The Journal of Chemical Physics

Generalized molecular solvation in non-aqueous solutions by a single parameter implicit solvation scheme

Cite as: J. Chem. Phys. **150**, 041710 (2019); https://doi.org/10.1063/1.5050938
Submitted: 03 August 2018 . Accepted: 18 October 2018 . Published Online: 05 December 2018

Christoph Hille, Stefan Ringe, Martin Deimel, Christian Kunkel, William E. Acree, Karsten Reuter, and Harald Oberhofer

AIP Publishing

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Generalized molecular solvation in non-aqueous solutions by a single parameter implicit solvation scheme

Cite as: J. Chem. Phys. 150, 041710 (2019); doi: 10.1063/1.5050938
Submitted: 3 August 2018 • Accepted: 18 October 2018 •
Published Online: 5 December 2018

Christoph Hille,[1,a)] Stefan Ringe,[2,a),b)] Martin Deimel,[1] Christian Kunkel,[1] William E. Acree,[3] Karsten Reuter,[1] and Harald Oberhofer[1]

AFFILIATIONS

[1] Chair for Theoretical Chemistry and Catalysis Research Center, Technische Universität München, Lichtenbergstr. 4, 85747 Garching, Germany
[2] SUNCAT Center for Interface Science and Catalysis, Department of Chemical Engineering, Stanford University, Stanford, California 94305, USA
[3] Department of Chemistry, University of North Texas, 1155 Union Circle Drive #305070, Denton, Texas 76203, USA

[a)] C. Hille and S. Ringe contributed equally to this work.
[b)] **Electronic mail:** sringe@stanford.edu

ABSTRACT

In computer simulations of solvation effects on chemical reactions, continuum modeling techniques regain popularity as a way to efficiently circumvent an otherwise costly sampling of solvent degrees of freedom. As effective techniques, such implicit solvation models always depend on a number of parameters that need to be determined earlier. In the past, the focus lay mostly on an accurate parametrization of water models. Yet, non-aqueous solvents have recently attracted increasing attention, in particular, for the design of battery materials. To this end, we present a systematic parametrization protocol for the Self-Consistent Continuum Solvation (SCCS) model resulting in optimized parameters for 67 non-aqueous solvents. Our parametrization is based on a collection of ≈6000 experimentally measured partition coefficients, which we collected in the Solv@TUM database presented here. The accuracy of our optimized SCCS model is comparable to the well-known universal continuum solvation model (SMx) family of methods, while relying on only a single fit parameter and thereby largely reducing statistical noise. Furthermore, slightly modifying the non-electrostatic terms of the model, we present the SCCS-P solvation model as a more accurate alternative, in particular, for aromatic solutes. Finally, we show that SCCS parameters can, to a good degree of accuracy, also be predicted for solvents outside the database using merely the dielectric bulk permittivity of the solvent of choice.

I. INTRODUCTION

Today, the importance of an accurate treatment of solvation effects in simulating chemical reactions in liquids is undisputed. To name one highly relevant example, simulations of electro-chemical systems cannot give a complete picture without considering the response of the solvent medium to the inherent transport of the charged species.[1–4] Unfortunately, the greatly increased computational cost of sampling the large number of–chemically not directly relevant–solvent degrees of freedom often restricts atomistic simulations.[5,6] This is particularly pronounced in first-principles based approaches, where only very small solvent boxes are tractable or where only less dynamic immediate solvation shells are considered.[7,8]

Circumventing such problems with the explicit treatment of solvation, the old idea of *implicit* solvation models[9,10] has been undergoing a renaissance in the last ten to fifteen years.[5,11–26] There, solvation is treated, e.g., within

Density Functional Theory (DFT), at the level of an effective continuum, which is polarizable and includes additional terms for non-electrostatic contributions such as dispersion interactions, cavity formation energy (related to the solvent's surface tension), or entropic contributions.[18,22,25] The effective nature of the models leads to a number of—solvent dependent—parameters which, in general, cannot be determined from first principles, but rather are determined by fitting the respective model to sufficiently large databases of experimental reference data—most commonly, the solvation free energies of diverse sets of molecules.[5,13,16,22,24,26–29]

Owing to its importance as a solvent, and the complexities involved in its explicit simulation from first principles, by far most of the solvation parameter sets published to date concentrate on water.[5,11,22,25–27,30] Non-aqueous solvents, on the other hand, play important roles in diverse fields such as crystal synthesis[31] and liquid-liquid interface chemistry,[32] or as stable electrolytes which also protect the electrodes in next-generation batteries.[33] Yet, the optimization of implicit solvation models for such non-aqueous solvents using statistically converged experimental training sets has to date rarely been achieved, leading to large errors in the prediction of solvation energies.[34] Notable exceptions to this are given in the Miertuš-Scrocco-Tomasi (MST) model by Miertuš and co-workers[11–13,16,35] and the well-known universal continuum solvation model (SMx) family of methods[20,21,24] based on the Minnesota (MNsol) database of reference data.[36] These models have been widely applied in various fields of science.[37]

In this work, we make use of the Self-Consistent Continuum Solvation (SCCS) model, first described and implemented in the pseudo-potential based QUANTUM ESPRESSO[38] DFT package by Andreussi *et al.*[22] and recently adopted by us in the full-potential numeric atomic orbital program FHI-aims (Fritz Haber Institute *ab initio* molecular simulations).[25,39] In this context, we also showed that the SCCS model can be efficiently combined with a Poisson-Boltzmann description of dissolved ion distributions,[25,28] which are of high relevance, in particular, in energy conversion and storage processes.[40–43] The main strength of the SCCS approach, however, lies in the description of solvation energies relying on merely four adjustable parameters, from which—as we show below—only a single parameter needs to be effectively adjusted to each solvent of choice. This dramatically reduces the parameter space compared to the roughly 30-100 parameters appearing, for example, in the SMx family of models.[21,44] This greatly improves the statistical stability of the fitting procedure, in particular, for small experimental datasets without losing accuracy in the prediction.

Here, we specifically present optimized SCCS parameters for neutral molecules in 67 non-aqueous solvents and water. The values were obtained by a rigorous fitting procedure reproducing solvation free energies from an extensive new database of ≈6000 experimentally measured solvation partition coefficients which we collected. This database, described in more detail in Sec. II D, can be obtained from Ref. 45. In order to allow the usage of the SCCS model for solvents outside of the current database, we also present a first parameter prediction model merely based on the dielectric bulk permittivity of the solvent.

Finally, in the course of fitting our parameter sets, we found a way to further improve the accuracy of the non-electrostatic contributions of the SCCS model by replacing its dependence on the volume of the solvation cavity with the solute's isotropic polarizability (cf. Sec. III C). Especially for aromatic solutes, our results show a marked improvement of this SCCS+polarizability (SCCS-P) model over standard SCCS. With the new SCCS parameter sets and the parameter prediction method, as well as the new SCCS-P model, we aim to open up non-aqueous continuum solvation calculations to a wider user-base.

This work is organized as follows: In order to place the significance of the fitted parameters into context, we first give an outline of the theoretical background of the model, the efficient reduction of the parameter space, and the fitting procedure in Sec. II. Then we describe the database used in the fitting procedure in Sec. II D. Our new parameter sets together with tests of their accuracy are given in Sec. III, where we also present our improved SCCS-P model as well as the parameter prediction scheme for solvents not contained in the database before giving some concluding remarks in Sec. IV.

II. METHODS

A. The SCCS implicit solvation model

The basis of all implicit solvation models is the formulation of a Generalized Poisson Equation (GPE), given here in atomic units as[5,25]

$$\nabla \cdot [\varepsilon[n_{\mathrm{el}}]\nabla v] = -4\pi n_{\mathrm{sol}}, \tag{1}$$

with v being the electrostatic potential, $n_{\mathrm{sol}} = n_{\mathrm{el}} + n_{\mathrm{nuc}}$ being the solute's charge density, comprising electronic and nuclear parts, and $\varepsilon[n_{\mathrm{el}}]$ representing a model function for the isotropic, static dielectric solvent permittivity. The latter has most often been parametrized in terms of the solute's electron density (provided by DFT).[22,25] In the SCCS model, it is defined as

$$\varepsilon_{\delta_n,n_c}[n_{\mathrm{el}}] = \begin{cases} 1 & -\ln(n_{\mathrm{el}}) < n_c - \frac{1}{2}\delta_n \\ \exp(t_{\delta_n,n_c}(\ln(n_{\mathrm{el}}))) & n_c - \frac{1}{2}\delta_n > -\ln(n_{\mathrm{el}}) > n_c + \frac{1}{2}\delta_n \\ \varepsilon^{\mathrm{s,bulk}} & -\ln(n_{\mathrm{el}}) > n_c + \frac{1}{2}\delta_n \end{cases}, \tag{2}$$

with the switching function

$$t_{\delta_n,n_c}(\ln(n_{\mathrm{el}})) = \frac{\ln(\varepsilon^{\mathrm{s,bulk}})}{2\pi}\left[2\pi\frac{\frac{1}{2}\delta_n - n_c - \ln(n_{\mathrm{el}})}{\delta_n} - \sin\left(2\pi\frac{\frac{1}{2}\delta_n - n_c - \ln(n_{\mathrm{el}})}{\delta_n}\right)\right]. \tag{3}$$

Here, $\varepsilon^{s,bulk}$ is the isotropic, static bulk permittivity of the solvent. In contrast to the original implementation of the SCCS model defining the dielectric transition region by means of minimum and maximum electron densities n_{min} and n_{max},[22,25] we switched above to a different representation by transforming the parameters via

$$\delta_n = -(\ln(n_{min}) - \ln(n_{max})), \tag{4}$$

$$n_c = -\frac{1}{2}(\ln(n_{min}) + \ln(n_{max})), \tag{5}$$

with n_c describing the electron density cutoff of the transition and δ_n representing the smoothness. Due to the exponential decay of the electron density, this logarithmic transformation leads to a more homogeneous grid density and therefore a more stable discrete grid representation of the parameter space. Ultimately, this enables a reduction of the parameter space to a single dimension.

