Revisiting the prediction of the partition coefficient using variables determined by the conductor-like polarizable continuum model

José G. Parra[1,*]

[1]Department of Chemistry, Faculty of Experimental Sciences and Technology, University of Carabobo, Laboratory of Molecular Simulation for Experimental Chemistry (SIMOLQUIMEX), Chemistry Building, Salvador Allende Avenue, Bárbula, Venezuela

*To whom correspondence should be addressed: José G. Parra,
Phone: +584128437640. email: jgparra2@uc.edu.ve.

ABSTRACT

Using the method of solvation Conductor-like Polarizable Continuum Model (C-PCM) and the method of Amovilli-Mennucci contained in the GAMESS-US package, were estimated the electrostatic, dispersion and repulsion free energies of organic molecules in 1-octanol and water. The calculations of solvation in water and 1-octanol were performed using the method C-PCM//PBE-VWN/6-31G (2p, 2d), to determine the electrostatic contribution to the solvation free energy. Consequently, using the concept of structure-property relationship was obtained a QSPR model to estimate the cavitation free energies of organic molecules with the experimental solvation free energies. Additionally, with the surface atomics (C, H, N and O) of the atoms that conform to the organic molecules calculated by the method GEPOL-GB, the electrostatic free energies, and non-bonding free energies, two QSPR models were constructed to predict the 1-octanol/water partition coefficient (logP) for neutral organic molecules containing a single functional group from the following classes: alcohols, ketones, aldehydes, carboxylic acids, esters, ethers, amines, alkanes (linear and cyclic), and aromatic compounds. In contrast, in the case of alcohol compounds, molecules of 1,2-ethanediol, 1,3-propanediol and 1,4-butanediol were used in the training set. Model A, which uses electrostatic free energies and atomic surface areas, achieved a correlation coefficient of $R^2 = 0.991$, while Model B, which incorporates non-bonding free energies, achieved $R^2 = 0.986$. The applicability domain of the models is explicitly defined for monofunctional compounds at infinite dilution and 298.15 K.

Keywords: electrostatic free energy; partition coefficient; solvation; computational chemistry; density functional theory.

1. INTRODUCTION

The equilibrium between liquid phases or between liquid and vapor is critical to understanding a wide range of chemical and biological phenomena. Indeed, it is imperative to understand and predict such equilibriums to optimize chemical products and processes (1)In general, researchers have employed these types of equilibrium to theoretically estimate a range of properties, including the aqueous solvation energy of polar and nonpolar species, the vapor pressure and boiling point of pure liquids, the equilibrium constant of carboxylic acids in the aqueous phase, and the partition coefficient of a solute between two immiscible liquids(2-8).

The partition coefficient of a solute in two immiscible liquids is one of the most significant properties employed. This property is particularly advantageous in the context of describing the behavior of amphiphilic molecules and species employed in the pharmaceutical industry. The distribution of a molecule between water and a lipid material is typically accomplished through the 1-octanol/water partition coefficient ($K_{o/w} = \log P$), which is regarded as the most significant physicochemical property for describing the affinity of chemical species in a specific solvent (9-11).

The 1-octanol is widely accepted as the material whose behavior is most similar to the lipids found in nature, and the partition coefficient is the most widely used descriptor for structure-property relationships QSPR (12). Concurrently, models of structure-property relationships (QSPR) emerge as a result of a process that begins with the description of molecular structures and culminates in the prediction of the behavior of molecules in different phases (13-16). Consequently, QSPR models were developed using properties calculated by quantum mechanical methods and molecular dynamics (17-20).

A variety of quantum mechanical solvation models, including C-PCM (Conductor-Like Polarizable Continuum Model) and COSMO (Conductor-Like Screening Model), have been employed to investigate and calculate the properties of chemical compounds in solvents (21-27). In these methods, the solvent is represented by a continuous dielectric model, which provides a satisfactory description of the properties of molecules with low computational cost and generates favorable results. Concurrently, numerous methods have been developed to predict the molecular properties of solutes in different solvent media. Some of these methods have been employed to determine the molecular volume of liquids, vapor pressure, and solvation energies (28-32). Consequently, in this study, the C-PCM solvation method was employed to calculate the structural and energetic properties of an organic molecules set. Finally, two structure-property models were obtained to estimate the 1-octanol/water partition coefficient ($K_{o/w} = \log P$) and the free energies of solvation of these groups of organic molecules. While solvation free energies are essential for understanding solute-solvent interactions, the partition coefficient is inherently a transfer property between two phases. The distinction between transfer and solvation free energies is critical because the former accounts for the competitive solvation of the solute by two immiscible solvents, including effects of activity coefficients, conformational changes, and specific solvent-solvent interactions that are not captured by single-solvent continuum models.

2. THEORETICAL FRAMEWORK

The solvation free energy (ΔG_{solv}) is free energy change necessary for transferring a molecule of the vacuum into the solvent[23] and can be defined by the equation 1:

$$\Delta G_{solv} = \Delta G_{elec} + \Delta G_{disp} + \Delta G_{rep} + \Delta G_{cav} + RT\ln\left(\frac{q_{rot,g}q_{vib,g}}{q_{rot,s}q_{vib,s}}\right) - RT\ln\left(\frac{n_{M,g}\Lambda_{M,g}^3}{n_{M,s}\Lambda_{M,s}^3}\right) \quad (1)$$

The term ΔG_{elec} is the electrostatic component due to the interaction between the solute and the solvent medium. The variables ΔG_{rep} and ΔG_{disp} are the free energies of repulsion and dispersion between the

solute and solvent. Finally, the term ΔG_{cav} is the free energy required to construct the cavity of the solute in the solvent and the last two terms are negligible.

2. 1. Electrostatic free energy

Under the influence of a solvent described by a dielectric continuum model, the Hamiltonian of the solute can be defined by equation 2 as:

$$\hat{H} = \hat{H}^o + \hat{V} \quad (2)$$

In which $\hat{H}^o$ is the Hamiltonian of the isolated solute, the operator $\hat{V}$ describes the solute-solvent interaction, which depends linearly on the state function Ψ of the solute. In this case, the electrostatic free energy ΔG_{elec} is determined by the variational minimization of the free energy of solute given by equation 3:

$$\Delta G_{elec} = \langle \Psi | \hat{H}^o | \Psi \rangle + \frac{1}{2} \langle \Psi | \hat{V} | \Psi \rangle \quad (3)$$

In general, the interaction operator $\hat{V}$ is written in terms of apparent polarization charges. In each tessera i, a charge q_i appears, according to the conductor-like boundary condition:

$$V(\vec{r}) + \sum_i^{segments} V_{q_i}(\vec{r}) = 0 \quad (4)$$

Where $V(\vec{r})$ and $V_{q_i}(\vec{r})$ are the electrostatic potentials due to the solute and to the polarization charges, and $\vec{r}$ is a point on the segment surface.

The vector of polarization charges Q can be defined by Equation 5:

$$\boldsymbol{AQ} = -V \quad (5)$$

Here, the vector Vcontains the electrostatic potential due to the solute on the surface segment. The elements of the matrix A are expressed by equations 6 and 7:

$$A_{ii} = 1.07\sqrt{\frac{4\pi}{S_i}} \quad (6)$$

$$A_{ij} = \frac{1}{|\vec{r}_i - \vec{r}_j|} \quad (7)$$

Where S_i is the area of tessera i. The equation 6 was defined by Klamt and Schüürmann for a simple sphere partitioned into a number of segments of identical surface area[21]. In C-PCM, the polarization charges are scaled so that the total polarization charge obeys Gauss's law and a very effective way to do this is to multiply each charge by the factor $\frac{\varepsilon-1}{\varepsilon}$ and new charges are given by equation 8:

$$q = \frac{\varepsilon-1}{\varepsilon} Q \quad (8)$$

Here ε corresponds to the dielectric constant of solvent. At the same time, the contribution ΔG_{elec} is important for polar and charged solutes due to the polarization of the solvent, which are modeled as a uniform medium of dielectric constant.

2. 2. Free energy required to the construction of a molecular cavity in the solvent

With the aim of determine the solute-solvent interaction, the molecules are placed in cavities formed by overlapping spheres centered on solute atoms. The construction of the cavity for the solute in the solvent requires the application of a certain amount of energy known as cavitation free energy whose concept was introduced by Uhlig (33) and the definition was applied initially to determine the solubility of gases in liquids. To determine the cavitation free energy, ΔG_{cav}, different models have been developed and the formulation given by Reiss and Pierotti (34-36). This theory describes the molecules of fluid using hard spheres to evaluate the ΔG_{cav}, which is determined by a polynomial depending of the radius R_{ms}, which corresponds to the sum of the radii of the solute and solvent:

$$\Delta G_{cav} = K_0 + K_1 R_{ms} + K_2 R_{ms}^2 + K_3 R_{ms}^3 \quad (9)$$

The coefficients K_i ($i = 0,1,2,3$) are expressed in terms of the properties of the solvent (molecular radius R_s, and density number n_s) and of the system (pressure P and temperature T). This can be calculated by the equations 10, 11, 12 and 13:

$$K_0 = RT\left[-\ln(1-y) + \frac{9}{2}\left(\frac{y}{1-y}\right)^2\right] - \frac{4\pi R_S^3 P}{3} \quad (10)$$

$$K_1 = -\frac{3RT}{R_S}\left[\frac{y}{1-y} + 3\left(\frac{y}{1-y}\right)^2\right] + 4\pi R_S^2 P \quad (11)$$

$$K_2 = \frac{3RT}{R_S^2}\left[\frac{y}{1-y} + \frac{3}{2}\left(\frac{y}{1-y}\right)^2\right] - 4\pi R_s P \quad (12)$$

$$K_3 = \frac{4\pi P}{3} \quad (13)$$

The term $y = \frac{4\pi R_S^3 n_s}{3}$ is associated to the microscopic terms. The cavitation free energy is a positive term which included the loss of entropy of the solvent due to the reorganization of these molecules by the presence of the solute.

2.3. Free energies of Dispersion and Repulsion

This section shows the models to calculate the dispersive and repulsive contributions to solvation free energy. Using the atom-atom interaction model (37), the dispersion free energy between two molecules A and S is described by equation 14:

$$E(A,S)_{disp} = -\sum_{n=6}^{\infty}\sum_{t\in S}\sum_{a\in A} d_{at}^{(n)} r_{at}^{-n} \quad (14)$$

Here, $d_{at}^{(n)}$ is the dispersion coefficient and r_{at} is the distance between atoms. In real solutions, the description of the relative distances between the atoms of the solute and solvent can be replaced by the distribution function, defined in equation 15:

$$\rho_{at}(r_{at}) = N_t \rho_s g_{at}(r_{at}) \quad (15)$$

In equation 15, N_t is the number of atoms of type t for each solvent molecule, ρ_s is the number density of the solvent and g_{at} is the correlation function, which is dependent on the interaction distance (r_{at}). The use of a continuous function allows the replacement of the summation over S, by an integral shown in the equation 16:

$$G_{disp}(AenS) = -\sum_{t\in S}\sum_{a\in A} d_{at}^{(6)} N_t \rho_s \times \int dr_t g_{at}(r_{at}) r_{at}^{-6} \quad (16)$$

With the equation 16 is possible to determine the dispersion free energy for a solute in a solvent. In turn Amovilli (38), proposed a model to calculate the free energy of dispersion through a quantum-mechanical treatment:

$$\Delta G_{disp} = \frac{1}{\pi}\int_0^{\infty} d\omega \sum_{K\neq 0} \frac{\omega_{0\mathrm{K}}^{A}}{\omega_{0\mathrm{K}}^{A^2}+\omega^2} \times \int d\vec{r}_1 \oint \frac{d\vec{r}_2}{r_{12}} P_A(0\mathrm{K}|\vec{r}_1)\, \sigma_B[B(i\omega), P_A(0\mathrm{K}|\vec{r})](\vec{r}_2) \quad (17)$$

