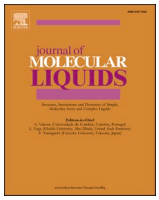
Journal of Molecular Liquids 422 (2025) 126907

Contents lists available at ScienceDirect

Journal of Molecular Liquids

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



Prediction of hydrogen-bonding interaction energies with new COSMO-based molecular descriptors

I. Zuburtikudis [a], W.E. Acree Jr [b], C. Panayiotou [c,*]

[a] Department of Chemical Engineering Abu Dhabi University Abu Dhabi United Arab Emirates
[b] Department of Chemistry, 1155 Union Circle # 305070, University of North Texas, Denton, TX 76203-5017, USA
[c] Department of Chemical Engineering, Aristotle University of Thessaloniki 54624 Thessaloniki, Greece

ARTICLE INFO

Keywords:
Solvation thermodynamics
Hydrogen-bonding
COSMO-RS
Sigma-profiles
Intramolecular association

ABSTRACT

Quantum chemical (QC) calculations and the Linear Solvation Energy Relationship (LSER) approach have been combined for the development of a simple method for the prediction of hydrogen-bonding interaction energies. Each hydrogen-bonded molecule is characterized by an acidity or proton donor capacity, α, and/or a basicity or proton acceptor capacity, β. When two molecules, 1 and 2, interact, their overall hydrogen-bonding interaction energy is $c(\alpha_1\beta_2 + \alpha_2\beta_1)$, where c is a universal constant equal to $2.303RT = 5.71$ kJ/mol at 25 ^{0}C. When the two molecules are identical, the self-association energy is $2c\alpha\beta$ and this is particularly useful for the development of the method. Molecular descriptors α and β are reported for several common hydrogen-bonded molecules but they may be obtained in a straightforward manner for any hydrogen-bonded molecule, even for those not synthesized yet. A high number of hydrogen-bonding interaction energies may be predicted already out of the reported descriptors. The descriptors α and β are heavily based on the molecular surface charge distributions already available or easily obtained via relatively cheap DFT / basis-set QC calculations. The new predictive scheme is particularly useful for solvation studies and equation-of-state developments in molecular thermodynamics. The capacity of the method to address the role of conformational changes on hydrogen bonding is discussed. The limitations of the method, especially its application to systems of complex solvent molecules with many distant hydrogen-bonding sites, are also discussed.

1. Introduction

We live in the era of artificial intelligence and machine – learning tools, with an unpreceded abundance of experimental and scientific information, databases and ever-increasing computational speed. Their impact on science is enormous as they are reshaping many branches and deepen our understanding of our environment and its material content. Protein structures are predicted much faster now than before (Nobel Prize in Chemistry 2024) with important implications in biomedical science and drug discovery [1]. Quantum chemical calculations and molecular simulations are becoming integral parts of molecular thermodynamics in practical applications [2]. Efficient parameterization of force fields based not on experimental data but on ab initio quantum chemical calculations [3] tend to become the game – changing practice in molecular simulations of thermophysical properties and processes. Thus, molecular thermodynamics itself (not the edifice of classical thermodynamics of laws) is one of the vividly reshaping branches of science today with a beneficial impact to a broad range of domains from industry to biology and everyday life.

Our understanding of the physicochemical behavior of pure substances or their mixtures is heavily based on our understanding and handling the intermolecular interactions. There are various ways interactions are perceived, handled, and classified. Intermolecular potentials of fixed or variable range are used to describe the van der Waals attractive and repulsive interactions [4,5]. Stronger interactions are treated in molecular dynamics simulations, mainly, as electrostatic interactions via polarizable or non-polarizable (fixed charges and pairwise additive interaction) force fields. The physical picture of the studied system emerges using these intermolecular potentials / force-fields in the radial distribution function (RDF) approach or the structure factors often determined experimentally via neutron or X-ray diffraction measurements [5,6]. Strong specific interactions of a directional character, however, like hydrogen bonding, are both important and peculiar [7–14].

* Corresponding author.
E-mail address: cpanayio@auth.gr (C. Panayiotou).

https://doi.org/10.1016/j.molliq.2025.126907
Received 10 October 2024; Received in revised form 22 December 2024; Accepted 8 January 2025
Available online 17 January 2025

The role of hydrogen bonding (HB) in numerous physicochemical processes in industry, in biology, in drug and xenobiotic interaction with biota, in aquatic environments or in life itself cannot be overemphasized [6–16]. HB is studied experimentally and theoretically, at least, since the time of Dolezalek [17] for more than one century now. HB interaction energies are obtained indirectly via modelling their contribution to measurable thermodynamic properties and, thus, they are attributed a quasi-thermodynamic character. Many attempts were made to model and quantify HB interactions for use in quantitative structure–property and structure – activity relationships (QSPR / QSAR) and in the development of molecular thermodynamic models of non-ideal mixtures [15,16,18–28].

Hydrogen bonding is, primarily, an electrostatic interaction between proton-donor and proton acceptor sites of opposite molecular surface charges within a range of geometrical constraints of distance and orientation / angle. Often, however, it is interpreted as a Lewis acid – base interaction or charge – transfer interaction or as electron donation from a highest occupied molecular orbital (HOMO) to a lowest unoccupied molecular orbital (LUMO). Thus, a variety of molecular descriptors from various levels of molecular theory or quantum chemical calculations have been used for the description of the hydrogen-bonding capacity of molecules, such as polar surface areas, surface charge densities, molecular electrostatic potentials, acidity and basicity descriptors, HOMO and LUMO, polarizabilities, etc. Such descriptors have been extensively used especially in QSPR/QSAR approaches [18–27,29–31]. In this regard, one of the most successful approaches for solvation, solubility, solute partitioning studies is Abraham's LSER (Linear Solvation Energy Relationships) model [32–39].

In Abraham's LSER model, the HB contribution to solvation energy of a solute 1 by solvent 2 is modelled as the sum, $a_2A_1 + b_2B_1$, where A_1 and B_1 are the acidity and basicity molecular descriptors of the solute while a_2 and b_2 are the complementary basicity and acidity solvent-specific coefficients, respectively. There are two main limitations of this otherwise highly successful model. The descriptors A and B and the coefficients a and b are essentially obtained from extensive experimental data correlations and depend on their availability. In fact, for just a few decades of solvents are a and b coefficients available [37–39]. The second limitation arises from the very way this LSER model treats hydrogen bonding. On self-solvation, where solute and solvent become identical, one would expect the acid – base (aA) interaction to be identical to the base – acid (bB) interaction between the very same pair of donor – acceptor sites. However, in LSER the product aA is in general different from bB [40–42]. In fact, this is the case for other popular QSPR – type models, including Raevski's HYBOT model [31]. This restricts severely the transfer of this HB information into other models, especially in Molecular Thermodynamics modeling of HB.

In Molecular Thermodynamics, various approaches are used to account for the typically significant contribution of HB interactions to mixture non-idealities. In the "Chemical Theory" approach, the hydrogen bond formation is treated as a chemical reaction and is attributed an energy and a free energy of formation and the mass law is applied at equilibrium [43,44]. The family of SAFT (statistical associating – fluid theory) models [45–49] uses Wertheim's first order thermodynamic perturbation theory (TPT1) [50,51], and treats HB with a square-well intermolecular potential, the depth of which reflects the strength of the HB interaction energy. The quantum – mechanics based COSMO-RS (conductor screening model for real solvents or realistic solvation) model [52–57] treats HB as an electrostatic interaction, which is handled as a nearest-neighbor pair interaction energy in a statistical – thermodynamic framework similar to Guggenheim's Quasi-Chemical theory [42,58–60]. In the NRHB (non-randomness and hydrogen bonding) equation-of-state approach, the hydrogen bonding interaction is also attributed an energy and a free energy of (bond) formation and is handled as a nearest neighbor pair interaction energy with Veytsman's combinatorial statistics [61–65]. In all these approaches a given HB interaction is attributed a given strength independent of mixture composition.

In contrast to this practice, very often, the HB contribution to solvation energy is considered composition dependent. This is particularly pronounced in weaker hydrogen bonds where polarizability / induction effects are more difficult to separate from a "net" HB electrostatic interaction. This difficulty is one of the main reasons for the observed discrepancies in HB interaction energies in the literature [7,8,13,14,18–49]. Thus, significant effort is required for the reliable transfer of thermodynamically consistent information between the various HB approaches and their acidity / basicity scales. In this regard, much effort is focused nowadays on the extraction of valuable information on HB interactions from advanced quantum chemical calculations and molecular simulations of HB dependent measurable thermodynamic / thermophysical properties, phase equilibria or solvation quantities. Out of these older and newer experimental and theoretical studies, numerous extensive compilations and databases of HB-related quantities have been produced and are already available [7,8,14,24,25,28,31,39,66–69].

Yet, despite the immense effort devoted to their study, HB interactions resist a universally accepted precise definition and quantitatively accurate measurement. It may sound surprising but there is not even a single HB interaction energy whose accurate value is widely considered to be known. In fact, the discrepancies in literature values of energies even for very common hydrogen bonds, such as the self-association of water or ethanol, are often larger than various thermal energy units. It is easy to find in the open literature interaction energies for a given hydrogen bond with relative discrepancies well over 50 % [13,14,24–28,31,38–43]. Even in advanced thermodynamic models like the SAFT models, the relevant discrepancies in the depth, ε, of square-well HB interaction potential are quite significant and often well over 50 % [28].

