Comment on the Correct Use of Continuum Solvent Models

3 AUTHORS:

Junming Ho
Institute Of High Performance Computing

27 PUBLICATIONS **545** CITATIONS

Andreas Klamt
Universität Regensburg

113 PUBLICATIONS **10,702** CITATIONS

Michelle L Coote
Australian National University

214 PUBLICATIONS **5,687** CITATIONS

Comment on the Correct Use of Continuum Solvent Models

Junming Ho,[†] Andreas Klamt,[‡,§] and
Michelle L. Coote*,[†]

*ARC Centre of Excellence for Free-Radical Chemistry and
Biotechnology, Research School of Chemistry, Australian
National University, Canberra ACT 0200, Australia,
COSMOlogic GmbH&CoKG, Burscheider Str. 515,
51381 Leverkusen, Germany, and Institute of Physical and
Theoretical Chemistry, University of Regensburg,
93040 Regensburg, Germany*

The development of dielectric continuum solvent models[1,2] (CSMs) has facilitated the study of chemical reactions in the condensed phase in a computationally efficient manner. These methods have been parametrized to deliver accurate values of the free energies of solvation, which can be added to accurate values of the free energies in the gas-phase, to obtain the corresponding solution-phase free energy:

$$G_{soln} = G_{gas} + \Delta G_{solv} + RT\ln\left(\frac{RT}{P}\right) \qquad (1)$$

where the final term converts from the gas-phase standard state (defined by T and P) to the solution-phase standard state of 1 M. However, there are a growing number of recent studies that employ alternative approaches to evaluating G_{soln}. This comment aims to highlight some of their shortcomings as well as clarify some potential points of confusion concerning the usage of these models.

1. CSM Free Energy of Solvation Is Not Calculated As the Difference in Solution- and Gas-Phase Free Energies

The standard textbook definition of solvation free energy is the difference in free energy of the solute in the solution and gas phases

$$\Delta G_{solv} = G_{soln} - G_{gas} \qquad (2)$$

However, in most continuum solvent models such as the PCM and its variants (e.g., CPCM[3,4]) as well as the SM*x* models (e.g., $x = 6$[5]), solvation free energy is in fact defined

$$\Delta G_{solv} = (E_{soln} + G_{nes}) - E_{gas} \qquad (3)$$

where E_{soln} and E_{gas} refer to the electronic energy of the solute in the presence and absence of the continuum solvent field, and G_{nes} denotes the sum of any nonelectrostatic contributions (e.g., cavitation and dispersion−repulsion interactions) to the solvation free energy. As such, eq 3 implicitly assumes that the thermal

* Corresponding author.
[†] Australian National University.
[‡] COSMOlogic GmbH&CoKG.
[§] University of Regensburg.

contributions from nuclear motion of the system (i.e., translation, rotation, and vibration) are identical or very similar in both the gas and the solution phase such that appreciable cancellation is expected. However, for the various reasons highlighted in point 2 of this commentary, including the breakdown of the ideal gas assumption in solution, this assumption is unlikely to be correct. As a result, any noncanceling differences in the thermal contributions in the two phases (e.g., entropic changes associated with structural reorganization of the solvent) are indirectly incorporated into ($E_{soln} + G_{nes}$) through parametrization of eq 3 to reproduce experimental solvation free energies.

In this light, it follows from comparison of eqs 3 and 2 that the expression for the free energy in solution is, rather counterintuitively, given by eq 4.

$$G_{soln} = E_{solv} + G_{nes} + \Delta G_{corr\_gas} \qquad (4)$$

where ΔG_{corr_gas} refers to the thermal correction to the free energy of the solute in the gas phase, not the solution phase. Furthermore, when one needs to calculate the solvation energy by combining the results of separate gas- and solution-phase calculations, one must use its definition within the CSM framework, eq 3, rather than the standard textbook definition, eq 2. These results will hold whenever a CSM is parametrized according to eq 3, irrespective of the software used to calculate $E_{soln} + G_{nes}$.

