.31	
β-carotene [29–31]	$C_{40}H_{56}$	536.89	
betulinic acid [35,36]	$C_{30}H_{48}O_3$	456.71	
coumarin [34]	$C_9H_6O_2$	146.14	
hesperidin [37–39]	$C_{28}H_{34}O_{15}$	610.57	
rutin [40,41]	$C_{27}H_{30}O_{16}$	610.52	
sclareol [42,43]	$C_{20}H_{36}O_2$	308.5	
ferulic acid [44–46]	$C_{10}H_{10}O_4$	194.18	

Table 2b. Basic information of API compounds used for procedure validation (right)

Source	Uses
Artemesia annua L	Antimalarial
Vanilla bean	Anticancer, antidiabetic, antibacterial, anti-inflammatory and antioxidant effects
Rosmarinus officinalis, Origanum vulgare	antioxidant and antiflammantory
rich algae as *dunaliella salina*	precursor to vitamin A
white-barked birch trees	anticancer, antiretroviral, antimalarial and anti-inflammatory
tonka beans, cinnamon	Treatment of prostate cancer, renal cell carcinoma and leukemia
citrus fruits	for hemorrhoids, varicose vein and venous stasis
Passion flower, bugkwheat, tea and apple	strengthen blood vessels, improve blood circulation, lower cholesterol
Salvia sclarea	cytostatic effects against human leukemic
Is found in commelinid plants (rice, wheat, oats, and pineapple)	anti-inflammatory, antioxidant, antimicrobial activity, anticancer, and antidiabetic effect

Table 3. Melting temperature, experimental and estimated values of ΔH_{fus} by Joback and Reid method

Compound	T_f (K)	ΔH_{fus} exp. (kJ·mol^{-1}) Ref.	ΔH_{fus} est (kJ·mol^{-1})	Abs. error
artemisinin	429.60	24.30 [24]	26.72	9.96
vanillic acid	484.90	32.80 [32]	27.24	16.96
rosmarinic acid	444.52	51.27 [26]	56.45	10.11
β-carotene	456.00	56.00 [31]	49.80	11.07
betulinic acid	588.45	42.23 [36]	43.26	2.44
coumarin	344.40	16.04 [34]	16.39	2.18
hesperidin	535.00	63.86 [39]	102.40	60.35
rutin	450.15	82.30 [40]	100.96	22.67
sclareol	376.45	28.70 [42]	19.23	33.00
ferulic acid	444.60	33.34 [46]	30.54	8.40

In the case of the flavonoids hesperidin and rutin, the estimates for ΔH_{fus} are higher than those determined experimentally, as in the case of sclareol. In this work were used, only compounds in which experimental data of ΔH_{fus} were reported, with the purpose of comparing the results obtained, but if this information is not available, the procedure must be applied with the value of the estimated properties.

In the case of rosmarinic acid, the determination of HSP by various estimation methods is also contradictory, as reported by Cher Haan Lau and Lee Suan Chua [28], who finally adopted the estimates by Hoy's method, due to its similarity with results obtained with the COSMO software. It is then evident that, although the procedure developed by Van Krevelen has been used in this work, other estimation methods for HSP can be implemented, if desired.

Tables 4 and 5 provide the results of the prediction of good solvents for the compounds evaluated with the use of experimental and estimated values of ΔH_{fus}. Table 4 reports the results of the estimation of good solvents for various compounds, which match with those reported as good in literature, according to the methodology described above and considering the use of experimental values of enthalpy of fusion. It reports the Hansen solubility parameters determined by the Yamamoto molecular rupture method, H. (Y-MB) [27], the Van Krevelen functional group contribution method [17], those reported experimentally by the literature and the determined using Microsoft Excel [23] with their R_o and fit for the data used in the "fit" column.

Table 4. Results of good solvents estimation with the use of experimental values of fusion heat

