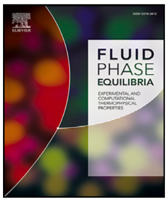
Fluid Phase Equilibria 589 (2025) 114250

Contents lists available at ScienceDirect

Fluid Phase Equilibria

journal homepage: www.sciencedirect.com/journal/fluid-phase-equilibria



Predicting solvation free energies for neutral molecules in any solvent with openCOSMO-RS

Simon Müller [a,*], Thomas Nevolianis [b], Miquel Garcia-Ratés [c], Christoph Riplinger [c], Kai Leonhard [b], Irina Smirnova [a]

[a] *Institute of Thermal Separation Processes, Hamburg University of Technology, Hamburg, 21073, Germany*
[b] *Institute of Technical Thermodynamics, RWTH Aachen University, Aachen, 52062, Germany*
[c] *FAccTs GmbH, Cologne, 50677, Germany*

ARTICLE INFO

Keywords:
Solvation free energies
Activity coefficients
Partition coefficients
Phase equilibrium
openCOSMO-RS
COSMO-RS

ABSTRACT

In this study, we introduce openCOSMO-RS 24a, an improved version of the open-source COSMO-RS model parameterized using quantum chemical calculations from ORCA 6.0, leveraging a comprehensive dataset that includes solvation free energies, partition coefficients, and infinite dilution activity coefficients for various solutes and solvents mainly at 25 °C. This is the first version of the model also capable of predicting solvation free energies based on ORCA calculations. Additionally, we develop a Quantitative Structure-Property Relationships model to predict molar volumes of the solvents, an essential requirement for predicting solvation free energies and partition coefficients from structure alone. Our results show that openCOSMO-RS 24a achieves an average absolute deviation of 0.45 kcal mol^{1} for solvation free energies, 0.76 for the logarithm of the partition coefficients, and 0.51 for the logarithm of infinite dilution activity coefficients, demonstrating improvements over the previous openCOSMO-RS 22 parameterization and comparable results to COSMOtherm 24 BP-TZVP. The user interface was extended to be able to use it as solvation model directly from within ORCA 6.0 or from the command line to provide researchers with a robust tool for applications in chemical and materials science.

1. Introduction

The accurate prediction of solvation free energies of solutes ΔG_{solv} is crucial for understanding various phenomena in the liquid phase, including reaction rates, equilibrium constants, activity coefficients, and partition coefficients. Despite extensive research, precise prediction of solvation free energies remains challenging. From this quantity, one can determine various thermodynamic and kinetic properties such as reaction rates, equilibrium constants, activity coefficients, dissociation acidity constants, partition coefficients. Consequently, solvation free energy plays a pivotal role in chemical reactions [1–5] and design of materials with novel properties [6–11]. Despite significant efforts over recent decades, precise prediction of ΔG_{solv} remains a challenge.

In the last decade, various explicit [12,13], implicit [14–19], and data-driven [20–26] approaches have been used for solution phase property prediction. Explicit approaches such as Molecular Dynamics (MD) are less common methods as they are quite computational time expensive as one needs to dissolve a solute in thousands of solvent molecules. Implicit approaches are more common in solution phase property prediction since they are less computational demanding. These approaches accurately predict the solvation free energies of neutral solutes with an uncertainty ranging from $0.4\,\mathrm{kcal\,mol^{-1}}$ to $1.1\,\mathrm{kcal\,mol^{-1}}$ [15,27–30]. Among others, the Conductor-like screening model for realistic solvation (COSMO-RS) is a frequently used fully predictive implicit model with an uncertainty of $0.40\,\mathrm{kcal\,mol^{-1}}$ to $0.45\,\mathrm{kcal\,mol^{-1}}$ for predicting the solvation free energy of neutral solutes [27,31,32]. The COSMO-SAC implementation has been applied in the literature to calculate solvation energies for the purpose of predicting vapor pressures [33,34] and pure component parameters for cubic equations of state [35]. Additionally, it has been used to indirectly calculate solvation energies through infinite dilution activity coefficients [36] and for solvation free energy calculation as mixing rule for the Peng-Robinson equation of state [37]. The basic principle of COSMO-RS is based on the approximation of molecular interactions by the interactions of surface segments from the molecular cavities. This makes the calculations less demanding than MD calculations as the required input information only needs to be calculated once for each molecule from Quantum Mechanics (QM). Data-driven models have shown great potential in liquid phase property prediction mainly because many

---

[*] Corresponding author.
E-mail address: simon.mueller@tuhh.de (S. Müller).

https://doi.org/10.1016/j.fluid.2024.114250
Received 26 July 2024; Received in revised form 1 October 2024; Accepted 7 October 2024

well established experimental databases are available. For example, Machine Learning (ML) methods have been quite promising in predicting the solvation free energies of neutral solutes [38–40]. Vermeire et al. [32], trained and fine tuned a Graphical Neural Network (GNN) model for predicting solvation free energies of neutral molecules reporting an uncertainty of 0.24 kcal mol^{-1}. While these GNN perform well on training datasets, their ability to generalize to new structures with different atoms and functional groups remains challenging. Empirical evidence suggests that their embeddings can generalize across different molecular spaces, but achieving robust out-of-distribution performance is still difficult [41,42].

Recently, we published an open source version of the COSMO-RS model, which we will call openCOSMO-RS 22 [43] in the following. This implementation of COSMO-RS is the first open source version introducing additional descriptors besides the screening charge density. Having additional descriptors allowed the model to be modified for electrolytes with great success in the past [44–49]. For neutral molecules, openCOSMO-RS 22 performs quite well for predicting the infinite dilution activity coefficient (IDAC) with a Root Mean Square Deviation (RMSD) of 0.76 based on TURBOMOLE 6.6 parameterization and 0.65 based on ORCA 5.0.3 [50–53]. Although openCOSMO-RS 22 was able to predict equilibrium properties between two or more solvents, it was not capable of predicting properties between gas and liquid phase such as the solvation free energy.

In this study, we perform a new parameterization of the model based on quantum chemical calculations from the software ORCA 6.0. This will be called openCOSMO-RS 24a. To do so, initially, we compile experimental data on solvation free energies, partition coefficients, and activity coefficients for a representative range of solutes and solvents. Since the molar volume of the solvent is required to predict the solvation free energy, we develop a Quantitative Structure-Property Relationships (QSPR) model to predict the molar volume of the solvent at 25 °C based on experimental data available in the literature [54]. We modified the openCOSMO-RS conformer workflow compared to that of our previous work [43] by adding quantum chemical calculations of gas phase energies as these are needed to calculate the solvation free energies. Leveraging the experimental data together with the QSPR model, we parameterize openCOSMO-RS 24a for predicting the solvation free energies for a wide range of solutes and solvents. During this work, we found that the Gaussian charge scheme, used within CPCM [55,56] in ORCA [57] produced very small segments leading to unusually large screening charge densities. This was addressed by rejecting the addition of segments smaller than a specified threshold (0.01 Å^2) with minimum effect on the calculated energies. Furthermore, in ORCA, a Lagrangian-based algorithm is used to calculate the outlying charge correction [58]. Although this is not a treatment as advanced as the method proposed by Klamt [59] as it neglects the spacial distribution of the outlying charge, it should be enough for the neutral molecules tested. In the future a more thorough analysis of this is planned as it becomes especially important for anions. Finally, we report the performance of openCOSMO-RS 24a, which as of now can directly be used within the ORCA 6.0 software as additional solvation model enabling the user to access a variety of liquid phase properties which previously was not possible.