Where $P_A(0\mathrm{K}|\vec{r})$ and $\omega_{0\mathrm{K}}^{A}$, are the densities and energies of transition of a state K of the solute. The term σ_B is the surface charge density induced by the electric field of the charge distribution$P_A(0\mathrm{K}|\vec{r})$, depending on a dielectric constant calculated at imaginary frequencies. Additionally, σ_B is always smaller than the square of the refractive index η_B^2 estimated to zero frequency and measured of the electronic transitions in the visible spectrum. However, Mennucci and Amovilli [39] defined these terms using the equation 18:

$$\sigma_B[B(i\omega), P_A(0\mathrm{K})](\vec{r}) = \frac{\Omega_B^2}{\Omega_B^2+\omega^2}\frac{\eta_B^2-1}{4\pi\eta_B^2} E_{0\mathrm{K}}^{A}(\vec{r}) \quad (18)$$

Where $\Omega_B = \eta_B I_B$ and depends on the ionization potential I_B of the solvent. Here $E_{0\mathrm{K}}^{A}(\vec{r})$ is the component of the electric field, which is orthogonal to the surface of the solute cavity at the point $\vec{r}$ that is associated with the density of transition $P_A(0\mathrm{K}|\vec{r})$. By using the equation 18, equation 17 can be defined as:

$$\Delta G_{disp} = -\frac{\eta_B^2-1}{8\pi\eta_B\left(\eta_B+\frac{\omega_A}{I_B}\right)} \sum_{K\neq 0} \oint d\vec{r}\, V_{0\mathrm{K}}^{A}(\vec{r}) E_{0\mathrm{K}}^{A}(\vec{r}) \quad (19)$$

The term $V_{0\mathrm{K}}^{A}$ is the potential associated with the electric field $E_{0\mathrm{K}}^{A}$. All terms associated with the equation of dispersion free energy are calculated with the restricted Hartree-Fock method.

To calculate the repulsion free energy in the solvation process, Amovilli and Mennucci [39] proposed a model based on the method of Floris [37] and the equation 20:

$$\Delta G_{rep} = \rho_B \int d\vec{R} U_{rep}^{AB}(\vec{R}) g_{AB}(\vec{R}) \quad (20)$$

Here A and B, corresponds to the solute and the solvent. $\vec{R}$, is an appropriate set of coordinates that define the internal geometry of the complex AB, ρ_B is the number density of the solvent and g_{AB} is a correlation function whose value is zero inside the solute cavity and one outside the cavity. This model is based on that the repulsive force between two interacting molecules is given by the Pauli Exclusion Principle. Based on this principle, the free energy of repulsion increases with the overlap of the two distributions and is related to the density of electrons with the same electron spin. Thus, the repulsion free energy in the process of solvation can be calculated by equation 21:

$$\Delta G_{rep} = \rho_B n_{par}^B \int \vec{R} \frac{1}{2} \int \frac{d\vec{r}_1 d\vec{r}_2}{r_{12}} P_A(\vec{r}_1, \vec{r}_2)\, P_{par}\left(\vec{R}|\vec{r}_2, \vec{r}_1\right) \quad (21)$$

Here n_{par}^B is the number of pairs of valence electrons of the solvent, $P_{par}\left(\vec{R}|\vec{r}_2, \vec{r}_1\right)$ and $P_A(\vec{r}_1, \vec{r}_2)$ correspond to the pair density matrices adapted. Here $\vec{R}$ are the coordinates of the center of the localized orbital which contain the pairs of free electrons or bonded to the solute molecule.

2. 4. Distribution of a solute in two immiscible liquids

Suppose two immiscible liquids in contact, which is incorporated a solute A non dissociable whose concentration is in the range of a dilute solution (9, 40). The solute A to be distributed between the two phases for form a solution. If both phases are in equilibrium, the chemical potentials of the solute A for both phases are equal and can be written as:

$$\mu_{(soluto,\alpha)} = \mu_{(soluto,\beta)} \quad (22)$$

In turn, the chemical potential of solute A in a phase can be written as:

$$\mu_{(soluto,i)} = \mu^0_{(soluto,i)} + RT\ln x_{(soluto,i)} \quad (23)$$

Where $x_{(soluto,i)}$ correspond to the fraction of solute in equilibrium with one of the phases and $\mu^0_{(soluto,i)}$ is the standard potential of solute A in one of the phases. By using the equation 22 and 23 can be established a relationship to calculate the solute distribution in the two phases:

$$\mu^0_{(soluto,\alpha)} + RT\ln x_{(soluto,\alpha)} = \mu^0_{(soluto,\beta)} + RT\ln x_{(soluto,\beta)} \quad (24)$$

Where the concentration ratio is given by:

$$\left(\frac{x_{(soluto,\alpha)}}{x_{(soluto,\beta)}}\right) = K_{distribucion} = EXP\left(\frac{\mu^0_{(soluto,\beta)} - \mu^0_{(soluto,\alpha)}}{RT}\right) \quad (25)$$

Assuming that $\mu^0_{(soluto,i)} = \Delta G^0_{(solv,i)} + \Delta G^0_{vap(A)}$, where $\Delta G^0_{vap(A)}$ is the vaporization free energy of the solute, we can calculate the solvation of A in a given solvent. In this case, the standard chemical potential of solute A in a phase $\mu^0_{(soluto,i)}$ is equivalent to the dissolution free energy ΔG^0_{sol}. Therefore, we can write the distribution constant of the solute in both phases as follows:

$$K_{o/w} = EXP\left(\frac{\Delta G^o_{vap(A)} + \Delta G^0_{(solv,\beta)} - \left(\Delta G^0_{vap(A)} + \Delta G^0_{(solv,\alpha)}\right)}{RT}\right) \quad (26)$$

Here $\Delta G^0_{(solv,\beta)}$ and $\Delta G^0_{(solv,\alpha)}$ correspond to the solvation free energies of solute in two immiscible liquids. Simplifying Equation 26, then we have an expression using quantum mechanical methods of solvation to calculate the distribution constant $K_{\frac{o}{w}}$ of a solute A in two immiscible liquids:

$$K_{o/w} = EXP\left(\frac{\Delta G^0_{(solv,\beta)} - \Delta G^0_{(solv,\alpha)}}{RT}\right) \quad (27)$$

Generally, it is possible to calculate the free energy of the phase changes and the solvation free energy. However, the determination of the free energy of phase transfer of a solute of a liquid to other is a limitation of quantum mechanical methods used to estimate the chemical potential of a system. For the reason, the equation 26 is very appropriate to estimate this free energy associated with the phase transfer.

3. METHODOLOGY

3.1 Quantum mechanical calculations

All three-dimensional molecular structures were constructed using the wxMacMolPlt v7.4.5 software (41). Then, molecular structures were optimized in the vacuum using the Density Functional Theory (DFT). The functional PBE-VWN was used with the basis set 6-31G (2p, 2d) (42- 44).

With the coordinates obtained in the optimization, the calculations of solvation in water and 1-octanol were performed using the method C-PCM//PBE-VWN/6-31G(2p,2d), to determine the electrostatic contribution to the solvation free energy. These calculations were developed based on the theory of Ben-Naim (45), where it is considered that the molecule does not change its structural conformation when the molecule is transferred of the vacuum to the solvent medium. Figure 1, shows the process of solvation of the ethanol in a continuum dielectric medium using the Ben-Naim principle.

For the construction of the cavities of the solute molecules, we used the GEPOL-GB method proposed by Pascual-Ahuir and Tomasi (46, 47) using the van der Waals radii for the atoms of C, N, H and O. In Table 1, are showed the values of the radii used. For the study of solvation in water and 1-octanol, were used as input parameters the dielectric constant, molecular volume and surface tension (48, 49). The repulsion and dispersion free energies were calculated using the method Amovilli-Menucci, being necessary to define some parameters of solvents such as ionization potential, the square of the refractive index, density and number of electrons the valence shell of the solvent. The ionization potentials of water and 1-octanol was calculated by an optimization procedure SCF//UHF/PBE-VWN with a basis set 6-31G (2p, 2d). All calculations were performed with the GAMESS-US software (50). The used parameters are shown in Table 2.

3.2 Training Set Selection and Model Validation

The training set of 42 organic molecules used for parameterization of the cavitation free energy model (see Table 3) and the 50 molecules used for the QSPR models (Models A and B) were selected based on three primary criteria: (i) availability of high-quality experimental solvation free energies and logP values

in the literature; (ii) structural and functional diversity, including alcohols, ketones, aldehydes, carboxylic acids, esters, ethers, amines, alkanes (linear and cyclic), and aromatic compounds; and (iii) variation in molecular size and alkyl chain length to capture hydrophobic contributions systematically. The selection of a diverse training set is essential for developing QSPR models with broad applicability and reliable predictions for unseen compounds. As an exception, in cavitation free energy model, 1,2-ethanediol alcohol was included in the training set. In contrast, in logP models, alcohol molecules such as 1,2-ethanediol, 1,3-propanediol, and 1,4-butanediol were included in the training set with a view to extending the model's applicability. Glycerol was used to validate this extension of the logP models.

To prevent overfitting and evaluate the predictive power of the models, we employed two complementary validation strategies. First, leave-one-out cross-validation was performed on the 42 molecules used for parameterization of the cavitation free energy model, yielding correlation coefficients of R^2=0.985 for water and R^2=0.982 for 1-octanol, confirming that the models are not overfitted to the training data. Second, external validation was conducted using an independent set of ten molecules (discussed in section 4.3) not included in the training set, including 2-pentanol, n-hexanal, hexylamine, dibutyl ether, propyl acetate, n-butylbenzene, dioxane, anisole, glycerol, and 3-pentanone. The excellent agreement between predicted and experimental logP values for these molecules (discussed in Section 4) provides strong evidence that our models possess genuine predictive capability and are not the result of overfitting. Additionally, all coefficients in the models (see sections 4.1 and 4.2 in Results and Discussion) were examined for chemical consistency, with positive coefficients for hydrophobic contributions (carbon surfaces, alkyl groups) and negative coefficients for polar contributions (oxygen surfaces, hydrogen-bonding groups), confirming that the models capture meaningful structure-property relationships rather than spurious correlations.

4. RESULTS AND DISCUSSION

4.1. Linear regression model to estimate the cavitation free energy

In general, the solvation free energy is calculated by quantum mechanical methods by means of the equation 1. In this model, the cavitation free energy is not determined accurately by various theoretical models, as for example, the Claverie-Pierotti method (34-36). Therefore, this work estimates a model based on a structure-property relationship using the experimental solvation free energies of 42 liquid organic compounds, allowing for a more accurate calculation of the cavitation free energy. Table 3 shows the values of electrostatic, dispersive and repulsive free energies calculated using the methods C-PCM and Amovilli-Menucci for this group of 42 organic molecules. Using the approximation of Amovilli-Menucci, the free energy for the construction of the cavity, ΔG_{cav} , was calculated by the equation 28:

$$\Delta G_{cav} = \Delta G_{solv(exp)} - \Delta G_{elec} - \Delta G_{rep} - \Delta G_{disp} \quad (28)$$

The term $\Delta G_{solv(exp)}$ corresponds to the experimental solvation free energy reported for these organic molecules in 1-octanol and water. The equation 28 was used by Amovilli and Menucci to calculate the cavitation free energy. For this, they plotted the free energy associated to the construction of the cavity in function of the cavity areas (A_{cavity}) of the molecules, obtaining a linear relationship of type,

$$\Delta G_{cav} = \beta \times A_{cavity} + \alpha \quad (29)$$

In equation 29, the α and β terms are parameters obtained from linear regression. This model has been proposed in other studies of solvation (51, 52). In this case, we propose an equation to calculate the free energy for the construction of the cavity based on the use of molecular descriptors calculated with the C-PCM and GEPOL-GB methods included in the GAMESS-US package. The linear regression model is shown in equation 30,

$$\Delta G_{cav(estimate)} = \sum_{i=1}^{n} \sigma_i A_i + \sum_{j=1}^{m} k_j X + b \quad (30)$$

Where $n = 4$ and $m = 9$. The term i corresponds to the different atoms in the molecule of the solute (i =C, H, N and O). The factor σ_i is associated with each atom and k_j is a factor related to each functional group present in the molecule. This equation obtained by linear regression includes the area of the atoms that conform to the molecule of the solute and a factor associated with the functional groups present in the molecules. Furthermore, the term X is equal to one if the functional group is present and zero otherwise.