The model presented so far depicts only the purely electrostatic interactions between the solute and solvent, based on the mean-field approximation. Additional non-electrostatic, non-mean-field correction terms can be effectively added to the energy functional. They comprise cavity-formation, Pauli-repulsion, and dispersion interaction parts

$$G^{non\text{-}mf} = G^{cav} + G^{rep} + G^{dis}. \tag{6}$$

In the SCCS model, $G^{non\text{-}mf}$ is expressed in terms of the quantum surface $S[n_{el}]$ and volume $V[n_{el}]$ of the solvation cavity which are evaluated from a finite difference scheme using the switching function ϑ_{δ_n,n_c}

$$G^{non\text{-}mf}(\boldsymbol{p})[n_{el}] = (\alpha+\gamma)S_{\delta_n,n_c}[n_{el}] + \beta V_{\delta_n,n_c}[n_{el}], \tag{7}$$

where γ represents the solvent's surface tension and α and β are effective solvent-specific parameters.

In total, the SCCS model therefore gives rise to a four-dimensional parameter space represented by the vector

$$\boldsymbol{p} = \begin{pmatrix} \delta_n \\ n_c \\ (\alpha+\gamma) \\ \beta \end{pmatrix} \in \mathbb{R}^4. \tag{8}$$

The non-mean-field energy correction is added to the Kohn-Sham (KS) energy functional yielding the general SCCS energy functional which needs to be minimized. This minimization results in the GPE and the KS equation having to be solved in a self-consistent manner. In order to solve the GPE, both QUANTUM ESPRESSO and FHI-aims utilize a self-consistent relaxation scheme using the vacuum Poisson Green's function as a preconditioner. QUANTUM ESPRESSO additionally supports a conjugate gradient scheme which uses a preconditioner representing the Poisson equation in the limit of a slowly varying dielectric constant.[29] For a more detailed description, the reader is referred to the respective publications.[22,25,29]

B. Technical remarks

The self-consistent treatment of energy corrections in the form of Eq. (7) is known to sometimes lead to numerical instabilities.[46] In addition, it dramatically increases the computational costs of any parameter optimization scheme as it requires DFT-SCCS calculations on the full four-dimensional parameter space. In our earlier work,[25] however, we found that Eq. (7) can also be added as a post self-consistent field (SCF) correction with only a sub-meV effect on the calculated solvation energies. This implies that the functional derivatives $\frac{\delta G^{non\text{-}mf}(\boldsymbol{p})[n_{el}]}{\delta n_{el}}$ added to the KS-Hamiltonian play only a minor role for calculated solvation energies. In the present work, we verified this result also for a low-dielectric solvent (chloroform) (cf. Fig. S1 in the supplementary material). Due to this property of the DFT-SCCS approach, the parameter space to be explored numerically can be reduced to only two dimensions (n_c and δ_n). $G^{non\text{-}mf}(\boldsymbol{p})[n_{el}]$ is then always added as an analytic extension of the parameter space being coupled to the DFT-SCCS grid via S_{δ_n,n_c} and V_{δ_n,n_c}.

A further important question for the computational efficiency of the optimization procedure is to which degrees the geometries change due to the presence of the solvent. Our database consists of small neutral organic molecules, for which we generally found no major changes in geometries under the influence of the implicit solvent. Due to this fact, we perform all DFT-SCCS calculations as single-point calculations using optimized geometries in a vacuum. We note, however, that our scheme is trained for predicting the variation of solvation energies among solutes differing in geometry and functionality. From this, we expect that some degree of geometry dependence of the solvation energy is also effectively included. Some transferability to larger solutes with stronger rearrangements in solvents, e.g., amino acids, can therefore be assumed, although this assumption has to be validated in future work.

All DFT-SCCS calculations in this paper are performed using the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional[47] using the all-electron program package FHI-aims as well as "tight" convergence settings.[39,48] Systematic test calculations using water as a solvent indicated that such settings yield a numerical convergence of solvation energies in the sub-meV region.[25] In Fig. S3 of the supplementary material, we show that this result also holds for non-aqueous solvents, here tested for the low-dielectric solvent chloroform using our optimized parameters described in Table I. The parameter optimization scheme (cf. Sec. III B) was implemented in the Python programming language using common scientific packages such as Scipy, Numpy, and Pandas.

C. Solvation free energies

Following the previous definitions, solvation energies are then calculated via

$$\Delta G^{sol}(\boldsymbol{p}) = \Delta G^{el}_{\delta_n,n_c} + (\alpha+\gamma)S^{\circ}_{\delta_n,n_c} + \beta V^{\circ}_{\delta_n,n_c}, \tag{9}$$

TABLE I. Optimized ($\alpha + \gamma$) parameters (in dyn/cm) for the SCCS model for 67 non-aqueous solvents. A corresponding value for water evaluated on the Shivakumar training set[132] is also given for comparison. For all solvents, $\delta_n = 2.0$, $n_c = 7.6$, and $\beta = -0.5$ GPa. "ID" is the unique molecular solute identifier in the database (only for solvents, which also occur as solutes in the database).

	ID	$(\alpha+\gamma)$ (dyn/cm)	RMSD (meV)
1,4-dioxane[53,123]	055	36.9	64.2
1-chlorobutane[138]	125	41.0	30.4
1-hexadecene[60]	...	40.1	25.2
1-methyl-2-piperidinone[55]	...	46.4	48.6
1-octanol[51]	088	44.6	42.1
1-propanol[134,137]	083	45.4	33.8
2,2,4-trimethylpentane[145]	118	38.4	25.6
2-ethoxyethanol[130]	257	43.9	35.6
4-formylmorpholine[55]	620	46.5	28.6
Acetic acid[146]	150	43.1	46.0
Acetone[57]	059	43.7	41.2
Acetophenone[61]	263	44.5	29.5
Benzonitrile[59]	219	43.8	42.9
Bromobenzene[56]	231	41.8	37.7
Butanone[148]	060	44.0	38.2
Butyl acetate[135]	067	40.4	45.1
Carbon tetrachloride[106,138]	041	36.5	38.6
Chlorobenzene[56]	102	42.3	49.0
Chloroform[138]	040	37.7	34.7
Cyclohexanone[148]	251	43.8	30.4
Decan-1-ol[137]	131	42.4	42.5
Decane[144]	104	40.2	27.3
Dibutyl ether[51,58]	050	39.3	29.1
Dichloroethane[136]	146	43.1	42.5
Dichloromethane[138]	039	41.9	29.5
Diethyl ether[53,58]	047	40.1	18.2
Diiodomethane[50]	423	44.8	52.5
Dimethylformamide[55]	117	46.3	39.0
Dodecane[54]	246	41.4	35.6
Ethanol[137]	082	44.6	34.3
Ethyl acetate[136]	116	40.9	27.3
Ethylbenzene[142]	096	37.4	47.7
Ethylene glycol[58]	...	51.5	47.3
Formamide[55,124]	625	53.2	46.4
Heptane[105,106,144]	022	38.2	26.9
Hexadecane[49,80]	105	39.8	32.5
Hexane[144]	020	38.8	29.1
Iodobenzene[56]	283	40.3	42.9
Isobutanol[139]	151	43.8	38.6
Isopropanol[134,139]	112	45.4	29.5
m-xylene[143]	217	38.2	27.3
Methanol[137]	081	46.4	39.5
Methoxyethanol[91]	464	44.9	34.3
Methyl acetate[135]	065	41.9	26.5
Methyl-cyclohexane[63,96,97,99,118,120,155]	159	39.2	36.0
n,*n*-dibutylformamide[55]	...	44.8	37.3
n,*n*-dimethylacetamide[55]	485	45.5	46.4
n-butanol[137]	084	44.8	35.1
n-ethylacetamide[55]	...	47.0	41.6
n-methylacetamide[55,134]	621	47.1	44.7
n-methylformamide[55]	466	48.7	25.6
n-methylpyrrolidone[55,108]	615	46.9	36.0
o-xylene[143]	158	37.8	47.3
Octane[144]	024	40.4	24.7
p-xylene[143]	114	37.9	44.2
Pentane[79,94,98,100,107,109,119,151]	018	38.5	20.4
Pentanol[137]	085	44.6	41.2
Phenylamine[61]	232	43.7	36.9
Propylene carbonate[58,108]	478	49.1	38.6
Pyridine[129]	115	42.7	32.1
Sulfolane[141]	...	47.8	45.5
tert-butanol[139]	153	43.5	29.5
Tetraethylene glycol dimethyl ether[92]	...	43.0	36.9
Tetrahydrofuran[53,123]	054	40.3	31.2
Toluene[142]	095	38.4	33.4
Tributyl phosphate[68]	335	42.5	24.7
Undecane[54]	245	37.7	39.0
Water[132]	080	57.2	79.4
Average all			37.1 meV (0.856 kcal/mol)
Average non-aqueous			36.4 meV (0.839 kcal/mol)

with the electrostatic energy difference

$$\Delta G^{\mathrm{el}}_{\delta_n,n_c} = E^{\circ}_{\mathrm{sol},\delta_n,n_c} - E^{\circ}_{\mathrm{vac}}. \tag{10}$$

We use the superscript $\circ$ to indicate a property calculated from the fully self-consistent electron density and electrostatic potential. $E^{\circ}_{\mathrm{sol},\delta_n,n_c}$ is the respective total energy resulting from a DFT-SCCS calculation, while E°_{vac} is the same energy in the absence of SCCS described solvent (a regular vacuum DFT calculation).

The experimental values in our database are not the solvation free energies themselves, but rather given as decadic logarithm values of the partition coefficients K, which were determined from experimental solvation data (cf. Sec. II D). The partition coefficient K is defined by

$$K_{\mathrm{solvent/air}} = \frac{[\mathrm{solute}]_{\mathrm{solvent}}}{[\mathrm{solute}]_{\mathrm{air}}}, \tag{11}$$

where $[\mathrm{solute}]_{\mathrm{solvent}}$ is the equilibrium concentration of the solute in the organic solvent phase and $[\mathrm{solute}]_{\mathrm{air}}$ is the equilibrium concentration in the gas phase. Hence, the partition coefficient describes the distribution of a substance between two phases, in our case the organic and the gas phase. Using the concentration dependence of the chemical potential, the standard-state solvation free energy for neutral solutes can then be computed from the partition coefficient as

$$\Delta G^{\mathrm{sol,exp}} = -RT \ln(K_{\mathrm{solvent/air}}), \tag{12}$$

where R is the ideal gas constant and T is the temperature.

