# Calculation of Solvation Free Energies with DCOSMO-RS

Andreas Klamt[1,2],Michael Diedenhofen[1]

[1] COSMOlogic GmbH&Co.KG, Imbacher Weg 46, 51379 Leverkusen, Germany

[2] Institute of Physical and Theoretical Chemistry, University of Regensburg, Germany

e-mail: klamt@cosmologic.de

**Abstract**

The concept of dielectric continuum models has turned out to be very fruitful for the qualitative description of solvation effects in quantum chemical calculations, although from a theoretical perspective its basis is questionable, at least if applied to polar solvents, since the electrostatic nearest neighbour interactions in polar solvents are much too strong to be described by macroscopic dielectric continuum theory. Based on this insight the Conductorlike Screening Model for Realistic Solvation (COSMO-RS) had been developed, which gives a thermodynamically consistent, quantitative description of solvation effects in polar and non-polar solvents, even in mixtures and at variable temperature, starting from quantum chemical calculations of solute and solvent molecules embedded in a virtual conductor (COSMO). While COSMO-RS usually only requires quantum chemical calculations in the conductor and thus does not allow for studying of the concrete solvent influence on the solute electron density, the direct COSMO-RS (DCOSMO-RS) has been introduced, which uses the $\sigma$-potential, i.e. a solvent specific response function provided by COSMO-RS, as a replacement of the conductor or dielectric response employed in continuum solvation models. In this article we describe the current status of DCOSMO-RS and demonstrate the performance of the DCOSMO-RS approach for the prediction of free energies of solvation.

**Keywords**

dielectric continuum model, COSMO, COSMO-RS, solvation, DCOSMO-RS

**Introduction**

Quantum chemical methods have been developed since more than 50 years, and the combination of many methodological developments with the exponential growth of available compute power has made them indispensible tools for many areas of chemistry. But quantum chemistry by itself is only suitable for individual molecules or smaller clusters of molecules, and it is less applicable to

condensed systems with the exception of periodic structures, which can be reduced to small systems by making use of the periodic boundary conditions. Quantum chemistry is not well suited for treating molecules in the liquid state, in which each molecule is interacting with a large number of fluctuating neighbouring atoms. But since at least 90% of industrial chemistry and almost 100% of biochemistry takes place in solution, the restriction of quantum chemistry to gas phase or ordered crystalline states constitutes a serious limitation of the practical applicability and acceptance of quantum chemistry. Unfortunately, a fundamental treatment of liquid phases would require sufficiently long ab initio molecular dynamics simulations of sufficiently large ensembles of molecules, which is currently and in foreseeable future not feasible, at least not as a routine tool. Therefore the inclusion of solvation effects into quantum chemical calculations inevitably requires the introduction of some empiricism. One route is the combination of higher level quantum chemistry for a core region of interest with lower level, often force field representations for the surrounding solvent molecules, the so called QM/MM approach.[1] Another, computationally much more efficient approach is the dielectric continuum solvation approach, in which the entire solvent is represented by a simple dielectric continuum, supplemented by a few surface proportional non-dielectric contributions. Despite or because of its simplicity, this approach has become widely accepted and used in the quantum chemistry community.

The polarizable continuum model PCM, developed by Tomasi and many co-workers over more than three decades, comprises the most developed implementation of the dielectric continuum solvation approach.[2] It has been extended to many levels of ab initio quantum theory and many kinds of property calculations. Another variant of the dielectric continuum models is the conductor-like screening model COSMO,[3] which was originally independently developed, but shows a high degree of similarity to PCM. The novelty in COSMO was the usage of a scaled conductor boundary condition, instead of the exact dielectric boundary condition. This slightly approximate COSMO boundary condition was simpler to handle and computationally more efficient than standard PCM, and it turned out to strongly reduce the problem of outlying charge.[4] Therefore COSMO and its PCM implementation C-PCM have become very popular.[5] If used with the original COSMO scaling function $f(\varepsilon)=(\varepsilon-1)/(\varepsilon+0.5)$, which unfortunately is not the default in C-PCM, COSMO is as accurate as PCM, at least for neutral solutes, even at low dielectric constants[6].

The most empirical part of all dielectric continuum solvation models (DCSMs) is the construction of the interface between the solute and dielectric continuum. In the vast majority of DCSMs these cavities are constructed based on atom centred spheres with radii that are roughly 20% larger than typical Bondi radii. The number of empirical parameters introduced in this way varies strongly, since some models as COSMO just use element specific radii, i.e. one parameter per element, while many others employ atom-type specific radii, often even depending on the charge of the atom, and

sometimes even on the solvent. Our own finding is that element specific radii are good enough, at least for neutral compounds.

One big problem of the DCSMs results from the non-electrostatic short range interactions. While the dispersive interactions appear to be reasonably well approximated by element or atom type specific and surface proportional terms, which introduce one more element specific parameter, hydrogen bonding (hb) is completely out of the concept of DCSA. While the hb interactions between solvent molecules may be somehow hidden in the empirical, solvent specific, and mostly solute surface proportional cavity formation energy, there is no way in which the variation of the polarity of the hb donors and acceptors of the solute enters into the interaction energy expressions of the available DCSMs.