Name	Method	δ_D	δ_P	δ_H	Fit (R_0)	Mean abs error (MAE) HSPs est. (%)	Good Solv. Lit.	% Good Solv. Match	Solvents do not match	Model fitted	Ref.
artemisinin	Y-MB	17.20	5.20	5.90		17.05	18	100	none	Hansen	11,24
	Van Krevelen	22.00	4.92	7.84	100	20.95					
	Excel	16.20	9.38	9.45	(13.58)	20.44					
	Laboukhi-Khorsi	17.40	7.40	7.40							
vanillic acid	Y-MB	19.80	8.60	15.30		0.00	7	100	none	Hansen	32
	Van Krevelen	22.40	6.44	16.58	100	15.52					
	Excel	16.84	8.24	11.73	(12.13)	14.14					
	HSPiP tables	19.80	8.60	15.30							
rosmarinic acid	Y-MB	20.80	6.30	12.40		48.70	5	40	DMSO, methyl acetate, ethyl acetate	Hansen	26,28,33
	Van Krevelen	20.76	4.64	19.37	100	82.50					
	Excel	16.45	8.26	18.36	(6.68)	62.97					
	Haan Lau, C. et al.[28]	16.74	15.65	7.65							
β-carotene	Y-MB	17.70	0.80	1.90		4.50	11	81.82	hexane, ethyl acetate	Hansen	29,30
	Van Krevelen	18.89	0.29	0.00	99	57.45					
	Excel	18.58	2.05	1.90	(7.33)	58.46					
	Lara A., G.[30]	17.40	0.80	1.70							
betulinic acid	Y-MB	18.00	2.90	4.50		36.50	15	80	methanol, ethanol hexane	Hildebrand	35,36
	Van Krevelen	21.40	2.04	5.63	87	42.23					
	Excel	20.75	10.18	13.22	(16.53)	44.27					
	B. Da Nóbre, A.[35]	16.80	7.40	7.70							
coumarin	Y-MB	19.30	10.50	5.50		39.83	11	100	none	Hansen	34
	Van Krevelen	12.86	5.60	5.66	100	35.51					
	Excel	16.61	9.19	9.62	(13.61)	29.01					
	HSPiP tables	20.00	12.50	6.70							
hesperidin	Y-MB	19.60	10.30	13.90		29.30	6	83.33	ethylene glycol	Hildebrand	37,38
	Van Krevelen	15.26	4.02	19.63	87	33.75					
	Excel	19.57	15.86	15.23	(12.68)	18.06					
	Milescu, R.A et al.[37]	17.30	21.20	18.10							
rutin	Y-MB	19.60	10.60	10.40		22.70	6	83.33	ethylene glycol	Hildebrand	37,41
	Van Krevelen	12.85	3.80	20.42	83	40.03					
	Excel	19.76	16.05	14.98	(13.05)	13.23					
	Milescu, R.A et al.[37]	20.30	13.50	18.30							
sclareol	Y-MB	17.20	3.00	4.50		45.00	12	91.67	hexane	Hansen	42,43
	Van Krevelen	22.39	2.48	11.71	100	42.27					
	Excel	18.15	6.57	11.98	(13.76)	16.68					
	Martínez Aguda, R.[42]	17.20	8.10	16.10							
ferulic acid	Y-MB	17.90	10.10	22.90		67.61	9	88.89	ethylene glycol	Hansen	44–46
	Van Krevelen	18.85	4.98	14.57	100	34.11					
	Excel	16.03	9.68	9.38	(13.56)	60.12					
	Zuorro, A.[45]	22.70	5.70	15.60							

Table 5. Results of good solvents estimation with the use of estimated values of fusion heat by the mean of Joback method

Name	Method	δ_D	δ_P	δ_H	Fit (R_0)	Mean abs error (MAE) HSPs est. (%)	Good Solv. Lit.	% Good Solv. Match	Solvents do not match	Model fitted	Ref.
artemisinin	Y-MB	17.20	5.20	5.90	100.00 (13.55)	20.63	23	100	none	Hansen	11,24
	Van Krevelen	22.00	4.92	7.84							
	Excel	16.34	9.42	9.51							
	Laboukhi-Khorsi [11]	17.40	7.40	7.40							
vanillic acid	Y-MB	19.80	8.60	15.30	100.00 (15.56)	29.91	7	100	none	Hansen	32
	Van Krevelen	22.40	6.44	16.58							
	Excel	18.19	12.27	9.35							
	HSPiP tables	19.80	8.60	15.30							
rosmarinic acid	Y-MB	20.80	6.30	12.40	100 (7.15)	66.59	5	40	DMSO, methyl acetate, ethyl acetate	Hansen	26,28,33
	Van Krevelen	20.76	4.64	19.37							
	Excel	17.25	8.42	19.16							
	Haan Lau, C. et al. [28]	16.74	15.65	7.65							
β-carotene	Y-MB	17.70	0.80	1.90	99 (7.43)	168.98	11	90.91	hexane	Hansen	29,30
	Van Krevelen	18.89	0.29	0.00							
	Excel	17.55	4.63	1.23							
	Lara Alvarenga, G [30]	17.40	0.80	1.70							
betulinic acid	Y-MB	18.00	2.90	4.50	68 (16.29)	48.75	15	80	methanol, ethanol hexane	Hildebrand	35,36
	Van Krevelen	21.40	2.04	5.63							
	Excel	21.49	10.55	13.53							
	B. Da Nóbrega, A. [35]	16.80	7.40	7.70							
coumarin	Y-MB	12.86	5.60	5.66	100 (13.61)	29.07	11	100	none	Hansen	34
	Van Krevelen	20.41	5.60	5.66							
	Excel	16.61	9.19	9.63							
	HSPiP tables	20.00	12.50	6.70							
hesperidin	Y-MB	19.60	10.30	13.90	Determi-nation fail						
	Van Krevelen	14.67	4.02	19.63							
	Excel										
	M., R.A et al. [37]	17.30	21.20	18.10							
rutin	Y-MB	19.60	10.30	13.90	84 (11.21)	19.74	6	66.67	ethylene glycol, N,N-dimethyl-acetamine	Hildebrand	37,41
	Van Krevelen	12.85	3.80	20.42							
	Excel	13.90	16.04	16.68							
	M., R.A et al. [37]	20.30	13.50	18.30							
sclareol	Y-MB	17.20	3.00	4.50	100 (12.96)	17.90	12	91.67	hexane	Hansen	42,43
	Van Krevelen	16.69	6.29	11.53							
	Excel	15.87	6.93	10.81							
	Martínez A, R. [42]	17.20	8.10	16.10							
ferulic acid	Y-MB	17.90	10.10	22.90	100 (13.56)	19.49	9	88.89	ethylene glicol	Hansen	44–46
	Van Krevelen	18.85	4.98	14.57							
	Excel	18.16	6.57	11.98							
	Zuorro, A. [45]	22.70	5.70	15.60							