D. Solv@TUM, the non-aqueous solvation energy database

All of the parameters presented in this work were fitted to training data collected by us in the "Solv@TUM"

database, hosted for download at the Technical University of Munich. All of the experimental partition coefficients used for the determination of the reference free energies were taken from compilations in publications of Abraham, Acree, and co-workers (cf. Table I for specific listing of references).[49–157] The authors calculated the partition coefficients from published experimental infinite dilution activity coefficient, Henry's law constant, and mole fraction solubility data using standard thermodynamic relationships. For those organic solvents that were almost completely immiscible with water, partition coefficients were calculated from published water-to-organic solvent partition coefficient data. For the latter water-organic solvent partitioning systems, the slight degree of mutual solvent miscibility has little effect on the partitioning behavior of the solute.

The different origin of the partition coefficients makes it difficult to assess the general accuracy of the data. Nevertheless, based on those solute/solvent combinations measured independently by a number of research groups, we estimate the worst case errors of our log K values as ±0.08. This implies an experimental accuracy of about ±5 meV for the solvation free energies contained in the Solv@TUM database. By comparing the MNsol database of solvation free energies with the Solv@TUM database (cf. Fig. S4 in the supplementary material), we found a mean absolute deviation of 10 meV, so we generally expect the experimental uncertainty to be within 5-10 meV. In total, 5952 partition coefficients of neutral molecules were measured, comprising 148 non-aqueous solvents and 658 solutes. These solutes and solvents range from small gaseous molecules to large, chemically diverse organic and inorganic molecules, mainly composed of the classic elements of organic chemistry, i.e., C, H, N, O, F, Cl, S, Br, and I, with a smaller subset of 13 molecules containing Si, Ge, Pb, Sn, Fe, and P. In addition, data on monoatomic species (i.e., Hg, as well as noble gases He, Ne, Ar, Kr, Xe, and Rn) are also included. This set thereby covers a variety of common functional groups[158] as illustrated in Fig. 1(a). The database contains 884 unique atomic environments, measured by considering all neighbors up to the 2nd neighbor shell.[159,160] As expected from this analysis, and most important for the fitting models based largely on electrostatics, the isotropic polarizabilities [evaluated by Density Functional Perturbation Theory (DFPT)[161]] and dipole moments of the database molecules cover wide ranges of 0.2–29.9 Å^3 (in cgs units) and 0.0–7.8 D, respectively, see Figs. 1(b) and 1(c). Previous studies on water as a solvent suggested that the optimized parameters derived from the reference data of purely neutral solutes are not directly applicable to charged systems.[23] A future goal is therefore the extension of Solv@TUM to the case of charged solutes that will allow the derivation of parameters particularly tailored for the modeling of charged systems as of importance in many electrochemical systems.

In the final database, we made sure that each solute entry is unique and has a unique name identifier. The partition coefficients, vacuum optimized geometries, and polarizabilities and dipole moments of all solutes are provided as a single structure data file (sdf). To ensure a fast and straightforward access, a database interface on the basis of Pybel[162] and OpenBabel[163] was implemented. This enables a simple, fast, and intuitive filtered search through the database. Further features of the database interface and the exact usage are described in more detail in the manual. The database and the interface are provided free-of-charge in Ref. 45.

E. Cost function

The accuracy of the computationally predicted solvation free energies with respect to the experimental reference data has to be quantified by an appropriate error function. Most common choices are the Mean Absolute Error (MAE)

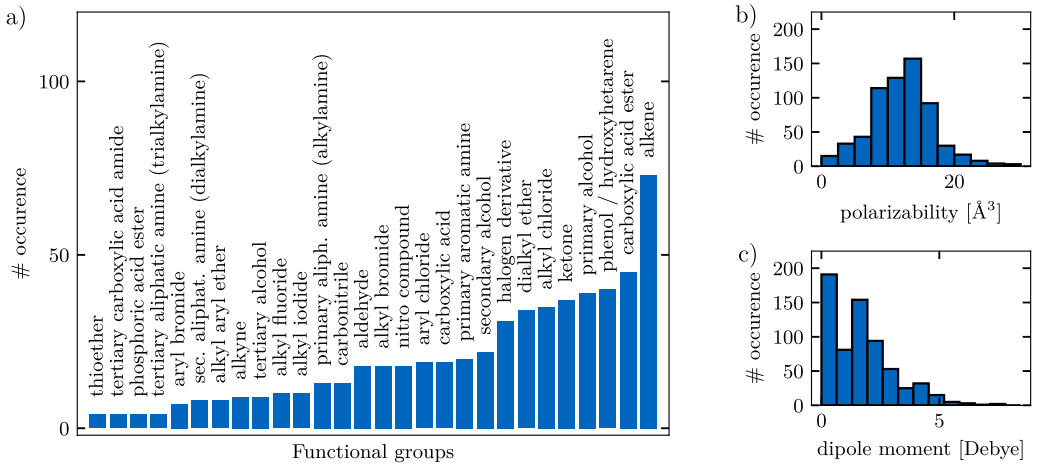

FIG. 1. (a) Chemical diversity of the generated solvation free energy database (Solv@TUM). The histogram shows the occurrence of the different functional groups of the solutes. Only functional groups occurring more than three times are shown in this diagram. [(b) and (c)] Molecular polarizabilities and dipole moments of all solutes.

$$\mathrm{MAE}(\boldsymbol{p}) = \frac{1}{N}\sum_{i=1}^{N}\left|\Delta G_i^{\mathrm{sol}}(\boldsymbol{p}) - \Delta G_i^{\mathrm{sol,exp}}\right|, \tag{13}$$

and the Root Mean Square Deviation (RMSD) functions

$$\mathrm{RMSD}(\boldsymbol{p}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\Delta G_i^{\mathrm{sol}}(\boldsymbol{p}) - \Delta G_i^{\mathrm{sol,exp}})^2}, \tag{14}$$

or even both. Here, $\boldsymbol{p}$ denotes a given choice of parameters. The sums in the above equations go over all N solutes for which reference data for the solvent of choice is available. Note that thereby we fit the parameters for each solvent independently.

While the MAE was extensively used to assess the quality of implicit solvent parametrizations,[21,22,24] we note here that it represents a uniform weighting of the errors for each data point. The RMSD, in contrast, has a tendency of being dominated by the (few) points with the largest deviation. It is therefore more suitable for validating the transferability of an effective model which we ultimately want to show here. In order to demonstrate this, all following results–if not mentioned otherwise–are further evaluated with respect to an independent test set instead of a training set of reference data.

Use of the RMSD is also the better choice for the parameter optimization cost function. Being defined as a sum over absolute values, the MAE is only a piecewise-defined function consisting of linear, affine functions. This makes it non-differentiable and therefore less suitable as a cost function for most optimization protocols. Instead, we thus minimize here the residual sum of squares (RSS) function

$$\min_{\boldsymbol{p}} \mathrm{RSS}(\boldsymbol{p}), \tag{15}$$

which is defined as

$$\mathrm{RSS}(\boldsymbol{p}) = N \cdot \mathrm{RMSD}^2(\boldsymbol{p}), \tag{16}$$

and has its minima at the same positions in parameter space as the RMSD.

III. RESULTS AND DISCUSSION

A. Reduction of the parameter space

By treating the non-electrostatic part of the solvation free energy as a post-SCF correction, DFT calculations only need to be performed on the two dimensional $\{\delta_n, n_c\}$-grid. In principle, parameter optimization techniques could directly be applied to these grid data, ideally leading to a global minimum. Before applying such a scheme it is, however, useful to first assess the shape of the cost function in parameter space in order to gain a rough overview of the optimization problem.

We therefore started to evaluate solvation free energies for N = 30 solutes and the solvent chloroform as a first example on an $n \times n$ parameter space, with $n = 5$ being the number of grid points per axis. The non-electrostatic parameters $(\alpha + \gamma)$ and β were optimized (as described in Sec. III B), at each single $\{\delta_n, n_c\}$ point. This results in an RMSD surface for chloroform, spanned by δ_n and n_c, as depicted in Fig. 2. The therein plotted RMSD surface does not show a well separated minimum in the δ_n dimension. The very flat parameter space loosens the impact of a global minimum and suggests that the δ_n dimension does not add any physical value to the model. On the twelve widely spread points in Fig. 2 indicated by black crosses, the RMSD is minimal with values between 36.7 and 37.7 meV. This removes the need for an explicit optimization of δ_n and suggests to simply fix it to a constant value. We decided to set $\delta_n = 2.0$ which is large enough to prevent numerical instabilities due to very sharp transitions in the dielectric function $\varepsilon(\mathbf{r})$ and yet small enough for a good computational efficiency. This value for δ_n will be used for all subsequent calculations. We note that Andreussi *et al.* suggested that the smoothness of the dielectric function may be chosen arbitrarily[22] which we confirm here for non-aqueous solutions also. Furthermore, by transforming the parameter space from $n_{\min}$ and $n_{\max}$ to n_c and δ_n, we are able to directly utilize this observation and reduce the number of free parameters.

Figure 2 also shows that the RMSD surface is shallow not only in δ_n but also in the n_c-dimension. We therefore investigated the sensitivity of the RMSD to the choice of n_c. To this end, we determined the optimal n_c values for 17 solvents again on a set of 30 solutes per solvent, but now choosing $n = 25$ grid points in the remaining n_c dimension. Figure 3 depicts the optimal values for each solvent as well as the uncertainty interval of n_c for RMSD variations of 10 meV around the minimum. Errors of 10 meV are usually acceptable in terms of solvation energy predictions, in particular, considering the accuracy of the experimental reference data and uncertainties inherent to DFT calculations. Furthermore, Fig. 3 shows that a value of $n_c = 7.6$ yields errors well within the 10 meV range for all solvents, indicating that the optimization of n_c can be omitted as well. For all following results, we therefore set $\delta_n = 2.0$ and $n_c = 7.6$, meaning that only a single DFT calculation is required for any solvent-solute calculation in the optimization process.

