On the performance of continuum solvation methods.
A comment on "Universal approaches to solvation modeling"

Andreas Klamt,[*,1,2] *Benedetta Mennucci,* [3] *Jacopo Tomasi,* [3] *Vincenzo Barone,* [4] *Carles Curutchet,*[5]
Modesto Orozco[6] *and F. Javier Luque*[7]

1 COSMOlogic GmbH&COKG, Burscheider Str. 515, 51381 Leverkusen, Germany

2 Department of Physical and Theoretical Chemistry, University of Regensburg, 93040 Regensburg, Germany

3 Department of Chemistry, University of Pisa, Via Risorgimento 35, 56126 Pisa, Italy

4 IPCF-CNR, via Moruzzi 1, 56100 Pisa and Dipartimento di Chimica Università Federico II, Napoli.

5 present affiliation: Department of Chemistry, University of Toronto, 80 St. George Street, Toronto, Ontario, M5S 3H6 CANADA

6 Molecular Modeling and Bioinformatics Unit, Institut de Recerca Biomèdica. Barcelona Scientific Park, Josep Samitier 1-6, 08028 Barcelona, Spain, Department of Life Sciences, Barcelona Supercomputing Centre, Jordi Girona 29, 08034 Barcelona, Spain and Departament de Bioquímica, Facultat de Biologia, Avgda Diagonal 647, Barcelona 08028, Spain

7 Department of Physical Chemistry and Institute of Biomedicine (IBUB), Faculty of Pharmacy, University of Barcelona, Avda. Diagonal 643, 08028 Barcelona, Spain

RECEIVED DATE (to be automatically inserted after your manuscript is accepted if required according to the journal that you are submitting your paper to)

* Corresponding Author: Tel.: +49-2171-731681. E-mail: klamt@cosmologic.de

CONSPECTUS   In a recent Account Cramer and Truhlar present a comparison between the SM8 method and other continuum solvation models showing that the former leads to "considerably smaller errors for aqueous and nonaqueous free energies of solvation for neutrals, cations, and anions, with particularly good performance for nonaqueous data". In this Commentary we demonstrate that such a

comparison is not well founded and that competing solvation methods such as C-PCM, IEF-PCM, or COSMO-RS are indeed as accurate as SM8 method, if they are applied with the same rigor.

In their recent Account[1] "A Universal Approach to Solvation Modeling" Cramer and Truhlar provide a concise description of the performances of the SM8 solvation model, which have been described in detail in a recent paper.[2] The SM8 model is the latest version of the series of SMx continuum methods and has been parameterized using the experimental solvation free energies of a large set of compounds containing a variety of functional groups in different solvents. These data, which consist of 940 entries for neutral compounds and 332 for ions, have been used to optimize a large number of empirical model parameters, including solvent-dependent Coulomb radii, solvent radii, atomic surface tensions, and other parameters contained in the charge models and in the generalized Born model. The SM8 model is described as the first continuum solvation model universally applicable to general solvents, for which a small number of physicochemical descriptors are known either experimentally or from estimation.

In Table 1 of their Account the authors demonstrate that the SM8 model succeeds in reproducing the solvation free energies of the training data set with a mean unsigned error (MUE) of 0.59 kcal/mol for neutral solutes and 4.31 kcal/mol for ions. A comparison with four other solvation models is also presented in Table 1. One of them is the Poisson-Boltzmann method implemented in the Jaguar program[3,4] and the other three, which pertain to the class of self-consistent reaction field (SCRF) continuum solvation models, are

1) the integral equation formalism of the Polarizable Continuum Model, IEF-PCM[5], which is the default version of the PCM family of methods[6] in Gaussian03[7] program,

2) the C-PCM method,[8] which is an implementation of the conductor-like screening ansatz COSMO[9] in the framework of the PCM model (the most effective version of the method is also available in Gaussian03[10]),

3) and the COSMO implementation[11] in the NWCHEM program,[12] which is inadequately referenced as GCOSMO by Cramer and Truhlar[1,2].

For the same set of training compounds, MUEs of 5.66, 2.43 and 4.29 kcal/mol are reported for the solvation free energies predicted for neutral solutes (and even larger MUEs for ionic species) using IEF-PCM, C-PCM and NWCHEM/COSMO models, respectively. In our view the conclusions that a reader can get from such numbers are not well founded as they are based on an unfair comparison of the predicted solvation free energies determined by using the highly trained SM8 model with non-optimized versions of the IEF-PCM, C-PCM and COSMO models. This point must be stressed in order to avoid a wrong impression about "non-SMx models" for non-expert readers in SCRF models.