The Tables 4 and 5 also report the average absolute percentage errors (MAE) between the estimates reported by Y-MB, Van Krevelen, and Excel with those reported by the literature according to Equation (8).

$$MAE = \left(\sum_{i=1}^{N} \left(100 \cdot \frac{ABS(\delta_{il} - \delta_{ie})}{\delta_{il}} \right) \right) / N \tag{8}$$

where "*il*" are the subscripts of the *N* Hansen parameters reported in the literature, and "*ie*" are the subscripts of the estimated parameters.

In Table 4, the average errors of the three parameters are less than 30% for the determinations with Microsoft Excel, with the exception of the cases of rosmarinic acid, β-carotene. betulinic acid and ferulic acid, which are also high for the estimates of rosmarinic acid, betulinic acid, coumarin, sclareol and ferulic acid with the Y-MB method, which calls into question the validity of the HSPs reported in the literature for some of these compounds. This result is similar to that obtained in the estimation of Tables 5, where rosmarinic acid and β-carotene also showed a poor adjustment of HSP.

Table 4 shows that, except for rosmarinic acid, more than 80% of the solvents reported as good solvents by the literature are confirmed by the methodology developed when experimental values of fusion enthalpy are used, therefore, they fall within the estimated Hansen sphere. Something similar occurs when estimates of ΔH_{fus} are used (Tables 5) except in the cases of rosmarinic acid, hesperidin and rutin; specially in the case of hesperidin where the great difference in ΔH_{fus} experimental and estimated (66%) did the Excell estimation fail. In the case of Rosmarinic acid various factors affect, where the low number of good solvents used as a reference and the difference between the estimated HSP and those reported in the literature (*MAE*) leads to low estimates.

Unfortunately, only three articles presented reliable experimental information on Hansen's sphere: artemisinin [11], hesperidin and rutin [37]; another compound that provides information about the Hansen solubility region is betulinic acid, but the authors use only five good solvents to estimate the Hansen sphere [35], so it is not reliable.

In Figure 2–4 are represented the Hansen spheres experimental and estimated in this work respectively with the good solvents used in the validation procedure for these compounds. In Case of artemisinin, methanol was considered bad solvent in the work developed by Laboukhi-Khorsi, S et al. [11], but the solubility of artemisinin in pure methanol was calculated by Nti-Gyabaah, J. et al. [24], therefore it was included in our Hansen sphere estimation and fell inside the green sphere and outside the blue one.

The sphere radius for hesperidin and rutin were calculated with HSPiP and the supplementary information provided by the authors in Tables S1 and S2 [37], respectively, by the classical Hansen method and keeping the HSPs fitted by the authors [37].