Thus, the need arises for the development of a reliable and robust predictive scheme for HB interaction energies, which could be used not only in QSAR/QSPR-type approaches but also in Molecular Thermodynamic modeling of hydrogen-bonded fluids and their mixtures. From the above exposition, it is clear that such a task is not trivial, and some compromises have to be made, especially, in the case of weak HB interactions. Focusing then on the stronger electrostatic HB interactions, significant help may be obtained from quantum chemical calculations, especially when implemented in a statistical thermodynamic framework, as is the case of COSMO-RS model [52–57].

There are four important features, which make COSMO-RS a good basis for an insightful discussion of HB and the development of a robust predictive scheme for HB interaction energies. The first is its very quantum-chemical basis, namely, the COSMO solvation model, out of which the molecular surface charge distribution or *sigma (σ)- profile* is obtained for each molecule at a sufficiently high level of density – functional theory (DFT) / basis-set. The second is its solid statistical-thermodynamic basis, which is in essence the Guggenheim's Quasi-Chemical theory [42,58,59]. The third is the clear way it separates and treats HB as an electrostatic interaction, which permits thermodynamic calculations in a consistent and straightforward manner. Consequently, COSMO-RS does provide a "net" HB contribution to solvation energy. The fourth is its quantum – mechanics – based account of conformer population for each molecule. The prevailing conformer population (dictated by free energy minimization and a Boltzmann weighing of conformer contributions) may change with mixture composition and all properties based on conformer structure, such as the strength of HB interaction, may also change with composition. What does not change with composition is the HB strength of interaction between a given conformer of the solute with a given conformer of the solvent.

This HB picture from COSMO-RS may not be ideally suited for the very weak hydrogen bonds mentioned above but this may not be a significant handicap in practical applications of the thermodynamics of non-ideal systems: They may be treated sufficiently well as the rest of the

weaker polarity / polarizability interactions [28]. On the other hand, COSMO-RS is not a perfect model. It is also parameterized against experimental data (though with very few parameters) and some empiricism does exist in it as well [52–57]. Thus, the COSMO-RS estimated HB contribution to solvation energy could be corroborated or even substituted by the corresponding estimation by some other sufficiently reliable predictive method, such as the Abraham's LSER model [32–39]. Estimations, then, from the developed predictive scheme could be compared with corresponding estimations from, both, the COSMO-RS and the LSER model. Such a predictive scheme would require, of course, an appropriate framework and/or HB descriptors for its development.

Recently [70], a novel set of molecular descriptors, based on molecular surface charge densities or σ – profiles, was reported, which is well suited for the development of the sought predictive scheme. Two of these descriptors, the HB acidity, A_h, and the HB basicity, B_h, are particularly suited for the sought predictive scheme due to their sound basis and insightful character [70].

Thus, the central objective of the present work is the development of a reliable, simple and robust predictive method for the HB interaction energies based on relatively cheap COSMO-type quantum chemical (QC) calculations. The new predictive scheme is discussed in the next section after recalling the essentials of the new QC-LSER molecular descriptors [70]. Extensive calculations with the new method for a variety of solute / solvent systems are reported in the Supplementary Information (SI) file. The role of conformers in HB calculations is discussed. The limitations and the perspectives for extension of the new method for the prediction of HB interaction free energies and entropies are also discussed.

2. The new predictive method for HB interaction energies

This section is divided in two sub-sections. In the first sub-section, the essentials and the definitions of the new QC-LSER descriptors are recalled. Details may be found in ref. [70]. Newly derived QC-LSER descriptors for a number of hydrogen-bonded solutes are reported in Table S1 of the Supplementary Information (SI) file. In the second subsection, the new predictive method is developed. The rationale for the derivation of the acidity, α, and basicity, β, parameters from the corresponding A_h and B_h QC-LSER descriptors of the molecule and their implementation into a simple predictive scheme are presented.

2.1. The quantum-chemistry based LSER (QC-LSER) molecular descriptors

As mentioned in the Introduction and as discussed in ref. [70], the new QC-LSER descriptors were, primarily, derived in order to remove some of the limitations of Abraham's LSER descriptors S_1, A_1, B_1, for the polarity/polarizability, acidity and basicity, respectively, of solute (1) molecules. These LSER descriptors, although attributed a more or less clear physical meaning, are determined, in essence, via global optimization of experimental data correlations and, thus, they depend on data availability [32–39]. In contrast, the new QC-LSER descriptors were obtained solely from the molecular structure of the solute and the COSMO-based molecular surface charge distributions or σ-profiles [52–57,69]. The derivation of the new descriptors was part of a broader goal for the consistent reformulation of the widely used linear relationships that would permit the extraction of reliable information on intermolecular interactions, especially, hydrogen bonding interactions [41,42].

It is recalled that in the quantum chemical COSMO solvation approach, a σ-profile is the distribution of probabilities $P_i(\sigma_i) = A_i/A$ to find a surface charge density σ_i on the surface (COSMO solvation / screening cavity) of the studied molecule [52–57]. The sum of the partial surface areas A_i gives the total surface area, A, of the molecular cavity, while the sum of the partial charges $\sigma_i A_i$ gives the total charge of the molecule or ion. Molecular descriptors or *COSMOments* M_i^X of type i, were derived by Klamt [53] from the σ-profile $A^X(\sigma)$ of a solute X, as the integral $\int A^X(\sigma) f_i(\sigma) d\sigma$, for various functional forms $f_i(\sigma)$. The new QC-LSER descriptors were derived as such COSMOments with functional forms that would produce descriptors reflecting the strength of intermolecular forces. In this regard, the product of the surface charge density σ_i by the charge $\sigma_i A_i$ or the product $\sigma_i^2 A_i$ (being charge by charge over unit area, with units C^2m^{-2} or $e^2Å^{-2}$) appears appropriate for reflecting an intermolecular interaction or an electrostatic force. An example of the distribution function of this product is shown in Fig. 1 derived from the σ-profile of methanol [69]. As expected, the usual central dome above zero charge density of the σ-profiles (cf. Fig. 2), becomes a valley in Fig. 1 touching the ground zero of electrostatic interactions.

In COSMO-RS [53], (screening) charge densities lower than −0.01 $eÅ^{-2}$ (far left blue region in Fig. 1) correspond to acidic or donor sites able to participate in strong specific or hydrogen-bonding interactions (HBD sites). Similarly, charge densities higher than 0.01 $eÅ^{-2}$ (far right red region in Fig. 1) correspond to basic or acceptor (HBA) sites. The molecular sites with charge densities between these two cutoff lines (−0.01 and 0.01 $eÅ^{-2}$) interact with much weaker forces, primarily, dispersion / induction / polarizability forces. They may also interact via polarity/ polarizability forces with sites beyond the two cutoff lines when the latter are not involved in hydrogen bonding interactions. The two cutoff lines and the zero-charge density divide the total area under the distribution curve of Fig. 1 in four sub-areas. The two areas pertaining to hydrogen bonding sites are marked by the subscript h (A_h the acidic or HBD sites and B_h the basic or HBA sites). The two central "polarizability" sub-areas are marked by the subscript p (A_p the sites to the left with screening charge sign negative and B_p the sites towards the basic sites of positive screening charge sign). The four new QC-LSER descriptors are just the areas of these four sub-sections in Fig. 1. Thus, they may be obtained via a simple integration as follows (the factor 10^6 accounts for the units of σ and A in σ-profiles):

$$A_h = 10^6 \int_{-\infty}^{-0.01} A_i \sigma_i^2 d\sigma_i \quad (1)$$

$$A_p = 10^6 \int_{-0.01}^{0} A_i \sigma_i^2 d\sigma_i \quad (2)$$

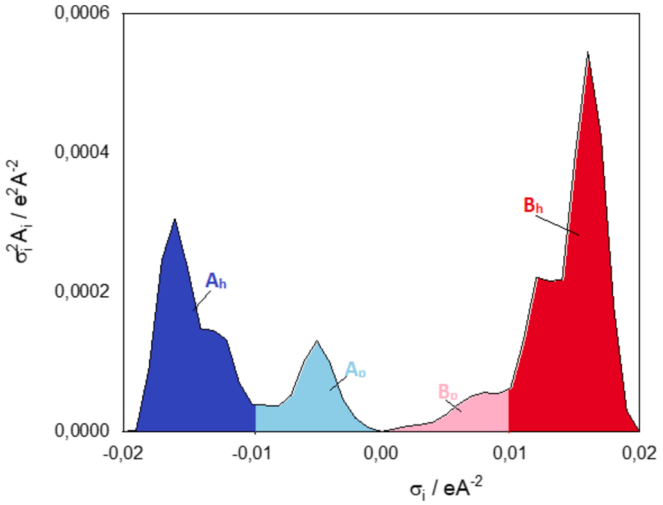


Fig. 1. The distribution function of the product $\sigma_i^2 A_i$ of charge densities σ_i with the molecular charges $\sigma_i A_i$ of Methanol and the definition areas of the QC-LSER molecular descriptors. The four surface areas under the distribution curve, multiplied by 10^6, give the four QC-LSER descriptors of Methanol (cf. Table S1): $A_h = 1.39$, $A_p = 0.55$, $B_p = 0.29$, $B_h = 2.39$.