In order to examine the ability of IEF-PCM, C-PCM and COSMO models to estimate solvation free energies, we have determined the solvation free energy for the compounds included in the SM8 training set with the IEF-PCM and COSMO-RS[13,14] models using the MST (Miertus-Scrocco-Tomasi)[15,16] and COSMOtherm_C2.1_0107[17] versions, respectively, which represent in our view the state-of-the-art for these methods when applied to the calculation of solvation free energies (due to the close similarity between IEF-PCM and C-PCM results for a common set of parameters, the latter method will not be discussed here).

With this aim, we retrieved the chemical structures of all solutes from the chemical names given in the supplementary material of the original SM8 paper.[2] Then we used the standard procedures established for IEF-PCM/MST and COSMO-RS to determine the solvation free energies. In essence,

IEF-PCM/MST computations were performed at both HF and DFT (B3LYP) levels of theory using the 6-31G(d) basis and molecular geometries optimized in the gas phase, whereas COSMO-RS results were obtained at the DFT level (BP functional with TZVP basis set) and geometries optimized in vacuum and at $\varepsilon = \infty$. Multiple conformations were taken into account self-consistently where necessary. Computations were performed using Gaussian03 for IEF-PCM/MST and TURBOMOLE[18] and COSMOtherm programs for COSMO-RS, respectively, with no additional fitting of any parameters.

The statistical performance of COSMO-RS can be examined from the results given in Table 1, which also reports the corresponding SM8 results taken from Table 1 in the reference paper.[1] Since COSMOtherm does not require individual quantum chemical calculations of each solute in each solvent, but only one per solute in a perfect conductor, COSMOtherm was applied to all 91 solvents and 2346 solute/solvent pairs given in the supplementary material of the SM8 paper. The results show that the performance of COSMO-RS is identical or slightly better than that of the SM8 method, even though the set of compounds was used as training data for SM8. It must be noted that no experimental or adjustable solvent parameters have been used in COSMOtherm calculations, because the solvents are described on the same DFT/COSMO basis as the solutes within the COSMO-RS theory. The largest deviations in COSMOtherm calculations are found for thiophosphate compounds, for which the structures are slightly questionable and the experimental data show at least some strange trends. The straight applicability and good performance of COSMO-RS for all 91 solvents also proves that COSMO-RS is a universal solvation model. Indeed, the "universal" nature of COSMO-RS is further underlined by its ability to treat mixtures of solvents without the need of any experimental solvent data.

Table 1: Statistical performance (mean unsigned error in kcal/mol) of COSMO*therm* and SM8 on solvation free energies of neutral compounds.

data set	N	COSMOtherm	SM8
all neutral	2346	0.48	0.59
subset of 17 solvents	960	0.56	0.61
aqueous solvation	284	0.58	0.55

Since the MST version of the IEF-PCM has been parameterized for water, octanol, chloroform and carbon tetrachloride, the statistical analysis reported in Table 2 is limited to the subset of experimental data available for these solvents. In addition, compounds containing P, Br and Si were eliminated from the SM8 data set as these elements were not considered in the MST parameterization.

Table 2: Statistical performance (mean unsigned error in kcal/mol) of IEFPCM-MST and SM8 on solvation free energies of neutral compounds.

data set	N	IEF-PCM/MST HF	IEF-PCM/MST B3LYP	SM8
nonaqueous solvation	960*	0.64	0.64	0.61
aqueous solvation	284**	1.01	1.01	0.55

* IEF-PCM/MST calculations performed on 350 compounds and 3 solvents (octanol, $CHCl_3$, CCl_4).
** IEF-PCM/MST calculations performed on 237 compounds.

The results show that the MUEs of IEF-PCM/MST and SM8 results are nearly the same for compounds in octanol, chloroform and carbon tetrachloride. Moreover, the MUE resembles that obtained in the original parameterization of the IEF-PCM/MST method for each solvent using a smaller set of compounds (root mean-square deviation, RMSD, of 0.6/0.6, 0.5/0.5 and 0.3/0.3 kcal/mol in octanol, chloroform and carbon tetrachloride at the HF/B3LYP level for 63, 56 and 48 compounds).[15,16] A larger MUE is found in water, also slightly larger than that obtained in the original MST parameterization (RMSD of 0.8/0.8 for 72 compounds). This trend is not surprising and can be mainly attributed to the limited number of optimized empirical parameters used in the MST implementation in order to maintain the general applicability of the IEF-PCM formalism.

