94   10   28

Multiscale Computer Simulations for Prediction of Thermophysical Properties and Phase Equilibrium

Shiang-Tai Lin

Department of Chemical Engineering, National Taiwan University

Abstract

In this project, we propose to utilize the novel concept of computational hierarchy for the prediction of the thermophysical properties of chemicals and complicated materials and their phase equilibria. The method employed is based on the computation of solvation free energy for a chemical species in solution. It is shown that both the vapor pressure and activity coefficient, two important properties for making phase equilibrium predictions, can be determined from the knowledge of solvation free energy.

The focus in this project is the development of a new approach that significantly enhances the efficiency and accuracy in calculation of the long-range electrostatic interaction in implicit solvation calculations using the Polarizable Continuum Model (PCM) and its variants, C-PCM/COSMO and IEF-PCM, where the solvent electrostatics effects are represented by discrete apparent charges distributed on tesserae of the molecular cavity surface embedding the solute. In principle, the accuracy of these methods is improved if the cavity surface is tessellated to finer tesserae; however, the computational time is increased rapidly.

We show that such undesired dependency between accuracy and efficiency is a result of inaccurate treatment the apparent charge self-contribution to the potential and/or electric field. By taking into account of the full effects due to the size and curvature of the segment occupied by each apparent charge, the error in calculated electrostatic solvation free energy is essentially zero for ions (point charge at the center of a sphere) regardless of the degree of tessellation for the solute. For amphiphilic molecules (approximated by a point charge located at a position away from the center of a sphere) the effects from the gradient of apparent charges become important. For such cases, we propose a multiple-sampling technique which lowers the calculated error by at least one order of magnitude compared to the original PCM methods.

Background and Introduction

The knowledge of thermophysical properties and phase behavior for both pure and mixture fluids is of great importance to scientists and engineers in various fields.[1,2] Chemical engineers need such knowledge in the design, development, and optimization of chemical processes.[3] The impact of releasing chemicals into the environment, and an appropriate strategy for removing hazardous chemicals can be established based on this information.[4] Recently, the pharmaceutical industry has begun to use these properties (e.g., the octanol-water partition coefficient) to assist in the design of new drugs.[5-9] As new compounds are being created daily, it is not feasible to measure all the properties of each new compound and their mixtures. An accurate and efficient prediction methodology will satisfy the growing demand for these data.[10,11]

The utilization of modern computational chemistry in solving industrial problems has becoming an imperious issue in the chemical engineering industry. Computational approaches provides a "bottom-up" route, i.e., solving a problem based on electronic and molecular interactions, and can serve as a complementary for most "top-down" experimental measurements, from which fundamental principles are deduced. Moreover, computational chemistry may be the only option when experiments are prohibited by the large expense in cost and time, the limitations of the apparatus, or the hazardous nature of the chemicals.

One promising approach based on molecular solvation has recently been proposed for thermodynamic properties. Solvation theories, as revised by Ben-Naim[12,13] based on statistical mechanics, provide a unique perspective to traditional thermodynamic problems. In contrast to classical thermodynamics where macroscopic properties of a system are concerned, solvation thermodynamics places its focus on the properties of one molecule solvated by others in the system. The transferring of a solute molecule from a fixed position in the ideal gas phase to a fixed position in a solution is defined as the solvation process, with its corresponding free energy charge referred to as the solvation free energy, $\Delta \underline{G}^{*sol}$. Many important thermodynamic quantities can be determined from this free energy, such as the activity coefficient γ and vapor pressure P^{vap}:

$$\ln \gamma_{1/2} = \frac{\Delta \underline{G}_{1/2}^{*sol} - \Delta \underline{G}_{1/1}^{*sol}}{RT} + \ln \frac{c_1}{x_1 c_1^0} \qquad (1)$$

$$\ln P_1^{vap} = \frac{\Delta \underline{G}_{1/1}^{*sol}}{RT} - \ln c_1^0 RT \qquad (2)$$

where c_i is the molar concentration of component i, x_i is the mole fraction of i, superscript 0 indicates a pure liquid, and subscript i/j denotes the property of molecule i in solution j. Based on these equations, many other interesting properties such as the infinite dilution activity coefficient, Henry's law constant, solubility, partition coefficients (e.g. the octanol-water partition coefficient), VLE, and LLE can also be determined.[7,14-18]