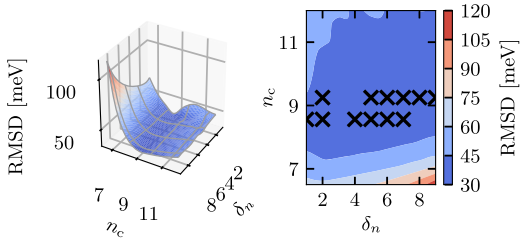

FIG. 2. "Lowest RMSD"-surface as a function of the two DFT-SCCS parameters $\{\delta_n, n_c\}$ shown as 3D (left) and contour (right) plot. Both plots are reconstructed from a 5×5-grid in $\{\delta_n, n_c\}$-space. RMSD values within 1 meV of the minimum are depicted as black crosses in the contour plot. The plot illustrates the flatness of the parameter space and the possibility of fixing δ_n to a value of choice.

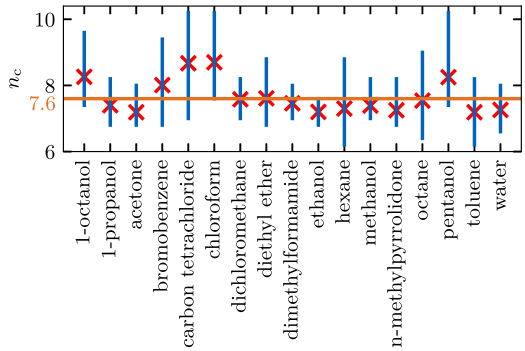

FIG. 3. Optimal n_c values for a range of non-aqueous solvents and water plotted as red crosses. The blue lines indicate the n_c interval corresponding to a 10 meV uncertainty in the RMSD around the minimum. A value of $n_c = 7.6$, indicated by the orange horizontal line, lies well within the 10 meV error range for all solvents.

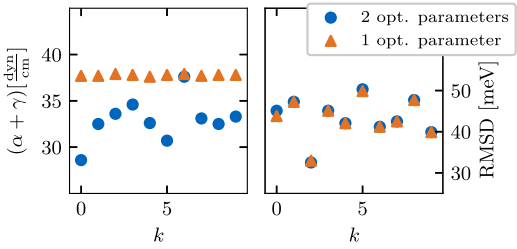

FIG. 4. Results of the 9-fold cross validation on the 222 entries of the solvent chloroform. Shown on the left are the optimized $(\alpha + \gamma)$ values for the different cross validation runs k (see the text). While large variations are observed for the fitting procedure with two variable non-electrostatic parameters (blue), the values are almost constant for a fixed $\beta = -0.5$ GPa (orange). This means that optimizing both non-electrostatic parameters does not match the requirement of solute independent parameters. On the other hand, the errors on the test set, depicted on the right, are equal in both cases. This implies that fixing β does not cause a loss of accuracy.

By treating the non-electrostatic part of the solvation free energy as post-correction and by setting δ_n and n_c to fixed values, the RSS cost function can be evaluated for all of the remaining parameter space from just one single DFT-SCF calculation. It is easy to show that the resulting RSS function is strictly convex, which implies that a stationary point of the function is the unique global minimum. Therefore, the optimal values of the non-electrostatic parameters $(\alpha + \gamma)$ and β can be determined by solving the linear equation system

$$\nabla \mathrm{RSS}((\alpha+\gamma),\beta) = 0 \tag{17}$$

(for a detailed derivation, cf. the supplementary material).

Analyzing the optimal values of the parameters $(\alpha + \gamma)$ and β for a range of solvents, we discovered that the optimal values were strongly dependent on the choice of the training set. This effect indicates that the model might be overfitting the training set data. Cross validation is an efficient technique for quantifying such transferability issues of the model. In order to assess the transferability of the model, we therefore applied k-fold cross validation with $k = 9$ as an example of the solvent chloroform and a set of 222 solutes. For this purpose, we split the reference data randomly into 9 parts (with six of them comprising 25 solutes and three 24 solutes). Each of the 9 parts is treated once as a disjoint validation set while the training set is built up from the respective remaining 8 parts. The resulting optimized $(\alpha + \gamma)$ and β parameters of the 9 unique cross-validation runs are shown in Fig. 4. As indicated in the plot, the optimization of both parameters at the same time leads to a large variation of the optimal $(\alpha + \gamma)$-parameter in a range between 28.0 and 37.5 dyn/cm. This variation can be explained by the shallow global minimum of the RMSD function with respect to $(\alpha + \gamma)$ and β (cf. the supplementary material).[22,26] Only a few changes in the training set and thereby changes in the quantum surface S and volume V are enough for a noticeable shift of the strict global minimum. These changes in the parameters, though, lead to only minor changes in the fitting error, due to the shallowness of the RMSD surface. Thus, for all solvents we considered in this work (also water), we can fix β to a value of −0.5 GPa without a loss of accuracy.

In addition to the optimization of both parameters, Fig. 4 depicts the optimized $(\alpha + \gamma)$ value for β being fixed at $\beta = -0.5$ GPa. While the minimal RMSD values are nearly unchanged, the optimized $(\alpha + \gamma)$ parameter is now stable at $(\alpha + \gamma) = 37.5$ dyn/cm with respect to the definition of the training set. Furthermore, if the training sizes would be much smaller, the two-parameter approach could be even less stable in the optimized parameters which could also result in a worse predictability. For the remainder of this work, we therefore follow the one-parameter approach, which then removes the need for running cross-validation. Test sets are accordingly simply selected randomly from the list of data. The one-parameter approach also has the advantage that it simplifies the derivation of a parameter prediction scheme in which one now merely has to correlate a single parameter with solvent properties (cf. Sec. III D).

B. Parameter optimization

The parameter reduction effectively reduces the need for DFT-SCCS calculations to a minimum of one per solute-solvent combination. We therefore decided to consider the complete set of solutes that was available for a particular solvent in the database. Solvents with less than 30 solvation energy entries were not considered due to a possible overfitting to such small training sets. A list of these 67 solvents with appropriate dielectric permittivity values and the number of solutes in the dataset is given in Table S1 of the supplementary material.

The parameter values $\delta_n = 2.0$, $n_c = 7.6$, and $\beta = -0.5$ GPa were used for all solutes, and only the $(\alpha + \gamma)$ was optimized for each solvent. 20% of the available solutes for each solvent were selected randomly as a test set, and the remainder was

used for training (cf. Table S1 of the supplementary material for exact numbers). The optimal ($\alpha + \gamma$) values lie in a range between 36.5 and 57.2 dyn/cm and are listed explicitly for each solvent in Table I. Additionally, the RMSD on the independent test set is listed. If not specified otherwise, all RMSD errors in this work are errors on an independent test set. The average RMSD over all solvents is 37.1 meV, and the average RMSD over all non-aqueous solvents is as low as 36.4 meV. MAEs and errors on the training sets are listed for each solvent in Table S1 of the supplementary material. Furthermore, Table S2 of the supplementary material provides the parameters for a different fit using β = 0.0 GPa which is slightly less accurate for molecular solutes, but could be of possible use, e.g., for extended slab calculations where the cavity volume is not well defined.

In order to assess the accuracy of the optimized SCCS model, we compared the resulting errors to the widely used SM8[44] and SMD ("D" stands for "density") models.[21] SMx models, such as SM8, have been developed since the early 1990s in the group of Cramer and Truhlar,[20] with the SMD model being the successor of the SM8 model. Both the SMD and the SM8 models showed the best performance and accuracy in comparison with other continuum solvation models based on polarizable continuum model (PCM) or **Co**nductor-like **S**creening **Mo**del (COSMO) approaches[44] and have been used extensively in many areas of science.[164–166] They are among the most systematically and extensively optimized models for non-aqueous solvents based on a large collection of experimental reference data.[34] Both models were trained on the Minnesota Solvation Database, which, among others, contains 2140 solvation free energies for 321 neutral solutes in 90 non-aqueous solvents. From the 67 non-aqueous solvents of which we determined the SCCS parameters, we identified 40 as also being part of the Minnesota Database.

Figure 5 compares the accuracy of the SMD and SM8 model (training set error) with that of the SCCS model (test set error) using our newly derived optimized parameter sets. For a fair comparison, we calculated the SCCS errors both on test sets derived from our new database and on the same Minnesota sets for which the MAEs of SM8 and SMD were reported.[21,44] Most importantly, Fig. 5 clearly shows that the SCCS and SMx models can reach similar accuracy. Comparing in detail the errors of all schemes on the same test set (MNsol), the SCCS approach shows, however, also some larger deviations. These can be attributed to the composition of the MNsol test set: While the Solv@TUM derived sets are chemically very diverse (cf. Fig. 1), the selected test sets from the Minnesota database are dominated by a few functional groups, for which we know that the accuracy of the SCCS model tends to be worse (e.g., alcohols). Due to the higher flexibility of the SM8 model using 64 independent parameters and the SMD model using 37 independent parameters, these approaches partially show a higher accuracy resulting from the specific training to these datasets. In contrast, our single-parameter SCCS approach shows similar accuracy for most systems even without specific training to these sets. The results therefore show the high transferability of our scheme which was already indicated before by the independence of the training set (cf. Fig. 4). The single-parameter approach also partly averages out experimental error fluctuations which could result in an erroneous prediction of solvation effects. On the other hand, the SMx models are universal models, which means that they can be applied to every solvent, while the SCCS model is only applicable to solvents for which parameters have been calculated. In Sec. III D, we will present an approach with which the SCCS model gets universal but at the expense of accuracy.