2. Methods

2.1. openCOSMO-RS

The theory of openCOSMO-RS has been discussed in previous studies [43,60] and only the equation related to solvation free energy is briefly summarized here. The solvation free energy can then be calculated similarly to Klamt et al. [17,61] from

$$\Delta G_{\mathrm{solv}} = E_{\mathrm{diel}} + RT \ln \gamma^{\infty} - \sum_{\alpha} \tau_{\alpha} A_{\alpha} - \omega_{\mathrm{ring}} n_{\mathrm{ring\_atoms}} - RT \ln \frac{v_{\mathrm{IG}}}{v_{\mathrm{liquid}}} - \eta \tag{1}$$

The term E_{diel} represents the dielectric energy, which is the energy involved in transferring the solute from the gas phase to an ideal conductor. The second term refers to the chemical potential at infinite dilution in the liquid phase, using the ideal conductor as the reference state and it is directly predicted by openCOSMO-RS. The third term encompasses the energy required for cavity formation and includes a van der Waals-like dispersion contribution to the solvation free energy, calculated by summing the product of each atom's area A_{α} on the solute molecule and a factor τ_{α} that depends on the underlying atom $\alpha = \{\mathrm{H, C, N, O, F, Si, P, S, Br, Cl}\}$. The fourth term provides a correction for molecules containing rings, determined by multiplying a general parameter ω_{ring} by the number of atoms in a ring $n_{\mathrm{ring\_atoms}}$ in the solute structure, independent of the type of ring (aromatic, non-aromatic, single, fused). The fifth term accounts for the change in units of the reference states from mole fraction (units of the calculation) to molar concentration (units of the experimental data) with v_{IG} and v_{liquid} representing the molar volumes of the ideal gas and the liquid phase, respectively. The final term η is an adjustable parameter which is described by Klamt et al. [17] to account for the entropy of the molecule in the gas phase.

2.2. Computational details

In the following, we describe the openCOSMO-RS 24a conformer workflow for searching and calculating all necessary input data for the gas and CPCM phase. This is the updated overview of the pipeline also available on github [62]:

- Gas phase calculations
  - Molecular mechanics-based conformer generation using RDKit [63,64].
  - Filter conformers by an energy window of 6 kcal mol^{-1}.
  - Cluster conformers by an RMSD window of 1 Å and save these for CPCM[57] calculations.
  - Geometry optimizations at DFT/BP86/def2-TZVP(-f) [65–67] level using ORCA.
  - Single point energy calculation using DFT/BP86/def2-TZVPD level in ORCA for the conformer with the lowest energy.
- CPCM calculations
  - Optimize geometries in water using ALPB [68] with GFN2-xTB [69] calculations from within ORCA, starting from saved conformers.
  - Filter conformers by an energy window of 6 kcal mol^{-1}.
  - Cluster conformers by an RMSD window of 1 Å and select the conformers with the three lowest energies.
  - CPCM geometry optimizations at the DFT/BP86/def2-TZVP(-f) level in ORCA.
  - Filter conformers by an energy window of 6 kcal mol^{-1}.
  - Cluster conformers by an RMSD window of 1 ∀ and select the conformer with the lowest energy.
  - CPCM geometry optimizations of DFT/BP86/def2-TZVP level in ORCA.
  - CPCM single point energy calculation using DFT/BP86/def2-TZVPD level in ORCA.

To search a large parameter space, the global solver differential evolution as implemented in SciPy [70] is used. Similar to our previous studies [43,60], the following objective function is minimized for all optimizations:

$$\mathrm{OF} = \frac{1}{N_p} \sum_{i} \left(Y^{\mathrm{calc}} - Y^{\mathrm{exp}} \right)^2 \tag{2}$$

The average absolute deviation is calculated from:

$$\mathrm{AAD_Y} = \frac{1}{N_\mathrm{p}} \sum_i \left| Y^\mathrm{calc} - Y^\mathrm{exp} \right| \tag{3}$$

whereby Y is either $\ln \gamma_\infty$, $\ln \mathrm{K_{AB}}$ or ΔG_solv. Employing the objective function described above and evaluating the solvation free energies in $\mathrm{kcal\,mol^{-1}}$, all three datasets span roughly the range between zero and twenty. This balances the effect of different data types as all errors are in the same order of magnitude.

2.3. Dataset overview

The dataset used in this work is comprised of three data types mainly at 25 °C: (i) infinite dilution activity coefficients, (ii) partition coefficients and (iii) solvation free energies. The 800+ infinite dilution activity coefficients are taken from Parcher et al. [71], Voutsas & Tassios [72], Kontogeorgis et al. [73], Kato et al. [74] and He & Zhong [75]. Although the temperature range spans from 262 to 393 K, the majority of data is composed of temperatures around 25 °C and alkane/alkane mixtures (athermal mixtures) at 373 K. The partition coefficients for the following solvent combinations: octanol + water, benzene + water, hexane + water, and diethyl ether + water are collected by Klamt et al. [17]. All partition coefficients were measured at 25 °C. The 2000+ solvation free energies at 25 °C including a variety of solvents, functional groups, molecules with and without (fused) rings are taken from Marenich et al. [76]. Xylene is excluded from the calculations as it is a mixture of constitutional isomers. Additionally, values for water as a solute in all three data types were excluded due to their known issues when solvated in non-polar solvents within the COSMO-RS framework without further model improvements [17,77]. Even for molecular simulations treating mixtures of water and alkanes over the complete concentration range is challenging for most models [78–80]. To calculate the solvation free energy, the molar volume of the pure solvent is required (see Eq. (1)). Thus, we develop a QSPR model in this study to predict the molar volume of the solvent at 25 °C based on experimental data available from Mathieu & Bouteloup [54]. The same model was also applied when converting log P values (based on molar volume) into log K values (based on mole fraction). The complete dataset is cleaned and normalized: isotopes and explicit hydrogens are deleted, duplicates are merged, and only the first value in the original data for each component is retained. The complete dataset is available on zenodo [81].

3. Results and discussion

3.1. Predictive QSPR model for molar volumes of the solvent at 25 °C

To enable the fully predictive calculation of solvation energies, a model is developed to predict the only quantity not calculated within openCOSMO-RS; the molar volume of the pure solvent. The model is based on a linear combination of descriptors, represented by the following equation:

$$\begin{aligned} v_{pure} &= 0.6977 A_\mathrm{CPCM} - 0.3161 M_2 + 0.03244 M_4 \\ &\quad + 0.9431 n_\mathrm{atoms} + 8.113 n_\mathrm{Si,atoms} - 0.07067 \end{aligned} \tag{4}$$

where A_{CPCM} is the area of the surface segments on the cavity of the solute, M_i are the respective sigma moments, n_atoms is the number of atoms and $n_\mathrm{Si,atoms}$ is the number of silicon atoms in the molecule. In general, the zeroth sigma moment M_0 corresponds to the area of the molecule (A_{CPCM}), the first sigma moment M_1 is the charge of the molecule (i.e. zero for neutral molecules), the second moment M_2 describes the polarity of the molecules, the third moment M_3 relates to the skewness of the sigma profile and the fourth moment M_4 measures the effect of strong local charge densities. For more information on sigma moments the reader is referred to the literature (e.g. [82–84]). All descriptor combinations were systematically evaluated. Notably, A_{CPCM} offers a more effective representation of the molar volume of the solvent compared to the volume of the cavity while utilizing fewer descriptors. The significant effect of the number of silicon atoms on the model's accuracy might suggest that the silicon radius might not be optimal. Fig. 1 shows the predicted molar volumes of the solvents using the QSPR molar volume of the solvent model against the experimental molar volumes of the solvents at 25 °C based on experimental data described more in detail in the previous section. Overall, the predicted molar volumes of the solvents agree well with the experimental ones with Average Absolute Deviation (AAD) of $3.48\,\mathrm{cm^3\,mol^{-1}}$ and $\mathrm{R^2}$ of 0.995. Mathieu and Bouteloup report a model for predicting the standard density with an average relative error <1.7%. For density prediction, our QSPR model achieves a relative error of 2.2%, which is an accuracy similar to that of other group contribution methods [85–87].