At the same time, the parameter b is a constant term obtained by linear regression, and A_i is the surface area of the atom i obtained by the GEPOL-GB method (46). Models like equation 29 and 30 using statistical mechanics have been reported by Marenich, Cramer, Truhlar and Kollman (53-55).

The values of the parameters obtained by means of linear regression in order to estimate the free energy required for the construction of the cavity can be found in Table 4. This model was parameterized using the root mean square error ($RMSE$) defined by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left|\Delta G_{cav} - \Delta G_{cav(estimate)}\right|^2}{n}} \quad (31)$$

The parameterization was done for water and 1-octanol whose linear correlation coefficients were $R^2 = 0.990$ and $R^2 = 0.986$ for 1-octanol and water, respectively. In the case of 1-octanol, the coefficient k_j was not determined for the acid functional group.

As illustrated in Table 5, the values of solvation free energy in water and 1-octanol were obtained by means of the free energy for the construction of the cavity, estimated by equation 30. The results are acceptable for the 42 molecules used in the parameterization process.

As illustrated in Figures 2 and 3, the estimated values of solvation free energy are plotted with respect to the experimental solvation free energy in water and 1-octanol. For the multivariate model, a linear curve for regression type demonstrates a strong correlation for the solvation free energies in water and 1-octanol.

The errors are generally larger for 1-octanol solvation energies than for water solvation energies. This is evident from the RMSE values in Table 4. For water, RMSE = 1.25 kJ mol^{-1}, while for 1-octanol, RMSE = 0.67 kJ mol^{-1}. Although the RMSE for 1-octanol is smaller, the relative errors (see values in Table 5) show that the deviations in 1-octanol are more systematic and depend on the functional group. This indicate that the use of macroscopic solvent parameters for 1-octanol introduces systematic errors because this solvent exhibits a complex molecular structure that is not adequately represented by a uniform dielectric continuum. 1-octanol forms dynamic aggregates through hydrogen bonding and hydrophobic interactions, creating a heterogeneous environment that affects solute partitioning in ways that simple continuum models cannot capture. This is particularly evident for molecules with polar functional groups (e.g., alcohols, ketones) where specific interactions with the 1-octanol hydroxyl group are important.

The lowest correlation was obtained for the determination of the solvation free energy in water. This is due to the basis set 6-31G(2p,2d) used to estimate the electrostatic contribution and the minimum basis set used in the Amovilli-Menucci method to determine the dispersion and repulsion free energies. In other papers, the estimation of these solvation free energies has been performed using other molecular descriptors as the index acceptor or hydrogen donor. This has enhanced the estimation of these properties (56, 57).

4. 2. Prediction of the partition coefficient $K_{o/w}$

Using the estimated solvation free energies and the equation 27, the determination of the partition coefficient was performed. Table 6 shows the estimated values of partition coefficient, which vary considerably with the experimental values reported.

The correlation coefficient obtained was $R^2 = 0.974$ using the values of solvation free energies calculated in this work (see correlation in Figure 4). However, these results have a better correlation than those reported by Amovilli-Menucci using their model to estimate the cavitation free energy.

At the same time, Table 6 presents the partition coefficients with their respective relative error where the greatest error in the estimate was obtained for 2-propanol and molecules that contain in their structure the carbonyl functional group. Additionally, we find that the amines used have a high magnitude of the relative error in the estimation of partition coefficient. However, we emphasize that the results obtained for the partition coefficient using solvation free energies estimated with the model for the free energy of cavitation, are better than the values that were obtained by Amovilli-Menucci. At the same time, it is important to take account that the mechanism for the distribution of a solute in two immiscible solvents is a process of transfer and therefore to use the solvation free energies is not entirely suitable for estimating the partition coefficients, which limits the study and this problem should be solved by using a model to determine the transfer free energy. Therefore, the basis set 6-31G(2p, 2d) used to calculate the electrostatic contribution to solvation free energy is appropriate. A systematic source of error is the use of a minimal basis set by the Amovilli-Menucci method for calculating dispersion and repulsion free energies. As shown in Table 6, molecules with significant van der Waals contributions to their solvation in 1-octanol, such as 2-propanol and dimethyl ether, exhibit the largest deviations. In this case, the minimal basis set STO-3G is limited to describe the anisotropic electron density distributions necessary for accurate dispersion interactions. This limitation is particularly severe for molecules with secondary alcohol groups, where the steric environment around the hydroxyl group modulates dispersion interactions with the solvent in ways that require a more flexible basis set. In future works, basis sets of at least triple-zeta quality with polarization functions, such as 6-311+G(d,p) will be employed to improve the accuracy of these non-electrostatic contributions. It is important to note that the deviations observed in Table 6 are not

random but reflect a fundamental limitation of the approach based on the difference in solvation energies (see equation 27 in manuscript). This method assumes that the transfer free energy between phases is exactly equal to the difference in solvation energies, which is a simplification that ignores activity effects, conformational changes, and molecular association in the organic phase. Additionally, the GEPOL-GB method for cavity construction, while computationally efficient, relies on fixed van der Waals radii and a single molecular conformation optimized in vacuum. This approach does not account for conformational flexibility, which is particularly problematic for molecules that can adopt multiple low-energy conformations with different solvent-accessible surface areas. For example, 2-propanol can rotate around its C-O bond, altering the exposure of its hydroxyl group to the solvent. For the 1,2-ethanediol shows an excellent prediction because its conformation is relatively constrained by intramolecular hydrogen bonding. In contrast, flexible molecules like 2-propanol and methyl butanoate can adopt multiple conformations, and using a single optimized structure does not capture the conformational ensemble that contributes to solvation.

Additionally, in this paper we propose two models of structure-property (QSPR) to estimate the partition coefficient using properties obtained with the method C-PCM. In these models QSPR, were included electrostatic free energy calculated in 1-octanol and water, the atomic surfaces of the atoms that form the molecules determined using the GEPOL-GB method and associated factors with functional groups present in the molecule. In this case, 50 molecules were used for the parameterization. The first model QSPR which we define as model A is shown in Equation 32:

$$K_{o/w} = \log P = a \times \left(\Delta G_{elec(wate\ )} - \Delta G_{elec(1-octanol)}\right) + \sum_{i=1}^{n} \sigma_i A_i + \sum_{j=1}^{m} k_j X + b \quad (32)$$

where $n = 4$ and $m = 9$. The term i corresponds to the different atoms in the molecule of the solute (i =C, H, N and O). The factor σ_i is the atomic tensor associated with each atom and k_j is a factor related to each functional group present in the molecule. The terms $\Delta G_{elec(water)}$ and $\Delta G_{elec(1-octanol)}$

correspond to the electrostatic free energies of solvation and the term X is equivalent to one if the functional group is present and zero otherwise.

The model A presents a correlation coefficient R^2=0.991, being simpler than the model of equation 30, which requires more computational cost for the estimation of the free energy of dispersion and repulsion. In turn, this model is more accessible and suitable for large molecular size molecules.

All parameters of the equation 32 were obtained using the root mean square error ($RMSE$) defined by equation 33:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left|\log P_{(\exp)} - \log P_{(estimate)}\right|^2}{n}} \quad (33)$$

Table 7 contains the values obtained by multivariate regression of the coefficients associated with the model A. Unlike the model that uses the cavitation free energy, in this model A, we obtain directly the distribution coefficient. The RMSE was equal to 0.16 and the statistical adjustment shows that the most influential factors in the property correspond to the presence of aromatic and amines functional groups. This Model A is parameterized directly using experimental logP values, which involves the transfer free energy between phases and avoiding the systematic errors inherent in the solvation energy difference approach. The chemical significance of the functional group coefficients in Model A reflects the distinct solvation environments in water versus 1-octanol, where negative coefficients for aromatic groups (-3.168) and amines (-3.789) indicate their strong preference for the organic phase due to π-π stacking and hydrophobic effects, while positive coefficients for acids (0.189) reflect their greater affinity for water through hydrogen bonding. The ketone coefficient (-1.457) captures the balance between carbonyl hydrogen bonding with water and dispersion interactions with the octanol alkyl chain, while ether (-1.208) and ester (-0.783) coefficients reflect their intermediate polarity and ability to act as hydrogen bond acceptors in both phases. The atomic surface coefficients enhance the performance of the Model A by accounting for molecular size and shape effects, with the positive carbon surface coefficient (0.086)

capturing hydrophobic contributions and the negative oxygen surface coefficient (-0.063) reflecting polar solvation effects. This combination of direct transfer free energy modeling, empirically calibrated functional group contributions, and atomic surface descriptors allows Model A to achieve superior correlation (R^2=0.991, RMSE=0.16) while maintaining computational simplicity and chemical interpretability. In turn, Table 8 shows the estimated values for the partition coefficient and its absolute relative errors using the model A where the molecules of dimethyl ether and 2-propanol have the greatest relative error. In Figure 5, shows the estimated values of partition coefficient for the model A with respect to experimental partition coefficient.

Finally, we present a second model to determine the partition coefficient of the 50 organic molecules which we define as model B. In this model B, we use properties like non-bonding free energy, which is given by equation 34:

$$\Delta G_{non-bond} = \Delta G_{elec} + \Delta G_{disp} + \Delta G_{rep} \quad (34)$$

This model of multivariate regression includes atomic surfaces A_i of the atoms that conform to the molecules and the coefficients associated with the functional groups of these molecules. The model B is shown in Equation 35:

$$\log P = a \times \Delta G_{non-bond(1-octanol)} + b \times \Delta G_{non-bond(water)} + \sum_{i=1}^{n} \sigma_i A_i + \sum_{j=1}^{n} k_j X + c \quad (35)$$

Similarly, the coefficients in the equation 35 are obtained using the experimental values of partition coefficient in the multivariate regression. Figure 6 shows the estimated values for the model B versus the experimental values of partition coefficient. This model B present a correlation coefficient equal to 0.986 and the Root Mean Square Error (RMSE) obtained was of 0.17 with respect to the experimental data. Table 9 shows the values obtained for the associated parameters with model B. In this model, there are no coefficients associated with aldehydes, alcohols, esters and cycloalkanes functional groups which make it a most simple model, but the determination of the coefficients of these functional groups would

have improved the multivariate correlation. However, the model B requires more computational cost to involve the dispersive and repulsive free energies. The systematic deviations observed for alcohols (2-propanol, relative error 6.14), amines (diethylamine, relative error 0.55), and carbonyl-containing compounds (acetone, relative error 0.84; methyl acetate, relative error 1.64) indicate that hydrogen bonding and specific intermolecular interactions are not fully captured by our current models, reflecting the inherent limitations of continuum solvation methods that represent solvents as uniform dielectric media. The inclusion of additional descriptors such as hydrogen-bond donor/acceptor ability, molecular dipole moment, or polar surface area would likely improve predictive capability by explicitly encoding information about the solute's capacity for specific interactions. While the atomic surface coefficients in models partially capture polar solvation effects, they do not distinguish between hydrogen bond donors and acceptors, nor do they account for interaction strength or directionality.

Additionally, Table 10 shows the logP values for 50 organic molecules using this model B where relative errors are largest with molecules that have the ester functional group, hydroxyl functional group and aldehyde functional group whose parameters were not included in this analysis. The two QSPR models are limited to the presence of a single functional group present in the structures of molecules.