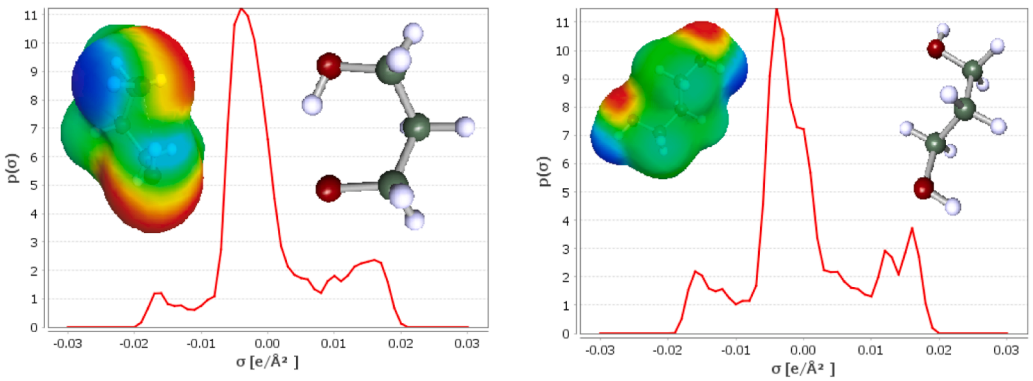


Fig. 2. The σ-profiles, σ-surfaces and HB descriptors of conformers c0 and c3 of 1,3-propylene glycol. Data from [69].

$$B_p = 10^6 \int_0^{0.01} A_i \sigma_i^2 d\sigma_i \quad (3)$$

$$B_h = 10^6 \int_{0.01}^{\infty} A_i \sigma_i^2 d\sigma_i \quad (4)$$

σ- Profiles are available, free of charge, for thousands of molecules in the open literature, as an example in ref. [71]. They may, of course, be obtained also by using appropriate quantum-chemical calculation suites, such as the TURBOMOLE, DMol^3 of BIOVIA's MATERIALS STUDIO suite, or the SCM suite [72–74]. In this work, the σ-profiles from COSMObase [69] at the DFT / TZPVD-Fine level of quantum chemical calculations are used and the corresponding QC- LSER descriptors for common hydrogen bonded solutes are reported in Table S1.

The descriptors A_h and B_h are the analogues of the A and B descriptors of the Abraham LSER model [32–39] and are considered reflecting the HBD and the HBA capacity, respectively, of the molecule. The descriptors A_p and B_p do not have direct analogues in the Abraham's LSER model but they may be used to substitute the LSER descriptor S [70].

One significant feature of the new QC-LSER descriptors, not encountered in Abraham's LSER descriptors, is their capacity to account for conformational changes of the solute upon transfer to the solvent. As seen in Table S1 and in Table 1 below, several conformers are reported for some solutes. As an example, four conformers are reported for 1,3 – propylene glycol (more conformers are reported in ref. [69]). The σ-profiles and sigma-surfaces of two of these conformers are shown in Fig. 2. As seen in the figure, the conformer c0 of 1,3-propylene glycol has a significantly reduced HBD capacity compared to conformer c3 (acidity peaks on the left-hand side of σ-profiles). Conformer c0 is hydrogen-bonded intra-molecularly and, thus, the proton donor is not much available for intermolecular association. There is an energy cost in changing conformation from c0 to c3, reported in Table 1 and S1. However, this cost is offset by the reduction in energy when conformer c3 participates in intermolecular hydrogen bonding. Having the QC-LSER descriptors, we may now focus on their use for the prediction of HB interaction energies.

2.2. The predictive method for HB interaction energies

In ref. [70] the following equation was examined for the contribution of hydrogen bonding to the solvation energy of solute 1 by solvent 2:

$$\frac{-\Delta E_{12}^{hb}}{2.303RT} = c_{h12}(A_{h1}B_{h2} + B_{h1}A_{h2}) \quad (5)$$

where, ΔE_{12}^{hb} is the hydrogen-bonding interaction energy of solute 1 by solvent 2, and c_{h12} is a characteristic constant of the 1 – 2 pair. This equation meets the consistency criterion when the linear relationship is applied to self-solvation (when solute and solvent become identical, the right-hand side of equation (5) becomes $2c_{h12}A_hB_h$ as it should do). Equation (5) was applied to a variety of systems for which the corresponding LSER and COSMO-RS estimations were known. It was observed that the c_{h12} coefficient takes nearly common values for homologous series of solutes and "distinguishes" their type. This observation was the starting point of the present work. Before proceeding, we should point out that in the present work we confine ourselves to calculations at ambient conditions where interaction "energy" and interaction "enthalpy" are practically equivalent and the corresponding terms may be used interchangeably.

In order to draw safer conclusions, a large variety of solute types in a variety of solvents were examined and the values for the c_{h12} coefficients were obtained. On self-solvation, a nearly unique value c_h was obtained for each solute from the various sources of experimental or theoretical estimations with a minimal deviation. This c_h coefficient was a combined contribution from both the acidity and the basicity descriptors. In general, then, the following equation could be written for self-solvation:

$$2c_hA_hB_h = 2(f_AA_h)(f_BB_h) \quad \text{(self - solvation)} \quad (6)$$

The key question now is whether the factors f_A and f_B are unique parameters of the solute and independent of the solvent. If this were the case, then, the HB contribution data for the self- solvation of solute 1 would permit the expression of f_A in terms of f_B. or $f_A = g(f_B)$. In this way, each solute could be characterized uniquely by one factor (either f_A or f_B). Combining this with the above-mentioned observation in ref. [70], one single pair of f_A and f_B were expected for a homologous series, say, 1-alkanols. Indeed, this was verified by numerous data in solvents ethanol, 1-butanol and 1-octanol. From these data, the pair, $f_A = 0.78$ and $f_B = 0.93 \pm 0.01$ was obtained for all 1-alkanols. The uniqueness of the pair was further verified with solvation data of 1-alkanols by water. From this exercise, the pair $f_A = 0.70$ and $f_B = 0.59 \pm 0.01$ was obtained for water. With these two pairs of factors known, it was rather straightforward to determine the corresponding factors of all other solutes using their solvation data.

But, if a solute has a unique set of f_A and A_h values and / or a unique set of f_B and B_h values, it is more convenient to report their unique products and use them as the effective acidity, α, and basicity, β, descriptors, or

Table 1
The QC-LSER Molecular Descriptors of Common Hydrogen-Bonded Solutes. Conformational energies are given in kJ/mol.

SOLUTE	A_h	B_h	f_A	f_B	α	β
1-PENTYNE	0.49	0.16	0.78	1.00	0.38	0.16
1-HEXYNE	0.48	0.17	0.78	1.00	0.38	0.17
3-HEXYNE	0.00	0.34	0.00	1.00	0.00	0.34
DICHLOROMETHANE	0.94	0.00	0.50	0.00	0.47	0.00
CHLOROFORM	1.14	0.00	0.50	0.00	0.57	0.00
DIETHYL ETHER	0.00	1.81	0.00	1.00	0.00	1.81
DI-n-PROPYL ETHER	0.00	1.79	0.00	1.00	0.00	1.79
DIISOPROPYL ETHER	0.00	1.74	0.00	1.00	0.00	1.74
DI-n-BUTYL ETHER	0.00	1.79	0.00	1.00	0.00	1.79
DI-n-PENTYL ETHER	0.00	1.71	0.00	1.00	0.00	1.71
12-CROWN-4	0.00	5.44	0.00	0.93	0.00	5.06
15-CROWN-5	0.00	7.89	0.00	0.93	0.00	7.34
DIGLYME	0.00	5.13	0.00	0.93	0.00	4.77
TRIGLYME	0.00	5.24	0.00	0.93	0.00	4.87
FURAN	0.24	0.13	0.78	1.00	0.18	0.13
TETRAHYDROFURAN	0.00	2.15	0.00	1.00	0.00	2.15
1,3-DIOXANE	0.00	3.12	0.00	1.00	0.00	3.12
1,4-DIOXANE	0.00	3.12	0.00	1.00	0.00	3.12
TETRAHYDROPYRAN	0.00	1.97	0.00	1.00	0.00	1.97
METHYL FORMATE	0.19	2.21	0.00	0.55	0.00	1.22
ETHYL FORMATE	0.15	2.36	0.00	0.55	0.00	1.30
ETHYL ACETATE	0.00	2.81	0.00	0.55	0.00	1.55
n-PROPYL ACETATE	0.00	2.80	0.00	0.55	0.00	1.54
ISOPROPYL ACETATE	0.00	2.68	0.00	0.55	0.00	1.47
n-BUTYL ACETATE	0.00	2.81	0.00	0.55	0.00	1.55
METHYL PROPIONATE	0.00	2.55	0.00	0.55	0.00	1.40
ETHYL PROPIONATE	0.00	2.66	0.00	0.55	0.00	1.46
ETHYL n-BUTYRATE	0.00	2.65	0.00	0.55	0.00	1.46
n-PROPYL PROPIONATE	0.00	2.69	0.00	0.55	0.00	1.48

(*continued on next page*)

$$\alpha = f_A A_h$$
$$\beta = f_B B_h \quad (7)$$

The factors f_A and f_B and the descriptors α and β are reported in Table 1 for a number of common hydrogen-bonded solutes. With these descriptors, the HB interaction energy for any solute 1 – solvent 2 pair may be obtained at 25 ^{0}C by the simple equation:

$$-\Delta E_{12}^{hb} = 2.303RT(\alpha_1\beta_2 + \beta_1\alpha_2) = 2.303 \times 8.314 \times 298.15(\alpha_1\beta_2 + \beta_1\alpha_2) = \\ = 5.71(\alpha_1\beta_2 + \beta_1\alpha_2) \ \ kJ/mol \quad (8)$$

Equation (8) was extensively tested as a predictive tool and the results are discussed in the next section. Before discussing the predictions, it is worth making one comment on Table 1.

One of the striking features of Table 1 is the value of f_A, which seems to be the same for most of the studied solutes with proton-donor capacity. It is rather precarious to attribute a physical meaning to this most common value of $f_A = 0.78$ of the proton-donor sites. In any case, this common value has facilitated the preparation of Table 1 very much.