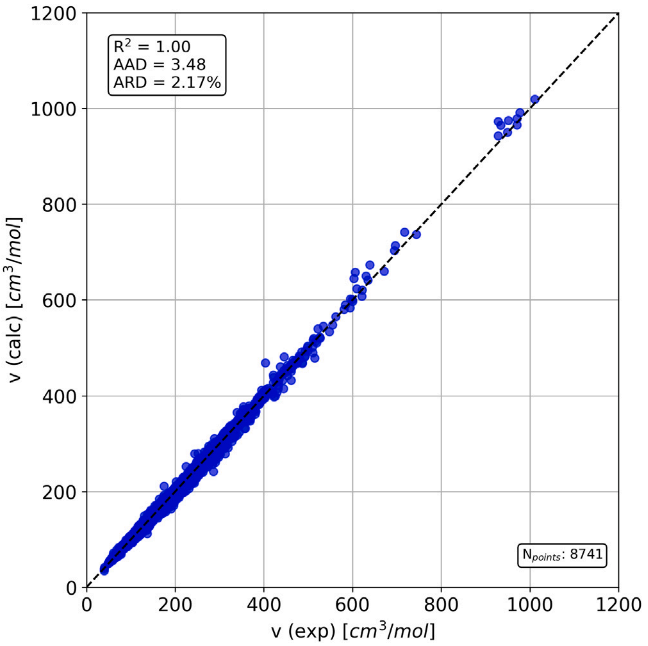


Fig. 1. Parity plot for the prediction of molar volume of the solvent model based on Eq. (4).

3.2. Parameterization

All non-fixed parameters in Table 1 are simultaneously adjusted using the differential evolution algorithm implemented in SciPy [70]. Following the approach used in openCOSMO-RS 22 [43], we incorporate the improved misfit term, which includes the additional descriptor $\sigma^\perp$ to recover some of the lost 3D information. All data are included in the regression of the parameters.

3.3. Model performance

In this work, infinite dilution activity coefficients, partition coefficients, and solvation free energies, all at 25 °C are used to parameterize openCOSMO-RS 24a. Table 2 provides an overview of all calculations for openCOSMO-RS 24a, COSMOtherm 24 BP-TZVP and COSMOtherm 24 BP-TZVPD-FINE, calculated with the lowest energy conformer. Figs. 2, 3, and 4 show the predicted values obtained from openCOSMO-RS 24a against the experimental values compiled from the literature for the activity coefficients, solvation free energies, and partition coefficients, respectively. In all Figures, black represents solutes without hydrogen bonds, red represents solutes that are hydrogen bond donors, and blue represents solutes that are hydrogen bond acceptors. Whether or not a solute is considered hydrogen bonding depends on the existence of area having a screening charge density larger than the threshold hydrogen bonding parameter σ_HB.

Table 1
Parameterization of openCOSMO-RS 24a based with gas and CPCM geometry optimizations at DFT/BP86/def2-TZVP level and gas and CPCM single point calculations at DFT/BP86/def2-TZVPD level in ORCA 6.0. [*] denotes the parameter was fixed. In this table, all parameters were converted to SI units. For actual use, the parameters have to be transformed into other units as shown on the example on github [62].

Parameter	Value	Parameter	Value
r^*_{av} [Å]	0.5	$\tau_1 \left[\frac{\text{kJ}}{\text{mol Å}^2}\right]$	0.123
a_{eff} [Å^2]	5.925	$\tau_6 \left[\frac{\text{kJ}}{\text{mol Å}^2}\right]$	0.096
$\alpha_{mf} \left[\frac{\text{kJ Å}^2}{\text{mol e}^2}\right]$	7281	$\tau_7 \left[\frac{\text{kJ}}{\text{mol Å}^2}\right]$	0.003
f^*_{corr} [−]	2.4	$\tau_8 \left[\frac{\text{kJ}}{\text{mol Å}^2}\right]$	0.015
$c_{hb} \left[\frac{\text{kJ Å}^2}{\text{mol e}^2}\right]$	43 327	$\tau_9 \left[\frac{\text{kJ}}{\text{mol Å}^2}\right]$	0.023
c^{T*}_{hb} [−]	1.5	$\tau_{17} \left[\frac{\text{kJ}}{\text{mol Å}^2}\right]$	0.143
σ_{hb} [e/Å^2]	0.00961	$\tau_{35} \left[\frac{\text{kJ}}{\text{mol Å}^2}\right]$	0.171
A_{std} [Å^2]	41.624	$\tau_{14} \left[\frac{\text{kJ}}{\text{mol Å}^2}\right]$	0.018
$\eta \left[\frac{\text{kJ}}{\text{mol}}\right]$	−18.61	$\tau_{15} \left[\frac{\text{kJ}}{\text{mol Å}^2}\right]$	0.015
$\omega_{ring} \left[\frac{\text{kJ}}{\text{mol}}\right]$	1.100	$\tau_{16} \left[\frac{\text{kJ}}{\text{mol Å}^2}\right]$	0.146

For the logarithm of the infinite dilution activity coefficients, a total of 882 data points are used (see Fig. 2). openCOSMO-RS 24a achieves an AAD of 0.51 and R^2 of 0.98, showing an improvement compared to our previous work openCOSMO-RS 22 [43], which had an AAD of 0.65 Overall, it can be observed, that for the less polar solutes, the infinite dilution activity coefficients are slightly underestimated while for the more polar ones are somewhat overestimated. The infinite dilution activity coefficient represents the energy required for transferring one molecule from pure component to being infinitely dilute in a second solvent. Hence, this systematic shift in model performance based on solute polarity might be either due to the overestimation of attractive hydrogen bonding for more polar molecules in the reference state (i.e. pure solute) or due to an overestimation of the repulsive misfit energy at infinite dilution in the solvent. This will be investigated further in future work.

For the logarithm of the partition coefficients, the dataset includes 296 data points (see Fig. 3). The openCOSMO-RS 24a models achieves an AAD of 0.76 and R^2 of 0.92, indicating good agreement between the calculated and the experimental data. Similar to IDAC, the model tends to overestimate partition coefficients for more polar solutes and slightly underestimate them for less polar ones. The partition coefficient measures the energy required to transfer a solute at infinite dilution from water to another solvent, representing the relative interaction energy of the solute with the other solvent compared to water. The data suggests that the greater the polarity difference is between the solute and the other solvent, the larger the deviation in calculated values by openCOSMO-RS 24a. As mentioned earlier, systems with water as a solute are excluded from the dataset because the usual COSMO-RS theory struggles to handle water at infinite dilution in very non-polar solvents [17,77]. This issue appears to extend to other polar molecules in non-polar solvents, though less pronounced than with water. This suggests a general systematic issue that could potentially be addressed to improve the model.