Overall, the preceding discussion allows us to conclude that IEF-PCM, C-PCM, and COSMO-RS methods are capable to predict the experimental solvation free energies of (bio)organic compounds in a variety of solvents. The only premise is that the user must keep the specific details of the computational protocol adopted in the parameterization of each continuum model.[19] In particular, the default implementations of IEF-PCM (and C-PCM) in Gaussian03, which have been used in the reference paper,[1] have never been quantitatively parameterized for the reproduction of solvation free energies. As a result, the default set of radii (known with the UA0 acronym) used to build up the solute/solvent interface is not recommended for the quantitative prediction of solvation free energies. As a matter of fact such a cavity represents a compromise to gain qualitative insight into the role played by solvation on a variety of properties of the solute without the need to change the cavity parameters each time. At this point, we should emphasize that the accurate prediction of solvation free energies is not the only requirement for a solvation method, which must also reliably predict solvent effects on geometrical, vibrational, electronic and magnetic properties, as well as on many response properties of any solvated system.[6,20,21]

CONCLUSIONS

Understanding of the solvent effect on the properties of molecules and chemical processes still represents a challenge for the community of theoretical chemists, which explains the variety of computational strategies developed for the study of chemistry in condensed phases. Among these strategies, SCRF methods, in spite of the rather crude approximation of representing the solvent as a polarizable continuum medium, have a number of advantages, such as the definition of computational algorithms based on well established physical laws, the simplicity of the mathematical formalism, and the reduced computational cost. The progress made in the last decades in SCRF methods have crystallized in a series of optimized continuum methods, like SM8, PCM (either IEFPCM or CPCM) or COSMO-RS, which yield accurate solvation free energies provided that the user of these methods pay

attention to the specific details of the parameterization made for each model.

This is an encouraging achievement that should stimulate further studies aimed at the calibration of the most advanced implementations of continuum models for understanding the physical and chemical properties of solutes in solution. In this context, we think that the concept of "universal" solvation models should not be limited exclusively to the ability to estimate reliably the solvation free energy of chemical species in a large number of solvents, but also to the most fundamental problem of providing useful information about the response of the solute to the condensed phase environment and its influence on a variety of physical and chemical properties.

References

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

2 Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Self-Consistent Reaction Field Model for Aqueous and Nonaqueous Solutions Based on Accurate Polarized Partial Charges *J. Chem. Theory Comput.* 2007, 3, 2011–2033.

3 Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A., III; Honig, B. Accurate First Principles Calculation of Molecular Charge Distributions and Solvation Energies from Ab Initio Quantum Mechanics and Continuum Dielectric Theory. J. Am. Chem. Soc. 1994, 116, 11875–11882.

4 Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. New Model for Calculation of Solvation Free Energies: Correction of Self-Consistent Reaction Field Continuum Dielectric Theory for Short-Range Hydrogen- Bonding Effects. J. Phys. Chem. 1996, 100, 11775–11788.

5 Cances, E.; Mennucci, B.; Tomasi, J. A New Integral Equation Formalism for the Polarizable Continuum Model: Theoretical Background and Applications to Isotropic and Anisotropic Dielectrics. J. Chem. Phys. 1997, 107, 3032–3041.

6 Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* 2005, *105*, 2999–3093.

7 Frisch, M.J.; et al. Gaussian 03, revision D.01 Gaussian, Inc. : Pittsburgh, PA, 2003

8 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

9 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. Perkin Trans. 2*, 1993, 799-805.

10 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.

11 Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. Improved Methods for Semiempirical Solvation Models. J. Comput. Chem. 1995, 16, 422–440.

12 Bylaska, E. J.; et al. NWChem, A Computational Chemistry Package for Parallel Computers, Version 4.7; Pacific Northwest National Laboratory: Richland, WA, 2006.

13 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

14 Klamt, A. COSMO-RS from quantum chemistry to fluid phase thermodynamics and drug design *Elsevier, Amsterdam*, 2005.

15 Curutchet, C.; Orozco, M.; Luque, F.J. Solvation in Octanol: Parametrization of the Continuum MST Model *J. Comput. Chem.* 2001, 22, 1180-1193.

16 Soteras, I.; Curutchet, C.; Bidon-Chanal, A.; Orozco, M.; Luque, F.J. Extension of the MST model to the IEF formalism: HF and B3LYP parametrizations *J. Mol. Struct. (THEOCHEM)* 2005, 727, 29-40.

17 F. Eckert, A. Klamt, COSMOtherm, Version C2.1-Revision 01.07; COSMOlogic GmbH&CoKG, Leverkusen, Germany (2007).

18 Turbomole version 5.10 (2008), http://www.turbomole.com

19 Curutchet, C.; Cramer, C. J.; Truhlar, D. G.; Ruiz Lopez, M.; Rinaldi, D.; Orozco, M.; Luque, F.J. Electrostatic Component of Solvation: Comparison of SCRF Continuum Models J. Comput. Chem. 2003, 24, 284-297.

20 Barone, V.; Improta, R.; Rega, N. Quantum mechanical computations and spectroscopy: from small rigid molecules in the gas phase to large flexible molecules in solution, Acc. Chem. Res. 2008, 41, 605-616.

21 Mennucci, B.; Cammi, R. (Eds.), Continuum solvation models in chemical physics, from theory to applications, Wiley (2007)