To an extent, Table 1 is structured in groups of homologous series. It is useful and, probably, insightful to notice the common f_A and f_B factors for the members of each homologous series. Any eventual departure from this rule may point to the impact of some structural or other factors. The rule, however, may be used for a first guess of the f_A and f_B factors for a new member of a homologous series. As an example, a good guess for the f_B factor of di-n-hexyl ether would be $f_B = 1.00$, while a

Table 1 (*continued*)

ACETONE	0.02	2.95	0.00	0.55	**0.00**	**1.62**
METHYL ETHYL KETONE	0.00	2.78	0.00	0.55	**0.00**	**1.53**
2-PENTANONE	0.00	2.81	0.00	0.55	**0.00**	**1.55**
3-PENTANONE	0.00	2.58	0.00	0.55	**0.00**	**1.42**
CYCLOPENTANONE	0.00	3.02	0.00	0.55	**0.00**	**1.66**
2-HEXANONE	0.00	2.82	0.00	0.55	**0.00**	**1.55**
3-HEXANONE	0.00	2.62	0.00	0.55	**0.00**	**1.44**
2-HEPTANONE	0.00	2.84	0.00	0.55	**0.00**	**1.56**
2-OCTANONE	0.00	2.82	0.00	0.55	**0.00**	**1.55**
2-NONANONE	0.00	2.80	0.00	0.55	**0.00**	**1.54**
ACETOPHENONE	0.01	2.44	0.00	0.55	**0.00**	**1.34**
FORMALDEHYDE	0.00	1.55	0.00	0.50	**0.00**	**0.78**
ACETALDEHYDE	0.00	2.36	0.00	0.50	**0.00**	**1.18**
PROPANAL	0.00	2.33	0.00	0.50	**0.00**	**1.17**
BUTANAL	0.00	2.32	0.00	0.50	**0.00**	**1.16**
PENTANAL	0.00	2.33	0.00	0.50	**0.00**	**1.17**
HEXANAL	0.00	2.32	0.00	0.50	**0.00**	**1.16**
OCTANAL	0.00	2.33	0.00	0.50	**0.00**	**1.17**
BENZALDEHYDE	0.01	2.20	0.00	0.50	**0.00**	**1.10**
METHANOL	1.39	2.39	0.78	0.93	**1.08**	**2.22**
ETHANOL	1.23	2.46	0.78	0.93	**0.96**	**2.29**
1-PROPANOL	1.23	2.44	0.78	0.93	**0.96**	**2.27**
1-BUTANOL	1.22	2.47	0.78	0.93	**0.95**	**2.30**
1-PENTANOL	1.22	2.46	0.78	0.93	**0.95**	**2.29**
1-HEXANOL	1.22	2.48	0.78	0.93	**0.95**	**2.31**
1-HEPTANOL	1.21	2.48	0.78	0.93	**0.94**	**2.31**
1-OCTANOL	1.19	2.46	0.78	0.93	**0.93**	**2.29**
1-NONANOL	1.21	2.47	0.78	0.93	**0.94**	**2.30**
1-DECANOL	1.23	2.47	0.78	0.93	**0.96**	**2.30**
ISOPROPANOL	1.18	2.52	0.78	0.95	**0.92**	**2.39**
2-METHYL-1-PROPANOL	1.20	2.22	0.78	0.93	**0.94**	**2.06**
2-METHYL-2-PROPANOL	1.04	2.49	0.78	0.95	**0.81**	**2.37**
2-BUTANOL	0.91	2.46	0.78	1.00	**0.90**	**2.21**
2-METHYL-1-BUTANOL	1.19	2.19	0.78	0.93	**0.93**	**2.04**
3-METHYL-1-BUTANOL	1.20	2.47	0.78	0.93	**0.94**	**2.30**
2-METHYL-2-BUTANOL	1.00	2.24	0.78	1.00	**0.78**	**2.24**
2-PENTANOL	1.14	2.22	0.78	1.00	**0.89**	**2.22**
4-METHYL-2-PENTANOL	1.15	2.17	0.78	1.00	**0.90**	**2.17**

(*continued on next page*)

good guess for the same factor for heptanal would be the value of 0.50. If then, the σ-profiles of these new members are available, good guesses for their basicity descriptors β will become also available.

On the other hand, it is rather the 3D structure of the molecule which dictates the accessibility or availability of its HB sites and has an impact on the values of its f_A and f_B factors. It is not certain if closer inspection of the graphical representation of the exposure of the HB loci, like Figs. 2 and 3, could help visualize this impact. Nevertheless, in contrast to Fig. 2, the proton-acceptor loci in Fig. 3 seem to be more accessible for hydrogen bonding and the corresponding f_B factor is 1.00 like in ethers. Apparently, the presence of the ether oxygen in the structure of Fig. 3 has an influence on its f_B factor and has raised its value in comparison to the f_B factor of the structure in Fig. 2 (f_B equal to 0.93). Safe conclusions, however, relating the values of f_A and f_B factors to the molecular structure requires much deeper analysis and exceeds the scope of the present work.

Table 1 (*continued*)

2-HEXANOL	1.15	2.24	0.78	1.00	**0.90**	**2.24**
2-ETHYL-1-HEXANOL	1.18	2.18	0.78	0.93	**0.92**	**2.03**
1-ADAMANTANOL	1.00	2.56	0.78	1.00	**0.78**	**2.56**
CYCLOPENTANOL	1.16	2.29	0.78	0.93	**0.90**	**2.13**
CYCLOHEXANOL	1.13	2.55	0.78	0.93	**0.88**	**2.37**
1-NAPHTHOL	2.12	0.50	0.78	0.93	**1.65**	**0.47**
GLYCEROL	2.98	5.26	0.78	0.93	**2.32**	**4.89**
ETHYLENE GLYCOL, c0	2.13	3.87	0.78	0.87	**1.66**	**3.37**
ETHYLENE GLYCOL c1	2.16	3.68	0.78	0.87	**1.68**	**3.20**
ETHYLENE GLYCOL c3	2.83	4.89	0.78	0.87	**2.21**	**4.25**
ETHYLENE GLYCOL, c6 E_{conf} = 8.98	2.97	4.11	0.78	0.87	**2.32**	**3.57**
1,3-PROPYLENE GLYCOL, c0	1.58	3.92	0.78	0.93	**1.23**	**3.65**
1,3-PROPYLENE GLYCOL c1	1.62	3.73	0.78	0.93	**1.26**	**3.47**
1,3-PROPYLENE GLYCOL, c3 E_{conf} = 5.54	2.63	4.38	0.78	0.93	**2.05**	**4.07**
1,3-PROPYLENE GLYCOL c6	2.38	4.31	0.78	0.93	**1.86**	**4.01**
DIETHYLENE GLYCOL	1.47	4.55	0.78	1.00	**1.15**	**4.55**
1,2-BUTANEDIOL	1.73	3.66	0.78	0.93	**1.35**	**3.41**
1,3-BUTANEDIOL	1.60	3.96	0.78	0.93	**1.25**	**3.68**
1,4-BUTANEDIOL	1.63	3.98	0.78	0.93	**1.27**	**3.70**
2-METHOXYETHANOL c0	0.77	3.13	0.78	1.00	**0.60**	**3.13**
2-METHOXYETHANOL c7 E_{conf} = 10.97	1.39	3.19	0.78	1.00	**1.08**	**3.19**
2-ETHOXYETHANOL c0	0.63	3.13	0.78	1.00	**0.49**	**3.13**
2-ETHOXYETHANOL c1 E_{conf} = 2.59	1.35	4.23	0.78	1.00	**1.05**	**4.23**
2-ETHOXYETHANOL c9 E_{conf} = 10.5	1.39	3.28	0.78	1.00	**1.08**	**3.28**
2-BUTOXYETHANOL	0.67	3.11	0.78	1.00	**0.52**	**3.11**
PHENOL	2.03	0.88	0.78	0.93	**1.58**	**0.82**
o-CRESOL	1.99	0.77	0.78	0.93	**1.55**	**0.71**
m-CRESOL	1.98	0.93	0.78	0.93	**1.54**	**0.87**
p-CRESOL	1.96	0.96	0.78	0.93	**1.53**	**0.90**
BENZYL ALCOHOL	1.23	2.07	0.78	0.93	**0.96**	**1.93**
2-METHOXYPHENOL C0	0.90	0.91	0.78	1.00	**0.70**	**0.91**
2-METHOXYPHENOL C1	2.07	1.96	0.78	1.00	**1.61**	**1.96**
4-METHOXYPHENOL C0	1.95	1.70	0.78	1.00	**1.52**	**1.70**
1,3-DIHYDROXYBENZENE	4.03	1.73	0.78	0.93	**3.14**	**1.61**
4-CHLOROPHENOL	2.28	0.72	0.78	0.93	**1.78**	**0.67**

(*continued on next page*)

3. Results

Extensive calculations were conducted to support the findings summarized in Table 1 and the results are reported in Tables S2 to S11 of the Supplementary Information (SI) file. Each of these Tables correspond to one of the following solvents: Water (S2), Ethanol (S3), 1-Octanol (S4), 2-Butanol (S5), Ethyl Acetate (S6), Acetone (S7), Di-n-Butyl Ether (S8), Tetrahydrofuran (S9), Chloroform (S10) and Acetic Acid (S11). The choice of the solvents was not arbitrary. The selection was made from the list of solvents with known LSER (estimation of) HB contribution to solvation energy [75]. These estimations were part of the overall solvation energy calculations which were remarkably close to the experimental solvation data [75]. Although this does not guarantee that the estimations of the separate HB contributions are correct, they have been adopted as a basis for comparison because of their general validity and systematic character. In fact, there was not much choice for such a