The linear regression models assume additive contributions from atomic surfaces and functional groups. However, the results indicate that this additivity assumption breaks down for molecules with intermediate alkyl chain lengths (e.g., 2-butanone) where the balance between hydrophilic and hydrophobic contributions is most sensitive to the local molecular environment. This suggests that non-linear effects, such as synergistic interactions between the carbonyl group and the alkyl chain, are not fully captured by the current model. Future work should explore non-linear regression techniques, such as random forests or neural networks, with a larger and more diverse training set to capture these complex structure-property relationships. Furthermore, the $K_{o/w}$ describes the equilibrium distribution of a solute between two

immiscible liquid phases and is fundamentally a property of the transfer process between solvents, not a property of solvation from vacuum. While solvation free energies (ΔG_{solv}) describe the transfer from vacuum to a single solvent, the transfer free energy ($\Delta G_{transfer}$) describes the transfer between solvents and is directly related to logP. The use of solvation free energies to estimate logP (Equation 27) introduces systematic errors because it assumes that the difference in solvation free energies (both referenced to vacuum) equals the transfer free energy between phases. This assumption neglects the activity coefficient differences between solvents, conformational changes induced by different solvent environments, and specific solute-solvent and solvent-solvent interactions that are not captured by continuum models. These limitations explain the variations of the solvation energy difference approach in Table 6 showing larger deviations (R^2=0.974) compared to Model A (R^2=0.991), which is parameterized directly with the experimental logP values and effectively models the transfer free energy.

Additionally, the dispersive and repulsive free energies in Model B account for van der Waals interactions and Pauli exclusion effects that are particularly significant in 1-octanol, where the long alkyl chain provides extensive dispersion interactions with solute hydrophobic regions, while repulsive terms capture excluded volume effects governing molecular packing and cavity formation in both phases. However, these contributions are calculated using a minimal basis set (STO-3G) by the Amovilli-Menucci method, which systematically underestimates dispersion interactions for molecules with diffuse electron densities and fails to capture crucial effects for accurate logP prediction, limiting their overall contribution to predictive accuracy. The inclusion of these terms in Model B only marginal improvement in correlation (R^2=0.986) compared to Model A (R^2=0.991), while introducing additional computational cost and potential error propagation, demonstrating that the dominant factors for logP prediction are electrostatic interactions and molecular surface descriptors rather than explicit dispersion-repulsion calculations. The systematic errors in dispersion-repulsion terms also explain why the solvation energy difference approach

448 (see values in Table 6) shows larger deviations for molecules where van der Waals interactions dominate
449 solvation in 1-octanol, such as alkanes and ethers, while Model A effectively circumvents this limitation
450 through direct parameterization against experimental data. Consequently, despite their theoretical
451 importance, the practical contribution of dispersive and repulsive free energies to predictive accuracy is
452 limited by current methodological constraints, making Model A the most computationally efficient and
453 accurate choice for routine logP prediction.
454
455 **4. 3. Validation of QSPR Models with Additional Molecules**
456 The results presented in Table 11 demonstrate the predictive performance of Models A and B for ten
457 additional organic molecules not included in the training set, encompassing a chemically diverse range of
458 functional groups including secondary alcohols (2-pentanol), aldehydes (n-hexanal), amines
459 (hexylamine), ethers (dibutyl ether, dioxane, anisole), esters (propyl acetate), aromatics (n-butylbenzene),
460 and polyols (glycerol). Both models show excellent agreement with experimental logP values, with Model
461 A achieving a correlation coefficient of R^2=0.991 and Model B achieving R^2=0.986, confirming the
462 robustness and general applicability of our QSPR approach across different chemical families. For most
463 compounds, the predicted logP values are within 0.2-0.3 log units of the experimental values, with n-
464 butylbenzene showing nearly perfect prediction by both models (Model A: 4.26, Model B: 4.38, Exp:
465 4.38) and glycerol showing the largest deviations (Model A: -2.02, Model B: -2.59, Exp: -1.76), which
466 can be attributed to the strong hydrogen bonding capabilities of the triol that are not fully captured by
467 continuum solvation models. The excellent prediction for hexylamine by Model A (4.98 with respect to
468 2.06 experimental) reveals a significant overestimation, suggesting that the amine functional group
469 coefficient in Model A may require further refinement for primary amines with longer alkyl chains, while
470 Model B provides a much more accurate prediction (2.08 with respect to 2.06), demonstrating the value

of including dispersive and repulsive contributions for molecules where van der Waals interactions dominate solvation in the organic phase. For ethers and esters, both models perform well, with dibutyl ether predicted at 3.27 (Model A) and 4.01 (Model B) versus 3.21 experimental, and propyl acetate at 1.12 (Model A) and 1.14 (Model B) versus 1.24 experimental, indicating that the atomic surface descriptors effectively capture the hydrophobic contributions of alkyl chains. The prediction for anisole (Model A: 2.03, Model B: 1.88, Exp: 2.11) confirms that the aromatic coefficient (-3.168 in Model A) appropriately accounts for the π-π interactions and hydrophobic character of the methoxybenzene structure. For dioxane, both models predict the negative logP value correctly (Model A: -0.14, Model B: -0.67, Exp: -0.27), reflecting the hydrophilic nature of the cyclic ether due to its two oxygen atoms, although Model A provides better accuracy due to its direct parametrization against experimental data. The 3-pentanone prediction (Model A: 1.96, Model B: 2.06, Exp: 2.03) demonstrates that the ketone functional group coefficient successfully captures the balance between carbonyl hydrogen bonding and hydrophobic alkyl chain effects for symmetrical ketones. These results collectively validate that Model A, despite its computational simplicity, provides predictions comparable to or better than Model B for most compound classes, with the exception of long-chain amines where dispersion contributions become important, highlighting the complementary nature of both approaches for different chemical spaces.

4. 4 Deviations and Limitations

The deviations observed in Table 6 between estimated and experimental logP values are systematic rather than random, and can be attributed to different factors. Firstly, it is important to note a fundamental limitation of the solvation energy difference approach, as illustrated in equation 27. This approach assumes that the free energy of transfer is exactly equal to the difference in solvation energies. However, this assumption fails to consider activity effects, conformational changes, and molecular association in

494 the organic phase. This limitation is most severe for molecules like 2-propanol and dimethyl ether, which
495 can adopt different conformations or interact specifically with the solvent. Secondly, the utilization of a
496 minimal basis set for the dispersion and repulsion terms is imperative. The Amovilli-Menucci method
497 utilizes a minimal basis set (STO-3G) as a default setting. However, this set is inadequate in adequately
498 describing the anisotropic electron density distributions necessary for accurate dispersion interactions.
499 This phenomenon pertains to molecules where van der Waals interactions predominate over other forces
500 in the solvation process of 1-octanol. Thirdly, the GEPOL-GB cavity model, which utilizes fixed van der
501 Waals radii and a singular vacuum-optimized conformation, fails to account for conformational flexibility
502 or conformational averaging. This phenomenon presents a significant challenge, particularly in the
503 context of molecules that exhibit multiple low-energy conformations. The fourth issue pertains to the
504 assumptions underlying linear regression models. Specifically, the additivity assumption in linear
505 regression becomes invalid for molecules with intermediate alkyl chain lengths, due to the sensitivity of
506 their hydrophilic and hydrophobic contributions to the local molecular environment. Also, macroscopic
507 solvent parameters for 1-octanol, in where, the micro heterogeneous structure of 1-octanol (monomers,
508 dimers, aggregates) is not captured by uniform dielectric models, leading to systematic errors for
509 molecules with polar functional groups. Finally, it is important to emphasize that Model A (in equation
510 32) significantly mitigates these limitations by being parameterized directly against experimental logP
511 values, achieving a substantially better correlation (R^2=0.991, RMSE=0.16) than the approach based on
512 solvation energy differences (R^2=0.974). In contrast, the Amovilli-Menucci method, which systematically
513 underestimates dispersion interactions for molecules with diffuse electron densities and fails to capture
514 crucial effects for accurate logP prediction, limiting their overall contribution to predictive accuracy. The
515 inclusion of these terms in Model B only marginal improvement with a correlation coefficient of
516 R^2=0.986. Additionally, a limitation of the current QSPR models is the use of binary indicators for

functional groups, where a value of 1 is assigned if the functional group is present and 0 otherwise, regardless of the number of identical groups in the molecule. For example, 1,2-ethanediol receives the same alcohol indicator as methanol or ethanol, despite having two hydroxyl groups that significantly increase its hydrogen bonding capacity and hydrophilicity. This simplification is particularly problematic for polyols such as glycerol, which contains three hydroxyl groups and shows larger deviations (Model A: -2.02, Model B: -2.59, Exp: -1.76) compared to monohydric alcohols, suggesting that the model underestimates the hydrophilicity of polyfunctional compounds. The use of functional group counts (e.g., number of OH groups, number of NH groups, number of carbonyl groups) rather than binary presence would allow the model to capture the additive contributions of multiple identical groups to the overall partitioning behavior, improving predictions for polyfunctional molecules. Furthermore, for compounds containing different functional groups simultaneously (e.g., amino alcohols, hydroxy acids, keto, esters), interaction terms between functional groups could be incorporated to account for non-additive effects such as intramolecular hydrogen bonding, which can significantly modify solvation properties compared to the sum of individual group contributions. The binary indicator approach was chosen in this work for its simplicity and interpretability, as well as the limited number of polyfunctional compounds in the training set.

5. CONCLUSIONS

In this work was possible to estimate the partition coefficient of some organic molecules using models of structure-property relationship (QSPR). For the construction of these models was sufficient the electrostatic free energy obtained by the method C-PCM//PBE-VWN/6-31G (2p, 2d) and the dispersive and repulsive free energies calculated by the method of Amovilli-Menucci. Of the three proposed models that contain parameters calculated with the C-PCM method, we find that model A has the best correlation

coefficient which corresponds to 0.991 and a RMSE equal to 0.16. This model related the electrostatic free energy and the terms associated to the functional groups present in the molecule with the partition coefficient. Additionally, the QSPR models obtained in this work to estimate the free energy for the construction of the cavity in water and 1-octanol were acceptable. In this case, the correlation coefficients obtained for the solvation free energies were R^2=0.990 and R^2=0.986, for 1-octanol and water, respectively.

While a direct comparison with alternative solvation models such as SMD, COSMO-RS, or IEFPCM would further validate the robustness of our conclusions, the primary objective of this work is to demonstrate that C-PCM-derived descriptors, when combined with QSPR modeling, can achieve high predictive accuracy for logP (R^2=0.991) as a revision and extension of the original Amovilli-Menucci methodology. The C-PCM//PBE-VWN/6-31G(2p,2d) approach was selected for its favorable balance between accuracy and computational cost, as well as its wide availability in the GAMESS-US package, ensuring accessibility for the broader research community. Although SMD and COSMO-RS incorporate more refined parametrization of non-electrostatic contributions, the excellent correlation achieved by our model suggests that the electrostatic and surface descriptors derived from C-PCM capture the essential physics governing the partitioning process. Future work will explore the application of SMD, COSMO-RS, and IEFPCM to further refine the predictive capability of our QSPR models and establish their general applicability across different solvation methodologies. Finally, the predictive accuracy of the proposed models would be substantially enhanced by expanding the training dataset to include a wider range of chemical functionalities, particularly molecules with multiple functional groups, heterocycles, and halogenated compounds, as well as by employing rigorous external validation on independent datasets not used during parameterization.