C. The SCCS-P model: An improved model for aromatic solutes

So far, we have optimized the SCCS model for non-aqueous solvents and demonstrated its performance using

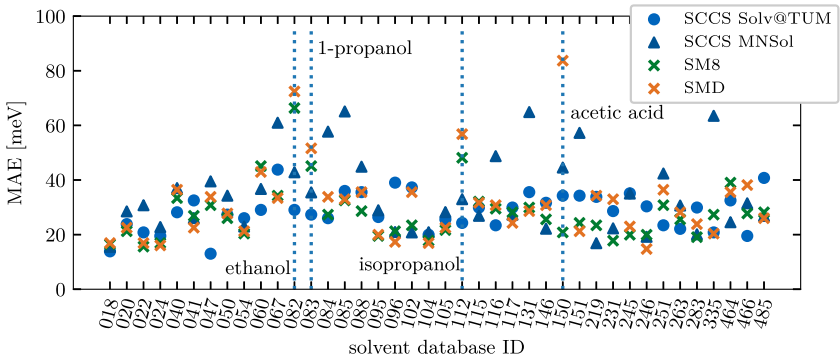

FIG. 5. Mean Absolute Error (MAE) for the prediction of solvation free energies of 40 non-aqueous solvents for the SM8 and SMD[21,44] and our optimized SCCS model. Errors for the SCCS model are given as errors calculated on an independent Solv@TUM sub-set and errors calculated on the SMx sets (MNSol). Solvents for which the difference between our and the SMx errors are large are marked with dotted lines. The difference between the SCCS Solv@TUM and the SCCS MNSol errors for some solvents can be explained by the composition of these sets. While our sets are very diverse, the sets of these solvents in the Minnesota database are dominated by a few functional groups, for which we know that the accuracy of the SCCS model tends to be worse (e.g., alcohols). Here we plot MAEs instead of RMSDs as only the former is given in the available SMx literature. The SMD solvation energies were calculated at the M05-2X/6-31G* and the SM8 at the SM8/mPW1PW/6-31G(d) level of theory.

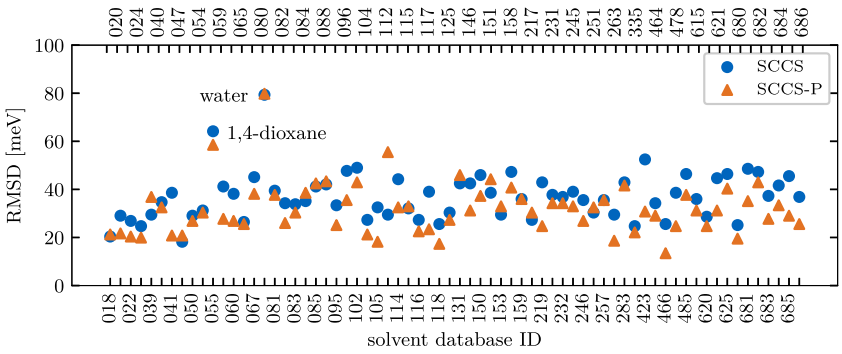

FIG. 6. Comparison of the RMSDs of the SCCS and the polarizability modified variant, the SCCS-P model on random test sets for a range of solvents from the Solv@TUM database.

only a single fitting parameter. The non-electrostatic part of the SCCS model is, however, a relatively coarse approximation to the physical reality. As an example, dispersion interactions are generally proportional to the polarizabilities of the interacting systems. Yet, SCCS approximates the polarizability of the solute as a linear function of the solvation cavity volume. A straightforward way to improve on the SCCS model could therefore be to utilize isotropic solute polarizabilities instead of solvation cavity volumes in the non-electrostatic energy expression. The energy expression of this new SCCS-P model then reads

$$\Delta G^{\mathrm{sol}}(\boldsymbol{p}) = \Delta G^{\mathrm{el}}_{\delta_n,n_c} + (\alpha_P + \gamma)S^{\circ}_{\delta_n,n_c} + \beta_P \alpha^{\mathrm{iso},\circ}. \tag{18}$$

The subscript P of α and β here highlights that their values are different from those in the standard SCCS model. This also means that we have to apply a different constant value for β_P. In similar tests to the ones performed for standard SCCS, we found that $\beta_P = -4.2$ GPa was a suitable choice yielding low errors for all solvents. The polarizability $\alpha^{\mathrm{iso},\circ}$ can either be found in extended experimental databases[167] or directly be calculated with Density-Functional Perturbation Theory (DFPT) in FHI-aims.[161] In order not to rely on the presence of experimental data, we decided here to follow the latter approach. Nevertheless, we also compared to experimental values and found an overall perfect agreement (cf. Fig. S2 in the supplementary material).

Figure 6 shows the RMSDs obtained by optimizing the $(\alpha_P + \gamma)$ parameter and using the same fitting strategy as before for the original SCCS model (all values are listed in Table S3 in the supplementary material). The averaged RMSD reduces to 31.8 meV (0.733 kcal/mol), which corresponds to an increase in accuracy of about 15%. By correlating the isotropic polarizability with the solvation cavity volume, we found a good linear correlation, but also some distinct outliers, in particular, for more voluminous solutes (cf. Fig. S6 in the supplementary material). We therefore checked the sensitivity of the error to the solute chemical composition. From this, we found that larger solutes as well as solutes with conjugated systems lead to more drastically reduced errors than the complementary small solutes and non-conjugated groups of solutes in our database. By further separating the RMSD into errors related to aromatic and aliphatic solutes, we found that most of the improvement in accuracy is in fact related to a better description of aromatic solute solvation (cf. Fig. S7 in the supplementary material). The RMSD of 688 calculated solvation free energies of aromatic solutes decreases by 14.3 meV from 43.4 meV (SCCS) to 29.1 meV (SCCS-P), while the RMSD of 4357 calculated solvation free energies of aliphatic solutes only decreases by 4.5 meV from 37.3 meV (SCCS) to 32.8 meV (SCCS-P). In conclusion, the SCCS-P model requires the calculation of the polarizability for each solute of interest (or the usage of experimental reference data), but then yields an increased accuracy in the prediction of solvation effects, in particular, for larger aromatic solutes.

D. A simple parameter prediction scheme

Not all solvents which could be of potential interest, e.g., in battery modeling or electro-chemistry, have already been experimentally studied in terms of solvation free energies. It is therefore important to have an approximate method to predict SCCS parameters also for solvents for which reference data are missing. For this purpose, we studied the correlations of the optimized SCCS parameters with the intrinsic solvent properties. The properties then used in the prediction scheme should ideally be available for nearly every existing solvent in order for the scheme to be useful in practice. Examples of such widely available properties could be the molecular dipole moment or the solvent's bulk permittivity. By correlating our optimized $(\alpha + \gamma)$ parameters for all non-aqueous

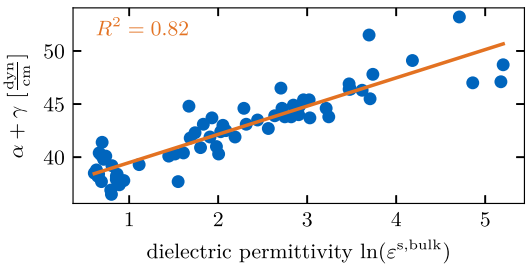

FIG. 7. Correlation between the optimal parameter values $(\alpha + \gamma)$ and the logarithm of the dielectric permittivity. A simple linear regression (orange line) yields a coefficient of determination (R^2) of 82%. This suggests our linear parameter prediction scheme as a first guess for solvents where there are no enough training data available.

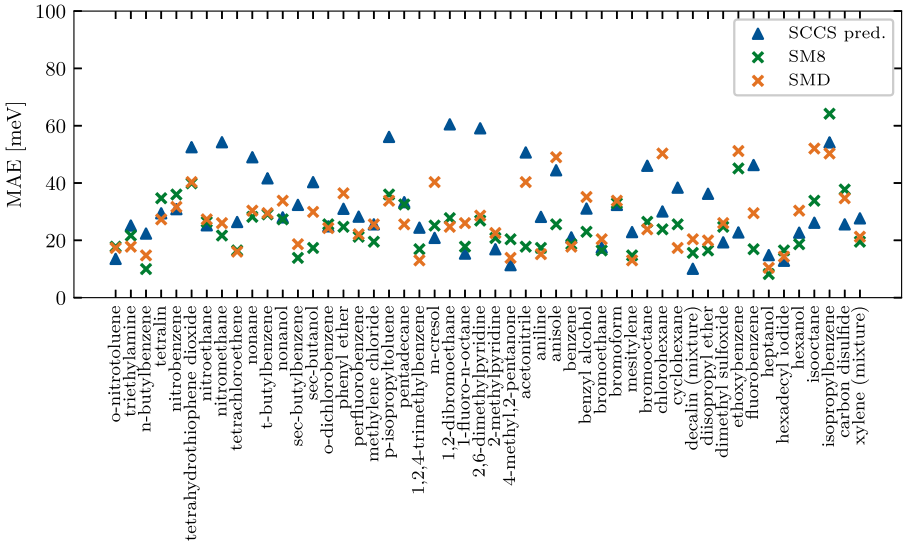

FIG. 8. Mean Absolute Error (MAE) for the prediction of solvation free energies of 50 non-aqueous solvents out of the Minnesota Solvation Database for the SM8 and SMD[21,44] and the SCCS model with our parameter prediction scheme. We here plot MAEs instead of RMSDs as only the former are given in the available SMx literature.

solvents with these descriptors, we found a good logarithmic correlation with the bulk solvent permittivity, resulting in an R^2-coefficient of 0.82 (cf. Fig. 7). The prediction function for the optimal SCCS ($\alpha + \gamma$) is given by the regression line

$$(\alpha + \gamma) = 3.0\,\text{dyn/cm} \cdot \ln(\varepsilon^{\text{s,bulk}}) + 36.0\,\text{dyn/cm}. \quad (19)$$

Analogously, the same parameter in the SCCS-P variant can be predicted by the equation

$$(\alpha_{\text{P}} + \gamma) = 2.0\,\text{dyn/cm} \cdot \ln(\varepsilon^{\text{s,bulk}}) + 12.0\,\text{dyn/cm}. \quad (20)$$

In order to test the accuracy of the prediction method, we evaluated the MAE using the SCCS with the prediction parameters on 50 solvents out of the Minnesota Database, which were not part of the linear regression and compared them with the SM8 and SMD model. As depicted in Fig. 8, the generalized SCCS model performs impressively well, in particular, considering the fact that all parameters were predicted. In the case of non-optimized implicit solvation methods, the most straightforward way to estimate non-aqueous solvation trends would be a change in the bulk dielectric permittivity while leaving all solvent parameters at a fixed value. Here, we show that the simplicity of this coarse approach can be retained in our optimized scheme where the parameters are also predicted by the value of the bulk dielectric permittivity. This results in a much higher accuracy and motivates the use of the generalized SCCS scheme for the estimation of neutral, molecular solvation effects in arbitrary non-aqueous environments. A future goal will be the extension of this scheme to the case of charged solutes similar to what has been done before for the case of water as a solvent.[23]

IV. CONCLUSIONS

Modeling the influence of non-aqueous solvents on chemical reactions is of high importance, in particular, in the field of battery engineering. Implicit solvation models combined with density-functional theory are, in principle, able to simulate these effects. This, however, requires first a parameter optimization to the solvents of choice, which to date has been only performed for the SMx family of models.