Table 1 (*continued*)

ACETIC ACID	2.04	2.70	0.78	0.55	**1.59**	**1.49**
PROPIONIC ACID	1.99	2.62	0.78	0.55	**1.55**	**1.44**
n-BUTYRIC ACID	1.98	2.62	0.78	0.55	**1.54**	**1.44**
n-PENTANOIC ACID	1.97	2.62	0.78	0.55	**1.54**	**1.44**
ETHYLAMINE	0.43	2.73	0.78	0.91	**0.34**	**2.48**
n-PROPYLAMINE	0.44	2.74	0.78	0.91	**0.34**	**2.49**
ISOPROPYLAMINE	0.41	2.61	0.78	0.91	**0.32**	**2.38**
n-BUTYLAMINE	0.43	2.73	0.78	0.91	**0.33**	**2.48**
n-PENTYLAMINE	0.43	2.73	0.78	0.91	**0.34**	**2.48**
n-HEXYLAMINE	0.43	2.75	0.78	0.91	**0.34**	**2.50**
n-HEPTYLAMINE	0.43	2.75	0.78	0.91	**0.34**	**2.50**
n-OCTYLAMINE	0.43	2.76	0.78	0.91	**0.34**	**2.51**
DIETHYLAMINE	0.20	2.02	0.78	0.92	**0.16**	**1.86**
DI-n-PROPYLAMINE	0.19	2.02	0.78	0.92	**0.15**	**1.86**
TRIMETHYLAMINE	0.00	1.49	0.00	1.00	**0.00**	**1.49**
ANILINE	1.62	0.78	0.78	0.91	**1.26**	**0.71**
N-METHYLANILINE	0.87	0.28	0.78	0.91	**0.68**	**0.25**
N-ETHYLANILINE	0.67	0.31	0.78	0.91	**0.52**	**0.28**
PYRIDINE	0.00	2.11	0.00	0.91	**0.00**	**1.92**
2-PYRROLIDONE	1.26	4.92	0.78	0.40	**0.98**	**1.97**
PIPERIDINE	0.29	2.14	0.78	0.91	**0.23**	**1.95**
PYRAZOLE	2.19	2.34	0.78	0.50	**1.71**	**1.17**
PYRROLIDINE	0.28	2.45	0.78	0.91	**0.22**	**2.23**
PYRROLE	1.68	0.34	0.78	0.91	**1.31**	**0.31**
INDOLE	1.78	0.01	0.78	0.91	**1.39**	**0.01**
N-METHYLIMIDAZOLE	0.21	3.42	0.78	0.80	**0.16**	**2.74**
NITROMETHANE	0.74	1.14	0.78	0.80	**0.58**	**0.91**
NITROETHANE	0.39	1.21	0.78	0.80	**0.30**	**0.97**
NITROBENZENE	0.11	1.02	0.78	0.80	**0.09**	**0.82**
ACETONITRILE	0.31	2.00	0.78	0.45	**0.24**	**0.90**
ACRYLONITRILE	0.43	1.52	0.78	0.45	**0.34**	**0.68**
PROPIONITRILE	0.08	2.02	0.78	0.45	**0.06**	**0.91**
BUTYRONITRILE	0.06	2.04	0.78	0.45	**0.05**	**0.92**
HEXANENITRILE	0.06	2.06	0.78	0.45	**0.05**	**0.93**
OCTANENITRILE	0.06	2.06	0.78	0.45	**0.05**	**0.93**
DECANITRILE	0.06	2.06	0.78	0.45	**0.05**	**0.93**
DODECANITRILE	0.06	2.06	0.78	0.45	**0.05**	**0.93**
BENZONITRILE	0.11	1.58	0.78	0.45	**0.08**	**0.71**
2-NITROPHENOL	0.43	1.09	0.78	0.80	**0.33**	**0.87**
4-NITROPHENOL	2.97	1.80	0.78	0.80	**2.32**	**1.44**
FORMAMIDE	3.03	4.23	0.70	0.45	**2.12**	**1.90**
N-METHYL FORMAMIDE	1.58	4.20	0.70	0.45	**1.10**	**1.89**
N,N-DIMETHYLFORMAMIDE	0.00	4.17	0.00	0.45	**0.00**	**1.88**
ACETAMIDE	2.61	4.74	0.70	0.45	**1.83**	**2.13**
DIMETHYL SULFOXIDE	0.10	5.24	0.78	0.45	**0.08**	**2.36**
WATER	2.89	2.99	0.70	0.59	**2.02**	**1.76**

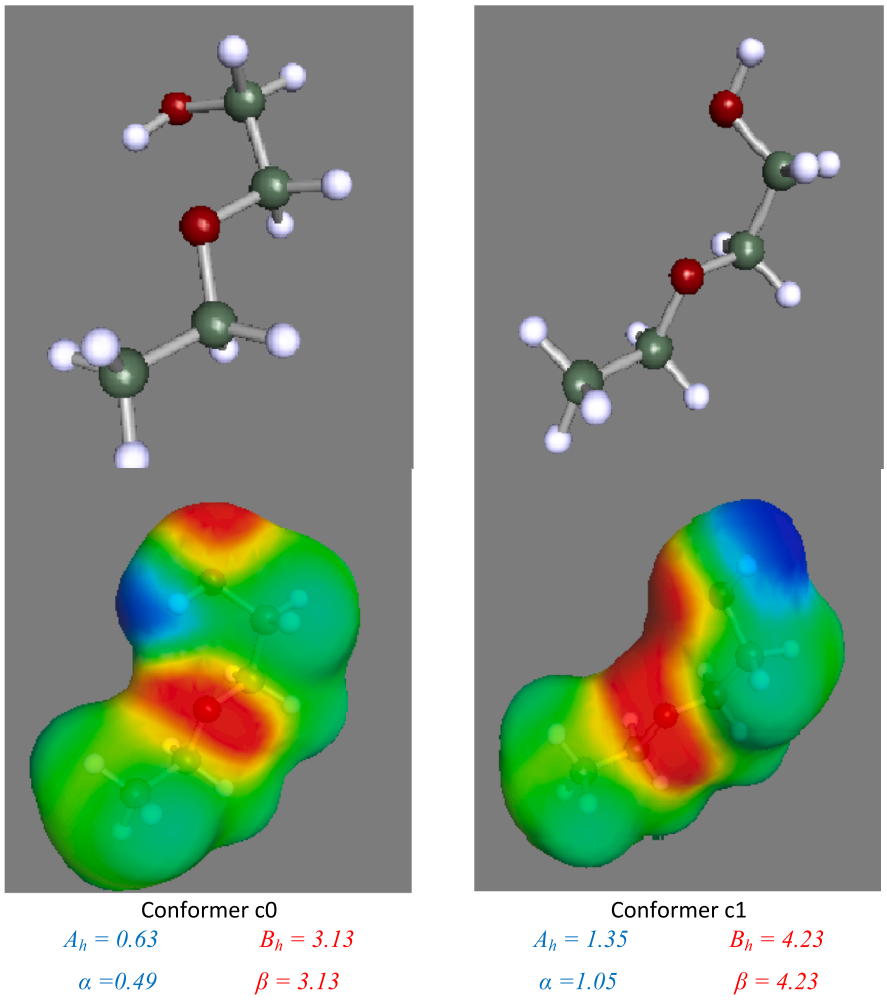


Fig. 3. The COSMO geometry (upper) and the Sigma surface (lower) of conformers c0 and c1 of 2-ethoxyethanol [69]. Proton donor sites are marked in blue and proton acceptor sites are marked in red. The HB descriptors are shown in the lower part. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
The QC-LSER molecular descriptors of ethanol obtained from three levels (dft – basis sets) of quantum chemical calculations.

Level	A_h	A_p	B_p	B_h
TZVPD-Fine	1,23	0,66	0,35	2,46
TZVP	1.37	0.66	0.36	2.76
DMol^3	1,32	0,57	0,31	2,46

comparison basis. The second base of comparison was COSMO-RS, which was also known to be one of the most successful models for solvation calculations [76,77]. The convenience with this model is the estimation of HB contribution to solvation energy, essentially, for any solute – solvent pair.

Thus, ten representative solvents were selected with a variety of hydrogen-bonding capacity for testing the predictions of equation (8). In each of the corresponding ten tables in SI file, details on the LSER and COSMO-RS estimations are reported and compared with the corresponding predictions of equation (8). We are using the term "predictions" here in a rather loose sense. In essence, the α and β descriptors in Table 1 were selected to give an optimum picture of all comparative calculations in Tables S2 to S11. This point will be further clarified by a closer inspection of these tables.

Table S2 for the aqueous solvation is complete in the sense that nearly all solutes of Table 1 were tested in it. Water is the most studied solvent and is a good basis for comparison. Before comparing predictions, it is worth to compare the LSER and the COSMO-RS calculations. As observed, there is a good agreement in the self-solvation of water, in the solvation of 1-alkanols and many other alcohols and fair agreement in the solvation of esters and alkanones. For the rest of the solutes there are disagreements of varying degrees, often exceeding various thermal energy units (RT = 2.48 kJ/mol at 298.15 K). This picture is rather common in the open literature but there is no other type of basis or some (widely recognized to be a) more accurate database to make our comparison. As a rule, much better agreement is observed for self-solvation estimations. This is why, the determination of α and β descriptors in Table 1 was based, primarily, on self-solvation data.