Furthermore, even a hypothetical liquid Lennard-Jones fluid of spherical molecules having permanent dipole moments, in the order of the dipole moment of water, would not behave like a dielectric continuum on a molecular length scale. If we assume a size of the solvent molecules similar to water, then nearest neighbour dipole interaction energy is about 4 kcal/mol, and the difference between aligned and anti-aligned dipoles is 8 kcal/mol.[7] But for applying the macroscopic, linear response dielectric theory, the dipole interaction energy needs to be small compared to the thermal energy, which is 0.6 kcal/mol at 298K. This simple numerical exercise proves that even the electrostatic interactions within the first solvation shell cannot at all be correctly described by the dielectric continuum approach. Given these deficiencies the considerable success of DCSMs is surprising. From a theoretical perspective the only solvents, for which the DCSMs should be applicable for systematic reasons are non-polar solvents as alkanes or noble gases, having no significant permanent local dipole moments or quadrupole moments.

The Conductorlike Screening Model for Realistic Solvation (COSMO-RS) was introduced in order to overcome these problems.[7-9] It starts from COSMO calculations for the solute and solvent molecules in the limit of a conductor, i.e. with $\varepsilon=\infty$. The resulting surface polarization charge densities $\sigma$ are used for a quantification of the electrostatic and hydrogen bond interactions of the molecules. By a statistical thermodynamics treatment of all possible pair-wise interactions of the surface segments of the solvent COSMO-RS generates a free energy function, the so-called solvent $\sigma$-potential $\mu_S(\sigma)$, which is the free energy of a surface of polarization charge density $\sigma$ in the solvent. Integration of this $\sigma$-potential over the surface of a solute X then leads to an expression for the chemical potential of the solute X in the solvent S. The full concept of COSMO-RS is described in more detail elsewhere.[7-9]

COSMO-RS was shown to provide currently the most accurate predictions $\Delta G_{solv}$ on a dataset of 2434 neutral compounds.[10] Meanwhile it is widely validated and used for many different kinds of fluid phase thermodynamics property predictions for chemical engineering, drug design and development, and many other purposes.[11-15]

Advantages of COSMO-RS over conventional DCSMs are the equal treatment of solutes and solvents, which avoids any special parameterization of solvents that was required in conventional DCSMs, and its applicability to mixtures and variable temperatures. Furthermore, COSMO-RS is very efficient, because the time consuming quantum chemical calculations are only required once per molecule, and not repeatedly in each solvent.

Nevertheless, the latter advantage also provides a certain disadvantage, because COSMO-RS thus does not yield the solvent specific response of the molecular properties, e.g. electron density, dipole moment, spectroscopic properties, etc., which can routinely be calculated with DCSMs. Therefore the direct COSMO-RS (DCOSMO-RS) approach has been introduced, [16,17] which replaces the empirical dielectric continuum solvent response functions by the solvent σ-potential, which results from regular COSMO-RS. Thus the solvent response includes hydrogen bonding and the more realistic electrostatic response function of COSMO-RS. This approach has already been introduced and applied to spectroscopic property prediction previously.[16] In this paper we present our latest, improved version of DCOSMO-RS and demonstrate its ability to predict solvation free energies of molecules in a broad variety of solvents.

**Theory**

In conventional DCSMs the solvation energy of a solute in a solvent is a quadratic expression with respect to the polarization charge densities, which can be most simply seen in the COSMO formalism, where the COSMO energy contribution of a molecule X in a solvent S with dielectric constant ε(S) is given as

$$E_S^X = -\frac{1}{2}\frac{\varepsilon(S)-1}{\varepsilon(S)+0.5}\sum_{i\in X}q_i\sum_j A_{ij}^X q_j = -\frac{1}{2}\frac{\varepsilon(S)-1}{\varepsilon(S)+0.5}\sum_{i\in X}\sigma_i s_i\sum_{j\in X}A_{ij}^X s_j\sigma_j \tag{1a}$$

$$= -\frac{1}{2}\sum_{i\in X}\sigma_i s_i\sum_{j\in X}A_{ij}^X s_j\sigma_j + \frac{3}{4}\frac{1}{\varepsilon(S)+0.5}\sum_{i\in X}\sigma_i s_i\sum_{j\in X}A_{ij}^X s_j\sigma_j \tag{1b}$$

where the $q_i$, $s_i$ and $\sigma_i$ are the charges, surface areas and polarization charge densities, respectively, on the surface segments of the solute X. $A_{ij}^X$ is the COSMO interaction matrix of the surface segments. The polarization charges densities are defined by

$$\sigma_i = \frac{q_i}{s_i} = -\sum_j \frac{A_{ij}^{X,inv}}{s_i}\Phi_j^X = -\sum_j \frac{A_{ij}^{X,inv}}{s_i}\left(\Phi_j^{X,nuc} + \int \frac{\rho_{el}^X(\underline{r})}{\|\underline{r}-\underline{r}_j\|}d\underline{r}\right) \tag{2}$$

where $A_{ij}^{X,inv}$ is the inverse of the COSMO interaction matrix and $\Phi_j^X$ is the electrostatic potential produced by the solute X, i.e. its nuclei and electron density, on segment j. In eq. 1b the expression has

been split into a conductor contribution, and an ε(S) dependent contribution, which reflects the deviation of the solvent from a perfect conductor. The COSMO operator required for HF or DFT calculations then is derived from the functional derivative of $E_S^X$ with respect to the electron density $\rho_{el}^X$.