List of Abbreviations

C-PCM — Conductor-like Polarizable Continuum Model
PCM — Polarizable Continuum Model
COSMO — Conductor-like Screening Model
COSMO-RS — Conductor-like Screening Model for Real Solvents
IEFPCM — Integral Equation Formalism Polarizable Continuum Model
SMD — Solvation Model based on Density
DFT — Density Functional Theory
RHF — Restricted Hartree-Fock
PBE — Perdew–Burke–Ernzerhof (exchange-correlation functional)
VWN — Vosko–Wilk–Nusair (correlation functional)
PBE-VWN — Perdew–Burke–Ernzerhof / Vosko–Wilk–Nusair functional
GEPOL-GB — GEneration of POLyhedral cavities – Gauss-Bonnet method
vdW — van der Waals
logP — Logarithm of the partition coefficient
QSPR — Quantitative Structure–Property Relationship
ΔGsolv — Solvation free energy
ΔGelec — Electrostatic free energy of solvation
ΔGcav — Cavitation free energy
ΔGdisp — Dispersion free energy
ΔGrep — Repulsion free energy
ΔGvap — Vaporization free energy
RMSE — Root Mean Square Error
R^2 — Coefficient of determination
r — Linear correlation coefficient
GAMESS-US — General Atomic and Molecular Electronic Structure System (US version)
wxMacMolPlt — Molecular visualization and input preparation software
C — Carbon
H — Hydrogen
N — Nitrogen
O — Oxygen

Author Contributions

The author (José G. Parra) is solely responsible for conceptualization, methodology, software, validation, formal analysis, funding acquisition, investigation, resources, data curation, writing-original draft preparation, writing-review and editing, visualization. The author has reviewed and approved the final version of the manuscript.

Declaration of Figure Originality

All figures included in the manuscript are original and not reproduced from previously published sources.

Conflicts of Interest

I declare that I have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. This research was conducted solely for academic purposes.

ORCID's

José G. Parra: 0000-0001-5991-3152

Availability of Data and Materials

The data supporting the findings of this study are available in the article and upon request.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgments

The author gratefully acknowledges the Universidad de Carabobo, Facultad Experimental de Ciencias y Tecnologia, and the Laboratorio de Simulación Molecular para la Química Experimental (SIMOLQUIMEX) for providing the computational resources and institutional support necessary to carry out this work.

AI Declaration

In accordance with the Committee on Publication Ethics (COPE) position on "Authorship and AI tools" (COPE Council, 2023), it is hereby declared that the AI-based tool DeepSeek (DeepSeek-V3 / DeepSeek-R1) was used exclusively for language polishing, grammatical correction, and stylistic refinement of the manuscript text. This comprised the correction of grammatical errors, the improvement of sentence structure, the rephrasing for enhanced clarity and readability, and the refining of the overall flow and coherence of the text. The AI tool was not used as an author or co-author of this work. It is important to note that all scientific content, research data, methodology, interpretations, conclusions, and all

intellectual contributions remain entirely the work of the author. All AI-assisted edits were performed under the direct supervision and control of the author, with each suggestion subjected to critical review, verification, and approval prior to incorporation into the manuscript. As the sole author, I hereby affirm that I assume full responsibility for the content, accuracy, integrity, and originality of the work; that all data, calculations, interpretations, and conclusions presented were derived from my own research and analysis; that no part of this work was generated or created by AI (the tool was used solely for editing existing text); that this work has not been published previously and is not under consideration for publication elsewhere; and that there are no conflicts of interest to declare.

REFERENCES:

(1) Drefahl, A.; Reinhard, M.; *Handbook for estimating physiochemical properties of organic compounds*; John Wiley & Sons, Inc: New York, 1999.

(2) Xiao, Z. J.; Chen, J. W.; Wang, Y.; Wang, Z. Y. **In silico package models for deriving values of solute parameters in linear solvation energy relationships.** *SAR and QSAR in Environmental Research* **2023**, *34*(1), 21–37. https://doi.org/10.1080/1062936X.2022.2162576

(3) Wang, S.; Shiang-Tai, L.; Chang, J.; Goddard, W. A., III; Sandler, S. I. **Application of the COSMO−SAC−BP Solvation Model to Predictions of Normal Boiling Temperatures for Environmentally Significant Substances.** *Industrial & Engineering Chemistry Research* ***2006***, *45*, 5426-5434. https://doi.org/10.1021/ie050352k

(4) Ulrich, N.; Ebert, A. Can deep learning algorithms enhance the prediction of solute descriptors for linear solvation energy relationship approaches?. *Fluid Phase Equilibria* **2022**, *555*, 113349. https://doi.org/10.1016/j.fluid.2021.113349

(5) **Zhurko, G. A.; Fedorova, A. A.** QSPR prediction of the acidities of carboxylic acids and phenols with different approaches. *Molecular Physics* **2025**, *123*(10), e2396535. https://doi.org/10.1080/00268976.2024.2396535

(6) Roy, D.; Patel, C. Revisiting the Use of Quantum Chemical Calculations in $LogP_{octanol-water}$ Prediction. *Molecules* **2023**, *28*(2), 801. https://doi.org/10.3390/molecules28020801

(7) Diaz, A. H.; Duque-Noreña, M.; Chamorro, E. Unveiling an electronic LogP analogue within the conceptual density functional theory framework. *Chemical Physics* **2024**, *584*, 112346. https://doi.org/10.1016/j.chemphys.2024.112346

(8) Thompson, J.; Cramer, C.; Truhlar, D. Predicting aqueous solubilities from aqueous free energies of solvation and experimental or calculated vapor pressures of pure substances. *Journal of Chemical Physics* ***2003***, *119* (3), 1661-1670. https://doi.org/10.1063/1.1579474

(9) Mecozzi, S.; Connors, K.; *Thermodynamics of pharmaceutical system*; John Wiley & Sons, Inc: New Jersey, 2010.

(10) Letcher, T. M.; *Thermodynamics, Solubility and Issues*: Elsevier Science and Technology Books, 2007.

(11) Chung, Y.; Vermeire, F. H.; Wu, H.; Walker, P. J.; Abraham, M. H.; Green, W. H. Extension of the Abraham Solvation Model to Ionic Solute Partitioning. *Journal of Chemical Information and Modeling* **2022**, *62*(3), 433–446. https://doi.org/10.1021/acs.jcim.1c01103

(12) Delgado, E. J.; Jaña, G. A. Quantitative Prediction of Solvation Free Energy in Octanol of Organic Compounds. *International Journal of Molecular Sciences* ***2009***, *10*, 1031-1044. https://doi.org/10.3390/i10031031

669 (13) Yu, X.; Wang, H.; Acree Jr., W.; Deng, J. QSPR models for solvation enthalpy based on quantum
670 chemical descriptors. *Journal of Molecular Liquids* ***2023***, *389*, 122884.
671 https://doi.org/10.1016/j.molliq.2023.122884

672 (14) Sharma, P.; Ranjan, P.; Chakraborty, T. Applications of conceptual density functional theory in
673 reference to quantitative structure–activity / property relationship. *Molecular Physics* ***2024***, *122(23)*,
674 e2331620. doi: 10.1080/00268976.2024.2331620.

675 (15) Consonni, V.; Todeschini, R.; *Molecular Descriptors for Chemo informatics* (Volume I y II),
676 WILEY-VCH Verlag GmbH & Co. KGa: Weinheim, 2009.

677 (16) Sodaei, Z.; Ekrami, S.; Hashemianzadeh, S.M. Machine learning analysis of molecular dynamics
678 properties influencing drug solubility. *Science Report* ***2025***, *15*, 26955. https://doi.org/10.1038/s41598-
679 025-11392-1

680 (17) Egert, T.; Langowski, H-C. Linear solvation energy relationships (LSERs) for robust prediction of
681 partition coefficients between low density polyethylene and water. Part II: Model evaluation and
682 benchmarking. *European Journal of Pharmaceutical Sciences* ***2022***, *172*, 106137.
683 https://doi.org/10.1016/j.ejps.2022.106137

684 (18) Sun, Y.; Hou, T.; He, X.; Man, V. H.; Wang, J. Development and test of highly accurate endpoint
685 free energy methods. 2: Prediction of logarithm of n-octanol–water partition coefficient (logP) for
686 druglike molecules using MM-PBSA method. *J. Comput. Chem.* ***2023***, *44(13)*, 1300.
687 https://doi.org/10.1002/jcc.27086

(19) Brown, T. N. QSPRs for Predicting Equilibrium Partitioning in Solvent-Air Systems from the Chemical Structures of Solutes and Solvents. *J. Solution Chem.* ***2022****, 51,* 1101–1132. https://doi.org/10.1007/s10953-022-01162-2

(20) Isert, C.; Kromann, J.; Stiefl, N.; Schneider, G.; Lewis, R. Machine Learning for Fast, Quantum Mechanics-Based Approximation of Drug Lipophilicity. *ACS Omega* ***2023***, *8 (2)*, 2046-2056. https://doi.org/10.1021/acsomega.2c05607

(21) Klamt A.; Schüürmann G. First principles implementation of solvent effects without outlying charge error. *J. Chem. Phys.* ***1997****, 106,* 6622–6633. https://doi.org/10.1063/1.473662

(22) Tshilande N.; Mammino L.; Bilonda M. K. The Study of Molecules and Processes in Solution: An Overview of Questions, Approaches and Applications. *Computation* ***2024***, *12(4)*, 78. https://doi.org/10.3390/computation12040078

(23) Si-Cong, L.; Xin-Rui, Z.; Dan-Yang, L.; De-Cai, F. DFT calculations in solution systems: solvation energy, dispersion energy and entropy. *Physical Chemistry Chemical Physics* ***2023***, *25 (2)*, 913-931. https://doi.org/10.1039/d2cp04720a

(24) Tomasi, J.; Persico, M. Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. *Chemical Reviews* ***1994***, 2027-2094. https://pubs.acs.org/doi/pdf/10.1021/cr00031a013

(25) Barone, V.; Tomasi, J. A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. *Journal of Chemical Physics* ***1997***, 107, 3210-3221. https://doi.org/10.1063/1.474671

(26) Mennucci, B.; Tomasi, J. Continuum solvation models: A new approach to the problem of solute's charge distribution and cavity boundaries. *Journal of Chemical Physics* ***1997****, 106*, 5151-5158. https://doi.org/10.1063/1.473558

(27) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *Journal Physical Chemistry A* ***1998***, 102(11), 1995-2001.

(28) Wittmann L.; Selzer C.; Grimme S. A diverse and chemically relevant solvation model benchmark set with flexible molecules and conformer ensembles. *Chemical Science* ***2025***, *16 (48)*, 22976-22995. https://doi.org/10.1039/d5sc06406f

(29) Thompson, J.; Cramer, C.; Truhlar, D. New Universal Solvation Model and Comparison of the Accuracy of the SM5.42R, SM5.43R, C-PCM, D-PCM, and IEF-PCM Continuum Solvation Models for Aqueous and Organic Solvation Free Energies and for Vapor Pressures. *Journal of Physical Chemistry A* ***2004****, 108*, 6532-6542.