There are two kinds of comparisons that can be made between the present predictions and the LSER and COSMO-RS calculations. The first is the overall HB contributions to solvation energy which both LSER and COSMO-RS provide. The second is the comparison of the separate acid – base interactions of solute – solvent and solvent – solute, which are provided only by LSER and, of course, by the present predictions. Regarding the first kind of comparison, as observed in Table S2, in the overwhelming majority of cases, the predicted values are either between the corresponding LSER and COSMO-RS values or close to one of them. This is particularly important because it is a common feature throughout all calculations for all solvents reported in SI file. Regarding the second kind of comparison, there is no or rare agreement, although in some

cases there is some qualitative agreement. There is no agreement even on self-solvation and, in fact, this makes it nearly impossible to produce a predictive equation like equation (8), based exclusively on LSER calculations.

It is tempting to consider the quantum-mechanics based COSMO-RS calculations as an exclusive basis for comparison. As mentioned already, COSMO-RS is also a parameterized model against solvation data, not separate hydrogen-bonding ones. More importantly, however, one should keep in mind that COSMO-RS calculations are average calculations from conformer populations, and this may have a significant impact especially when conformers may be stabilized either by intermolecular or by intramolecular hydrogen bonding. As observed in Table S2, different conformers have different contributions, and it is their appropriate combination through a Boltzmann weighing process which gives the reported COSMO-RS contribution to HB interaction energy [53,57]. We do not report all conformers in these tables, but the ones reported suffice to make this point clear. As seen in Table S2, none of the four reported conformers of ethylene glycol has a contribution equal to the COSMO-RS value of 63.62 kJ/mol. The prevailing conformer population in the (infinite dilution) limit of this solvation must include lower contribution conformers, like conformers c0 or c1 as well as higher contribution conformers like conformer c3 or c6. On the other hand, it is the conformer c1 whose contribution is closer to the LSER estimation of 51.18 kJ/mol.

Quite similar are the observations on the calculations reported in Table S3 of solvent ethanol. The comparison between LSER and COSMO-RS calculations gives a similar picture to the previous one in aqueous solvation. The large discrepancies in the case of phenols are rather difficult to explain. Neither the high LSER values nor the low COSMO-RS values are easy to explain. The predictions, however, with equation (8) fall in between these LSER and COSMO-RS values. The discrepancies of the present predictions from the corresponding LSER calculations arise, primarily, from the relatively high LSER value for the acid (phenol)–base (ethanol) interaction. This is persistently observed for all calculations involving phenols for all solvents in SI file. Apparently, this arises from the relatively high values of the LSER descriptors of phenols [39]. This HB behavior of phenols will be further discussed in the extension of the predictive method to HB interaction free − energies.

The comparisons in Table S4 for 1-octanol solvent corroborate those of Table S3. A slightly better picture could have been observed if the descriptor α for 1-octanol were set equal to 0.77 (instead of 0.78) or if the β descriptor were set equal to 0.92 (instead of 0.93). In view, however, of the scatter in the corresponding literature data, such refinements would not add much at this stage of development of the predictive method and have been disregarded throughout this work.

Table S5 for 2-butanol solvent gives also a similar overall picture to the one of the other two alkanols. The difference is in the basicity descriptor β of the solvent, which is equal to 1, as is the case for 2-pentanol and 2-hexanol (cf. Table 1). Obviously, this difference is due to the different structures of 1- and 2-alkanols. The calculations for phenols show a similar trend as above. LSER and COSMO-RS calculations for carboxylic acids exhibit like phenols discrepancies in solvation with all studied alkanol solvents. Again, the predictions with equation (8) fall in between these two estimations.

Of interest are Tables S6 to S9 for the solvents ethylacetate, acetone, di-n-butylether and tetrahydrofuran, which have proton acceptor capacity only and, thus, they are well suited for testing the α descriptor values of the studied solutes. As observed, although there is some scatter, the predictions follow nearly the same trend as with the other self-associated solvents. In some cases, a slightly higher α value and in some other a slightly lower value would give better results. Again, due to the rather significant scatter in the literature data, such refinements would not add much in this stage of development of this predictive model.

Table S10 of solvent chloroform would serve as a verification of the β descriptor values in Table 1. As observed, however, COSMO-RS estimations are always near zero for reasons not well understood. Naturally, here again the predictions of equation (8) fall between the LSER and the COSMO-RS estimations.

Table S11 of solvent acetic acid is interesting since in the previous tables it was tested as solute. It is worth observing the LSER and COSMO-RS estimations for the solvation of alkanol solutes by this solvent. As seen, it is the COSMO-RS estimations which are now remarkably higher than the LSER ones, in contrast to what was seen in the cases of solvation of acetic acid with alkanols. This reversal is rather difficult to explain. Again, the predictions of equation (8) fall between the LSER and the COSMO-RS estimations. As seen, the predicted values for the solvation of ethanol in acetic acid is 20.88 kJ/mol, which is identical to the corresponding value for the solvation of acetic acid by solvent ethanol reported in Table S3. Similarly, for water solvation by acetic acid as well as for the solvation of acetic acid by water, the predicted HB contribution is equal to 33.30 kJ/mol. The same holds true for all pairs of molecules, regardless of which is solute or solvent, and is verified throughout Tables S2 to S11 to within round-off errors. This is direct consequence of the form of equation (8) which remains the same by an exchange of indices 1 and 2. It is stressed that this is true when the conformer population in a mixture does not change with composition. When this population changes the predicted values change and, as seen in Tables S2 to S11, this change may be very large.

4. Discussion

The bulk of the reported QC-LSER descriptors in Table 1 and in Table S1 are based on the σ-profiles of the most stable conformers. This is sufficiently good for most of the cases of practical interest. Obviously, in the case of flexible molecules with capacity to form intramolecular and intermolecular bonds, such as the alkoxy-alkanols, more than one conformers should be taken into consideration and equation (8) should be used with much care.

All calculations in the present work were based on QC-LSER descriptors obtained from quantum chemical calculations at the DFT / TZVPD-Fine level [69,72]. Calculations could be done also at other levels. It is important, however, to keep in mind that the QC-LSER descriptors may vary by changing the level of quantum chemical calculations, as shown in Table 2. This implies that the availability fractions f_A and f_B will also change, accordingly, in order to lead to identical α and β descriptors of the compound for the practical calculations. Thus, for consistency, one should use QC-LSER descriptors and availability fractions obtained at the same level of quantum chemical calculations.

LSER estimations of ΔE_{12}^{hb} are available for a few decades of solvents [37–39]. Thus, the hydrogen-bonded solvents examined in the present work are relatively simple ones. More complex multi-sited solvents with many and distant HB sites have not been examined and some caution should be exercised when applying the above predictive method to such complex systems. This limitation of the predictive method does not arise only from the lack of LSER estimations. It also arises from the limitations of the COSMO-RS model itself. The information on the special or 3D distribution of HB sites in a complex molecule (cf. Fig. 3) is lost, to a significant degree, in its sigma-profile which is used for the COSMO-RS calculations. As a consequence, two distant proton donor or proton acceptor sites will increase the value of A_h or B_h descriptors of a solvent molecule but the information that a simple solute molecule (with one proton acceptor or proton donor site) can access one only of its HB sites is lost. In this way, such a complex multi-sited molecule may exhibit different availability fractions as solvent and as solute. Luckily, LSER estimations of the HB contribution to solvation free-energy are available for many more solvent systems, including complex multi-sited ones [37–39]. A more in depth discussion of this issue is, then, postponed to an extension of the present method for the prediction of HB contribution to solvation free energy.

In the present work we have confined ourselves to the HB contri-

bution to solvation energy (enthalpy). There are various ways these contributions could be used for the calculation of other thermodynamic quantities. First of all, this contribution could be used to replace the corresponding contribution, namely, the sum, $a_2A_1 + b_2B_1$, in the LSER equation for the overall solvation energy [32–39,75]. This would have a significant impact and could facilitate the expansion of the LSER method to many more solvents. The predictions of equation (8) could be used also for the estimation of the corresponding contribution to solvation free energy. One way to do this was discussed in ref. [70] where a method was proposed for the estimation of the HB contribution to solvation entropy. Combining this entropy estimation with the present energy (enthalpy) estimation, one would obtain the HB contribution to solvation free energy in the classical thermodynamic way:

$$\Delta G_{12}^{hb} = \Delta E_{12}^{hb} - T\Delta S_{12}^{hb}$$

The predictions of equation (8) could be combined also with information from partial solvation parameters (PSP) [41] or from Hansen solubility parameters (HSP) [68]. As discussed in ref [41], the hydrogen-bonding PSP of a self-associated compound i is given by

$$\sigma_{hb}^2 = \frac{-N_{hb}\Delta E_{ii}^{hb}}{V_m} \tag{9}$$

where, V_m is the molar volume, ΔE_{ii}^{hb} is the HB contribution to self-solvation energy obtained from equation (8), and the number of hydrogen bonds, N_{hb}, is given by

$$N_{hb} = 1 + \frac{1}{2K} - \sqrt{\frac{1}{K} + \frac{1}{4K^2}} \tag{10}$$

with

$$-RT\ln K = \Delta G_{ii}^{hb} \tag{11}$$

At 25 °C, the σ_{hb} is identical to Hansen's δ_{hb}. Thus, by knowing the latter [68], we could obtain the HB contribution to self-solvation free energy from equation (11). Various alternatives for equation (10) are also discussed in ref. [41].