In this work, we first presented a new, extended database of 5952 solvation energies of neutral, molecular solutes in non-aqueous solvents. We then used this database to train the SCCS model available in both QUANTUM ESPRESSO[22] and FHI-aims[25] for 68 solvents. In this process, we found that the effective parameter space can be reduced to a single dimension per solvent. This leads to statistically stable optimized parameter sets, which converge quickly with the training set size. At the same time, the accuracy of the optimized SCCS model reaches an average RMSD of 37.1 meV and an average MAE of 30.1 meV on independent test sets of all non-aqueous solvents. This performance is similar to the SMx series of models which, however, uses up to 54 effective parameters.

In the case of organic solutes, the description of dispersion interactions is essential for the accurate prediction of solvation energies. To this end, we presented an updated variant of the SCCS model that uses the solute's isotropic polarizability instead of the solvation cavity volume. The experimental value for the polarizability is available for a wide range of molecules. Alternatively, it can be calculated straightforwardly to a high accuracy via Density-Functional Perturbation Theory (DFPT), e.g., with FHI-aims.[161] The resulting SCCS-P model remarkably improves the description of aromatic solutes, for which dispersion interactions are more pronounced. The average MAE thereby reduces to 24.9 meV and the RMSD to 31.8 meV, making the SCCS-P model the preferred choice for more accurate predictions of non-aqueous solvation of aromatic solutes.

In order to also enable the calculation of solvation energies in solvents not included in our new database, we also

present a simple scaling relation of the optimized parameters with the bulk dielectric permittivity. Using this parameter-prediction scheme, we obtain an average MAE of 32.4 meV (RMSD 40.1 meV) for the SCCS model and an average MAE of 27.6 meV (RMSD 34.1 meV) for the SCCS-P model, being close to the non-predicted value. The mere knowledge of the solvent's bulk dielectric permittivity therefore enables an accurate modeling of neutral solutes in arbitrary non-aqueous solvent environments by means of the generalized SCCS scheme.

SUPPLEMENTARY MATERIAL

See supplementary material for a detailed mathematical background of the optimization scheme. We further provide convergence tests regarding the non-self-consistent treatment of non-electrostatic solute-solvent interactions as well as parameter correlation analysis. Convergence tests are performed exemplarily for chloroform as a solvent. The accuracy of the experimental reference data is assessed by comparing solvation energies from the MNsol and Solv@TUM database. We show the high accuracy of the polarizability calculations by comparing with the experimental reference data. Furthermore, a table is attached with all solvent bulk dielectric permittivities, optimized parameters, and all training and test set errors. Finally, three more plots visualize the improvement of the description of aromatic solute solvation by the SCCS-P method and again all results are tabulated.

ACKNOWLEDGMENTS

The authors gratefully acknowledge support from the Solar Technologies Go Hybrid Initiative of the State of Bavaria and the German Science Foundation DFG (Grant No. OB425/4-1).

REFERENCES

[1]J.-S. Filhol and M.-L. Doublet, Catal. Today **202**, 87–97 (2013).
[2]S. Siahrostami and A. Vojvodic, J. Phys. Chem. C **119**, 1032–1037 (2015).
[3]M. H. Hansen and J. Rossmeisl, J. Phys. Chem. C **120**, 29135–29143 (2016).
[4]H. Oberhofer, in *Handbook of Materials Modeling: Applications: Current and Emerging Materials*, edited by W. Andreoni and S. Yip (Springer International Publishing, Cham, 2018), pp. 1–33.
[5]J. Tomasi, B. Mennucci, and R. Cammi, Chem. Rev. **105**, 2999–3093 (2005).
[6]G. Monard and J.-L. Rivail, *Handbook of Computational Chemistry* (Springer, 2012), pp. 561–571.
[7]E. Skúlason, G. S. Karlberg, J. Rossmeisl, T. Bligaard, J. Greeley, H. Jónsson, and J. K. Nørskov, Phys. Chem. Chem. Phys. **9**, 3241–3250 (2007).
[8]K. Chan and J. K. Nørskov, J. Phys. Chem. Lett. **6**, 2663–2668 (2015).
[9]J. G. Kirkwood, J. Chem. Phys. **2**, 351–361 (1934).
[10]L. Onsager, J. Am. Chem. Soc. **58**, 1486–1493 (1936).
[11]M. Bachs, F. J. Luque, and M. Orozco, J. Comput. Chem. **15**, 446–454 (1994).
[12]F. Luque, Y. Zhang, C. Aleman, M. Bachs, J. Gao, and M. Orozco, J. Phys. Chem. **100**, 4269–4276 (1996).
[13]C. Curutchet, M. Orozco, and F. J. Luque, J. Comput. Chem. **22**, 1180–1193 (2001).
[14]D. M. Chipman, Theor. Chem. Acc. **107**, 80–89 (2002).
[15]D. M. Chipman and M. Dupuis, Theor. Chem. Acc. **107**, 90–102 (2002).
[16]F. J. Luque, C. Curutchet, J. Munoz-Muriedas, A. Bidon-Chanal, I. Soteras, A. Morreale, J. Gelpi, and M. Orozco, Phys. Chem. Chem. Phys. **5**, 3827–3836 (2003).
[17]J.-L. Fattebert and F. Gygi, Int. J. Quantum Chem. **93**, 139–147 (2003).
[18]D. A. Scherlis, J.-L. Fattebert, F. Gygi, M. Cococcioni, and N. Marzari, J. Chem. Phys. **124**, 074103 (2006).
[19]D. M. Chipman, J. Chem. Phys. **124**, 224111 (2006).
[20]C. J. Cramer and D. G. Truhlar, Acc. Chem. Res. **41**, 760–768 (2008).
[21]A. V. Marenich, C. J. Cramer, and D. G. Truhlar, J. Phys. Chem. B **113**, 6378–6396 (2009).
[22]O. Andreussi, I. Dabo, and N. Marzari, J. Chem. Phys. **136**, 064102 (2012).
[23]C. Dupont, O. Andreussi, and N. Marzari, J. Chem. Phys. **139**, 214110 (2013).
[24]A. V. Marenich, C. J. Cramer, and D. G. Truhlar, J. Chem. Theory Comput. **9**, 609–620 (2013).
[25]S. Ringe, H. Oberhofer, C. Hille, S. Matera, and K. Reuter, J. Chem. Theory Comput. **12**, 4052–4066 (2016).
[26]M. Sinstein, C. Scheurer, S. Matera, V. Blum, K. Reuter, and H. Oberhofer, J. Chem. Theory Comput. **13**, 5582–5603 (2017).
[27]I. Tuñón, M. F. Ruiz-López, D. Rinaldi, and J. Bertrán, J. Comput. Chem. **17**, 148–155 (1996).
[28]S. Ringe, H. Oberhofer, and K. Reuter, J. Chem. Phys. **146**, 134103 (2017).
[29]G. Fisicaro, L. Genovese, O. Andreussi, S. Mandal, N. N. Nair, N. Marzari, and S. Goedecker, J. Chem. Theory Comput. **13**, 3829–3845 (2017).
[30]L. R. Pratt, Annu. Rev. Phys. Chem. **53**, 409–436 (2002).
[31]A. Kuperman, S. Nadimi, S. Oliver, G. A. Ozin, J. M. Garcés, and M. M. Olken, Nature **365**, 239 (1993).
[32]Y. Lin, H. Skaff, T. Emrick, A. D. Dinsmore, and T. P. Russell, Science **299**, 226–229 (2003).
[33]X.-B. Cheng, R. Zhang, C.-Z. Zhao, and Q. Zhang, Chem. Rev. **117**, 10403–10473 (2017).
[34]J. Zhang, H. Zhang, T. Wu, Q. Wang, and D. van der Spoel, J. Chem. Theory Comput. **13**, 1034–1043 (2017).
[35]S. Miertuš, E. Scrocco, and J. Tomasi, Chem. Phys. **55**, 117–129 (1981).
[36]A. V. Marenich, C. P. Kelly, J. D. Thompson, G. D. Hawkins, C. C. Chambers, D. J. Giesen, P. Winget, C. J. Cramer, and D. G. Truhlar, Minnesota Solvation Database–Version 2012, 2012.
[37]Y. Zhao and D. G. Truhlar, Chem. Phys. Lett. **502**, 1–13 (2011).
[38]P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, J. Phys.: Condens. Matter **21**, 395502 (2009).
[39]V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, Comput. Phys. Commun. **180**, 2175–2196 (2009).
[40]Y.-H. Fang, G.-F. Wei, and Z.-P. Liu, Catal. Today **202**, 98–104 (2013).
[41]L. D. Chen, M. Urushihara, K. Chan, and J. K. Nørskov, ACS Catal. **6**, 7133–7139 (2016).
[42]A. V. Cresce, S. M. Russell, O. Borodin, J. A. Allen, M. A. Schroeder, M. Dai, J. Peng, M. P. Gobet, S. G. Greenbaum, R. E. Rogers, and X. Kang, Phys. Chem. Chem. Phys. **19**, 574–586 (2017).
[43]N. Keilbart, Y. Okada, A. Feehan, S. Higai, and I. Dabo, Phys. Rev. B **95**, 115423 (2017).
[44]A. V. Marenich, R. M. Olson, C. P. Kelly, C. J. Cramer, and D. G. Truhlar, J. Chem. Theory Comput. **3**, 2011–2033 (2007).
[45]C. Hille, S. Ringe, M. Deimel, C. Kunkel, W. E. Acree, K. Reuter, and H. Oberhofer, "Solv@TUM v 1.0," https://mediatum.ub.tum.de/1452571, last accessed 9 November 2018.
[46]S. N. Steinmann and P. Sautet, J. Phys. Chem. C **120**, 5619–5623 (2016).
[47]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865–3868 (1996).