In COSMO-RS the deviation of the solute energy of a solute X in a solvent S relative to a perfect conductor is given by

$$\mu_S^X(T) = \sum_i s_t \mu_S(\sigma_i, T) + \mu_C^X(S) \tag{3}$$

where $\mu_S(\sigma)$ is the σ-potential of the solvent, which can be calculated within the standard COSMO-RS formalism, and the last term is the combinatorial contribution, which takes into account the size ratio of solute and solvent, and which does not depend on the electron density. If we replace the last term in eq. 1b by $\mu_S^X(T)$ we get a new expression for the free energy of the solute in solvent S, which we consider as the DCOSMO-RS expression, being composed of the COSMO expression for the conductor limit and a COSMO-RS contribution.

$$E_S^X(T) = -\frac{1}{2}\sum_{i \in X} \sigma_i s_i \sum_{j \in X} A_{ij}^X s_j \sigma_j + f_{pol}\sum_i s_t \mu_S(\sigma_i, T) + \mu_C^X(S) \tag{4}$$

The factor $f_{pol}$ has been introduced to account for the missing solute-solvent back polarization. The expression in equation 4 is a combination of the conductor contribution ($\varepsilon = \infty$) of the COSMO model and the σ-potential depended part and the combinatorial contribution of the COSMO-RS model, which is the straight forward combination of the two step procedure, where the COSMO-RS calculation use the pre-calculated COSMO information.

Because the parabolic part of the σ-potential can be described well by the COSMO model,[18] the problem can be re-formulated using the ε dependent COSMO equation combined with a reduced σ-potential

$$E_S^X(T) = -\frac{1}{2}f(\varepsilon)\sum_{i \in X} \sigma_i s_i \sum_{j \in X} A_{ij}^X s_j \sigma_j + f_{pol}\sum_i s_t \tilde{\mu}_S(\sigma_i, T) + \mu_C^X(S) \tag{5a}$$

with

$$\tilde{\mu}_S(\sigma, T) = \mu_S(\sigma, T) - (1 - f(\varepsilon))c_0\sigma^2 \ , \ f(\varepsilon) = \frac{\varepsilon(S) - 1}{\varepsilon(S) + 0.5} \tag{5b}$$

Both models (eq. 4 and eq. 5) coincide in the limit of the perfect conductor ($\varepsilon = \infty$, $f(\varepsilon) = 1$).

Now functional variation with respect to the electron density leads to the required DCOSMO-RS operator required for HF or DFT calculations. This leads to

$$V_S^{DCRS}(\underline{r}) = \frac{\partial E_S^X(T)}{\partial \rho_{el}^X(\underline{r})} = -\sum_{i \in X} s_i \frac{\partial \sigma_i}{\partial \rho_{el}^X(\underline{r})} \sum_{j \in X} A_{ij}^X s_j \sigma_j + \sum_i s_t \frac{\partial \sigma_i}{\partial \rho_{el}^X(\underline{r})} f_{pol} \frac{d\tilde{\mu}_S(\sigma_i, T)}{d\sigma_i}$$

$$= \sum_{i \in X} \frac{\partial q_i}{\partial \rho_{el}^X(\underline{r})} \left[ -\sum_{j \in X} A_{ij}^X q_j + f_{pol} \tilde{\mu}_S^{'}(\sigma_i, T) \right] \tag{7}$$

with

$$\tilde{\mu}_S^{'}(\sigma_i, T) = \frac{d\tilde{\mu}_S(\sigma, T)}{d\sigma} \bigg|_{\sigma=\sigma_i} \tag{8}$$

being the derivative of the σ-potential. Obviously the operator $\dfrac{\partial q_i}{\partial \rho_{el}^X(\underline{r})}$ is already available in any

COSMO implementation. Therefore the additional DCOSMO-RS contribution to the solvation operator

can be programmed as a product of $\dfrac{\partial q_i}{\partial \rho_{el}^X(\underline{r})}$ and the derivative of the σ-potential. Using this operator

instead of the standard COSMO operator yields an effective description of the solvent influence on the

electron density of the solute X.

In the initial version of DCOSMO-RS [16] the σ-potential was represented in form of a generalized Taylor

expansion, and analytical expression for the derivative were available. Meanwhile we have found some

σ-potentials which cannot be fitted adequately by the generalized Taylor expansion. The two examples

given in figure 1 show deviations in the potentials and the derivatives in the range between -0.01 and

0.01 e/A$^2$ and for σ values smaller than -0.025 e/A$^2$. This deviations influence the interaction of less

polar molecular surface as well as the hydrogen bond donor interaction and it turned out as favorable to