While the use of eq 3 is computationally more efficient than eq 2 in that it dispenses with relatively expensive Hessian calculations in the gas phase, its use has given rise to confusion in the literature.[6,7] In particular, in popular quantum chemistry packages such as Gaussian03[8] and GAMESSPLUS,[9] the term ($E_{soln} + G_{nes}$) is misleadingly referred to as the total free energy in solution, including nonelectrostatic contributions (some sample output sections are provided in Table S5 the Supporting Information). As a result, this term has mistakenly been used as the solution-phase free energy in the literature.[7] To extract the solution-phase free energy from a continuum model calculation it is instead necessary to perform the gas-phase thermal corrections and use eq 4. However, even in this case, it should be recognized that CSMs have been parametrized at relatively low levels of theory (e.g HF and DFT with small basis sets) to reproduce experimental solvation energies, not total free energies in solution. Whereas a large portion of the SCF errors might be expected to cancel from the ($E_{soln} - E_{gas}$) term in eq 3, this cancellation is not possible in eq 4 and the resulting solution-phase free energies, calculated at such low levels of theory, may not be accurate enough for quantitative studies. For example, based on data provided in a recent assessment study,[10] use of the B3-LYP/6-31G(d) PCM-UAKS solution free energies to calculate the one-electron reduction potentials of nitroxide radicals would introduce an additional average error of over 5 kcal mol^{-1} when compared with using the high-level ab initio method, G3(MP2)-RAD, for the gas-phase free energies and B3-LYP/6-31G(d) merely for solvation energies. This is why the preferred approach in the literature is to use CSMs to calculate only the quantity for which they were parametrized (the solvation energy), which can then be combined with gas-phase free energies that are calculated at much higher levels of theory via eq 1.

The use of eq 3 when parametrizing CSMs also affects the meaning of the E_{soln} term. As the output of an SCF calculation,

Comments

J. Phys. Chem. A, Vol. 114, No. 51, 2010 **13443**

E_{soln} is often assumed to provide a measure of the solution-phase electronic energy, analogous to the use of E_{gas} as a measure of the gas-phase electronic energy. As such, in a growing number of studies, this term is used to calculate the solution-phase free energy "manually" through addition of separately calculated solution-phase thermal corrections.[6] Notwithstanding the accuracy of these corrections (see point 2 of this commentary), nor accuracy of E_{soln} at low levels of theory (see above), it also needs to be recognized that, unlike E_{gas}, the E_{soln} is dependent on parameters such as the atomic radii used to construct the solute cavity, which have been optimized so as to make the solvation energies from eq 3 reproduce experiment. As a consequence, some portion of the difference between the solution-phase and gas-phase thermal corrections is implicitly incorporated into the sum of the E_{soln} and any G_{nes} terms. Thus, unless one can clearly establish and quantify the extent to which the individual E_{soln} and G_{nes} terms partition the electronic energy and the remaining thermal corrections, using E_{soln} as a measure of the solution-phase electronic energy may incur some degree of double counting and introduce an additional source of error to the calculations.

2. Gas-Phase Partition Functions Do Not Yield Solution-Phase Free Energies

To calculate gas-phase partition functions and associated thermodynamic functions (entropy, enthalpy, free energy) from the geometries, frequencies, and energies derived from a standard ab initio calculation, there exist well established equations that are based on the statistical thermodynamics of an ideal gas, under the harmonic oscillator rigid rotor approximation. A number of studies[6] have recently applied the same approach to obtain solution-phase free energies:

$$G_{soln} = E_{soln} + \Delta G_{corr\_soln} \qquad (5)$$

The latter term in this equation is then obtained by applying the same ideal gas partition functions to the frequencies calculated in a dielectric continuum. As noted above, this approach is questionable because the parametrization of eq 3 to reproduce experimental solvation free energies implies that some of the contributions from thermal motions of the solute and the solvent are implicitly incorporated in E_{soln} (and G_{nes}). As such, explicitly accounting for these contributions would incur some degree of double counting. However, in addition, it should be stressed that the solute is not an ideal gas and therefore the ideal gas partition functions, particularly the translational and rotational contributions, are unlikely to be valid in solution.[11] Furthermore, the entropic contribution associated with the

structural reorganization of the solvent is completely neglected through this approach.