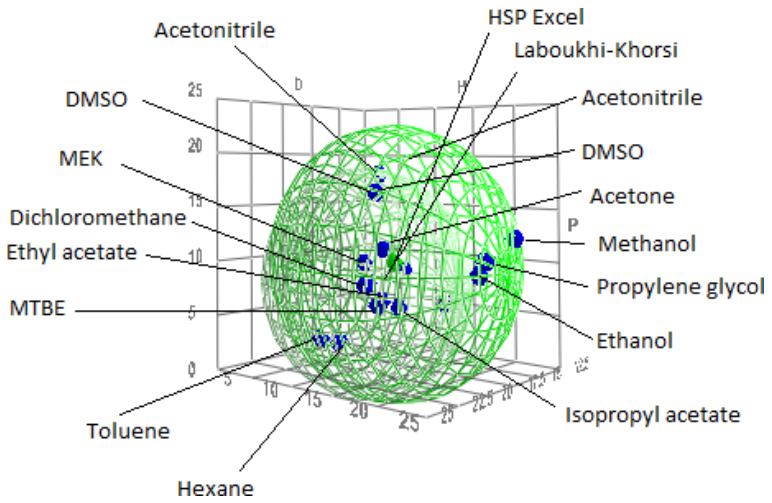


Figure 2. Hansen sphere for artemisinin reported in literature (light blue wireframe sphere) and determined (green wireframe sphere)

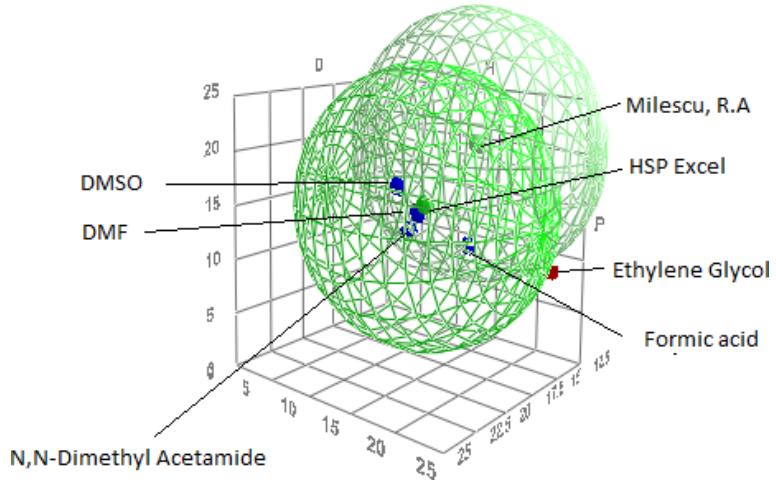


Figure 3. Hansen sphere for hesperidin reported in literature (light blue wireframe sphere) and determined (green wireframe sphere)

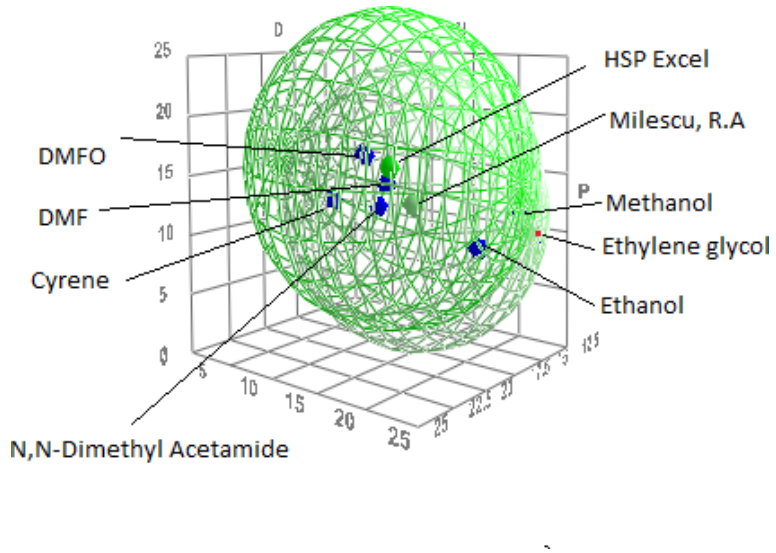


Figure 4. Hansen sphere for rutin reported in literature (light blue wireframe sphere) and determined (green wireframe sphere)

Ethylene glycol falls outside the Hansen sphere for hesperidine in both studies; with Excel and the published paper, although it was evaluated as good solvent by Milescu, R.A et al.[37], while in the case of rutin, ethylene glycol falls outside the Hansen sphere only in this study.

For artemisinin and rutin there are not important differences between the HSP estimated in this work and the values reported in literature; Hansen regions are similar, as can be seen in Figures 2 and 4.

It is not possible to formulate a general criterion on when the estimation of the solubility for the compound of interest better adjusts to Equations (6) or (7). Several factors intervene in this fit, such as the veracity of fusion enthalpy and of the HSP or their total value δt, but it is noteworthy that the three examples, which the Hildebrand model fits better, correspond to more complex chemical structures. It would be necessary to evaluate a greater number of compounds to arrive at more precise conclusions. The complexity that the evaluation of the activity coefficient for each of the solute/solvent interactions in the table used to determine the HSP and the Hansen sphere limits the possibility of using other models, usually reported in the literature.