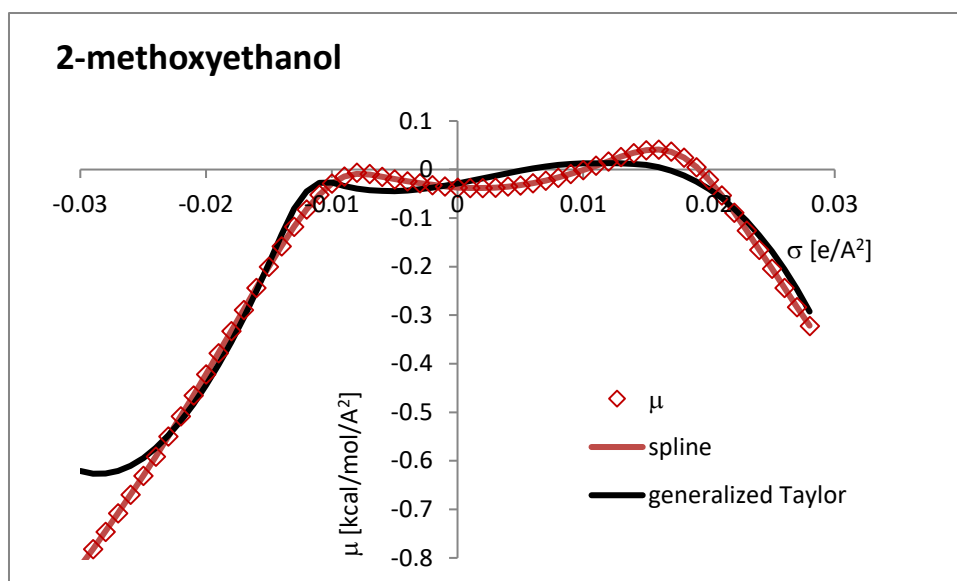represent the σ-potential as equally spaced cubic splines providing continuous derivatives.
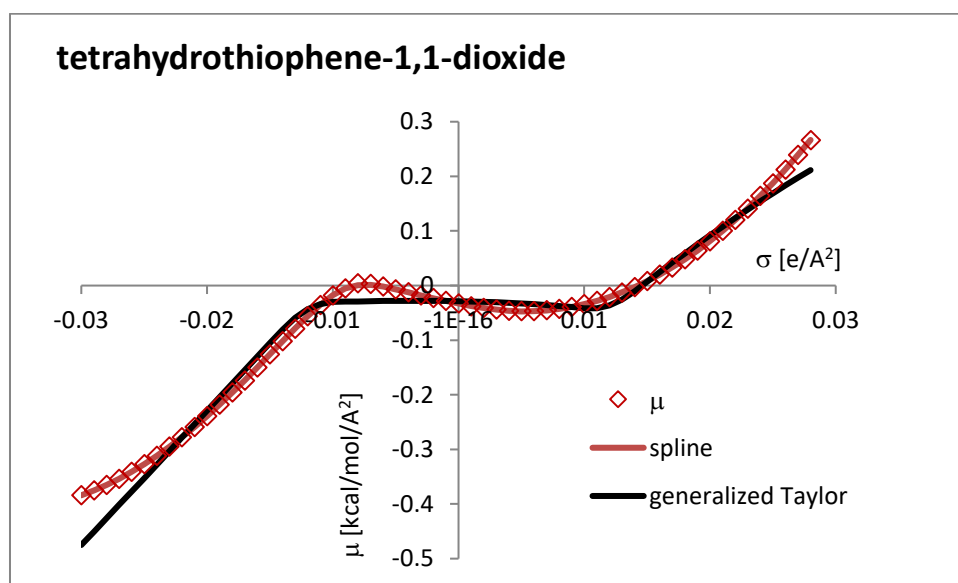
Figure 1: Different representations of the σ-potential (generalized Taylor expansions and splines) of tetrahydrothiophene-1,1-dioxide and 2-methoxyethanol (pure solvent, 25 °C)