In this light, it is interesting to note that many of the studies that employ this incorrect approach often deliver reasonably accurate results.[6] It turns out that, at least for the molecules in these studies, the thermal corrections obtained by applying the ideal gas partition functions to the solution-phase geometries and frequencies are actually very similar to those obtained when applying the same partition functions to the corresponding gas-phase quantities (see Table S1 in the Supporting Information). As a result, the approach adopted in these recent studies gives results that are similar to those that would be obtained using eq 4, simply because the authors' use of gas-phase partition functions to calculate solution-phase thermal corrections serendipitously provides a good approximation to the required gas-phase thermal corrections. However, this approach is not exact and it hardly justifies the correctness of using gas-phase partition functions to compute G_{corr_soln}, nor the use of eq 5 in place of the correct definition of G_{soln} within the CSM framework, eq 4.

3. CSMs Should Be Applied in the Manner Consistent with Their Parametrization

Most popular CSMs such as CPCM[3,4] and SM*x* (e.g., SM6[5]) contain parameters such as atomic radii and nonelectrostatic terms that have been optimized to fit experimental solvation energies at some specified temperature, usually 298 K. As discussed in a recent commentary,[12] the prerequisite for optimal performance of CSMs is that the user should adhere to the computational protocol used in the parametrization of that model and this includes using the appropriate level of theory and set of atomic radii. Thus, to obtain accurate solution-phase free energies, one should only use the high levels of theory for the gas-phase component of eq 1; for the solvation energies one should adhere to the level of theory at which the CSM was parametrized. What is often less well recognized is that CSMs are also only valid at the temperature(s) for which they was parametrized, which for most models is 298 K. When one wishes to stray beyond those temperatures, such as when decomposing the solution-phase free energies into their enthalpic and entropic components, one needs to use more advanced models such SM8T[13] or COSMO-RS[14] that explicitly incorporate the temperature dependence of the solvation energy.

4. Treatment of Multiple Molecular Conformations

In modern computational chemistry software packages, the solvation energy calculations triggered by keywords typically use the same geometry (either gas or solution phase according to the users preference) for both the solution-phase ($E_{soln} + G_{nes}$)
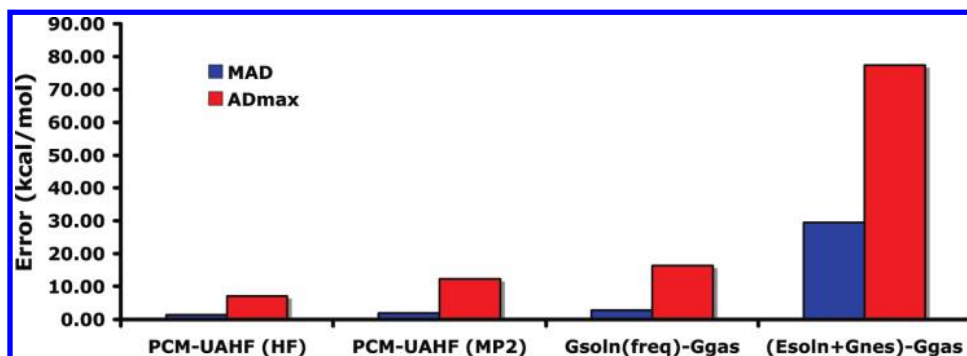
Figure 1. Comparison of mean and maximum absolute deviations (kcal mol^{-1}) of solvation free energies computed on 50 neutral and charged solutes using various definitions of solution free energies.

and gas-phase (E_{gas}) components of eq 3. The use of a single geometry might be reasonable for small rigid molecules with one dominating conformer, but is likely to lead to substantial errors in more complex systems (such as amino acids or other hydrogen bonding molecules) where conformer distributions and major structural features (such as formation of zwitterions) differ substantially between the gas and solution phases. In such cases, it is essential that the respective geometries/conformers are reoptimized in each phase and these separate calculations are combined in eq 3. However, in such cases, the contribution from changes in the thermal motions of the solute to ΔG_{solv} are likely to differ substantially from the test set of rigid molecules for which CSMs were originally parametrized. Thus, such systems may be beyond the limits for which current CSMs were parametrized. As highlighted in several recent contributions,[15,16] a more consistent approach would involve the independent treatment of gas- and solution-phase ensembles, where ΔG_{solv} is evaluated from the ratio of the partition functions of the conformational ensemble in each state. In this way, changes in solute conformational entropy and enthalpy are taken into account. This is a growing area of research and that will become increasingly important as CSMs are applied to study more complex molecular systems.