Table 6 presents other "green" solvents not reported in literature, that could be used for the extraction of each of the compounds evaluated, according to the HSP and RED estimates obtained by the methodology employed. Of course, the

solubility of each compound in the suggested solvent and the yield to be obtained must be confirmed experimentally, as well as evaluating their economic feasibility. Green or bio-based solvents are included in two main groups: agro-based raw materials (ABRM), such as lignocellulose, oils, sugars, starches and proteins, and Bio-based solvents (BBS), such as alcohols, esters, ethers and terpenes. The HSP of solvents included in Table 6 and its green relationship is described in Table 7.

Table 6. Green solvents suggested from the applied methodology

Compound	Green solvents (RED)
artemisinin	diacetone alcohol (0.14), diethylene glycol monobutyl Ether (0.20)
vanillic acid	ethyl lactate (0.16), tetrahydrofurfuryl alcohol (0.19), diacetone alcohol (0.19)
rosmarinic acid	1-propanol (0.29), dipropylene glycol (0.36)
β-carotene	anisole (0.78)
betulinic acid	tetrahydrofurfuryl alcohol (0.38), cyrene (0.46)
coumarin	diacetone alcohol (0.16), diethylene glycol monobutyl ether (0.20)
hesperidin	cyrene (0.67), dipropylene Glycol (0.69)
rutin	cyrene (0.67), dipropylene Glycol (0.68)
sclareol	tetrahydrofurfuryl alcohol (0.15), cyclohexanol (0.24)
ferulic acid	diacetone alcohol (0.16), acetone (0.20), diethylene glycol monobutyl ether (0.22)

It is observed that some solvents reported in Table 6 are common for the extraction of several of the polyphenolic compounds studied, such as diacetone alcohol, diethylene glycol monobutyl ether and tetrahydrofurfuryl alcohol.

Table 7. HSP of solvents related in Table 6

Compound	δ_D	δ_P	δ_H	Classification
1-propanol	16	6.8	17.4	BBS
acetone	15.5	10.4	7	ABRM
anisole	17.8	4.4	6.9	BBS
cyclohexanol	17.4	4.1	17.5	BBS
cyrene	13.8	12.4	7.1	ABRM
diacetone alcohol	15.8	8.2	10.8	BBS
diethylene glycol monobutyl Ether	16	7	10.6	BBS
dipropylene glycol	16.5	10.6	17.7	BBS
ethyl lactate	16	7.6	12.5	BBS
tetrahydrofurfuryl alcohol	17.8	8.2	12.9	ABRM

4. Conclusions

The present work demonstrates that, even when adequate computational tools, such as HSPiP and COSMO-RS are not available, it is possible to estimate the HSP of a solute of interest and its solubility region by using functional group contribution methods, with the help of Microsoft Excel, easily implementable in this platform, with an adequate level of precision.

Other green solvents not reported in the literature, are suggested as good solvents to evaluate for each study case, based on the results obtained by applying Hansen's methodology in ME, such as diacetone alcohol, diethylene glycol monobutyl ether anisole and tetrahydrofurfuryl alcohol. Once the estimation stage is completed, it is necessary to corroborate the results with experimental actions.

Conflict of interest

The authors declare no conflict of interest.

Supplementary materials

Supplementary information can be available at: https://ojs.wiserpub.com/index.php/UJGC/article/view/5505/2807.

References

[1] Li, Z.; Smith, K. H.; Stevens, G. W. The Use of Environmentally Sustainable Bio-Derived Solvents in Solvent Extraction Applications—A Review. *Chinese J. Chem. Eng.* **2016**, *24*, 215–220. https://doi.org/10.1016/j.cjche.2015.07.021.

[2] Chemat, F.; Vian, M. A.; Ravi, H. K.; Khadhraoui, B.; Hilali, S.; Perino, S.; Tixier, A.-S. F. Review of Alternative Solvents for Green Extraction of Food and Natural Products: Panorama, Principles, Applications and Prospects. *Molecules* **2019**, *24*, 3007. https://doi.org/10.3390/molecules24163007.

[3] More, P. R.; Jambrak, A. R.; Arya, S. S. Green, Environment-Friendly and Sustainable Techniques for Extraction of Food Bioactive Compounds and Waste Valorization. *Trends Food Sci. Technol.* **2022**, *128*, 296–315. https://doi.org/10.1016/j.tifs.2022.08.01.

[4] Baptiste, B.; Teychené, S.; Biscans, B. An Evaluation of Thermodynamic Models for the Prediction of Drug and Drug-Like Molecule Solubility in Organic Solvents. *Fluid Phase Equilibria* **2011**, *309*, 36–52. https://doi.org/10.1016/J.FLUID.2011.06.032.