Another change compared to the initial version of DCOSMO-RS has been introduced with respect to the element dependence of the σ-potential. Originally the σ-potentials had been identical for all segments, independent of the atom, to which the segment belongs. But due to the element-specific dispersion contributions, and due to the fact that the hydrogen bond donor interaction only applies to hydrogen atoms, the σ-potential is slightly element specific. Therefore the current version of DCOSMO-RS uses a different σ-potential for each of the parameterized elements. An example of σ-potentials of water and hexane for the solute elements H and O is given in figure 2.
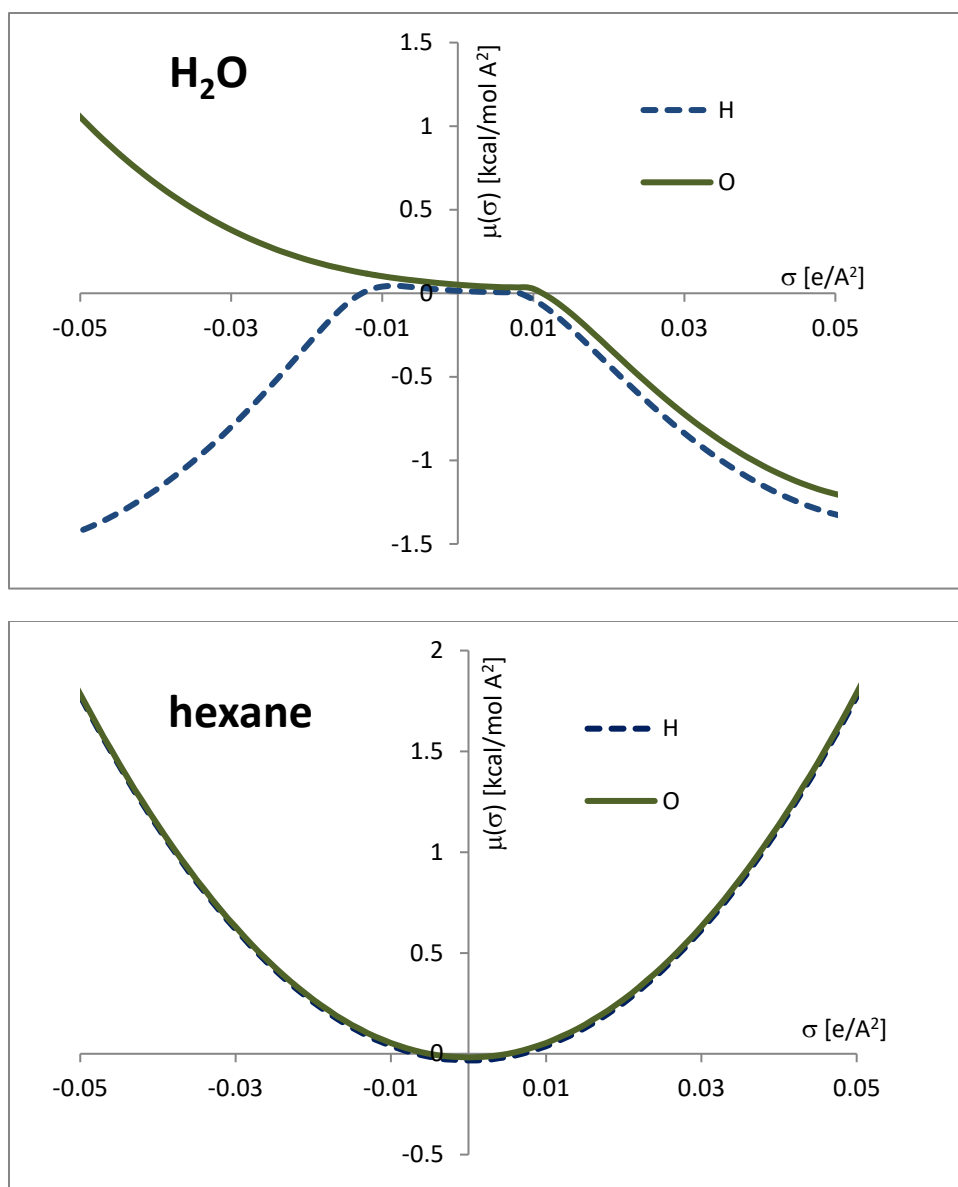
Figure 2: Element specific σ-potentials of water and hexane (pure solvent, 25 °C)

These modified σ-potentials now reflect closely the COSMO-RS procedure and thus it can be expected that solvation free energies calculated with the current version of DCOSMO-RS yield solvation free energies with nearly the quality achieved in COSMO-RS. Therefore we tested the accuracy of DCOSMO-RS in its TURBOMOLE implementation on the 2434 free energies of solvation of a broad collection of solutes in various solvents, which had been collected for the parameterization of the SM8 dataset.[19] Some differences still have to expected due to the fact that the polarization charge densities now are different from those in a perfect conductor due to increased backpolarization of the solute X in the DCOSMO-RS procedure.

## Technical Details

All quantum chemical calculation are performed with the TURBOMOLE V6.6 program package [20] using the BP86 functional[21-23] and the TZVP basis set [24,25]. The structures have been optimized for the perfect conductor (COSMO, $\varepsilon = \infty$) and the gas phase. The influence of the full DCOSMO-RS geometry optimization was found to be negligible. These structures have been used for the DCOSMO-RS and the COSMO calculations with finite $\varepsilon$.

The σ-potentials used in the DCOSMO-RS calculations and COSMO-RS free energies of solvation were calculated with the COSMO*therm* program using the BP_TZVP_C21_0111 parameterization and the standard conformer treatment.[9,26] All σ-potentials and dielectric constants used in this paper are provided as supplementary material.

The $c_0$ parameter of eq. 6 was obtained from the σ-potential of hexane ($c_0 = 1447.24$ kcal A$^2$/mol/e$^2$). The $f_{pol}$ factor, which was introduced in order to account for the potentially increased backpolarization in DCOSMO-RS compared to standard COSMO-RS, was set to 1. A variation of the factor did not improve the results. For the combinatorial contribution $\mu_C^X(S)$ we used the default combinatorial expression of COSMOtherm as implemented in TURBOMOLE.

The DCOSMO-RS free energies of transfer of a solute molecule from an ideal gas at molar (1 mol/l) concentration to an ideal solution at the same solute concentration were calculated as follows:

$$\Delta G_{solv} = E^{DCRS} - (E^{gas} + \omega_{ring} n_{ring} + \eta_{gas}) - RT \ln(\rho_S V_{iG} / M_S) \qquad (9)$$

$E^{DCRS}$ is the free energy on the DCOSMO-RS level as defined in eq. 4. The value has been corrected with the outlying charge correction of the COSMO model.[4]

$E^{gas}$ is the DFT gas phase energy. $V_{iG}$ and $M_S$ are the molar volume of the ideal gas and the molar weight of the solvent, and $\rho_S$ is the solvent density, which was taken from the COSMO*therm* prediction. Because the σ-potential contribution depends on the reference state of the COSMO-RS parameterization we need to introduce the $\eta_{gas}$ parameter providing the link between the reference states. During the COSMO-RS development an entropic ring correction term turned out to be important for the description of the gas phase. This term consists of the ring correction parameter $\omega_{ring}$ and the number of ring atoms of the solute ($n_{ring}$). The used corrections are tabulated in the supplementary material.