The solvation free energy consists of a van der Waals term, which includes cavity and dispersion interactions, and an electrostatic term, which contains the electrostatic interactions and polarization between the solute and solvent

$$\Delta \underline{G}^{*sol} = \Delta \underline{G}^{*vdw} + \Delta \underline{G}^{*el} \qquad (3)$$

In general, the long-range electrostatic interaction between the solute and solvent is the dominant component, especially for polar systems. In classical electrostatic theory,[19] all aspects of electrostatic response of the dielectric solvent due to the presence of a solute can be represented in terms of hypothetical charges q distributed on the boundary between the solute and solvent, that is, the surface of the molecular cavity embedding the solute. In practice, the cavity surface is tessellated into segments, called tesserae, each containing one point surface charge. With these approximations, the electrostatic component of solvation free energy ΔG^{*el} is calculated as

$$\Delta G^{*el} = \frac{1}{2} \underline{V}_S{}^t \underline{q} \qquad (4)$$

where $\underline{V}_S$ and $\underline{q}$ are vectors of rank n whose elements are the solute potential and apparent (screening) charge at each tessera and n being the total number of tesserae on the solute cavity. For a given solute (thus $\underline{V}_S$ is known), ΔG^{*el} is determined once apparent charges $\underline{q}$ at the cavity surface is known.

There are three most commonly used methods for evaluating $\underline{q}$: PCM,[20-23] C-PCM/COSMO,[24] and IEF-PCM.[25,26] In PCM, $\underline{q}$ is solved based on the proportionality of q to the normal component of electric field at the cavity surface

$$(\frac{4\pi\varepsilon}{\varepsilon-1}\underline{I} + \underline{\underline{se}})\underline{q} = -\underline{s}\underline{E}_S \qquad (5)$$

where ε is the dielectric of solvent, $\underline{I}$ is the identity matrix, $\underline{s}$ the tessera area matrix, $\underline{E}_S$ the normal component of electric field at the cavity surface due to the solute, and $\underline{e}$ (when multiplied by $\underline{q}$) produces the apparent charge contribution to the electric field normal to the cavity surface for each tessera. In the case of infinite solvent dielectric

constant, $\underline{q}$ can be more easily solved by the COSMO method

$$\underline{\underline{v}}\underline{q} = -\underline{V}_S \tag{6}$$

The ($n{\times}n$) matrix $\underline{\underline{v}}$ (when multiplied by $\underline{q}$) produces the potential due to the screening charges for each tessera. The IEF-PCM can be regarded as a generalization of COSMO to finite ε problems

$$\underline{\underline{v}}(4\pi\underline{\underline{I}} + \underline{\underline{se}})^{-1}(\frac{4\pi\varepsilon}{\varepsilon - 1}\underline{\underline{I}} + \underline{\underline{se}})\underline{q} = -\underline{V}_S \tag{7}$$

It can be seen that all three methods can be cast into the following general form

$$\underline{\underline{g}}\underline{q} = \underline{P} \tag{8}$$

that is, a set of n linear equations that gives a unique solution of $\underline{q}$ on each tessera for a given $\underline{P}$ and $\underline{\underline{g}}$. Matrix $\underline{P}$ contains the property (potential or electric field) at the molecular cavity due to the solute, and the product $\underline{\underline{g}}\underline{q}$ gives the corresponding property due to the apparent charges $\underline{q}$. To completely determine matrix $\underline{\underline{g}}$, we need the tessera self-contribution (v_{kk}, e_{kk}) to the potential (COSMO, IEF-PCM) or the electric field (PCM, IEF-PCM).

Both the calculation time and precision in calculated solvation energy depends on the number tesserae (n) used to describe the cavity surface. Although the precision can be improved by using more tesserae, the calculation time significantly increases at the same time (roughly proportional to n^3 for matrix inverse[27]). This is a result of discretization of the apparent charges to tessera, where the knowledge of self-contribution to electrostatic potential (v_{kk}) and electric field (e_{kk}) at a tessera from the charge locating on it is required. Due to the approximate treatment of size, curvature and charge gradient effects on such self-contribution, the solvation energy converges slowly with n in PCM, COSMO, and IEF-PCM models. The main achievement in this work is the development of new treatment of the apparent charge self-contribution so that the calculated error in ΔG^{*el} can be minimized, and in some cases eliminated, and the efficiency can be enhanced. We also show that the three methods are in fact identical (i.e., give the same numerical values of ΔG^{*el}) when the exact self-contribution can be determined.