For the solvation free energies, the dataset contains 2129 data points (see Fig. 4). The openCOSMO-RS 24a model achieves an AAD of 0.45 kcal mol^{-1} and R^2 of 0.91, showing a strong agreement between the calculated and experimental values, which is impressive considering that the molecules in this study are represented by only a single conformer. In comparison, the commercial software COSMOtherm reports a similar uncertainty of 0.40 kcal mol^{-1} to 0.45 kcal mol^{-1} [27,31,32] for predicting the solvation free energy of neutral solutes, though it uses an ensemble of conformers. However, the comparison may not be entirely fair, as the parameters of openCOSMO-RS 24a are directly adjusted to the data used in this study, whereas the accuracy reported by was likely based on a larger dataset.

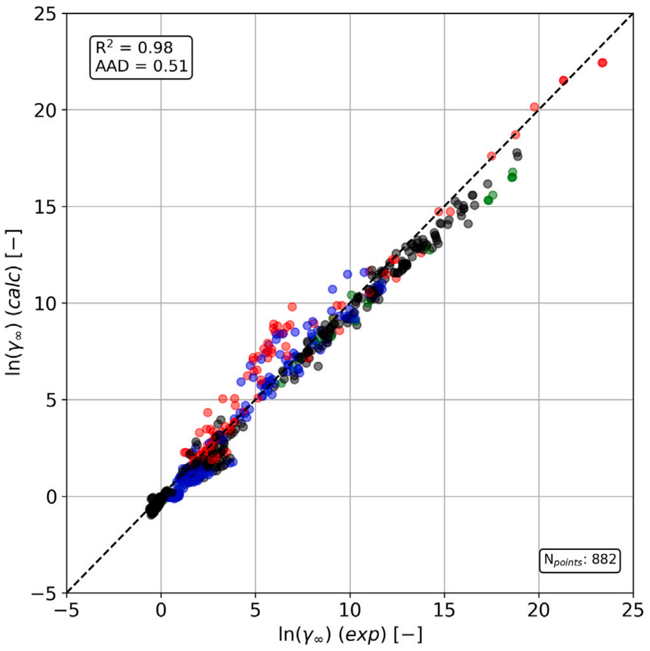


Fig. 2. Parity plot for infinite dilution activity coefficients calculated with openCOSMO-RS 24a. Colors represent different solute types: (●) non-HB, (●) HB acceptors, (●) HB donors and (●) HB donors/acceptors.

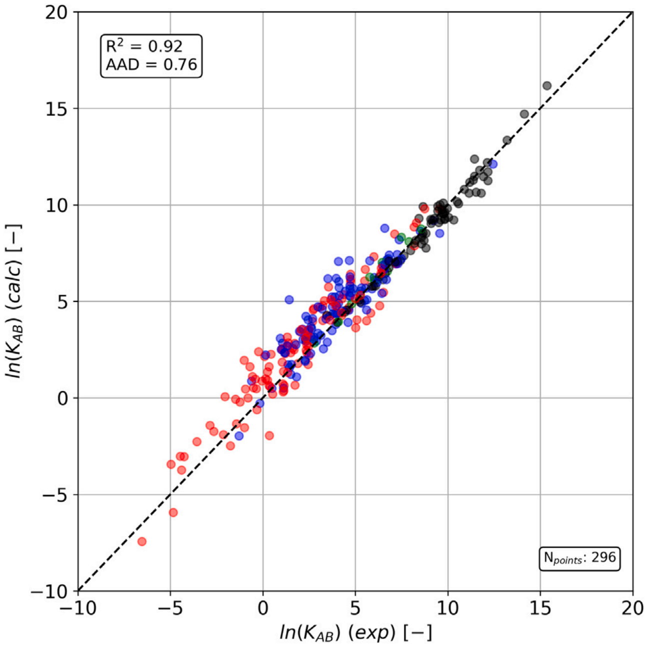


Fig. 3. Parity plot for partition coefficients at 25 °C calculated with openCOSMO-RS 24a whereby AB = [octanol/water, benzene/water, hexane/water, diethyl ether/water]. Colors represent different solute types: (●) non-HB, (●) HB acceptors, (●) HB donors and (●) HB donors/acceptors.

Table 2 presents a performance comparison of openCOSMO-RS 24a, COSMOtherm 24 BP-TZVP, and COSMOtherm 24 BP-TZVPD-FINE. All calculations are conducted using only the lowest energy conformer, providing a fair comparison since openCOSMO-RS 24a currently lacks the capability to integrate multiple conformers. Additional calculations

Table 2
Comparison of openCOSMO-RS 24a, COSMOtherm 24 BP-TZVP, and COSMOtherm BP-TZVPD-FINE for infinite dilution activity coefficients, partition coefficients, and solvation free energies. The calculations for all three models were performed only with the lowest energy conformer.

		openCOSMO-RS 24a	COSMOtherm 24 BP-TZVP	COSMOtherm 24 BP-TZVPD-FINE
IDAC/ln(γ_∞) [–]	$N_{\text{datapoints}}$	AAD	AAD	AAD
non-HB	568	0.41	0.43	0.40
HB acceptor	172	0.63	0.53	0.35
HB donor	35	0.64	0.93	0.40
HB acceptor/donor	107	0.78	0.66	0.33
Total	**882**	**0.51**	**0.50**	**0.38**
Partition coefficients/ln(K_{AB}) [–]				
non-HB	68	0.36	0.54	0.46
HB acceptor	104	0.84	0.63	0.50
HB donor	12	0.25	0.28	0.23
HB acceptor/donor	112	0.99	0.76	0.59
Total	**296**	**0.76**	**0.64**	**0.51**
Solvation free energies/ΔG_{solv} [kcal mol^{-1}]				
non-HB	434	0.36	0.34	0.32
HB acceptor	775	0.40	0.47	0.45
HB donor	69	0.52	0.39	0.32
HB acceptor/donor	851	0.54	0.53	0.40
Total	**2129**	**0.45**	**0.46**	**0.40**
Overall	**3307**	**0.49**	**0.50**	**0.41**

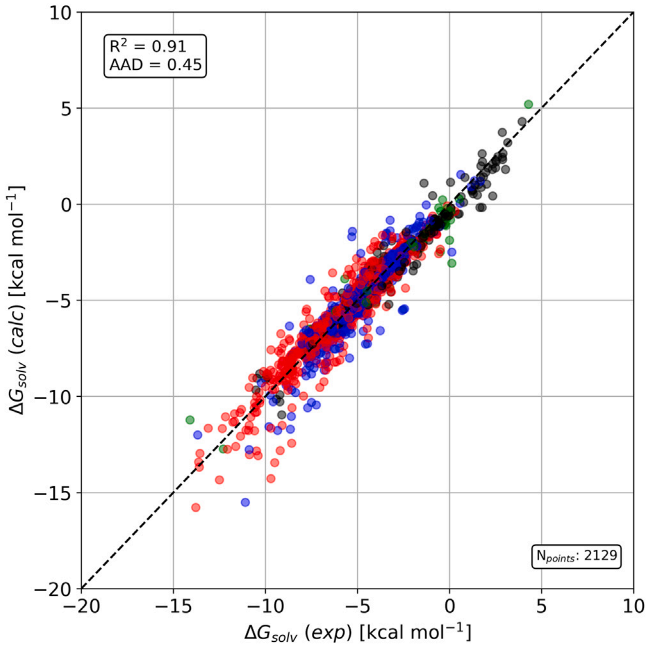


Fig. 4. Parity plot for solvation free energies at 25 °C calculated with openCOSMO-RS 24a. Colors represent different solute types: (●) non-HB, (●) HB acceptors, (●) HB donors and (●) HB donors/acceptors.