(30) Bell, I. H.; Mickoleit, E.; Chieh-Ming, H.; Shiang-Tai, L.; Vrabec, J.; Breitkopf, C.; Jäger, A. A Benchmark Open-Source Implementation of COSMO-SAC. *Journal of Chemical Theory and Computation* **2020**, *16 (4)*, 2635-2646. https://doi.org/10.1021/acs.jctc.9b01016

(31) Marenich, A.; Olson, R.; Kelly, C.; Cramer, C.; Truhlar, D. Self-Consistent Reaction Field Model for Aqueous and Nonaqueous Solutions Based on Accurate Polarized Partial Charges. *Journal of Chemical Theory Computational* ***2007****, 3*, 2011-2033. https://doi.org/10.1021/ct7001418

(32) Maldonado, A.; Hagiwara, S.; Choi, T. H.; Eckert, E.; Schwarz, K.; Sundararaman, R.; Otani, M.; Keith, J. A. *The Journal of Physical Chemistry A* ***2021***, *125 (1)*, 154-164. https://doi.org/10.1021/acs.jpca.0c08961

729 (33) Uhlig, H. The solubilities of gases and surface tension. *Journal of Physical Chemistry* ***1937****, 41*(9),
730 1215–1226. https://pubs.acs.org/doi/abs/10.1021/j150387a007

731 (34) Pierotti, R. A. The solubility of gases in liquids 1. *Journal of Physical Chemistry* ***1963****, 67*(9), 1840–
732 1845. https://pubs.acs.org/doi/pdf/10.1021/j100803a024

733 (35) Pierotti, R. A. Aqueous Solutions of Nonpolar Gases. *Journal of Physical Chemistry* ***1965****, 69*(1),
734 281–288. https://pubs.acs.org/doi/pdf/10.1021/j100885a043

735 (36) Pierotti, R. A. A Scaled Particle Theory of Aqueous and Nonaqueous Solutions. *Chemical Reviews*
736 ***1976****, 76*(6), 717. https://pubs.acs.org/doi/pdf/10.1021/cr60304a002

737 (37) Tomasi, J.; Floris, F. Evaluation of the dispersion contribution to the solvation energy. A simple
738 computational model in the continuum approximation. *Journal of Computational Chemistry* ***1989***, *10*(5),
739 616-627. https://doi.org/10.1002/jcc.540100504

740 (38) Amovilli, C. Calculation of the dispersion energy contribution to the solvation free energy. *Journal*
741 *of Chemical Physical Letters* ***1994***, *229*, 244-249. https://doi.org/10.1016/0009-2614(94)01044-7

742 (39) Amovilli, C.; Mennucci, B. Self-Consistent-Field Calculation of Pauli Repulsion and Dispersion
743 Contributions to the Solvation Free Energy in the Polarizable Continuum Model. *Journal of Physical*
744 *Chemistry B* ***1997****, 101(6)*, 1051-1057. https://doi.org/10.1021/jp9621991

745 (40) Sepassi, K.; Yalkowsky, S. Solubility prediction in octanol: A technical note. *AAPS Pharm. Sci.*
746 *Tech.* ***2006****, 7*(1), Article 26. https://doi.org/10.1208/pt070126

747 (41) Bode, B.; Gordon, M. Macmolplt: a graphical user interface for GAMESS. *Journal Molecular*
748 *Graphics and Modeling* ***1998****, 16(3)*, 133-138. https://doi.org/10.1016/S1093-3263(99)00002-9

(42) Vosko, S.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* ***1980***, *58*, 1200. https://doi.org/10.1139/p80-159

(43) Perdew, J.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical Review Letters* ***1996***, 77, 3865-3868. https://doi.org/10.1103/PhysRevLett.77.3865

(44) Šebesta, F.; Sovová, S.; Burda, J. V. Determination of Amino Acids' pKa: Importance of Cavity Scaling within Implicit Solvation Models and Choice of DFT Functionals. *The Journal of Physical Chemistry B* **2024,** *128 (7)*, 1627-1637. https://doi.org/10.1021/acs.jpcb.3c07007

(45) Ben-Naim, A.; Marcus, Y. Solvation thermodynamics of nonionic solutes. *Journal of Chemical Physics* ***1984****, 81*(4), 2016-2027. https://doi.org/10.1063/1.447824

(46) Silla1, E.; Tuñón, I.; Pascual-Ahuir, J. L. GEPOL: An improved description of molecular surfaces II. Computing the molecular area and volume. *Journal of Computational Chemistry* ***1991****, 12*(9), 1077–1088. https://doi.org/10.1002/jcc.540120905

(47) Pascual-Ahuir, J. L.; Tomasi, J.; Bonaccorsi, R. Electrostatic interaction of a solute with a continuum. Improved description of the cavity and of the surface cavity bound charge distribution. *Journal of Computational Chemistry* ***1986***, 778-787. https://doi.org/10.1002/jcc.540080605

(48) Lide, D.; *Handbook of chemistry and physics,* CRC press LLC: Boca ratón, Florida, 2000.

(49) Yaws, C.; Handbook of thermodynamic and properties of chemical compounds, McGraw-Hill: New York, 1997.

768 (50) Barca, G. M. J.; et al. Recent developments in the general atomic and molecular electronic structure
769 system. *The Journal of Chemical Physics* ***2020***, *152 (15)*, 154102. https://doi.org/10.1063/5.0005188

770 (51) Bensberg, M.; Türtscher, P.; Unsleber, J. P.; Reiher, M.; Neugebauer, J. *Journal of Chemical Theory*
771 *and Computation* ***2022*****,** 18 (2), 723-740. https://pubs.acs.org/doi/full/10.1021/acs.jctc.1c00864

772 (52) Rizzo, R. C.; Aynechi, T.; Case, D.; Kuntz, I. Estimation of Absolute Free Energies of Hydration
773 Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar
774 Contributions. *Journal of Chemical Theory and Computation* ***2006***, 2, 128-139.
775 https://doi.org/10.1021/ct050097l

776 (53) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Performance of SM6, SM8, and SMD on the SAMPL1
777 Test Set for the Prediction of Small-Molecule Solvation Free Energies. *Journal of Physical Chemistry B*
778 ***2009****, 113*(14), 4538–4543. https://doi.org/10.1021/jp809094y

779 (54) Wang, J.; Wang, W.; Huo, S.; Lee, M.; Kollman, P. A. Solvation Model Based on Weighted Solvent
780 Accessible Surface Area. *Journal of Physical Chemistry B* ***2001****, 105*, 5055-5067.
781 https://doi.org/10.1021/jp0102318

782 (55) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron
783 Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic
784 Surface Tensions. *Journal of Physical Chemistry B* ***2009,*** *113*(18), 6378–6396.
785 https://doi.org/10.1021/jp810292n

786 (56) Raevskya, O. A.; Grigoreva, V. J.; Raevskajaa, O. E.; Schaper, K. J. Physicochemical
787 properties/descriptors governing the solubility and partitioning of chemicals in water-solvent-gas systems.

Part 1. Partitioning between octanol and air. *SAR and QSAR in Environmental Research* ***2006****, 17*(3), 285-297. https://doi.org/10.1080/10659360600787742

(57) Hou, T. J.; Xu, X. J. ADME Evaluation in Drug Discovery. 2. Prediction of Partition Coefficient by Atom-Additive Approach Based on Atom-Weighted Solvent Accessible Surface Areas. *Journal Chemical Information and Computation Science* ***2003****, 43*, 1058-1067. https://doi.org/10.1021/ci034007m

TABLES

Table 1. vdW radii used in the GEPOL-GB method

Atom	vdW radii Å
C	1.70
N	1.60
O	1.50
H	1.20

Table 2. Parameter values used in the method Amovilli-Menucci and C-PCM

Parameter input	1-octanol	water
Dielectric constant (ε)	9.86	78.39
Molecular volume (cm^3/mol)	158.23	18.07
Surface tension (mN/m)	27.10	71.81
Ionization potential (Hartree)	0.332	0.396
Square of the refractive index	2.04	1.75

Table 3. Free energy electrostatic, repulsion and dispersion calculated using the methods C-PCM and Amovilli-Menucci.

	1-Octanol				Water			
	ΔGelec kJ mol^{-1}	ΔGdisp kJ mol^{-1}	ΔGrep kJ mol^{-1}	ΔGexp (solv.) kJ mol^{-1}	ΔGelec kJ mol^{-1}	ΔGdisp kJ mol^{-1}	ΔGrep kJ mol^{-1}	ΔGexp (solv.) kJ mol^{-1}
acetone	-16.32	-42.01	9.33	-13.18	-18.62	-37.99	11.63	-16.11
2-butanone	-15.02	-50.46	11.63	-15.82	-17.16	-45.61	14.52	-15.23
2-pentanone	-15.02	-59.46	14.14	-18.20	-17.16	-53.77	17.66	-14.77
3-pentanone	-13.93	-59.08	14.06	-18.24	-15.94	-53.43	17.57	-14.27
2-hexanone	-15.23	-68.29	16.53	-21.00	-17.41	-61.76	20.63	-13.77
1-propanal	-14.06	-42.26	9.41	-17.28	-16.03	-38.20	11.76	-14.39
1-butanal	-14.02	-51.17	11.97	-19.33	-15.94	-46.28	14.94	-13.31
acetic acid	-18.45	-38.49	6.61	-26.57	-20.71	-34.81	8.24	-28.03
propanoic acid	-17.74	-47.28	9.04	-28.70	-19.92	-42.76	11.26	-27.07
butanoic acid	-17.36	-56.11	11.38	-31.72	-19.50	-50.75	14.18	-26.61
pentanoic acid	-17.20	-64.98	13.98	-34.39	-19.29	-58.75	17.45	-25.77
hexanoic acid	-17.32	-73.73	16.19	-36.90	-19.41	-66.70	20.21	-25.98
benzene	-5.10	-51.42	11.59	-15.57	-5.73	-50.96	14.44	-3.64
toluene	-4.98	-59.96	13.93	-19.04	-5.56	-54.23	17.36	-3.72
ethyl benzene	-5.31	-67.91	16.40	-21.26	-5.94	-61.42	20.46	-3.35
methanol	-12.80	-26.57	5.61	-15.82	-14.35	-24.02	6.99	-21.38
ethanol	-12.39	-36.15	8.28	-18.24	-13.93	-32.68	10.33	-20.96
1-propanol	-11.97	-45.23	10.75	-21.00	-13.43	-40.88	13.43	-20.21
1-butanol	-12.26	-54.10	13.22	-23.89	-12.30	-53.60	16.53	-19.75
1-pentanol	-12.64	-63.01	15.57	-26.78	-14.18	-56.95	19.46	-18.70
1-hexanol	-12.68	-72.05	18.12	-29.54	-14.23	-65.15	22.59	-18.24
1-heptanol	-12.85	-80.96	20.54	-32.43	-14.44	-73.22	25.69	-17.74
2-propanol	-11.92	-44.44	10.88	-19.33	-13.43	-40.17	13.56	-19.92
1,2-ethanediol	-22.13	-42.26	7.99	-31.13	-24.85	-38.20	9.96	-38.91
dimethyl ether	-7.28	-35.65	8.66	-8.62	-8.20	-32.22	10.84	-8.03

diethyl ether	-6.28	-54.52	14.14	-12.09	-7.11	-49.33	17.66	-7.36

Table 3. (Continued)

	1-Octanol				Water			
	ΔGelec kJ mol^{-1}	ΔGdisp kJ mol^{-1}	ΔGrep kJ mol^{-1}	ΔGexp (solv.) kJ mol^{-1}	ΔGelec kJ mol^{-1}	ΔGdisp kJ mol^{-1}	ΔGrep kJ mol^{-1}	ΔGexp (solv.) kJ mol^{-1}
propyl methyl ether	-6.86	-54.06	13.93	-15.19	-7.74	-48.91	17.41	-6.95
methyl acetate	-14.94	-46.95	9.41	-14.81	-16.82	-42.47	11.72	-13.89
ethyl acetate	-15.06	-56.36	11.92	-16.99	-16.95	-50.96	14.90	-12.97
methyl propionate	-13.98	-55.61	11.80	-16.99	-15.73	-50.29	14.73	-12.26
methyl butanoate	-13.85	-64.86	14.31	-19.21	-15.61	-58.62	17.91	-11.84
cyclopentane	-0.54	-51.30	13.60	-11.09	-0.59	-46.40	16.95	5.02
cyclohexane	-0.46	-58.54	15.52	-14.48	-0.54	-52.93	19.37	5.15
methyl cyclohexane	-0.46	-67.11	18.28	-16.23	-0.50	-60.67	22.80	7.15
n-pentane	-0.67	-57.41	16.11	-10.25	-0.75	-51.93	20.13	9.75
n-hexane	-0.75	-66.11	18.28	-12.59	-0.84	-59.79	22.80	10.42
n-heptane	-0.84	-75.02	20.63	-15.65	-0.92	-67.83	25.73	10.96
n-octane	-0.79	-84.14	23.47	-17.49	-0.88	-76.11	29.29	12.09
n-propyl amine	-10.46	-46.65	11.97	-19.96	-11.76	-42.18	14.94	-17.53
n-butyl amine	-10.50	-55.52	14.48	-22.30	-11.80	-50.21	18.08	-17.95
diethyl amine	-6.40	-55.90	14.98	-19.87	-7.20	-50.55	18.70	-17.03
dipropyl amine	-6.44	-73.68	19.83	-25.19	-7.24	-66.65	24.77	-15.31

832 **Table 4.** σ_i and k_j coefficients obtained by multilinear regression for the estimation of the cavitation free

833 energy in 1-octanol and water.

k_j	1-octanol kJ mol^{-1}	water kJ mol^{-1}	σ_i	1-octanol kJ mol^{-1} Å^{-2}	water kJ mol^{-1} Å^{-2}
$k_1 = k_{aromatics}$	-10.45	-33.88	σ_{carbon}	0.44	1.08
$k_2 = k_{diol}$	0.20	7.83	σ_{oxygen}	0.12	-0.40
$k_3 = k_{ethers}$	2.43	-1.40	$\sigma_{hydrogen}$	0.07	-0.03
$k_4 = k_{ketona}$	8.04	1.22	$\sigma_{nitrogen}$	0.17	2.36
$k_5 = k_{cycloalkane}$	-2.64	-1.70			
$k_6 = k_{esters}$	9.67	8.75			
$k_7 = k_{aldehydes}$	3.12	2.29			
$k_8 = k_{amines}$	-2.56	-32.33			
$k_9 = k_{acids}$	-------	6.96			
b	9.02	3.84			
$RMSE$	0.67	1.25			

834

835

836

837

838

839

Table 5. Estimated free energies of solvation in water and 1-octanol for the 42 molecules studied.