[48]X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter, and M. Scheffler, New J. Phys. **14**, 053020 (2012).
[49]M. H. Abraham, G. S. Whiting, R. Fuchs, and E. J. Chambers, J. Chem. Soc. Perkin Trans. 2 **1990**, 291–300.
[50]M. H. Abraham, J. Andonian-Haftvan, J. P. Osei-Owusu, P. Sakellariou, J. S. Urieta, M. C. López, and R. Fuchs, J. Chem. Soc. Perkin Trans. 2 **1993**, 299–304.
[51]M. H. Abraham, A. M. Zissimos, and W. E. Acree, Jr., Phys. Chem. Chem. Phys. **3**, 3732–3736 (2001).
[52]M. H. Abraham, J. Le, W. E. Acree, P. W. Carr, and A. J. Dallas, Chemosphere **44**, 855–863 (2001).
[53]M. H. Abraham, A. M. Zissimos, and W. E. Acree, Jr., New J. Chem. **27**, 1041–1044 (2003).
[54]M. H. Abraham and W. E. Acree, Jr., New J. Chem. **28**, 1538–1543 (2004).
[55]M. H. Abraham, W. E. Acree, and J. E. Cometto-Muñiz, New J. Chem. **33**, 2034–2043 (2009).
[56]M. H. Abraham, W. E. Acree, Jr., A. J. Leo, and D. Hoekman, New J. Chem. **33**, 1685–1692 (2009).
[57]M. H. Abraham, W. E. Acree, A. J. Leo, and D. Hoekman, New J. Chem. **33**, 568–573 (2009).
[58]M. H. Abraham and W. E. Acree, Jr., New J. Chem. **34**, 2298–2305 (2010).
[59]M. H. Abraham and W. E. Acree, Thermochim. Acta **526**, 22–28 (2011).
[60]M. H. Abraham and W. E. Acree, New J. Chem. **36**, 1798–1806 (2012).
[61]M. H. Abraham, M. Zad, and W. E. Acree, J. Mol. Liq. **212**, 301–306 (2015).
[62]M. H. Abraham, W. E. Acree, and E. Matteoli, Fluid Phase Equilib. **421**, 59–66 (2016).
[63]M. Antosik and S. I. Sandler, J. Chem. Eng. Data **39**, 584–587 (1994).
[64]R. Battino, T. R. Rettich, and T. Tominaga, J. Phys. Chem. Ref. Data **12**, 163–178 (1983).
[65]R. Battino, T. R. Rettich, and T. Tominaga, J. Phys. Chem. Ref. Data **13**, 563–600 (1984).
[66]G. R. R. Bebahani, P. Hogan, and W. E. Waghorne, J. Chem. Eng. Data **47**, 1290–1292 (2002).
[67]S. Bo, R. Battino, and E. Wilhelm, J. Chem. Eng. Data **38**, 611–616 (1993).
[68]M. Brumfield, A. Wadawadigi, N. Kuprasertkul, S. Mehta, W. E. Acree, and M. H. Abraham, Phys. Chem. Liq. **53**, 10–24 (2015).
[69]M. Brumfield, W. E. Acree, and M. H. Abraham, Phys. Chem. Liq. **53**, 25–37 (2015).
[70]E. Brunner, Ber. Bunsengesell. Phys. Chem. **83**, 715–721 (1979).
[71]G. K. Bub and W. A. Hillebrand, J. Chem. Eng. Data **24**, 315–319 (1979).
[72]E. A. Campanella, Chem. Eng. Technol. **14**, 376–378 (1991).
[73]R. W. Cargill, "Ketones, acids, esters, ethers," in *IUPAC Solubility Data Series* (Pergamon Press Ltd., Oxford, England, 1990), Vol. 43, pp. 208–238.
[74]L. Červený, B. Vostrý, and V. Růžička, Collect. Czech. Chem. Commun. **46**, 1965–1969 (1981).
[75]A. del Río, B. Coto, C. Pando, and J. A. R. Renuncio, Ind. Eng. Chem. Res. **40**, 689–695 (2001).
[76]A. del Rıo, B. Coto, C. Pando, and J. A. R. Renuncio, Fluid Phase Equilib. **187-188**, 299–310 (2001).
[77]A. del Rıo, B. Coto, C. Pando, and J. A. R. Renuncio, Fluid Phase Equilib. **200**, 41–51 (2002).
[78]A. del Río, B. Coto, C. Pando, and J. A. R. Renuncio, Ind. Eng. Chem. Res. **41**, 1364–1369 (2002).
[79]D. I. Eikens, "Applicability of theoretical and semi-empirical models for predicting infinite dilution activity coefficients," Ph.D. thesis, University of Minnesota, Minneapolis, MN, USA, 1993.
[80]S. Endo and T. C. Schmidt, Fluid Phase Equilib. **246**, 143–152 (2006).
[81]F. D. Evans and R. Battino, J. Chem. Thermodyn. **3**, 753–760 (1971).
[82]R. Garriga, P. Pérez, and M. Gracia, Ber. Bunsengesell. Phys. Chem. **101**, 1466–1473 (1997).
[83]R. Garriga, F. Sánchez, P. Pérez, and M. Gracia, Fluid Phase Equilib. **138**, 131–144 (1997).
[84]R. Garriga, F. Sánchez, P. Pérez, and M. Gracia, J. Chem. Thermodyn. **29**, 649–659 (1997).
[85]R. Garriga, F. Sánchez, P. Fpérez, and M. Gracia, Ber. Bunsengesell. Phys. Chem. **102**, 14–24 (1998).
[86]R. Garriga, S. Martınez, P. Pérez, and M. Gracia, Fluid Phase Equilib. **147**, 195–206 (1998).
[87]F. Gibanel, M. C. Lpez, F. M. Royo, V. Rodrguez, and J. S. Urieta, J. Solution Chem. **23**, 1247–1256 (1994).
[88]R. Gonzalez, F. Murrieta-Guevara, O. Parra, and A. Trejo, Fluid Phase Equilib. **34**, 69–81 (1987).
[89]M. Gracia, *International DATA Series: Selected Data on Mixtures: Thermodynamic Properties of Non-Reacting Binary Systems of Organic Substances*, edited by Centre National de la Recherche Scientifique (France). Centre de Recherches de Microcalorimetrie et de Thermochemie, Texas Engineering Experiment Station. Thermodynamics Research Center, Université de Paris VII. Laboratoire de Chimie Organique Physique (Thermodynamics Research Center, Texas Engineering Experiment Station, Texas A & M University, 1999).
[90]L. M. Grubbs, M. Saifullah, N. E. De La Rosa, S. Ye, S. S. Achi, W. E. Acree, and M. H. Abraham, Fluid Phase Equilib. **298**, 48–53 (2010).
[91]E. Hart, D. Grover, H. Zettl, V. Koshevarova, S. Zhang, C. Dai, W. E. Acree, I. A. Sedov, M. A. Stolov, and M. H. Abraham, J. Mol. Liq. **209**, 738–744 (2015).
[92]E. Hart, S. Cheeran, G. E. Little, H. Singleton, W. E. Acree, and M. H. Abraham, Phys. Chem. Liq. **55**, 347–357 (2017).
[93]E. Hart, A. Klein, M. Barrera, M. Jodray, K. Rodriguez, W. E. Acree, and M. H. Abraham, Phys. Chem. Liq. **56**, 821–833 (2017).
[94]W. Hayduk, "Alkanes," in *IUPAC Solubility Data Series* (Pergamon Press, Inc., Elmsford, New York, USA, 1986), Vol. 24, pp. 130–195.
[95]P. J. Hesse, R. Battino, P. Scharlin, and E. Wilhelm, J. Chem. Eng. Data **41**, 195–201 (1996).
[96]D. V. S. Jain, S. B. Saini, and R. S. Sidhu, J. Chem. Thermodyn. **14**, 689–693 (1982).
[97]D. V. S. Jain and R. S. Sidhu, J. Chem. Thermodyn. **16**, 111–114 (1984).
[98]J. R. Khurma, O. Muthu, S. Munjal, and B. D. Smith, J. Chem. Eng. Data **28**, 93–99 (1983).
[99]I. Kikic, P. Alessi, P. Gozzi, R. Lapasin, and R. De Santis, J. Chem. Eng. Data **25**, 33–36 (1980).
[100]I. M. Korenman, N. Y. Gurevich, and T. G. Kulagina, Zh. Prikl. Khim. **46**, 683–684 (1973).
[101]M. Krummen, T. M. Letcher, and J. Gmehling, J. Chem. Eng. Data **47**, 906–910 (2002).
[102]L. Lepori, E. Matteoli, and B. Marongiu, Fluid Phase Equilib. **42**, 229–240 (1988).
[103]L. Lepori, E. Matteoli, and M. R. Tine, J. Chem. Eng. Data **35**, 179–182 (1990).
[104]L. Lepori, E. Matteoli, and M. R. Tine, J. Solution Chem. **20**, 57–70 (1991).
[105]L. Lepori, E. Matteoli, P. Gianni, and M. C. Righetti, Fluid Phase Equilib. **387**, 198–208 (2015).
[106]L. Lepori, P. Gianni, and E. Matteoli, J. Therm. Anal. Calorim. **124**, 1497–1509 (2016).
[107]J. Li, T. Zhu, G. D. Hawkins, P. Winget, D. A. Liotard, C. J. Cramer, and D. G. Truhlar, Theor. Chem. Acc. **103**, 9–63 (1999).
[108]X. Liang, S. Ye, Q. Xie, M. Lu, F. Xia, Y. Nie, Z. Pan, and J. Ji, J. Chem. Thermodyn. **125**, 11–16 (2018).
[109]J. Makranczy, K. Megyery-Balog, L. Rusz, and L. Patyi, Hung. J. Ind. Chem. **4**, 269–280 (1976).
[110]Y. Miyano, J. Chem. Eng. Data **50**, 211–215 (2005).
[111]Y. Miyano, J. Chem. Eng. Data **50**, 2045–2048 (2005).
[112]Y. Miyano, J. Chem. Thermodyn. **37**, 459–465 (2005).
[113]Y. Miyano, T. Kobashi, H. Shinjo, S. Kumada, Y. Watanabe, W. Niya, and Y. Tateishi, J. Chem. Thermodyn. **38**, 724–731 (2006).