using multiple conformers for BP-TZVP and TZVPD-FINE parameterizations are included in the supplementary information. Nevertheless, for the dataset examined in this study, incorporating multiple conformers does not significantly impact the results. The results are categorized by solute type regarding its hydrogen binding capabilities. There is no universally accepted definition of a hydrogen-bonding molecule in the context of COSMO-RS modeling. However, we classify molecules based on the presence of areas in the σ-profile with an absolute screening charge density larger than the σ_{HB} threshold. This method allows for a consistent categorization of molecule types, which directly relates to the terms in the interaction equations.

Overall, for IDAC and solvation free energies, in the observed dataset, openCOSMO-RS 24a performs comparable to COSMOtherm 24 BP-TZVP. Only for partition coefficients COSMOtherm 24 TZVP is more accurate. The COSMOtherm 24 BP-TZVPD-FINE model delivers more accurate results for the majority of the systems in the dataset, this is the most advanced model available to date including many effects that are currently not described by openCOSMO-RS yet. However, this is a first step to improving openCOSMO-RS in a more general fashion for neutral molecules with current work focusing on improvements like the combinatorial term [88], temperature dependency, dispersion interactions, multiple conformers and polarizability effects.

4. Conclusions

In this work, we extended openCOSMO-RS to be able to calculate solvation free energies. To do so, we developed a QSPR model to predict the molar volumes of the solvents required for calculating solvation free energies. By modifying the openCOSMO-RS conformer workflow and combining in total 3307 data points of activity coefficients, solvation free energies, and partition coefficients, we parameterized openCOSMO-RS 24a based on ORCA 6.0.

The openCOSMO-RS 24a parameterization based on ORCA 6.0 achieved an AAD of 0.45 kcal mol^{-1} for predicting solvation free energies, which is comparable to the uncertainty of 0.45 kcal mol^{-1} that is reported by the commercial software COSMOtherm. For predicting the logarithm of the partition coefficients, openCOSMO-RS 24a achieves an AAD of 0.76. Furthermore, openCOSMO-RS 24a showed an improvement in predicting the logarithm of the activity coefficients at infinite dilution with an AAD of 0.51 compared to an AAD of 0.65 with the previous openCOSMO-RS 22 parameterization.

While the openCOSMO-RS 24a parameterization performs well when representing each molecule with a single conformer, future work will focus on extending the model to handle conformer ensembles. Additionally, we plan to extend openCOSMO-RS to ionic solutes, as current models often struggle to accurately predict the solvation free energies of ionic compounds [60,89,90].

The performance of openCOSMO-RS in predicting liquid phase properties, even with a single conformer, demonstrates that the inclusion of additional chemical descriptors to surface charge improves the model's accuracy. Moreover, substantial efforts have been made to integrate openCOSMO-RS into ORCA 6.0, enabling users to directly access a variety of liquid phase properties; a capability that was previously unavailable. This integration marks a significant advancement, providing users with a powerful tool for comprehensive property prediction within a single software environment.

CRediT authorship contribution statement

Simon Müller: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Thomas Nevolianis:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Miquel Garcia-Ratés:** Writing – review & editing, Software. **Christoph Riplinger:** Writing – review & editing, Software. **Kai Leonhard:** Writing – review & editing, Supervision, Funding acquisition. **Irina Smirnova:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Christoph Riplinger and Miquel Garcia-Ratés report a relationship with FACCTs GmbH that includes: employment and equity or stocks. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.fluid.2024.114250. All data is available either on the supplementary information or zenodo [81].

References

[1] S. Mahalakshmi, V. Sathyanarayanamoorthi, V. Kannappan, Studies on free energy and its components of 2-Phenylindole and its derivatives, J. Mol. Liq. 139 (1–3) (2008) 43–47, http://dx.doi.org/10.1016/j.molliq.2007.10.012.

[2] L.C. Kröger, W.A. Kopp, K. Leonhard, Prediction of chain propagation rate constants of polymerization reactions in aqueous NIPAM/BIS and VCL/BIS systems, J. Phys. Chem. B 121 (13) (2017) 2887–2895, http://dx.doi.org/10.1021/acs.jpcb.6b09147, arXiv:http://dx.doi.org/10.1021/acs.jpcb.6b09147.

[3] M. Heyden, Disassembling solvation free energies into local contributions—Toward a microscopic understanding of solvation processes, WIREs Comput. Mol. Sci. 9 (2) (2018) http://dx.doi.org/10.1002/wcms.1390.

[4] X. Zhang, R.S. DeFever, S. Sarupria, R.B. Getman, Free energies of catalytic species adsorbed to Pt(111) surfaces under liquid solvent calculated using classical and quantum approaches, J. Chem. Inf. Model. 59 (5) (2019) 2190–2198, http://dx.doi.org/10.1021/ACS.JCIM.9B00089.

[5] T. Nevolianis, N. Wolter, L.F. Kaven, L. Krep, C. Huang, A. Mhamdi, A. Mitsos, A. Pich, K. Leonhard, Kinetic modeling of a poly(n-vinylcaprolactam-co-glycidyl methacrylate) microgel synthesis: A hybrid in silico and experimental approach, Ind. Eng. Chem. Res. 62 (2) (2023) 893–902, http://dx.doi.org/10.1021/acs.iecr.2c03291, arXiv:https://doi.org/10.1021/acs.iecr.2c03291.

[6] N.D. Austin, N.V. Sahinidis, D.W. Trahan, A COSMO-based approach to computer-aided mixture design, Chem. Eng. Sci. 159 (2017) 93–105.

[7] N.D. Austin, N.V. Sahinidis, I.A. Konstantinov, D.W. Trahan, COSMO-based computer-aided molecular/mixture design: A focus on reaction solvents, AIChE J. 64 (1) (2018) 104–122.

[8] C. Gertig, L. Kröger, L. Fleitmann, J. Scheffczyk, A. Bardow, K. Leonhard, Rx-COSMO-CAMD: Computer-aided molecular design of reaction solvents based on predictive kinetics from quantum chemistry, Ind. Eng. Chem. Res. 58 (51) (2019) 22835–22846, http://dx.doi.org/10.1021/acs.iecr.9b03232, arXiv:https://doi.org/10.1021/acs.iecr.9b03232.

[9] T. Zhou, K. McBride, S. Linke, Z. Song, K. Sundmacher, Computer-aided solvent selection and design for efficient chemical processes, Curr. Opin. Chem. Eng. 27 (2020) http://dx.doi.org/10.1016/j.coche.2019.10.007.

[10] L. Raßpe-Lange, A. Hoffmann, C.U. Gertig, J. Heck, K. Leonhard, S. Herres-Pawlis, Geometrical benchmarking and analysis of redox potentials of copper(I/II) guanidine-quinoline complexes: Comparison of semi-empirical tight-binding and DFT methods and the challenge of describing the entatic state (part III), J. Comput. Chem. 44 (3) (2023) 319–328, http://dx.doi.org/10.1002/JCC.26927.