	1-octanol			water		
	ΔGexp kJ mol^{-1}	ΔG(estimate) kJ mol^{-1}	relative error	ΔGexp kJ mol^{-1}	ΔG(estimate) kJ mol^{-1}	relative error
acetone	-13.18	-14.62	0.11	-16.11	-17.14	0.07
2-butanone	-15.82	-15.42	0.03	-15.23	-15.16	0.00
2-pentanone	-18.20	-18.26	0.00	-14.77	-14.67	0.01
3-pentanone	-18.24	-16.90	0.07	-14.27	-12.86	0.10
2-hexanone	-21.00	-21.24	0.01	-13.77	-14.32	0.03
1-propanal	-17.28	-16.98	0.02	-14.39	-14.25	0.01
1-butanal	-19.33	-19.63	0.02	-13.31	-13.45	0.01
acetic acid	-26.57	-27.04	0.02	-28.03	-28.68	0.03
propanoic acid	-28.70	-28.12	0.02	-27.07	-27.65	0.02
butanoic acid	-31.72	-31.58	0.00	-26.61	-26.38	0.01
pentanoic acid	-34.39	-34.11	0.01	-25.77	-25.51	0.01
hexanoic acid	-36.90	-37.10	0.01	-25.98	-25.26	0.03
benzene	-15.57	-16.22	0.04	-3.64	-7.35	1.03
toluene	-19.04	-18.73	0.02	-3.72	-2.19	0.41
ethyl benzene	-21.26	-20.91	0.01	-3.35	-1.16	0.67
methanol	-15.82	-15.70	0.02	-21.38	-20.55	0.02
ethanol	-18.24	-18.41	0.00	-20.96	-19.48	0.05
1-propanol	-21.00	-21.01	0.01	-20.21	-18.52	0.07
1-butanol	-23.89	-24.14	0.01	-19.75	-21.69	0.11
1-pentanol	-26.78	-26.40	0.02	-18.70	-18.79	0.01
1-hexanol	-29.54	-30.21	0.02	-18.24	-17.90	0.01
1-heptanol	-32.43	-33.21	0.02	-17.74	-17.54	0.01
2-propanol	-19.33	-19.96	0.03	-19.92	-17.95	0.10
1,2-ethanediol	-31.13	-31.13	0.00	-38.91	-38.90	0.01
dimethyl ether	-8.62	-13.40	0.01	-8.03	-6.05	0.14
diethyl ether	-12.09	-13.93	0.11	-7.36	-7.23	0.18

841 **Table 5.** (Continued)

	1-octanol			water		
	ΔGexp kJ mol^{-1}	ΔG(estimate) kJ mol^{-1}	relative error	ΔGexp kJ mol^{-1}	ΔG(estimate) kJ mol^{-1}	relative error
propyl methyl ether	-15.19	-14.27	0.08	-6.95	-14.93	0.04
methyl acetate	-14.81	-17.52	0.04	-13.89	-14.27	0.08
ethyl acetate	-16.99	-16.04	0.03	-12.97	-13.13	0.10
methyl propionate	-16.99	-20.16	0.06	-12.26	-8.64	0.07
methyl butanoate	-19.21	-11.01	0.05	-11.84	7.14	0.28
cyclopentane	-11.09	-14.15	0.01	5.02	4.86	0.41
cyclohexane	-14.48	-16.64	0.02	5.15	5.32	0.06
methyl cyclohexane	-16.23	-9.46	0.03	7.15	9.62	0.25
n-pentane	-10.25	-12.48	0.08	9.75	9.76	0.02
n-hexane	-12.59	-15.40	0.01	10.42	10.32	0.07
n-heptane	-15.65	-18.06	0.01	10.96	11.00	0.06
n-octane	-17.49	-21.31	0.04	12.09	-18.07	0.09
n-propyl amine	-19.96	-21.85	0.08	-17.53	-17.41	0.03
n-butyl amine	-22.30	-18.52	0.02	-17.95	-17.03	0.03
diethyl amine	-19.87	-14.62	0.05	-17.03	-17.14	0.02
dipropyl amine	-25.19	-15.42	0.02	-15.31	-15.16	0.02

842

843

844

845

846

847 **Table 6.** Estimated partition coefficients using the solvation free energies in water and 1-octanol.

	logP estimate	logP Exp.	relative error		logP estimate	logP Exp.	relative error
acetone	-0.44	-0.24	0.84	1-heptanol	2.75	2.62	0.05
2-butanone	0.05	0.29	0.84	2-propanol	0.36	0.05	6.14
2-pentanone	0.63	0.91	0.31	1,2-ethanediol	-1.36	-1.36	0.00
3-pentanone	0.71	0.99	0.28	dimethyl ether	-0.09	0.10	1.86
2-hexanone	1.21	1.38	0.12	ether diethyl	1.29	0.89	0.45
1-propanal	0.48	0.59	0.19	propyl methyl ether	1.17	1.21	0.03
1-butanal	1.08	0.88	0.23	methyl acetate	-0.11	0.18	1.64
acetic acid	-0.29	-0.17	0.68	ethyl acetate	0.57	0.73	0.22
propanoic acid	0.08	0.33	0.75	methyl propionate	0.51	0.82	0.38
butanoic acid	0.91	0.79	0.15	methyl butanoate	2.02	1.29	0.56
pentanoic acid	1.51	1.39	0.08	cyclopentane	3.18	3.00	0.06
hexanoic acid	2.08	1.92	0.08	cyclohexane	3.33	3.44	0.03
benzene	1.55	2.13	0.27	methylcyclohexane	3.85	3.61	0.07
toluene	2.90	2.73	0.06	n-pentane	3.34	3.39	0.01
ethyl benzene	3.46	3.15	0.10	n-hexane	3.90	3.90	0.00
methanol	-0.85	-0.77	0.10	n-heptane	4.51	4.66	0.03
ethanol	-0.19	-0.31	0.40	n-octane	5.09	5.18	0.02
1-propanol	0.44	0.25	0.74	n-propylamine	0.57	0.48	0.18
1-butanol	0.43	0.88	0.51	n-butylamine	0.78	0.97	0.20
1-pentanol	1.33	1.51	0.12	diethylamine	0.26	0.58	0.55
1-hexanol	2.16	2.03	0.06	dipropylamine	1.57	1.67	0.06

848

Table 7. σ_i and k_j coefficients obtained by multilinear regression for the estimation of the partition coefficients using the model A.

σ_i	$\dot{A}^{-2}$	k_j	
σ_{carbon}	0.0864	$k_{ketones}$	-1.4573
σ_{oxygen}	-0.0631	$k_{aldehydes}$	-0.9049
$\sigma_{nitrogen}$	0.1891	k_{acids}	0.1894
$\sigma_{hydrogen}$	-0.0028	$k_{aromatics}$	-3.1680
		$k_{cycloalkanes}$	-0.0152
Coefficients	*kJ/mol*	k_{ethers}	-1.2084
a	0.3286	k_{esters}	-0.7834
b	-0.0797	k_{amines}	-3.7888
		$k_{alcohols}$	-0.9449

865 **Table 8.** Estimated partition coefficients using the model A.

	logP Exp.	logP estimate	relative error		logP Exp.	logP estimate	relative error
acetone	-0.24	-0.24	0.01	2-propanol	0.05	0.25	4.09
2-butanone	0.29	0.35	0.19	2-butanol	0.61	0.81	0.33
2-pentanone	0.91	0.88	0.04	1,2-ethanediol	-1.36	-1.39	0.02
3-pentanone	0.99	0.93	0.06	1,3-propanediol	-1.04	-1.01	0.03
2-hexanone	1.38	1.41	0.02	1,4-butanediol	-0.83	-0.48	0.42
ethanal	-0.34	-0.21	0.39	dimethyl ether	0.10	-0.15	2.47
n-propanal	0.59	0.34	0.43	ether diethyl	0.89	1.02	0.15
n-butanal	0.88	0.88	0.00	ether dipropyl	2.03	2.24	0.10
n-pentanal	1.31	1.43	0.09	methyl propyl ether	1.21	1.11	0.08
acetic acid	-0.17	-0.23	0.37	methyl acetate	0.18	0.09	0.51
propanoic acid	0.33	0.35	0.06	ethyl acetate	0.73	0.67	0.08
butanoic acid	0.79	0.85	0.08	methyl propanoate	0.82	0.55	0.33
pentanoic acid	1.39	1.38	0.01	methyl butanoate	1.29	1.71	0.33
hexanoic acid	1.92	1.91	0.00	cyclopentane	3.00	2.99	0.00
benzene	2.13	2.14	0.00	cyclohexane	3.44	3.23	0.06
toluene	2.73	2.66	0.03	methylcyclohexane	3.61	3.83	0.06
ethylbenzene	3.15	3.19	0.01	n-pentane	3.39	3.48	0.03
propylbenzene	3.69	3.72	0.01	n-hexane	3.90	4.01	0.03
methanol	-0.77	-0.87	0.13	n-heptane	4.66	4.56	0.02
ethanol	-0.31	-0.30	0.02	n-octane	5.18	5.08	0.02
1-propanol	0.25	0.25	0.01	n-propylamine	0.48	0.47	0.01
1-butanol	0.88	0.87	0.01	n-butylamine	0.97	1.01	0.04
1-pentanol	1.51	1.33	0.12	n-pentylamine	1.49	1.45	0.02
1-hexanol	2.03	1.82	0.10	diethylamine	0.58	0.57	0.01

1-heptanol	2.62	2.35	0.10	dipropylamine	1.67	1.68	0.00

866

867 **Table 9.** σ_i and k_j coefficients obtained by multilinear regression for estimating the partition coefficients
868 using the model B.
869

σ_i	$\dot{A}^{-2}$	k_j	
σ_{carbon}	0.1067	$k_{ketones}$	-0.6333
σ_{oxygen}	-0.0347	$k_{aldehyds}$	--------
$\sigma_{nitrogen}$	0.2978	k_{acids}	1.0984
$\sigma_{hydrogen}$	0.0130	$k_{aromatics}$	-2.7048
coefficients	$\frac{mol}{Kcal}$	$k_{cycloalkanes}$	---------
		k_{ethers}	-0.8672
a	-0.0341	k_{esters}	---------
b	0.2667	k_{amines}	-4.0184
c	-0.2670	$k_{alcohols}$	--------

870

871 **Table 10.** Estimated partition coefficients using the model B.