Results and Discussion

I. Tessera self-contribution to potential and electric fields

Here we present the various model developed for the self-contribution to the potential ($v_{kk}q_k$) and electric field ($e_{kk}q_k$) at tessera k from the apparent charge q_k located on the same tessera. According to the Gauss's theorem, these two quantities

must satisfy the following condition as the segment size approaches zero (equivalent to $n \to \infty$)

$$
\left. \begin{array}{l} e_{kk} s_k = -2\pi \\ v_{kk} s_k = 0 \end{array} \right\} \quad \text{as} \quad s_k \to 0 \tag{9}
$$

For finite values of s_k, v_{kk} and e_{kk} should be functions of the area (s_k), the curvature (κ_k), and the spatial gradient of the apparent charge (∇q_k). However, the exact functional form of v_{kk} and e_{kk} are not yet known and different approximations are currently employed in the three models.
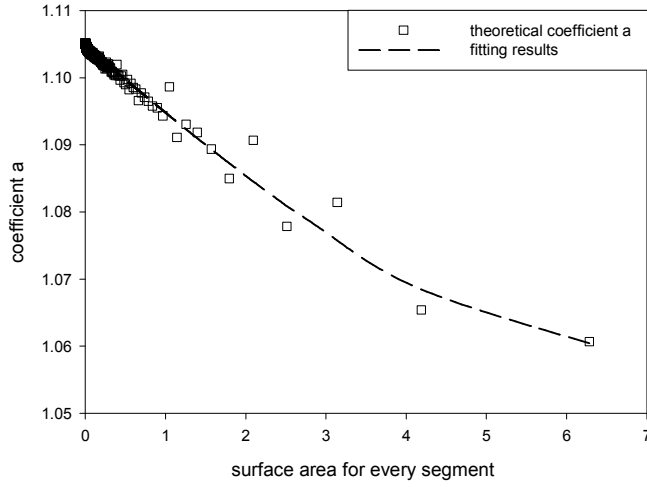

Figure 1. The data of a versus ($\kappa_k^2 s_k / 4\pi$) and the fitting result. The squares are the initial data, and the dash line is calculated from eq. 12.

By analyzing the solvation of an ion and ideal dipole, we found that v_{kk} and e_{kk} can be well approximated using the following general form

$$
v_{kk} = a \sqrt{\frac{4\pi}{s_k}} \tag{10}
$$

$$
e_{kk} = -\frac{2\pi}{s_k}\left(1 - \kappa_k a \sqrt{\frac{s_k}{4\pi}}\right) \tag{11}
$$

The values of the coefficient a are set to 1.0, 1.07, and 1.07 in PCM,[28] COSMO,[24] and IEF-PCM,[25] respectively. It is apparent that these methods ignore the effects from charge gradients, i.e., assuming a uniform apparent charge distribution within each tessera ($\nabla q_k = 0$). In fact, COSMO only considers (part of) tessera size effect in v_{kk}, and PCM and IEF-PCM considers (part of) the size and curvature effects in e_{kk}. By considering the solvation of ions, we find that the coefficient a can be well approximated in terms of a polynomial of $\kappa_k^2 s_k / 4\pi$ (shown in Figure 1)

$$a(\kappa_k, s_k) = \alpha + \beta\left(\frac{\kappa_k^2 s_k}{4\pi}\right) + \gamma\left(\frac{\kappa_k^2 s_k}{4\pi}\right)^2 \qquad (12)$$

with $\alpha = 1.10464$, $\beta = -9.92367 \times 10^{-3}$, and $\gamma = 7.32466 \times 10^{-5}$.

The use of eq. (12) for coefficient a provides a much better segment area and curvature correction; however, it is derived under the assumption that all the tesserae are of the same size on the molecular cavity. The local size inhomogeneity can be taken into account by an inclusion of the effects from neighboring segments as

$$a(s_k, \kappa_k) = a'(s_k', \kappa_k)\sqrt{\frac{s_k}{s_k'}} - \sum_{i=1, i \neq k, i \in \text{neighbor}}^{m} \frac{1}{\sqrt{4\pi s_k}} \frac{s_i}{|\vec{r}_i - \vec{r}_k|} \qquad (13)$$

where the summation is over the m neighboring segments of tessera k, s_k' is the total area of the m neighboring segments and s_k, and a' is calculated from eq. (12). We denote the usage of eqns. (10) to (13) as PCM-acc, COSMO-acc, and IEF-PCM-acc (area, curvature corrected) to emphasize the corresponding corrections.