[11] T. Nevolianis, R.A. Ahmed, A. Hellweg, M. Diedenhofen, K. Leonhard, Blind prediction of toluene/water partition coefficients using COSMO-RS: results from the SAMPL9 challenge, Phys. Chem. Chem. Phys. 25 (2023) 31683–31691, http://dx.doi.org/10.1039/D3CP04077A.

[12] E.J. Smith, T. Bryk, A.D.J. Haymet, Free energy of solvation of simple ions: Molecular-dynamics study of solvation of Cl- and Na+ in the ice/water interface, J. Chem. Phys. 123 (3) (2005) 034706, http://dx.doi.org/10.1063/1.1953578, arXiv:https://doi.org/10.1063/1.1953578.

[13] C. Xi, F. Zheng, G. Gao, Z. Song, B. Zhang, C. Dong, X.-W. Du, L.-W. Wang, Ion solvation free energy calculation based on ab initio molecular dynamics using a hybrid solvent model, J. Chem. Theory Comput. (2022) null, http://dx.doi.org/10.1021/acs.jctc.1c01298, PMID: 36253911. arXiv:https://doi.org/10.1021/acs.jctc.1c01298.

[14] J. Tomasi, B. Mennucci, R. Cammi, Quantum mechanical continuum solvation models, Chem. Rev. 105 (8) (2005) 2999–3094, http://dx.doi.org/10.1021/cr9904009, PMID: 16092826. arXiv:https://doi.org/10.1021/cr9904009.

[15] A.V. Marenich, C.J. Cramer, D.G. Truhlar, Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions, J. Phys. Chem. B 113 (18) (2009) 6378–6396, http://dx.doi.org/10.1021/jp810292n, PMID: 19366259. arXiv:https://doi.org/10.1021/jp810292n.

[16] A. Klamt, Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena, J. Phys. Chem. 99 (1995) 2224–2235, http://dx.doi.org/10.1021/j100007a062.

[17] A. Klamt, V. Jonas, T. Bürger, J.C.W. Lohrenz, Refinement and parametrization of COSMO-RS, J. Phys. Chem. A 102 (26) (1998) 5074–5085, http://dx.doi.org/10.1021/jp980017s.

[18] 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, http://dx.doi.org/10.1146/annurev-chembioeng-073009-100903.

[19] M. Stahn, S. Ehlert, S. Grimme, Extended conductor-like polarizable continuum solvation model (CPCM-X) for semiempirical methods, J. Phys. Chem. A 127 (33) (2023-08-24) 7036–7043, http://dx.doi.org/10.1021/acs.jpca.3c04382, Publisher: American Chemical Society.

[20] H. Kang, H. Choi, H. Park, Prediction of molecular solvation free energy based on the optimization of atomic solvation parameters with genetic algorithm, J. Chem. Inf. Model. 47 (2) (2007) 509–514, http://dx.doi.org/10.1021/CI600453B.

[21] J. Gilmer, S.S. Schoenholz, P.F. Riley, O. Vinyals, G.E. Dahl, Neural message passing for quantum chemistry, 2017, CoRR abs/1704.01212. arXiv:1704.01212.

[22] A.M. Schweidtmann, J.G. Rittig, A. König, M. Grohe, A. Mitsos, M. Dahmen, Graph neural networks for prediction of fuel ignition quality, Energy Fuels 34 (9) (2020) 11395–11407, http://dx.doi.org/10.1021/acs.energyfuels.0c01533.

[23] B. Winter, C. Winter, J. Schilling, A. Bardow, A smile is all you need: Predicting limiting activity coefficients from SMILES with natural language processing, 2022, http://dx.doi.org/10.48550/ARXIV.2206.07048, CoRR abs/2206.07048. arXiv:2206.07048.

[24] E.I. Sanchez Medina, S. Linke, M. Stoll, K. Sundmacher, Graph neural networks for the prediction of infinite dilution activity coefficients, Digit. Discov. 1 (3) (2022) 216–225, http://dx.doi.org/10.1039/d1dd00037c.

[25] K.C. Felton, H. Ben-Safar, A.A. Lapkin, DeepGamma : A deep learning model for activity coeffcient prediction, 2021, URL https://api.semanticscholar.org/CorpusID:252086004.

[26] J.G. Rittig, K.B. Hicham, A.M. Schweidtmann, M. Dahmen, A. Mitsos, Graph neural networks for temperature-dependent activity coefficient prediction of solutes in ionic liquids, 2022, http://dx.doi.org/10.48550/ARXIV.2206.11776, CoRR abs/2206.11776. arXiv:2206.11776.

[27] T.M. Letcher, Development and Applications in Solubility, Royal Society of Chemistry, 2007.

[28] A.V. Marenich, C.J. Cramer, D.G. Truhlar, Generalized born solvation model SM12, J. Chem. Theory Comput. 9 (1) (2013) 609–620.

[29] D. Qiu, P.S. Shenkin, F.P. Hollinger, W.C. Still, The GB/SA continuum model for solvation. a fast analytical method for the calculation of approximate born radii, J. Phys. Chem. A 101 (16) (1997) 3005–3014, http://dx.doi.org/10.1021/jp961992r, arXiv:https://doi.org/10.1021/jp961992r.

[30] B.A.C. Horta, P.T. Merz, P.F.J. Fuchs, J. Dolenc, S. Riniker, P.H. Hünenberger, A GROMOS-compatible force field for small organic molecules in the condensed phase: The 2016H66 parameter set, J. Chem. Theory Comput. 12 (8) (2016) 3825–3850, http://dx.doi.org/10.1021/acs.jctc.6b00187, PMID: 27248705. arXiv:https://doi.org/10.1021/acs.jctc.6b00187.

[31] A. Klamt, M. Diedenhofen, Calculation of solvation free energies with DCOSMO-RS, J. Phys. Chem. A 119 (21) (2015) 5439–5445, http://dx.doi.org/10.1021/jp511158y.

[32] F.H. Vermeire, W.H. Green, Transfer learning for solvation free energies: From quantum chemistry to experiments, Chem. Eng. J. 418 (2021) 129307, http://dx.doi.org/10.1016/j.cej.2021.129307.

[33] S.-T. Lin, J. Chang, S. Wang, W.A. Goddard, S.I. Sandler, Prediction of vapor pressures and enthalpies of vaporization using a COSMO solvation model, J. Phys. Chem. A 108 (36) (2004-09) 7429–7439, http://dx.doi.org/10.1021/jp048813n.

[34] C.-C. Tsai, S.-T. Lin, Improved vapor pressure prediction from PR + COSMOSAC EOS using normal boiling temperature, AIChE J. 69 (3) (2023) e17997, http://dx.doi.org/10.1002/aic.17997, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/aic.17997.

[35] C.-M. Hsieh, S.-T. Lin, Determination of cubic equation of state parameters for pure fluids from first principle solvation calculations, AIChE J. 54 (8) (2008) 2174–2181, http://dx.doi.org/10.1002/aic.11552.

[36] C. Nait Saidi, D.C. Mielczarek, P. Paricaud, Predictions of solvation Gibbs free energies with COSMO-SAC approaches, Fluid Phase Equilib. 517 (2020) 112614, http://dx.doi.org/10.1016/j.fluid.2020.112614.