[114]Y. Miyano, S. Uno, K. Tochigi, S. Kato, and H. Yasuda, J. Chem. Eng. Data **52**, 2245–2249 (2007).
[115]I. Mokbel, A. Blondel-Telouk, D. Vellut, and J. Jose, Fluid Phase Equilib. **149**, 287–308 (1998).
[116]F. Mößner, B. Coto, C. Pando, and J. A. R. Renuncio, Ber. Bunsengesell. Phys. Chem. **101**, 1146–1153 (1997).
[117]F. Mössner, B. Coto, C. Pando, R. G. Rubio, and J. A. R. Renuncio, J. Chem. Eng. Data **41**, 537–542 (1996).
[118]I. Nagata, *International DATA Series: Selected Data On Mixtures: Thermodynamic Properties of Non-Reacting Binary Systems of Organic Substances*, Series A (Thermodynamics Research Center, Texas Engineering Experiment Station, Texas A&M University, 1999).
[119]J. H. Park, A. Hussam, P. Couasnon, D. Fritz, and P. W. Carr, Anal. Chem. **59**, 1970–1976 (1987).
[120]K. A. Pividal, C. Sterner, S. I. Sandler, and H. Orbey, Fluid Phase Equilib. **72**, 227–250 (1992).
[121]G. L. Pollack, J. F. Himm, and J. J. Enyeart, J. Chem. Phys. **81**, 3239–3246 (1984).
[122]A. d. Río, B. Coto, J. A. R. Renuncio, and C. Pando, Fluid Phase Equilib. **221**, 1–6 (2004).
[123]M. Saifullah, S. Ye, L. M. Grubbs, N. E. De La Rosa, W. E. Acree, and M. H. Abraham, J. Solution Chem. **40**, 2082–2094 (2011).
[124]I. A. Sedov, M. A. Stolov, and B. N. Solomonov, J. Chem. Thermodyn. **64**, 120–125 (2013).
[125]I. A. Sedov, D. Khaibrakhmanova, E. Hart, D. Grover, H. Zettl, V. Koshevarova, C. Dai, S. Zhang, A. Schmidt, W. E. Acree, and M. H. Abraham, J. Mol. Liq. **212**, 833–840 (2015).
[126]I. A. Sedov, M. A. Stolov, E. Hart, D. Grover, H. Zettl, V. Koshevarova, C. Dai, S. Zhang, W. E. Acree, and M. H. Abraham, J. Mol. Liq. **209**, 196–202 (2015).
[127]I. A. Sedov, M. A. Stolov, E. Hart, D. Grover, H. Zettl, V. Koshevarova, W. E. Acree, and M. H. Abraham, J. Mol. Liq. **208**, 63–70 (2015).
[128]I. A. Sedov, T. Salikov, E. Hart, E. Higgins, W. E. Acree, and M. H. Abraham, Fluid Phase Equilib. **431**, 66–74 (2017).
[129]I. A. Sedov, T. I. Magsumov, E. Hart, A. M. Ramirez, S. Cheeran, M. Barrera, M. Y. Horton, A. Wadawadigi, O. Zha, X. Y. Tong, W. E. Acree, and M. H. Abraham, J. Solution Chem. **46**, 2249–2267 (2017).
[130]I. A. Sedov, T. M. Salikov, D. R. Khaibrakhmanova, A. Wadawadigi, O. Zha, E. Qian, E. Hart, M. Barrera, W. E. Acree, and M. H. Abraham, J. Solution Chem. **47**, 634–653 (2018).
[131]I. A. Sedov, T. M. Salikov, A. Wadawadigi, O. Zha, E. Qian, W. E. Acree, and M. H. Abraham, J. Chem. Thermodyn. **124**, 133–140 (2018).
[132]D. Shivakumar, J. Williams, Y. Wu, W. Damm, J. Shelley, and W. Sherman, J. Chem. Theory Comput. **6**, 1509–1519 (2010).
[133]M. Shokouhi, H. Farahani, M. Hosseini-Jenab, and A. H. Jalili, J. Chem. Eng. Data **60**, 499–508 (2015).
[134]M. Shokouhi, A. R. Rezaierad, S.-M. Zekordi, M. Abbasghorbani, and M. Vahidi, J. Chem. Eng. Data **61**, 512–524 (2016).
[135]L. M. Sprunger, A. Proctor, W. E. Acree, M. H. Abraham, and N. Benjelloun-Dakhama, Fluid Phase Equilib. **270**, 30–44 (2008).
[136]L. M. Sprunger, J. Gibbs, W. E. Acree, and M. H. Abraham, Fluid Phase Equilib. **273**, 78–86 (2008).
[137]L. M. Sprunger, S. S. Achi, R. Pointer, B. H. Blake-Taylor, W. E. Acree, and M. H. Abraham, Fluid Phase Equilib. **286**, 170–174 (2009).
[138]L. M. Sprunger, S. S. Achi, W. E. Acree, M. H. Abraham, A. J. Leo, and D. Hoekman, Fluid Phase Equilib. **281**, 144–162 (2009).
[139]L. M. Sprunger, S. S. Achi, R. Pointer, W. E. Acree, and M. H. Abraham, Fluid Phase Equilib. **288**, 121–127 (2010).
[140]L. M. Sprunger, S. S. Achi, W. E. Acree, and M. H. Abraham, Fluid Phase Equilib. **288**, 139–144 (2010).
[141]T. W. Stephens, N. E. De La Rosa, M. Saifullah, S. Ye, V. Chou, A. N. Quay, W. E. Acree, and M. H. Abraham, Fluid Phase Equilib. **309**, 30–35 (2011).
[142]T. W. Stephens, M. Loera, A. N. Quay, V. Chou, C. Shen, A. Wilson, W. E. Acree, Jr., and M. H. Abraham, Open Thermodyn. J. **5**, 104–121 (2011).
[143]T. W. Stephens, N. E. De La Rosa, M. Saifullah, S. Ye, V. Chou, A. N. Quay, W. E. Acree, and M. H. Abraham, Fluid Phase Equilib. **308**, 64–71 (2011).
[144]T. W. Stephens, A. N. Quay, V. Chou, M. Loera, C. Shen, A. Wilson, W. E. Acree, and M. H. Abraham, Global J. Phys. Chem. **3**, 1–42 (2012).
[145]T. W. Stephens, A. Wilson, N. Dabadge, A. Tian, M. Zimmerman, H. J. Hensley, W. E. Acree, and M. H. Abraham, Global J. Phys. Chem. **3**, 1–16 (2012).
[146]D. M. Stovall, A. Schmidt, C. Dai, S. Zhang, W. E. Acree, and M. H. Abraham, J. Mol. Liq. **212**, 16–22 (2015).
[147]D. M. Stovall, C. Dai, S. Zhang, W. E. Acree, and M. H. Abraham, Phys. Chem. Liq. **54**, 1–13 (2016).
[148]X. Tong, D. Woods, W. E. Acree, and M. H. Abraham, Phys. Chem. Liq. **56**, 571–583 (2017).
[149]M. Topphoff, D. Gruber, and J. Gmehling, J. Chem. Eng. Data **45**, 484–486 (2000).
[150]D. M. Trampe and C. A. Eckert, J. Chem. Eng. Data **35**, 156–162 (1990).
[151]H. C. Van Ness and B. D. Smith, *International DATA Series: Selected Data On Mixtures: Thermodynamic Properties of Non-Reacting Binary Systems of Organic Substances*, Series A (Thermodynamics Research Center, Texas Engineering Experiment Station, Texas A&M University, 1984).
[152]P. Vrbka, B. Hauge, L. Frydendal, and V. Dohnal, J. Chem. Eng. Data **47**, 1521–1525 (2002).
[153]M. S. Wainwright, T. Ahn, D. L. Trimm, and N. W. Cant, J. Chem. Eng. Data **32**, 22–24 (1987).
[154]R. J. Wilcock, R. Battino, and E. Wilhelm, J. Chem. Thermodyn. **9**, 111–115 (1977).
[155]E. Wilhelm and R. Battino, Chem. Rev. **73**, 1–9 (1973).
[156]C. L. Young, "Organic compounds containing oxygen," in *IUPAC Solubility Data Series* (Pergamon Press (Aust.) Pty. Ltd., Potts Point, N.S.W., Australia, 1981), Vol. 5/6, pp. 186–236.
[157]K. V. Zaitseva, M. A. Varfolomeev, and B. N. Solomonov, Russ. J. Gen. Chem. **83**, 438–444 (2013).
[158]N. Haider, The checkmol/matchmol homepage, http://merian.pch.univie.ac.at/~nhaider/cheminf/cmmm.html.
[159]N. Schneider, N. Fechner, G. A. Landrum, and N. Stiefl, J. Chem. Inf. Model. **57**, 1816–1831 (2017).
[160]G. Landrum, RDKit: Open-Source Cheminformatics, http://www.rdkit.org.
[161]H. Shang, C. Carbogno, P. Rinke, and M. Scheffler, Comput. Phys. Commun. **215**, 26–46 (2017).
[162]N. M. O'Boyle, C. Morley, and G. R. Hutchison, Chem. Cent. J. **2**, 5 (2008).
[163]N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, J. Cheminf. **3**, 33 (2011).
[164]N. S. Antonova, J. J. Carbo, U. Kortz, O. A. Kholdeeva, and J. M. Poblet, J. Am. Chem. Soc. **132**, 7488–7497 (2010).
[165]R. Bermejo-Deval, R. S. Assary, E. Nikolla, M. Moliner, Y. Román-Leshkov, S.-J. Hwang, A. Palsdottir, D. Silverman, R. F. Lobo, L. A. Curtiss, and M. E. Davis, Proc. Natl. Acad. Sci. U. S. A. **109**, 9727–9732 (2012).
[166]A. M. Suess, M. Z. Ertem, C. J. Cramer, and S. S. Stahl, J. Am. Chem. Soc. **135**, 9797–9804 (2013).
[167]R. Bosque and J. Sales, J. Chem. Inf. Comput. Sci. **42**, 1154–1163 (2002).