The above estimations of the HB contribution to solvation free energy could be compared with predictions of an equation analogous to equation (8), or

$$-\Delta G_{12}^{hb} = 5.71(\alpha_{G1}\beta_{G2} + \beta_{G1}\alpha_{G2}) \quad kJ/mol \tag{12}$$

where, α_G and β_G are the corresponding effective free-energy descriptors for acidity and basicity, respectively. It should be pointed out, however, that COSMO-RS does not provide estimations of this contribution and one has to rely on LSER estimations. But as mentioned above, LSER estimations alone are not highly suitable for the development of such a predictive method and the derived descriptor values should be treated as tentative. Work is in progress in our laboratories for the estimation of α_G and β_G descriptors through the combination of LSER estimations of ΔG_{12}^{hb} with the above information on ΔE_{12}^{hb} from COSMO-RS and LSER estimations.

Above all, however, the predictions with equation (8) are ideally suited for the direct use in equation-of-state models and other molecular thermodynamic models of hydrogen bonded mixtures. This was, in fact, one of the central incentives for the present work. Due to its simplicity, equation (8) may be incorporated also easily in process simulators. Table 1 may be further expanded and equation (8) may be used also in fast solvent screening calculations.

5. Conclusions

A simple method for the prediction of the HB contribution to solvation energy in simple hydrogen-bonded solvents is proposed. The method is based on two quantum-mechanics based molecular descriptors for the proton-donor and proton-acceptor capacity of the solute. HB interactions are treated as electrostatic interactions and the molecular surface-charge-density distributions of a given conformer are considered unchanged with mixture composition. Thus, the HB interaction energy for a given pair of molecules remains constant throughout the full composition range as long as there is no change in the prevailing conformer populations. The predictions of the proposed method have been compared with corresponding calculations of the Abraham's LSER model and of the COSMO-RS model for a variety of solute – solvent systems. In most cases, the predictions fall between these two calculations or are close to one of them. There are no limitations in the method, in principle, and it may be expanded to any hydrogen-bonded solute – compound, even to those not-synthesized yet. Caution must be exercised, however, when applying the method to systems of complex solvents with many distant HB sites, as these solvents may require a solvent – specific set of availability fractions. The proposed predictive method could be particularly useful in molecular thermodynamic modelling of hydrogen-bonded systems. Due to its simplicity, the use of the method does not require any expertise in quantum mechanics or in statistical thermodynamics.

CRediT authorship contribution statement

I. Zuburtikudis: Writing – review & editing, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation. **W.E. Acree:** Writing – review & editing, Validation, Methodology, Data curation. **C. Panayiotou:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is supported by ASPIRE, the technology program management pillar of Abu Dhabi's Advanced Technology Research Council (ATRC), via the ASPIRE "AARE (ASPIRE Awards for Research Excellence)" and through grant no. AARE20-246 to Ioannis Zuburtikudis of Abu Dhabi University.

Appendix A. Supplementary data

Supplementary Information File (SI): Table S1 is an updated table of the new QC-LSER descriptors and the calculations of HB interaction energies with the LSER and COSMO-RS models and predictions with equation (8) are reported in detail in Tables S2 to S11 corresponding to ten representative solvents. Supplementary data to this article can be found online at https://doi.org/10.1016/j.molliq.2025.126907.

Data availability

Data will be made available on request.

References

[1] X. Qiu, H. Li, G. Ver Steeg, A. Godzik, Advances in AI for protein structure prediction: implications for cancer drug discovery and development, Biomolecules 14 (3339) (2024), https://doi.org/10.3390/biom14030339.
[2] J.D. Moore, R.D. Mountain, R.B. Ross, V.K. Shen, D.W. Siderius, K.D. Smith, The Ninth Industrial Fluid Properties Simulation Challenge, Fluid Phase Equilib. 476 (2018) 1–5.

[3] L. Pereyaslavets, G. Kamath, O. Butin, A. Illarionov, M. Olevanov, I. Kurnikov, S. Sakipov, I. Leontyev, E. Voronina, T. Gannon, G. Nawrocki, M. Darkhovskiy, I. Ivahnenko, A. Kostikov, J. Scaranto, M.G. Kurnikova, S. Banik, H. Chan, M. G. Sternberg, S.K.R.S. Sankaranarayanan, B. Crawford, J. Potoff, M. Levitt, R. D. Kornberg, B. Fain, Accurate determination of solvation free energies of neutral organic compounds from first principles, Nat. Commun. (2022), https://doi.org/10.1038/s41467-022-28041-0.

[4] T.L. Hill, An Introduction to Statistical Thermodynamics, Addison-Wesley, Reading, MA, 1960.

[5] D. Chandler, Introduction to Modern Statistical Mechanics, Oxford University Press, Oxford, 1987.

[6] F. Hirata, Exploring Life Phenomena with Statistical Mechanics of Molecular Liquids, FL, CRC Press, Boca Raton, 2020.

[7] M.D. Joesten, L. Schaad, Hydrogen bonding, Marcel Dekker, New York, 1974.

[8] G. Gilli, P. Gilli, The Nature of the Hydrogen Bond, Oxford University Press, Oxford, 2009.

[9] A.K. Baev, Specific Intermolecular Interactions of Nitrogenated and Bioorganic Compounds, Springer-Verlag, Berlin Heidelberg, 2014.

[10] D.S. Goodsell, Bionanotechnology: Lessons from Nature, Wiley & Sons, New York, 2003.

[11] R. Car, M. Parrinello, Unified approach for molecular dynamics and density functional theory, Phys. Rev. Lett. 55 (22) (1985) 2471.

[12] P.L. Silvestrelli, M. Parrinello, Structural, electronic, and bonding properties of liquid water from first principles, J. Chem. Phys. 111 (8) (1999) 3572–3580.

[13] L. Sapir, D. Harries, Revisiting Hydrogen Bond Thermodynamics in Molecular Simulations, J. Chem. Theory Comput. 13 (2017) 2851–2857.

[14] G. Duarte Ramos Matos, D. Y. Kyu, H. H. Loeffler, J. D. Chodera, M. R. Shirts, D. L. Mobley 2017. Approaches for Calculating Solvation Free Energies and Enthalpies Demonstrated with an Update of the FreeSolv Database. J. Chem. Eng. Data 62, 5, 1559–1569.

[15] M.J. Tillotson, N.I. Diamantonis, C. Buda, L.W. Bolton, E.A. Muller, Molecular modelling of the thermophysical properties of fluids: expectations, limitations, gaps and opportunities, Phys. Chem. Chem. Phys. 25 (2023) 12607.

[16] J.M. Prausnitz, R.N. Lichtenthaler, E. Gomes de Azevedo, Molecular Thermodynamics of Fluid Phase Equilibria, 3rd ed, Prentice Hall, Upper Saddle River, NJ, 1999.

[17] F. Dolezalek, Zur theorie der binaren gemische und konzentrierten losungen, Z. Physick. Chem. 64 (1908) 727–747.

[18] I. Moriguchi, Quantitative Structure-Activity Studies I. Parameters Relating to Hydrophobicity, Chem. Pharm. Bull. 23 (1975) 247–257.

[19] M.J. Kamlet, J.L.M. Abboud, R.W. Taft, An Examination of Linear Solvation Energy Relationships, Proc. Phys. Org. Chem. 13 (1981) 485–630.

[20] M.J. Kamlet, R.M. Doherty, J.-L. Abboud, M.H. Abraham, R.W. Taft, Solubility: a new look, Chem. Tech. 16 (1986) 566–576.

[21] M.H. Abraham, R.M. Doherty, M.J. Kamlet, R.W. Taft, A new look at acids and bases, Chemical. Brit. 22 (1986) 551–554.

[22] M.J. Kamlet, R.M. Doherty, M.H. Abraham, G.D. Veith, D.J. Abraham, R.W. Taft, Solubility Properties in Polymers and Biological Media. 8. An Analysis of the Factors that Influence Toxicities of Organic Nonelectrolytes to the Golden Orfe Fish (Leuciscus idus melanotus), EnvIron. Sci. Technol. 21 (1987) 149–155.

[23] G. Schüürmann, Quantitative Structure-Property Relationships for the Polarizability, Solvatochromic Parameters and Lipophilicity. Quant. Struct.- Act, Relat. 9 (1990) 326–333.

[24] D. Katritzky, E. Fara, H. Yang, K. Tamm, T. Tamm, M. Karelson, Quantitative Measures of Solvent Polarity, Chem. Rev. 104 (2004) 175–198.

[25] C. Laurence, J.-F. Gal, J.-f., Lewis Basicity and Affinity Scales: Data and Measurements, Wiley, New York, 2010.

[26] L.A. Sedov, B.N. Solomonov, Hydrogen bonding in neat aliphatic alcohols: The Gibbs free energy of self-association and molar fraction of monomer, J. Mol. Liq. 167 (2012) 47–51.

[27] V. Dohnal, New QSPR molecular descriptors based on low-cost quantum chemistry computations using DFT/COSMO approach, J. Mol. Liq. 407 (2024) 125256, https://doi.org/10.1016/j.molliq.2024.125256.

[28] G.M. Kontogeorgis, G.K. Folas, Thermodynamic Models for Industrial Applications, From Classical and Advanced Mixing Rules to Association Theories, Chichester, U. K., John Wiley and Sons Ltd, 2010.

[29] L.Y. Wilson, G.R. Famini, Using Theoretical Descriptors in Quantitative Structure-Activity Relationships: Some Toxicological Indices, J. Med. Chem. 34 (1991) 1668–1674.