[5] Sheikholeslamzadeh, E.; Rohani, S. Solubility Prediction of Pharmaceutical and Chemical Compounds in Pure and Mixed Solvents Using Predictive Models. *Ind. Eng. Chem. Res.* **2012**, *51*, 464–473. https://doi.org/10.1021/ie201344k.

[6] Farajnezhad, A.; Afshar, O. A.; Khansary, M. A.; Shirazian, S.; Ghadiri, M. Correlation of Interaction Parameters in Wilson, NRTL and UNIQUAC Models Using Theoretical Methods. *Fluid Phase Equilibria* **2016**, *417*, 181–186. https://doi.org/10.1016/j.fluid.2016.02.041.

[7] Jouyban, A.; Mohammad, A. A.; Fakhree, M. A. A.; Shayanfar, A. Solubility Prediction Methods for Drug/Drug-Like Molecules. *Recent Patents Chem. Eng.* **2008**, *1*, 220–231.

[8] Zhuyifan, Y.; Ouyang, D. Prediction of Small-Molecule Compound Solubility in Organic Solvents by Machine Learning Algorithms. *J. Cheminformatics* **2021**, *13*, 98. https://doi.org/10.1186/s13321-021-00575-3.

[9] Gharagheizi, F.; Salehi, E. R. Prediction of Enthalpy of Fusion of Pure Compounds Using an Artificial Neural Network-Group Contribution Method. *Thermochim. Acta* **2011**, *521*, 37–40. https://doi.org/10.1016/j.tca.2011.04.001.

[10] Joback, K. G.; Reid, R. C. Estimation of Pure-Component Properties from Group Contributions. *Chem. Eng. Commun.* **1987**, *57*, 233–243. https://doi.org/10.1080/00986448708960487.

[11] Laboukhi-Khorsi, S.; Daoud, K.; Chemat, S. Efficient Solvent Selection Approach for High Solubility of Active Phytochemicals: Application for the Extraction of an Antimalarial Compound from Medicinal Plants. *ACS Sustain. Chem. Eng.* **2017**, *5*, 4332–4339. https://doi.org/10.1021/acssuschemeng.7b00384.

[12] Jabbari, M.; Lundin, M.; Hatamvand, M.; Skrifvars, M.; Taherzadeh, M. Computer-Aided Theoretical Solvent Selection Using the Simplex Method Based on Hansen Solubility Parameters (HSPs). *J. Inform. Tech. Softw. Eng.* **2018**, *8*, 4. https://doi.org/10.4172/2165-7866.1000242.

[13] Sánchez-Camargo, A.; Bueno, M.; Parada-Alfonso, F.; Cifuentes, A.; Ibáñez, E. Hansen Solubility Parameters for Selection of Green Extraction Solvents. *Trends Anal. Chem.* **2019**, *118*, 227–237. https://doi.org/10.1016/j.trac.2019.05.046.

[14] Klamt, A.; Ecker, F. COSMO-RS: A Novel and Efficient Method for the A Priori Prediction of Thermophysical Data of Liquids. *Fluid Phase Equilibria* **1995**, *172*, 43–72.

[15] Hansen, C. M. Solubility Parameters—An Introduction. In *Hansen Solubility Parameters: A User's Handbook*; CRC Press: Boca Raton, FL, USA, 2007; pp. 43–66. ISBN 978-0-9551220-2-6.

[16] Cunico, L. P.; Tula, A. K.; Ceriani, R.; Gani, R. Modeling and Prediction of Solid Solubility by GE MODELS. In *Computational Pharmaceutical Solid State Chemistry*, 1st ed.; Abramov, Y. A., Ed.; John Wiley & Sons: Hoboken, NJ, USA, 2016.

[17] van Krevelen, D. W. *Properties of Polymers. Their Correlation with Chemical Structure; Their Numerical Estimation and Prediction from Additive Group Contributions*, 4th ed.; Elsevier: Amsterdam, The Netherlands, 2009; ISBN 978-0-08-054819-7.