[37] F. Paes, R. Privat, J.-N. Jaubert, B. Sirjean, Predicting solvation energies of free radicals and their mixtures: A robust approach coupling the Peng-Robinson and COSMO-RS models, J. Mol. Liq. 401 (2024) 124641, http://dx.doi.org/10.1016/j.molliq.2024.124641.

[38] H. Lim, Y. Jung, MLSolvA: solvation free energy prediction from pairwise atomistic interactions by machine learning, J. Cheminform. 13 (1) (2021) 56, http://dx.doi.org/10.1186/S13321-021-00533-Z.

[39] Y. Chung, F.H. Vermeire, H. Wu, P.J. Walker, M.H. Abraham, W.H. Green Jr., Group contribution and machine learning approaches to predict Abraham solute parameters, solvation free energy, and solvation enthalpy, J. Chem. Inf. Model. 62 (3) (2022) 433–446, http://dx.doi.org/10.1021/ACS.JCIM.1C01103.

[40] J. Ferraz-Caetano, F. Teixeira, D.C. Santana, Explainable supervised machine learning model to predict solvation Gibbs energy, J. Chem. Inf. Model. 64 (7) (2024) 2250–2262, http://dx.doi.org/10.1021/ACS.JCIM.3C00544.

[41] K. Atz, F. Grisoni, G. Schneider, Geometric deep learning on molecular representations, 2021, CoRR abs/2107.12375. arXiv:2107.12375.

[42] H. Stärk, D. Beaini, G. Corso, P. Tossou, C. Dallago, S. Günnemann, P. Liò, 3D infomax improves GNNs for molecular property prediction, 2021, CoRR abs/2110.04126. arXiv:2110.04126.

[43] T. Gerlach, S. Müller, A.G. de Castilla, I. Smirnova, An open source COSMO-RS implementation and parameterization supporting the efficient implementation of multiple segment descriptors, Fluid Phase Equilib. 560 (2022) 113472, http://dx.doi.org/10.1016/j.fluid.2022.113472.

[44] T. Gerlach, S. Müller, I. Smirnova, Development of a COSMO-RS based model for the calculation of phase equilibria in electrolyte systems, AIChE J. 64 (1) (2018) 272–285, http://dx.doi.org/10.1002/aic.15875.

[45] S. Müller, A. González de Castilla, C. Taeschler, A. Klein, I. Smirnova, Evaluation and refinement of the novel predictive electrolyte model COSMO-RS-ES based on solid-liquid equilibria of salts and Gibbs free energies of transfer of ions, Fluid Phase Equilib. 483 (2019) 165–174, http://dx.doi.org/10.1016/j.fluid.2018.10.023.

[46] S. Müller, A. González de Castilla, C. Taeschler, A. Klein, I. Smirnova, Calculation of thermodynamic equilibria with the predictive electrolyte model COSMO-RS-ES: Improvements for low permittivity systems, Fluid Phase Equilib. 506 (2020) 112368, http://dx.doi.org/10.1016/j.fluid.2019.112368.

[47] A. González de Castilla, S. Müller, I. Smirnova, On the analogy between the restricted primitive model and capacitor circuits: Semi-empirical alternatives for over- and underscreening in the calculation of mean ionic activity coefficients, J. Mol. Liq. 326 (2021) 115204, http://dx.doi.org/10.1016/j.molliq.2020.115204.

[48] A. González de Castilla, S. Müller, I. Smirnova, On the analogy between the restricted primitive model and capacitor circuits. Part II: A generalized Gibbs-Duhem consistent extension of the Pitzer-Debye-Hückel term with corrections for low and variable relative permittivity, J. Mol. Liq. 360 (2022) 119398, http://dx.doi.org/10.1016/j.molliq.2022.119398.

[49] M. Arrad, K. Thomsen, S. Müller, I. Smirnova, Thermodynamic modeling using extended UNIQUAC and COSMO-RS-ES models: Case study of the cesium nitrate-water system over a large range of temperatures, Fluid Phase Equilib. 580 (2024) 114037, http://dx.doi.org/10.1016/j.fluid.2024.114037.

[50] F. Neese, The ORCA program system, WIREs Comput. Mol. Sci. 2 (1) (2012) 73–78, http://dx.doi.org/10.1002/wcms.81.

[51] F. Neese, Software update: the ORCA program system, version 4.0, WIREs Comput. Mol. Sci. 8 (1) (2018) http://dx.doi.org/10.1002/wcms.1327.

[52] F. Neese, F. Wennmohs, U. Becker, C. Riplinger, The ORCA quantum chemistry program package, J. Chem. Phys. 152 (22) (2020) 224108, http://dx.doi.org/10.1063/5.0004608.

[53] F. Neese, Software update: The ORCA program system—Version 5.0, WIREs Comput. Mol. Sci. 12 (5) (2022) e1606, http://dx.doi.org/10.1002/wcms.1606, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1606.

[54] D. Mathieu, R. Bouteloup, Reliable and versatile model for the density of liquids based on additive volume increments, Ind. Eng. Chem. Res. 55 (50) (2016) 12970–12980, http://dx.doi.org/10.1021/acs.iecr.6b03809.

[55] V. Barone, M. Cossi, Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model, J. Phys. Chem. A 102 (11) (1998) 1995–2001, http://dx.doi.org/10.1021/jp9716997.

[56] D.M. York, M. Karplus, A smooth solvation potential based on the conductor-like screening model, J. Phys. Chem. A 103 (50) (1999) 11060–11079, http://dx.doi.org/10.1021/jp992097l.

[57] M. Garcia-Ratés, F. Neese, Effect of the solute cavity on the solvation energy and its derivatives within the framework of the Gaussian charge scheme, J. Comput. Chem. 41 (9) (2020) 922–939, http://dx.doi.org/10.1002/JCC.26139.

[58] C.C. Pye, T. Ziegler, An implementation of the conductor-like screening model of solvation within the Amsterdam density functional package, Theor. Chem. Acc.: Theory Comput. Model. (Theor. Chim. Acta) 101 (6) (1999) 396–408, http://dx.doi.org/10.1007/s002140050457.

[59] A. Klamt, V. Jonas, Treatment of the outlying charge in continuum solvation models, J. Chem. Phys. 105 (22) (1996) 9972–9981, http://dx.doi.org/10.1063/1.472829.

[60] L.C. Kröger, S. Müller, I. Smirnova, K. Leonhard, Prediction of solvation free energies of ionic solutes in neutral solvents, J. Phys. Chem. A 124 (20) (2020) 4171–4181, http://dx.doi.org/10.1021/acs.jpca.0c01606, PMID: 32336096. arXiv:https://doi.org/10.1021/acs.jpca.0c01606.

[61] 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, http://dx.doi.org/10.1021/j100007a062, arXiv:https://doi.org/10.1021/j100007a062.

[62] openCOSMO-RS GitHub repository. URL https://github.com/TUHH-TVT.

[63] J. Ebejer, G.M. Morris, C.M. Deane, Freely available conformer generation methods: How good are they? J. Chem. Inf. Model. 52 (5) (2012) 1146–1158, http://dx.doi.org/10.1021/CI2004658.

[64] RDKit, version 2023.09.5. URL https://www.rdkit.org/.

[65] A.D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, Phys. Rev. A 38 (1988) 3098–3100, http://dx.doi.org/10.1103/PhysRevA.38.3098.