**Results**

The mean unsigned errors (MUEs) of the COSMO-RS and DCOSMO-RS free energies of solvation of the neutral compounds of the SM8 data set[19] are given in table 1. For reference we added the respective MUEs for when the bare COSMO to gas phase energy difference is used. The DCOSMO-RS calculations have been performed with the two models described in the theory section. The results obtained with finite (eq. 5) and infinite (eq. 4) dielectric constant are denoted as *eps* and *inf*, respectively.

Table1: Performance of DCOSMO-RS, COSMO-RS, and COSMO on Solvation Free Energies at 25°C of Neutral Compounds (Mean Unsigned Error in kcal/mol).

| data set | data points | DCOSMO-RS | | COSMO-RS | COSMO |
|---|---|---|---|---|---|
| | | *eps*[a] | *inf*[b] | | |
| All | 2346 | 0.66 | 0.67 | 0.42 | 2.14 |
| aqueous solvation | 274 | 0.94 | 1.05 | 0.52 | 2.19 |

[a] The dielectric constants of the solvents were used. $\eta_{gas}$ =-7.45 kcal/mol

[b] Calculated for $\varepsilon = \infty$. $\eta_{gas}$ =-6.85 kcal/mol

A correlation of the calculated and experimental free energies of solvation is given in figure 3.
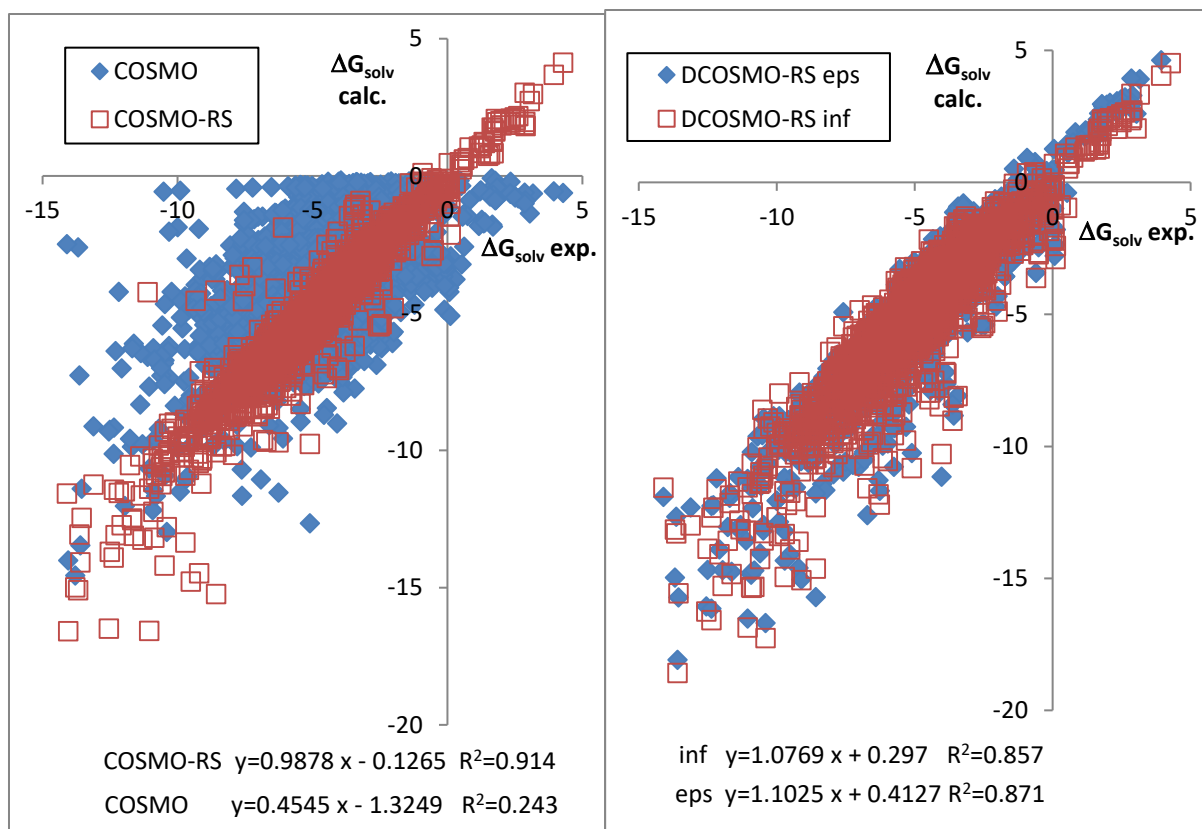
Figure 3: Correlation of calculated and experimental solvation free energies in kcal/mol (all 2346 data points).