[18] Lindvig, T.; Michelsen, M. L.; Kontogeorgis, G. M. A Flory—Huggins Model Based on the Hansen Solubility Parameters. *Fluid Phase Equilibria* **2002**, *203*, 247–260.
[19] Modarresi, H.; Conte, E.; Abildskov, J.; Gani, R.; Crafts, P. Model-Based Calculation of Solid Solubility for Solvent Selections: A Review. *Ind. Eng. Chem. Res.* **2008**, *47*, 5234–5242. https://doi.org/10.1021/ie0716363.
[20] Chiou, C. T.; Manes, M. Application of the Flory-Huggins Theory to the Solubility of Solids in Glyceryl Trioleate. *J. Chem. Soc., Faraday Trans. I* **1986**, *82*, 243–246.
[21] Kumoro, A. C.; Retnowati, D. S.; Ratnawati, R.; Widiyanti, M. Estimation of Aqueous Solubility of Starch from Various Botanical Sources Using Flory Huggins Theory Approach. *Chem. Eng. Commun.* **2021**, *20*, 5, 624–635. https://doi.org/10.1080/00986445.2019.1691539.
[22] The United States Pharmacopeia, USP 30-NF 25, 2007. Available online: https://www.uspnf.com/official-text/proposal-statuscommentary/usp-30-nf-25 (accessed on 8 December 2023).
[23] Díaz de los Ríos, M.; Hernández Ramos, E. Determination of the Hansen Solubility Parameters and the Hansen Sphere Radius with the Aid of the Solver Add-In of Microsoft Excel. *SN Appl. Sci.* **2020**, *2*, 676. https://doi.org/10.100/s42452-020-2512-y.
[24] Nti-Gyabaah, J.; Gbewonyo, K.; Chiew, Y. C. Solubility of Artemisinin in Different Single and Binary Solvent Mixtures Between (284.15 and 323.15) K and NRTL Interaction Parameters. *J. Chem. Eng. Data* **2010**, *55*, 3356–3363. https://doi.org/10.1021/je100125x.
[25] Aydi, A.; Claumann, C. A.; Wüst Zibetti, A.; Abderrabba, M. Differential Scanning Calorimetry Data and Solubility of Rosmarinic Acid in Different Pure Solvents and in Binary Mixtures (Methyl Acetate + Water) and (Ethyl Acetate + Water) from 293.2 to 313.2 K. *J. Chem. Eng. Data* **2016**, *61*, 3718–3723. https://doi.org/10.1021/acs.jced.6b00008.
[26] Santos Veras, K.; Silveira Fachel, F. N.; Pittol, V.; Ribeiro Garcia, K.; Linck Bassani, V.; dos Santos, V.; Henriques, A. T.; Ferreira Teixeira, H.; Scherer Koester, L. Compatibility Study of Rosmarinic Acid with Excipients Used in Pharmaceutical Solid Dosage Forms Using Thermal and Non-Thermal Techniques. *Saudi Pharm. J.* **2019**, *27*, 1138–1145. https://doi.org/10.1016/j.jsps.2019.09.010.
[27] Abbott, S.; Yamamoto, H. HSPiP Software, 5th Edition 5.4.03, 2015. Available online: https://www.hansen-solubility.com/HSPiP/ (accessed on 8 December 2023).
[28] Lau, H. C. C. H.; Chua, S. Solubility Models for the Recovery of Rosmarinic Acid from Orthosiphon Aristatus Extract Using Solid Phase Extraction. *ChemEngineering* **2019**, *3*, 64. https://doi.org/10.3390/chemengineering3030064.
[29] Yara-Varon, E.; Fabiano-Tixier, A. S.; Balcells, M.; Canela-Garayoa, R.; Bily, A.; Chemat, F. Is It Possible to Substitute Hexane with Green Solvents for Extraction of Carotenoids? A Theoretical Versus Experimental Solubility Study. *RSC Adv.* **2016**, *6*, 27750–27759. https://doi.org/10.1039/C6RA03016E.
[30] Lara Alvarenga, G.; Sarria Cuevas, M.; Carolina Capellini, M.; José Crevellin, E.; Alberto Beraldo de Moraes, L.; da Costa Rodrigues, C. E. Extraction of Carotenoid-Rich Palm Pressed Fiber Oil Using Mixtures of Hydrocarbons and Short Chain Alcohols. *Food Res. Int.* **2019**, *128*, 108810. https://doi.org/10.1016/j.foodres.2019.108810.
[31] Treszczanowicz, T.; Kasprzycka-Guttman, T.; Treszczanowicz, A. J. Solubility of β-Carotene in Binary Solvents Formed by Some Hydrocarbons with Dibutyl Ether and 1,2-Dimethoxyethane. *J. Chem. Eng. Data* **2003**, *48*, 1517–1520. https://doi.org/10.1021/je030165y.
[32] Zhang, Y.; Guo, F.; Cui, Q.; Lu, M.; Song, X.; Tang, H.; Li, Q. Measurement and Correlation of the Solubility of Vanillic Acid in Eight Pure and Water + Ethanol Mixed Solvents at Temperatures from 293.15 to 323.15 K. *J. Chem. Eng. Data* **2015**, *48*, 1517–1520. https://doi.org/10.1021/acs.jced.5b00619.
[33] Wüst Zibetti, A.; Aydi, A.; Claumann, C. A.; Eladeb, A.; Abderrabba, M. Correlation of Solubility and Prediction of the Mixing Properties of Rosmarinic Acid in Different Pure Solvents and in Binary Solvent Mixtures of Ethanol + Water and Methanol + Water from (293.2 to 318.2) K. *J. Mol. Liq.* **2016**, *216*, 370–376. http://dx.doi.org/10.1016/j.molliq.2016.01.061.
[34] Huang, X.; Wang, J.; Hao, H.; Ouyang, J.; Gao, Y.; Bao, Y.; Wang, Y.; Yin, Q. Determination and Correlation of Solubility and Solution Thermodynamics of Coumarin in Different Pure Solvents. *Fluid Phase Equilibria* **2015**, *394*, 148–155. http://dx.doi.org/10.1016/j.fluid.2015.03.022.
[35] Bezerra da Nóbrega, A.; da Motta Lessa, B.; Basil Bizarri, C. H.; Ribeiro de Palva, S.; Candido Alves Slana, G. B. Comparison of Betulinic Acid Extraction Techniques from Eugenia florida DC (Myrtaceae) Leaves. *J. Appl. Biotechnol. Bioeng.* **2022**, *9*, 120–125. http://dx.doi.org/10.15406/jabb.2022.09.00298.
[36] Cheng, Y.; Shao, Y.; Yan, W. Solubilities of Betulinic Acid in Thirteen Organic Solvents at Different Temperatures. *J. Chem. Eng. Data* **2011**, *56*, 4587–4591. http://dx.doi.org/10.1021/je200531k.