[30] Dearden, J. C.; Ghafourian, T. 1995. Investigation of Calculated Hydrogen Bonding Parameters for QSAR. In QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications; Sanz, F., Giraldo, J., Manaut, F., Eds.; Prous Science Publishers: Barcelona, pp 117-119.

[31] Raevsky, O. A.; Grigor'ev, V.; Mednikova, E. 1993. QSAR H-Bonding Descriptions. In Trends in QSAR and Molecular Modelling 92; Wermuth, C. G., Ed.; ESCOM: Leiden. pp 116-119.

[32] M.H. Abraham, J.C. McGowan, The use of characteristic volumes to measure cavity terms in reversed phase liquid chromatography, Chromatographia 23 (1987) 243–246.

[33] M.H. Abraham, Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes, Chem. Soc. Rev. 22 (1993) 73–83.

[34] M.H. Abraham, A. Ibrahim, A.M. Zissimos, Determination of sets of solute descriptors from chromatographic measurements, J. Chromatogr. A 1037 (2004) 29–47.

[35] M.H. Abraham, R.E. Smith, R. Luchtefeld, A.J. Boorem, R. Luo, W.E. Acree Jr., Prediction of solubility of drugs and other compounds in organic solvents, J. Pharm. Sci. 99 (2010) 1500–1515.

[36] K.-U. Goss, Predicting the equilibrium partitioning of organic compounds using just one linear solvation energy relationship (LSER), Fluid Phase Equilibr. 233 (2005) 19–22.

[37] S. Sinha, Ch. Yang, E. Wu, W.E. Acree, Abraham Solvation Parameter Model: Examination of Possible Intramolecular Hydrogen-Bonding using calculated solute descriptors, Liquids 2 (2022) 131–146.

[38] C. Mintz, T. Ladlie, K. Burton, M. Clark, W.E. Acree Jr, M.H. Abraham, Enthalpy of solvation correlations for gaseous solutes dissolved in alcohol solvents based on the Abraham model, QSAR Comb, Sci. 27 (2008) 627–635.

[39] Endo, S., Watanabe, N., Ulrich, N.., G. Bronner, K-U. Goss, UFZ-LSER database v 2.1 [Internet], Leipzig, Germany, Helmholtz Centre for Environmental Research-UFZ, 2015. [last accessed on 12.09.2024], available from https://www.ufz.de/index.php?en= 31698& contentonly=1&m=0&lserd_data[mvc]=Public/start.

[40] C. Panayiotou, I. Zuburtikudis, H. Abu Khalifeh, Linear Free-Energy Relationships (LFER) and Solvation Thermodynamics: The Thermodynamic Basis of LFER Linearity, Ind. Eng. Chem. Res. 62 (6) (2023) 2989–3000.

[41] C. Panayiotou, I. Zuburtikudis, H. Abu Khalifeh, Linear Solvation Energy Relationships (LSER) and Equation-of-State Thermodynamics: On the Extraction of Thermodynamic Information from LSER Database, Liquids 3 (2023) 66–89.

[42] Panayiotou, C., Acree Jr, W.E/, Zuburtikudis, I., 2023. COSMO-RS and LSER Models of Solution Thermodynamics: Towards a COSMOLSER Equation of State Model of Fluids. J. Molec. Liq. 390, 122992.

[43] I.G. Economou, M.D. Donohue, Equations of State for Hydrogen Bonding Systems, Fluid Phase Equilib. 116 (1996) 518–529.

[44] A. Anderko, Extension of the AEOS model to systems containing any number of associating and inert components, Fluid Phase Equilib. 50 (1–2) (1989) 21–52.

[45] W.G. Chapman, K.E. Gubbins, G. Jackson, M. Radosz, Ind. Eng.chem. Res. 29 (1990) 1709–1721.

[46] S.H. Huang, M. Radosz, M., Equation of State for Small, Large, Polydisperse, and Associating Molecules, Ind. Eng. Chem. Res 29 (1990) 2284–2294.

[47] E.A. Muller, K.E. Gubbins, Molecular-based equations of state for associating fluids: A review of SAFT and related approaches, Ind. Eng. Chem. Res. 40 (2001) 2193–2211.

[48] J. Gross, G. Sadowski, Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules, Ind. Eng. Chem. Res. 40 (2001) 1244–1260.

[49] L.F. Vega, F. Llovall, Review and new insights into the application of molecular-based equations of state to water and aqueous solutions, Fluid Phase Equil. 416 (2016) 150–173.

[50] M. Wertheim, Fluids with highly directional attractive forces. III. Multiple attraction sites, J. Statist. Phys. 42 (1986) 459–476.

[51] M. Wertheim, Fluids of dimerizing hard spheres, and fluid mixtures of hard spheres and dispheres, J. Chem. Phys. 85 (1986) 2929–2936.

[52] A. Klamt, Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena, J. Phys. Chem. 99 (7) (1995) 2224–2235.

[53] A. Klamt, COSMO-RS from Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design, Elsevier, Amsterdam, 2005.

[54] S.T. Lin, S.I. Sandler, A priori phase equilibrium prediction from a segment contribution solvation model, Ind. Eng. Chem. Res. 41 (2002) 899–913.

[55] H. Grensemann, J. Gmehling, Performance of a conductor-like screening model for real solvents model in comparison to classical group contribution methods, Ind. Eng. Chem. Res. 44 (2005) 1610–1624.

[56] C.C. Pye, T. Ziegler, E. van Lenthe, J.N. Louwen, An implementation of the conductor-like screening model of solvation within the Amsterdam density functional package, Part II. COSMO for real solvents, Can. J. Chem. 87 (2009) 790–797.

[57] A. Klamt, F. Eckert, W. Arlt, COSMO-RS: An alternative to simulation for calculating thermodynamic properties of liquid mixtures, Annu. Rev. Chem. Biomol. Eng. 1 (2010) 101–122.

[58] E.A. Guggenheim, Mixtures, Clarendon Press, Oxford, 1951.

[59] C. Panayiotou, J.H. Vera, The Quasi-Chemical Approach for Non-randomness in Liquid Mixtures. Expressions for Local Surfaces and Local Compositions with an Application to Polymer Solutions, Fluid Phase Equil. 5 (1980) 55–80.

[60] C.P. Panayiotou, J.H. Vera, Statistical Thermodynamics of r-Mer Fluids and their Mixtures, Polymer. J. 14 (1982) 681–694.

[61] C. Panayiotou, I.C. Sanchez, Hydrogen Bonding in Fluids: An Equation of-State Approach, J. Phys. Chem. 95 (1991) 10090–10097.

[62] C. Panayiotou, E. Stefanis, I. Tsivintzelis, M. Pantoula, I. Economou, Nonrandom Hydrogen-Bonding Model of Fluids and Their Mixtures 1, Pure Fluids. Ind. Eng. Chem. Res. 43 (2004) 6592–6606.

[63] C. Panayiotou, I. Tsivintzelis, I.G. Economou, Nonrandom Hydrogen-Bonding Model of Fluids and their Mixtures. 2. Multicomponent mixtures, Ind. Eng. Chem. Res. 46 (2007) 2628–2636.

[64] B.A. Veytsman, Are lattice models valid for fluids with hydrogen bonds? J. Phys. Chem. 94 (1990) 8499–8500.

[65] G. Mensitieri, G. Scherillo, C. Panayiotou, P. Musto, Towards a predictive thermodynamic description of sorption processes in polymers: The synergy between theoretical EoS models and vibrational spectroscopy, Materials Sci. & Eng. R 140 (2020) 100525.

[66] C. Hansch, A. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, John Wiley & Sons, New York, 1979.

[67] J. Li, T. Zhu, G.D. Hawkins, P. Winget, D.A. Liotard, C.J. Cramer, D.G. Truhlar, Extension of the platform of applicability of the SM5.42R universal solvation model, Theor. Chem. Acc. 103 (1999) 9–63.
[68] Abbott, S., Yamamoto, H., Hansen, C.M. 2010. Hansen Solubility Parameters in Practice, Complete with software, data and examples, third ed.-version 3.1.20. Book and Software published by Hansen-Solubility.com.
[69] COSMObase, ver. 2019, COSMOlogic GmbH &CoKG (now, BIOVIA Dassault Systemes).
[70] C. Panayiotou, Quantum Chemical (QC) Calculations and Linear Solvation Energy Relationships (LSER): Hydrogen-Bonding Calculations with New QC-LSER Molecular Descriptors, Liquids 3125552 (2024), https://doi.org/10.3390/liquids4040037.
[71] I.A. Bell, E. Mickoleit, C.-M. Hsieh, S.-T. Lin, J. Vrabec, C. Breitkopf, A. Jager, A Benchmark Open-Source Implementation of COSMO-SAC, J. Chem. Theory Comput. 16 (2020) 2635–2646.
[72] TURBOMOLE V7.5 2020, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.
[73] https://www.3ds.com/products/biovia/materials-studio.
[74] https://www.scm.com/product/cosmo-rs/.
[75] E. Hart, D. Grover, H. Zettl, V. Koshevarova, W.E. Acree Jr., M.H. Abraham, Development of Abraham model expressions for predicting the enthalpies of solvation of solutes dissolved in acetic acid, Phys. Chem. Liq. 54 (2) (2016) 141–154.
[76] A. Klamt, F. Eckert, J. Reinisch, K. Wichmann, Prediction of cyclohexane-water distribution coefficients with COSMO-RS on the SAMPL5 data set, J Comput Aided Mol Des. 30 (11) (2016) 959–967.
[77] J. Reinisch, A. Klamt, Prediction of free energies of hydration with COSMO-RS on the SAMPL4 data set, J Comput. Aided Mol. Des. 28 (3) (2014) 169–173.