II. Solvation of ions

The importance of inclusion of tessera curvature and area correction is shown in Fig. 2 where the solvation of an ion is considered. It can be seen that the error in calculated ΔG^{*el} from the original methods converges slowly with n. In contrast, with the inclusion of correction (eq. 10 to 13), the calculated error is essentially zero, regardless of the number of tessellation n used.
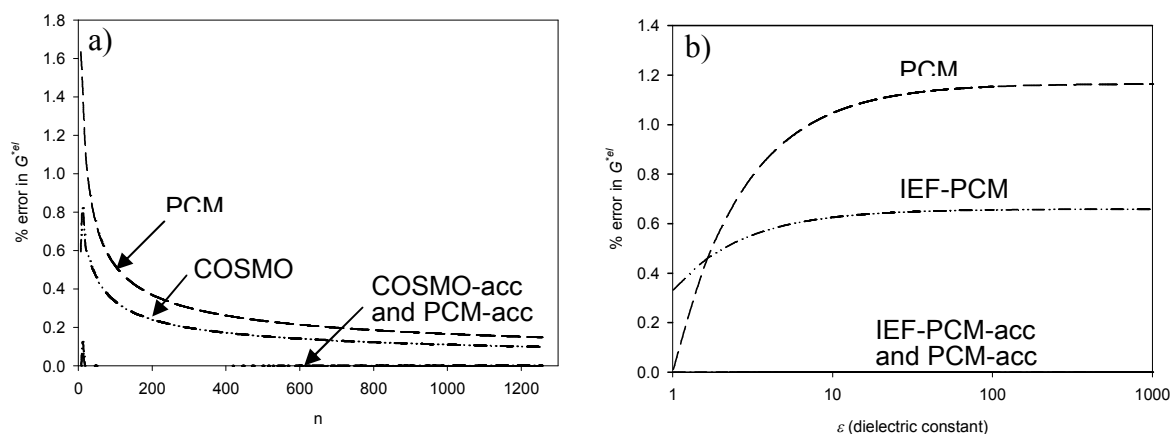

Figure 2. Comparison of error in calculated ΔG^{*el} with and without tessera area and curvature corrections with eqns. 12 and 13. (a) Variation of error with the number of tessellation. (b) Variation of error with the solvent dielectric constant ($n=18$).

III. Solvation of ionic amphiphilic molecules

In the above case of ion solvation the distribution of q on the cavity surface is

uniform. However, this is generally not true for solvation of molecules. Thus, representing the charges within a tessera using one single point charge introduces error. One may possibly find the dependence of tessera self-contribution on the gradient of $\underline{q}$ (spatial non-uniformity); however, he/she would end up with the need of solving a set of much more complicated nonlinear equations ($\underline{g}(\bigtriangledown\underline{q})\underline{q}=\underline{P}$). As an alternative, we propose to evaluate $\underline{q}$ and ΔG^{*el} using several slightly different tessellation while keeping the total number of tesserae (n) constant. We find that the average value of G^{*el} from 4 correlated tessellation samples converges much faster with the increase of n. Figure 3 shows the use of such averaging technique in the calculation of ΔG^{*el} of a unit point charge located at a distance 0.8 R away from the center of a spherical cavity (R is the radius of the cavity) using COSMO and PCM methods. It can be seen that due to the gradient of $\underline{q}$ at the cavity surface, original COSMO and PCM methods may give significant errors (more than 200%) unless a large amount of segments are used in tessellation (e.g. $n>20$). However, with the average of 4 correlated samples, the error rapidly reduces to less than 30% even for $n<20$.
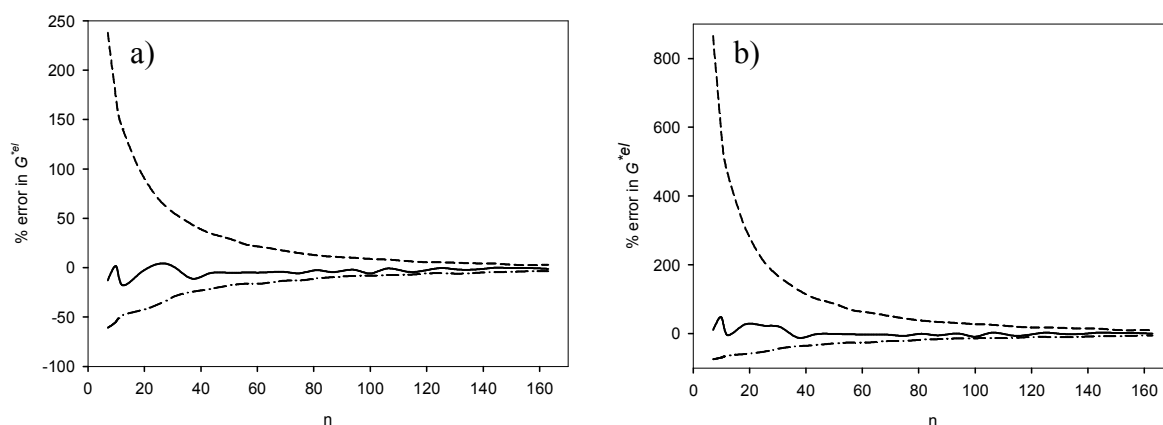