As expected, both DCOSMO-RS models exhibit MUEs which are in between the errors of the bare electrostatic COSMO model and the COSMO-RS predictions. For the considered data set there is no clear preference for one of them. For its slight advantage on the most important aqueous solvation we have chosen the *eps* model as default. Nevertheless the *inf* model may be useful, if no experimental dielectric constant is available. Both models can predict free energies of solution with the same accuracy. The plots in figure 3 show a poor correlation between the bare COSMO predictions and the experimental values, which is substantially improved by the two DCOSMO-RS models. The best correlation with respect to slope, intercept and coefficient of determination can be found for the COSMO-RS predictions. Because of the direct influence of the σ-potential on the SCF calculation, which does not exist in the pure COSMO-RS theory, we need to adjust the $\eta_{gas}$ parameter. The adjusted $\eta_{gas}$ parameter of the two DCOSMO-RS model turned out to be -7.45 kcal/mol for the *eps* model and -6.85 kcal/mol for the *inf* model. Both values are of the same order of magnitude as the $\eta_{gas}$ parameter of the used COSMO*therm* parameterization (-6.65 kcal/mol). If we use $\eta_{gas}$ from the COSMO*therm* parameterization for the *inf* model, which is closer to the pure COSMO-RS approach, we get a MUE of 0.67 kcal/mol for the full data set and MUE=1.17 kcal/mol for aqueous solvation subset. The performance on the aqueous data set is a bit worse, but this model is purely predictive without adjustment to the data set under consideration.

With a MUE of about 0.7 kcal/mol the DCOSMO-RS models provide a substantial improvement of the pure COSMO approach. The performance is only slightly worse than the one of the SM8 model (MUE=0.59 kcal/mol) which has been fitted on the data set under consideration.[19]

The predictions of the subset of the free energies of hydration show the same trends as for the full data set. The MUEs on the COSMO level are nearly the same whereas the COSMO-RS and DCOSMO-RS are less accurate for the aqueous solvation subset.

The achieved over all accuracy of DCOSMO-RS is almost equal to the accuracy of IEF-PCM/MST as reported on a subset of 960 data points for non-aqueous solvation (0.64 kcal/mol).[10] The performance on the subset for aqueous solvation is also similar.

The overall performance of the COSMO-RS model is slightly better than the value reported in ref. 10 (MUE=0.48 kcal/mol) which is mainly due to the newer parameterization that has been used in this work.

The two DCOSMO-RS models exhibit an asymmetrical error distribution, which can also be found for the COSMO-RS method albeit in general with smaller deviations. The largest positive deviations in the order of 2.7-3.0 kcal/mol appear for tertiary amines, which are known since long to cause problems in COSMO-RS, and for formaldehyde in alcohols. For the latter its tendency to react with solvents may put some question marks on the experimental data. For formaldehyde in 1-butanol a reaction free energy of ~1.4 kcal/mol can be derived from the work of Peschla et al..[27] Larger errors can be found at the negative deviation side. Here the largest outlier (-7.12 kcal/mol) is diethyl-4-nitrophenylthiophosphate, i.e. the pesticide parathion, in benzene. In this case we clearly doubt the experimental value, since on the one hand the predictions for parathion in water and octanol are both within an error range of less than 2 kcal/mol, and on the other hand the CRC Handbook of Pesticides[28] reports miscibility of parathion in benzene, which would correspond to a much more negative value of $\Delta G_{solv}$ than reported in the SM8 dataset. The same argument most likely also applies to dimethyl-4-nitrophenylthiophosphate in benzene and chloroform, which are both within the top 20 outliers. More serious are several solutes with electronically coupled donors and acceptors in hydrogen bonding solvents, as urea, water and hydrazine in water, which are reasonably well predicted by COSMO-RS but large negative outliers in D-COSMO-RS. Therefore these seem to be over-polarized by D-COSMO-RS.

**Conclusions**

The improved DCOSMO-RS method based on element specific σ-potentials provides a robust and widely applicable method for the quantitative treatment of solvation effects in a wide range of solvents. Solvation free energies for a broad variety of solutes in very different solvents can be predicted with an accuracy of ~0.7 kcal/mol (MUE) on a test set of 2434 data. This is a substantial improvement over the pure COSMO approach and only ~0.3 kcal/mol worse than the accuracy of the full COSMO-RS procedure. It is only slightly worse than the SM8 method, which had been parameterized exactly on the considered dataset.

Since COSMO-RS is applicable to a very wide range of pure and mixed solvents, including ionic liquids, at variable temperature without requiring any experimental information about the solvents, σ-potentials are easily available for all these solvents without additional parameterization. Thus DCOSMO-RS can be easily applied to almost any solvent, including mixtures. Its strength surely is not the bare prediction of solute free energies. For this purpose the COSMO-RS method is still more efficient and more accurate, and thermodynamically more consistent. But the presented improved DCOSMO-RS now can serve as an efficient tool which combines a robust quality solvation free energy prediction with the capability for response property calculations beyond the dielectric continuum limitations.

**Supporting Information Available:** The element-specific σ-potentials for all solvents used in this paper and the list of experimental and calculated solvation free energies for the 2346 compounds are available as supplementary material free of charge via the internet at http://pubs.acs.org.

**References**

(1) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme, *J. Mol. Bio.* **1976**, *103*, 227-249.

(2) Tomasi, J.; Persico, M. Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent, *Chem. Rev.* **1994**, *94*, 2027-2094.