	logP Exp.	logP estimate	relative error		logP Exp.	logP estimate	relative error
acetone	-0.24	-0.39	0.64	1,2-ethanediol	-1.36	-1.34	0.01
2-butanone	0.29	0.36	0.23	1,3-propanediol	-1.04	-1.50	0.44
2-pentanone	0.91	0.90	0.01	1,4-butanediol	-0.83	-0.93	0.13
3-pentanone	0.99	1.02	0.03	2-propanol	0.05	0.48	8.64
2-hexanone	1.38	1.45	0.05	2-butanol	0.61	0.96	0.57
ethanal	-0.34	-0.11	0.68	dimethyl ether	0.10	-0.12	2.21

n-propanal	0.59	0.46	0.22	ether diethyl	0.89	1.13	0.27
n-butanal	0.88	1.03	0.17	ether dipropyl	2.03	2.18	0.07
n-pentanal	1.31	1.61	0.23	methyl propyl ether	1.21	1.04	0.14
acetic acid	-0.17	-0.34	0.98	methyl acetate	0.18	0.01	0.94
propanoic acid	0.33	0.44	0.34	ethyl acetate	0.73	0.59	0.19
butanoic acid	0.79	0.83	0.05	methyl propanoate	0.82	0.63	0.23
pentanoic acid	1.39	1.39	0.00	methyl butanoate	1.29	1.53	0.19
hexanoic acid	1.92	1.93	0.00	cyclopentane	3.00	2.90	0.03
benzene	2.13	1.88	0.12	cyclohexane	3.44	3.13	0.09
toluene	2.73	2.72	0.00	methylcyclohexane	3.61	3.64	0.01
ethylbenzene	3.15	3.29	0.04	n-pentane	3.39	3.35	0.01
propylbenzene	3.69	3.81	0.03	n-hexane	3.90	3.88	0.01
methanol	-0.77	-0.74	0.04	n-heptane	4.66	4.44	0.05
ethanol	-0.31	-0.14	0.54	n-octane	5.18	4.99	0.04
1-propanol	0.25	0.43	0.73	n-propylamine	0.48	0.42	0.13
1-butanol	0.88	0.74	0.16	n-butylamine	0.97	0.98	0.01
1-pentanol	1.51	1.66	0.10	n-pentylamine	1.49	1.54	0.03
1-hexanol	2.03	2.02	0.00	diethylamine	0.58	0.58	0.00
1-heptanol	2.62	2.56	0.02	dipropylamine	1.67	1.67	0.00

872

873

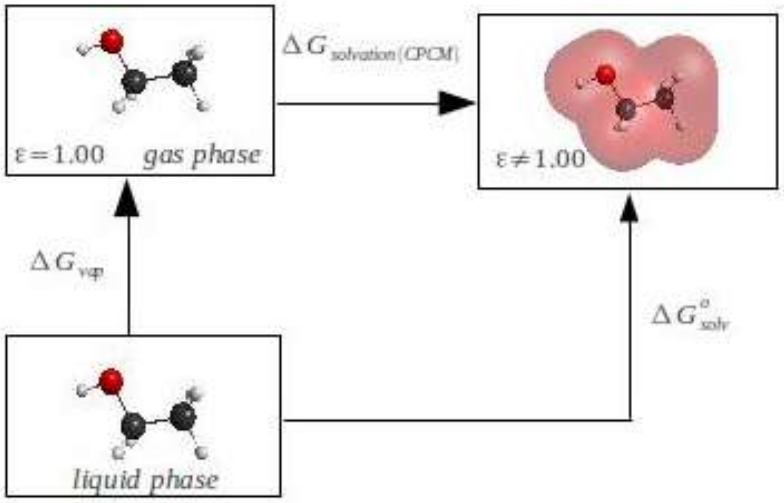
874

Table 11. Calculated electrostatic, dispersion, and repulsion free energies in 1-octanol and water, and predicted partition coefficients (logP) using Models A and B for ten additional organic molecules, compared with experimental values.

	1-octanol			water					
	ΔGelec (kJ mol⁻¹)	**ΔGdisp (kJ mol⁻¹)**	**ΔGrep (kJ mol⁻¹)**	**ΔGelec (kJ mol⁻¹)**	**ΔGdisp (kJ mol⁻¹)**	**ΔGrep (kJ mol⁻¹)**	**logP (Model A)**	**logP (Model B)**	**logP (Exp)**
2-pentanol	-11.05	-61.63	15.57	-12.43	-55.73	19.41	1.35	1.66	1.19
n-hexanal	-13.68	-68.54	16.74	-15.61	-61.97	20.92	1.96	2.17	1.78
hexylamine	-10.29	-73.43	19.37	-11.55	-66.40	24.18	4.98	2.08	2.06
dibutyl ether	-6.99	-90.04	23.68	-7.91	-81.42	29.54	3.27	4.01	3.21
propyl acetate	-14.90	-65.19	14.31	-16.82	-58.96	17.87	1.12	1.14	1.24
n-butylbenzene	-5.44	-85.65	21.13	-6.07	-77.45	26.40	4.26	4.38	4.38
dioxane	-13.10	-53.26	11.21	-14.69	-48.16	13.98	-0.14	-0.67	-0.27
anisole	-10.04	-64.81	14.02	-11.30	-58.62	17.49	2.03	1.88	2.11
glycerol	-27.11	-54.94	9.87	-30.34	-49.67	12.34	-2.02	-2.59	-1.76
3-pentanone	-14.31	-76.70	18.87	-16.40	-69.37	23.56	1.96	2.06	2.03

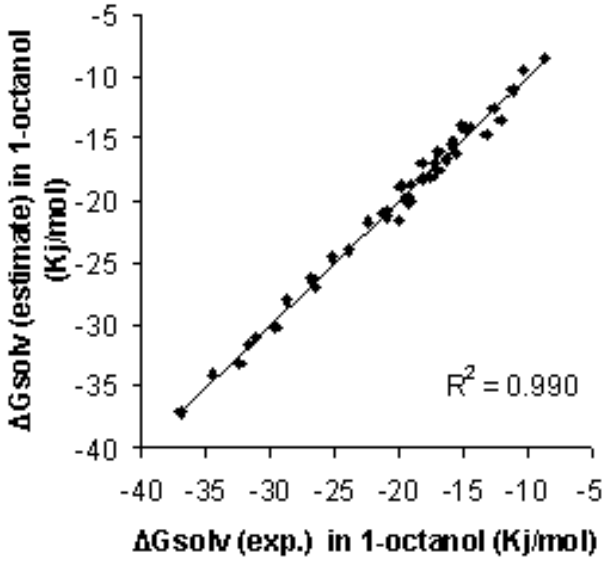
886 **FIGURES**

887



888

889 **Figure 1. Process of solvation of the ethanol in a continuum dielectric medium**

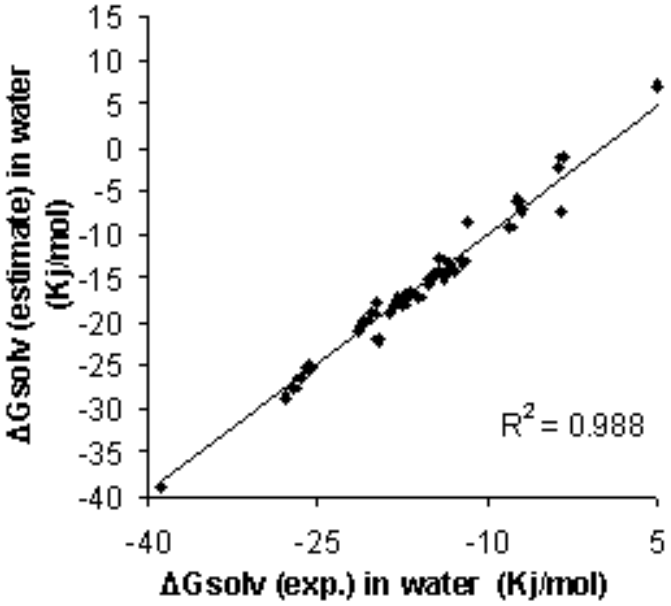


890

891 **Figure 2. Solvation free energy estimated in 1-octanol versus the experimental solvation free energy**

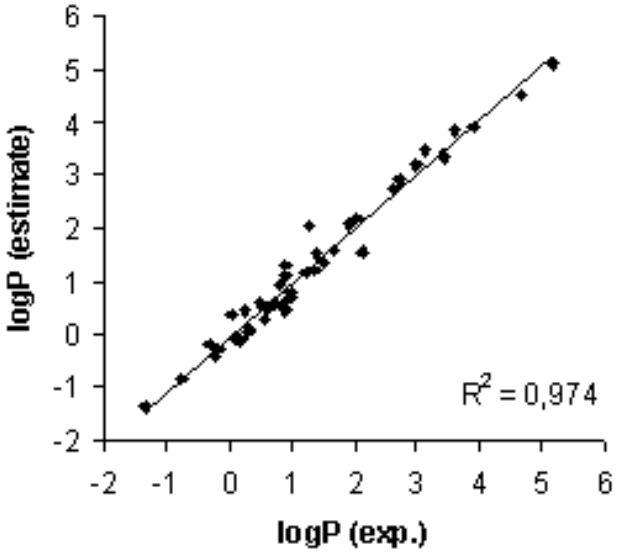
892

893



894

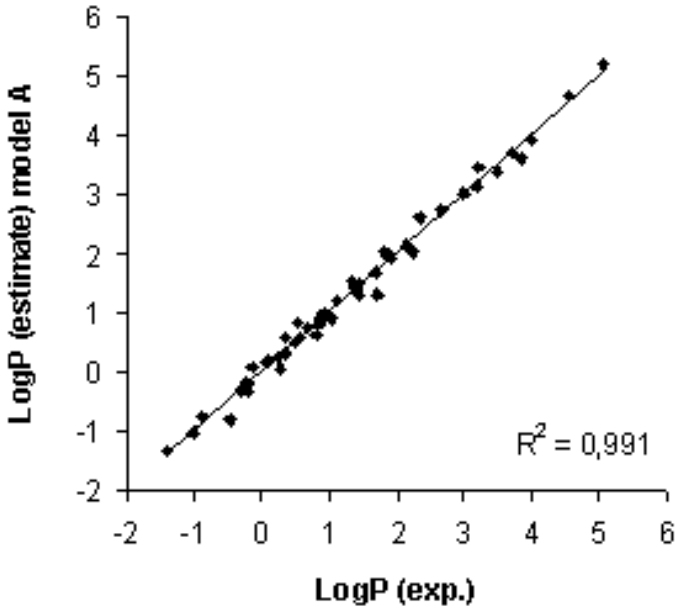
895 **Figure 3. Solvation free energy estimated in water versus the experimental solvation free energy**



896

897 **Figure 4. Partition coefficients estimated using the solvation free energy in 1-octanol and water**

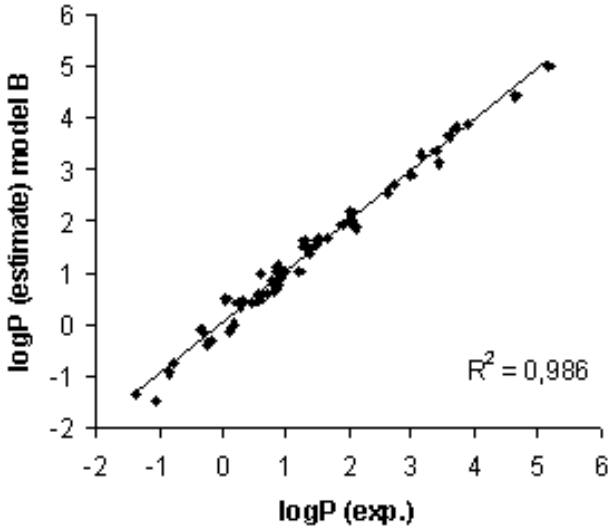
898



899

900 **Figure 5. Partition coefficients estimated using the model A**

901



902

903 **Figure 6. Partition coefficients estimated using the model B**

904