[66] J.P. Perdew, Density-functional approximation for the correlation energy of the inhomogeneous electron gas, Phys. Rev. B 33 (1986) 8822–8824, http://dx.doi.org/10.1103/PhysRevB.33.8822.

[67] F. Weigend, R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy, Phys. Chem. Chem. Phys. 7 (18) (2005) 3297–3305, http://dx.doi.org/10.1039/B508541A, Publisher: Royal Society of Chemistry.

[68] S. Ehlert, M. Stahn, S. Spicher, S. Grimme, Robust and efficient implicit solvation model for fast semiempirical methods, J. Chem. Theory Comput. 17 (7) (2021) 4250–4261, http://dx.doi.org/10.1021/acs.jctc.1c00471, Publisher: American Chemical Society.

[69] C. Bannwarth, S. Ehlert, S. Grimme, GFN2-xTB-An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions, J. Chem. Theory Comput. 15 (3) (2019) 1652–1671, http://dx.doi.org/10.1021/acs.jctc.8b01176.

[70] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, I. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy, SciPy 1.0-fundamental algorithms for scientific computing in Python, 2019, CoRR abs/1907.10121. arXiv:1907.10121.

[71] J.F. Parcher, P.H. Weiner, C.L. Hussey, T.N. Westlake, Specific retention volumes and limiting activity coefficients of C4-C8 alkane solutes in C22-C36 n-alkane solvents, J. Chem. Eng. Data 20 (2) (1975) 145–151, http://dx.doi.org/10.1021/je60065a015.

[72] E.C. Voutsas, D.P. Tassios, Prediction of infinite-dilution activity coefficients in binary mixtures with UNIFAC. A critical evaluation, Ind. Eng. Chem. Res. 35 (4) (1996) 1438–1445, http://dx.doi.org/10.1021/ie9503555, Number: 4.

[73] G.M. Kontogeorgis, G.I. Nikolopoulos, A. Fredenslund, D.P. Tassios, Improved models for the prediction of activity coefficients in nearly athermal mixtures Part II. A theoretically-based GE-model based on the van der Waals partition function, Fluid Phase Equilib. 127 (1) (1997) 103–121, http://dx.doi.org/10.1016/S0378-3812(96)03145-7, Number: 1.

[74] S. Kato, D. Hoshino, H. Noritomi, K. Nagahama, Infinite dilution activity coefficients of n-alkane solutes, butane to decane, in n-alkane solvents, heptane to hexatriacontane, Fluid Phase Equilib. 194–197 (2002) 641–652, http://dx.doi.org/10.1016/S0378-3812(01)00686-0.

[75] J. He, C. Zhong, A QSPR study of infinite dilution activity coefficients of organic compounds in aqueous solutions, Fluid Phase Equilib. 205 (2) (2003) 303–316, http://dx.doi.org/10.1016/S0378-3812(02)00296-0, Number: 2.

[76] A.V. Marenich, R.M. Olson, C.P. Kelly, C.J. Cramer, D.G. Truhlar, Self-consistent reaction field model for aqueous and nonaqueous solutions based on accurate polarized partial charges, J. Chem. Theory Comput. 3 (6) (2007) 2011–2033, http://dx.doi.org/10.1021/ct7001418, arXiv:https://doi.org/10.1021/ct7001418.

[77] A. Klamt, Prediction of the mutual solubilities of hydrocarbons and water with COSMO-RS, Fluid Phase Equilib. 206 (1) (2003) 223–235, http://dx.doi.org/10.1016/S0378-3812(02)00322-9.

[78] D. Ballal, P. Venkataraman, W.A. Fouad, K.R. Cox, W.G. Chapman, Isolating the non-polar contributions to the intermolecular potential for water-alkane interactions, J. Chem. Phys. 141 (6) (2014) 064905, http://dx.doi.org/10.1063/1.4892341.

[79] D. Ballal, D. Asthagiri, A.V. Parambathu, P. Venkataraman, W.A. Fouad, K.R. Cox, W.G. Chapman, Erratum: "Isolating the non-polar contributions to the intermolecular potential for water-alkane interactions" [J. Chem. Phys. 141, 064905 (2014)], J. Chem. Phys. 145 (11) (2016) 119901, http://dx.doi.org/10.1063/1.4962733.

[80] D. Asthagiri, A. Valiya Parambathu, D. Ballal, W.G. Chapman, Electrostatic and induction effects in the solubility of water in alkanes, J. Chem. Phys. 147 (7) (2017) 074506, http://dx.doi.org/10.1063/1.4997916.

[81] S. Müller, Parameterization dataset openCOSMO-RS 24a, 2024, http://dx.doi.org/10.5281/zenodo.13856201.

[82] A.M. Zissimos, M.H. Abraham, A. Klamt, F. Eckert, J. Wood, A comparison between the two general sets of linear free energy descriptors of Abraham and Klamt, J. Chem. Inf. Comput. Sci. 42 (6) (2002) 1320–1331, http://dx.doi.org/10.1021/ci025530o, Publisher: American Chemical Society.

[83] K. Wichmann, M. Diedenhofen, A. Klamt, Prediction of blood-βrain partitioning and human serum albumin binding based on COSMO-RS σ-moments, J. Chem. Inf. Model. 47 (1) (2007) 228–233, http://dx.doi.org/10.1021/ci600385w.

[84] K. Masuch, A. Fatemi, H. Murrenhoff, K. Leonhard, A COSMO-RS based QSPR model for the lubricity of biodiesel and petrodiesel components: Quantitative structure-property relationship lubricity, Lubr. Sci. 23 (6) (2011) 249–262, http://dx.doi.org/10.1002/ls.153.

[85] H.S. Elbro, A. Fredenslund, P. Rasmussen, Group contribution method for the prediction of liquid densities as a function of temperature for solvents, oligomers, and polymers, Ind. Eng. Chem. Res. 30 (12) (1991) 2576–2582, http://dx.doi.org/10.1021/ie00060a011.

[86] E.C. Ihmels, J. Gmehling, Extension and revision of the group contribution method GCVOL for the prediction of pure compound liquid densities, Ind. Eng. Chem. Res. 42 (2) (2003) 408–412, http://dx.doi.org/10.1021/ie020492j, arXiv:https://doi.org/10.1021/ie020492j.

[87] A.S. Hukkerikar, B. Sarup, A. Ten Kate, J. Abildskov, G. Sin, R. Gani, Group-contribution+ (GC+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis, Fluid Phase Equilib. 321 (2012) 25–43, http://dx.doi.org/10.1016/j.fluid.2012.02.010.

[88] G.J.P. Krooshof, G. de With, Gibbs probability entropy and its implication to combinatorial entropy models, Fluid Phase Equilib. 584 (2024) 114146, http://dx.doi.org/10.1016/j.fluid.2024.114146.

[89] T. Nevolianis, M. Baumann, N. Viswanathan, W.A. Kopp, K. Leonhard, DISSOLVE: Database of ionic solutes' solvation free energies, Fluid Phase Equilib. 571 (2023) 113801, http://dx.doi.org/10.1016/j.fluid.2023.113801.

[90] J.W. Zheng, W.H. Green, Experimental compilation and computation of hydration free energies for ionic solutes, J. Phys. Chem. A 127 (48) (2023) 10268–10281, http://dx.doi.org/10.1021/acs.jpca.3c05514.