Error Assessment. To illustrate some of the above points, Figure 1 compares the quality of PCM-UAHF solvation free energies computed using the various definitions of solution free energies. PCM-UAHF(HF) is the solvation free energy calculated based on eq 3 at the level of theory that it was parametrized[17] (HF/6-31+G*). PCM-UAHF(MP2) uses the same equation but employs a higher level of theory (MP2/6-311+G(3df,2p)). The other two approaches use eq 2 but adopt ($E_{soln} + G_{nes}$) or the G_{soln} obtained via eq 5 as solution free energies. As shown, the resulting solvation free energies are all inferior to the PCM-UAHF(HF) model and this reinforces the fact that CSMs perform best only when they are used in a manner that is consistent with their parametrization.

References and Notes

(1) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.

(2) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.

(3) Klamt, A.; Schurmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799.

(4) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669.

(5) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133.

(6) See for example: (a) Brown, T. N.; Mora-Diez, N. *J. Phys. Chem. B* **2006**, *110*, 9270. (b) Brown, T. N.; Mora-Diez, N. *J. Phys. Chem. B* **2006**, *110*, 20546. (c) Nagy, P. I.; Alagona, G.; Ghio, C. *J. Chem. Theor. Comput.* **2007**, *3*, 1249. (d) Govender, K. K.; Cukrowski, I. *J. Phys. Chem. A* **2009**, *113*, 3639. (e) Govender, K. K.; Cukrowski, I. *J. Phys. Chem. A* **2010**, *114*, 1868–1878. (f) Przybylski, J. L.; Wetmore, S. D. *J. Phys. Chem. B* **2009**, *113*, 6533. (g) Przybylski, J. L.; Wetmore, S. D. *J. Phys. Chem. B* **2010**, *114*, 1104. (h) Cukrowski, I.; Govender, K. K. *Inorg. Chem.* **2010**, *49*, 6931.

(7) See for example: (a) Martinez, A. G. T. V. E.; Osio Barcino, J.; de la Moya Cerero, S. *J. Am. Chem. Soc.* **2002**, *124*, 6676. (b) Martinez, A. G.; de la Moya Cerero, S.; Teso Vilar, E.; Fraile, A. G.; Maroto, B. L.; Morillo, C. D. *J. Org. Chem.* **2008**, *73*, 6607.

(8) Frisch, M. J. *Gaussian 03*, revision E.01; Gaussian, Inc.: Wallingford, CT, 2004.

(9) Higashi, M. *GAMESSPLUS*, version 2008−2; University of Minnesota: MN, 2008.

(10) Hodgson, J. L.; Namazian, M.; Bottle, S. E.; Coote, M. L. *J. Phys. Chem. A* **2007**, *111*, 13595.

(11) Ben-Nun, M.; Levine, R. D. *Int. Rev. Phys. Chem.* **1995**, *14*, 215.

(12) Klamt, A.; Mennucci, B.; Tomasi, J.; Barone, V.; Curutchet, C.; Orozco, M.; Luque, F. J. *Acc. Chem. Res.* **2009**, *42*, 489.

(13) Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2008**, *112*, 3024.

(14) Klamt, A. *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*; Elsevier Science Ltd.: Amsterdam, The Netherlands, 2005.

(15) Klamt, A.; Eckert, F.; Diedenhofen, M. *J. Phys. Chem. B* **2009**, *113*, 4508.

(16) Mobley, D. L.; Dill, K. A.; Chodera, J. D. *J. Phys. Chem. B* **2008**, *112*, 938.

(17) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210.

JP107136J