(3) Klamt, A.; Schüürmann, G. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. *J. Chem. Soc. Perk. T.* **1993**, *2*, 799-805.

(4) Klamt, A.; Jonas, V. Klamt, A., Jonas, V. Treatment of the Outlying Charge in Continuum Solvation Models. *J. Chem. Phys.* **1996**, *105*, 9972-9981.

(5) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995-2001.

(6) Cossi, M.; Rega, N.; Scalmani, G.; Barone V. Energies, Structures, and Electronic Properties of Molecules in Solution with the C-PCM Solvation Model. *J. Comput. Chem.* **2003**, *24*, 669-681.

(7) Klamt A. *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*, Elsevier: 2005

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

(9) Eckert, F.; Klamt, A. Fast Solvent Screening via Quantum Chemistry: COSMO-RS Approach. *AIChE Journal* **2002**, *48*, 369-385.

(10) Klamt, A.; Mennucci, B.; Tomasi, J.; Barone, V.; Curutchet, C.; Orozco, M.; Luque, F. J. On the Performance of Continuum Solvation Methods. A Comment on "Universal Approaches to Solvation Modeling". *Acc. Chem. Res.* **2009**, *42*, 489-492.

(11) Jork, C.; Kristen. C.; Pieraccini, D.; Stark, A.; Chiappe, C.; Beste, Y.A.; Arlt, W. Tailor-made Ionic Liquids. *J. Chem. Thermodynamics* **2005**, *37*, 537-558.

(12) Deglmann, P.; Schenk, S. Thermodynamics of Chemical Reactions with COSMO-RS: The Extreme Case of Charge Separation or Recombination. *J. Comp. Chem.* **2012**, *33*, 1304-1320.

(13) Oleszek-Kudlak, S.; Grabda, M.; Shibata, E.; Eckert, F.; Nakamura, T. Application of the Conductor-like Screening Model for Real Solvents for Prediction of the Aqueous Solubility of Chlorobenzenes Depending on Temperature and Salinity. *Environ Toxicol Chem.* **2005**, *24*, 1368-1375.

(14) Andersson, M. P.; Bennetzen, M. V.; Klamt, A.; Stipp, S. L. S. First-Principles Prediction of Liquid/Liquid Interfacial Tension. *J. Chem. Theory Comput.* **2014**, *10*, 3401-3408.

(15) Paloncýová, M.; DeVane, R; Murch, B.; Berka, K.; Otyepka, M. Amphiphilic Drug-Like Molecules Accumulate in a Membrane below the Head Group Region. *J. Phys. Chem. B* **2014**, *118*, 1030-1039.

(16) Sinnecker, S.; Rajendran, A.; Klamt, A.; Diedenhofen, M.; Neese, F. Calculation of Solvent Shifts on Electronic g-Tensors with the Conductor-Like Screening Model (COSMO) and Its Self-Consistent Generalization to Real Solvents (Direct COSMO-RS). *J. Phys. Chem. A* **2006**, *110*, 2235-2245.

(17) Renz, M.; Kess, M.; Diedenhofen, M.; Klamt, A.; Kaupp, M. Reliable Quantum Chemical Prediction of the Localized/Delocalized Character of Organic Mixed-Valence Radical Anions. From Continuum Solvent Models to Direct-COSMO-RS. *J. Chem. Theory Comput.* **2012**, *8*, 4189-4203.

(18) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. Refinement and Parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, 5074-5085.

(19) Cramer, C. J.; Truhlar, D. G. A Universal Approach to Solvation Modeling. *Acc. Chem. Res.* **2008**, *41*, 760-768.

(20) TURBOMOLE V6.6 2014, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007

(21) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098-3100.

(22) Vosko, S. H.; Wilk, L.; Nussair, M. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis. *Can. J. Phys.* **1980**, 58, 1200-1211.

(23) Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev. B* **1986,** *33*, 8822-8824.

(24) Schäfer, A.; Huber, C.; Ahlrichs, R. Fully Optimized Contracted Gaussian Basis Sets of Triple Zeta Valence Quality for Atoms Li to Kr. *J. Chem. Phys.* **1994**, *100*, 5829-5835.

(25) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. Auxiliary Basis Sets for Main Row Atoms and Transition Metals and Their Use to Approximate Coulomb Potentials. *Theor. Chem. Acc.* **1997**, *97*, 119-127.

(26) Eckert, F.; Klamt, A.; COSMO*therm*, Version C2.1, Release 01.10; COSMOlogic GmbH & Co. KG, Leverkusen, Germany, 2010.

(27) Peschla,R.; Garcia, B.C.; Albert, M.; Kreiter, C.; Maurer, G. Chemical Equilibrium and Liquid-Liquid Equilibrium in Aqueous Solutions of Formaldehyde and 1-Butanol *Ind. Eng. Chem. Res.* **2003**, *42*, 1508-1516.

(28) Milne, G.W.A., CRC Handbook of Pesticides, CRC Press, **1995**.

**TOC-graphics**



$$E_S^X(T) = -\frac{1}{2}f(\varepsilon)\sum_{i \in X}\sigma_i s_i \sum_{j \in X}A_{ij}^X s_j \sigma_j + f_{pol}\sum_i s_t \widetilde{\mu}_S(\sigma_i, T) + \mu_C^X(S)$$