[37] Milescu, R. A.; Segatto, M. L.; Stahl, A.; McElroy, C. R.; Farmer, T. J.; Clark, J. H.; Gomes Zuin, V. Sustainable Single-Stage Solid-Liquid Extraction of Hesperidin and Rutin from Agro-Products Using Cyrene. *Sustain. Chem. Eng.* **2020**, *8*, 18245–18257. http://dx.doi.org/10.1021/acssuschemeng.0c06751.

[38] Anwer, M. K.; Al-Shdefat, R.; Jamil, S.; Alam, P.; Abdel-Kader, M. S.; Shakeel, F. Solubility of Bioactive Compound Hesperidin in Six Pure Solvents at (298.15 to 333.15) K. *J. Chem. Eng. Data* **2014**, *59*, 2065–206. http://dx.doi.org/10.1021/je500206w.

[39] Mendez Sevillano, D. On the Prediction of Polyphenol Properties. Ph.D. Thesis, Technische Universiteit Delft, Delft, The Netherlands, 2015.

[40] Chebil, L.; Humeau, C.; Anthoni, J.; Dehez, F.; Engasser, J.-M.; Ghoul, M. Solubility of Flavonoids in Organic Solvents. *J. Chem. Eng. Data* **2007**, *52*, 1552–1556. http://dx.doi.org/10.1021/je7001094.

[41] Zi, J.; Peng, B.; Yan, W. Solubilities of Rutin in Eight Solvents at T = 283.15, 298.15, 313.15, 323.15, and 333.15 K. *Fluid Phase Equilibria* **2007**, *261*, 111–114.

[42] Martínez Aguda, R. Modeling the Solubility of Sclareol in Organic Solvents Using Solubility Parameters. Master's Thesis, North Carolina State University, Raleigh, NC, USA, 2007.

[43] Remil Aguda, R.; Chen, C.-C. Solubility of Nutraceutical Compounds in Generally Recognized as Safe Solvents at 298 K. *Int. J. Chem. Eng. Appl.* **2016**, *7*, 289–294. http://dx.doi.org/10.18178/ijcea.2016.7.5.591.

[44] Faiyaz, S.; Salem-Bekhit, M. M.; Haq, N.; Siddiqui, N. A. Solubility and Thermodynamics of Ferulic Acid in Different Neat Solvents: Measurement, Correlation and Molecular Interactions. *J. Mol. Liq.* **2017**, *236*, 144–150. http://dx.doi.org/10.1016/j.molliq.2017.04.014.

[45] Zuorro, A.; Iannone, A.; Lavecchia, R. Water—Organic Solvent Extraction of Phenolic Antioxidants from Brewers' Spent Grain. *Processes* **2019**, *7*, 126. http://dx.doi.org/10.3390/pr7030126.

[46] Mota, F. L.; Queimada, A. J.; Pinho, S. P.; Macedo, E. A. Aqueous Solubility of Some Natural Phenolic Compounds. *Ind. Eng. Chem. Res.* **2008**, *47*, 5182–5189. http://dx.doi.org/10.1021/ie071452o.