Figure 3. Error in calculated ΔG^{*el} for a unit point charge located 0.8 R away from the center of a spherical cavity from the (a) COSMO and (b) PCM model. The dashed lines are the maximum (dashed) and minimum (dot-dash) error from one arbitrary tessellation. The solid lines are averages over 4 tessellation samples.

Conclusions

In this project, we investigated the source of error in common solvation calculations and developed a new method that leads to highly accurate solvation free energy with much less computational time. This work presents an important step towards a more efficient and accurate calculation of solvation free energy. Our future plan is to generalize the presently developed method to describe the phase behavior of complex problems such as supercritical extraction.

References

1   S. I. Sandler, Int. J. Thermophys. **15** (6), 1013 (1994).

2   S. I. Sandler, J. Chem. Thermodyn. **31** (1), 3 (1999).

3   S. I. Sandler, *Chemical and Engineering Thermodynamics*, Third edition ed. (John Wiley & Sons, Inc., New York, 1999).

4   S. I. Sandler, Fluid Phase Equilib. **116** (1-2), 343 (1996).

5   J. Schulte, J. Durr, S. Ritter, W. H. Hauthal, K. Quitzsch, and G. Maurer, J. Chem. Eng. Data **43** (1), 69 (1998).

6   T. Harner and T. F. Bidleman, Environ. Sci. Technol. **32** (10), 1494 (1998).

7   S. T. Lin and S. I. Sandler, Ind. Eng. Chem. Res. **38** (10), 4081 (1999).

8   R. Seth, D. Mackay, and J. Muncke, Environ. Sci. Technol. **33** (14), 2390 (1999).

9   J. Sangster, *Octanol-Water Partition Coefficients: Fundamentals and Physical Chemistry*. (Wiley and Sons, New York, 1997).

10   S. I. Sandler, A. K. Sum, and S.-T. Lin,   **28**, 313 (2001).

11   S. I. Sandler, S. T. Lin, and A. K. Sum, Fluid Phase Equilib. **194**, 61 (2002).

12   A. Ben-Naim, J. Phys. Chem. **82** (7), 792 (1978).

13   A. Ben-Naim, *Solvation Thermodynamics*. (Plenum Press, New York, 1987).

14   S. T. Lin and S. I. Sandler, Aiche J. **45** (12), 2606 (1999).

15   S. T. Lin and S. I. Sandler, J. Phys. Chem. A **104** (30), 7099 (2000).

16   S. T. Lin and S. I. Sandler, Chem. Eng. Sci. **57** (14), 2727 (2002).

17   S. T. Lin, J. Chang, S. Wang, W. A. Goddard, and S. I. Sandler, J. Phys. Chem. A **108** (36), 7429 (2004).

18   S. T. Lin and S. I. Sandler, Ind. Eng. Chem. Res. **41** (5), 899 (2002).

19   C. J. F. Bötcher, *Theory of electric polarisation*. (Elsevier, Amsterdam, 1952).

20   S. Miertuš, E. Scrocco, and J. Tomasi, Chem. Phys. **55** (1), 117 (1981).

21   S. Miertus and J. Tomasi,   **65** (2), 239 (1982).

22   J. L. Pascualahuir, E. Silla, J. Tomasi, and R. Bonaccorsi,   **8** (6), 778 (1987).

23   J. Tomasi and M. Persico, Chem. Rev. **94** (7), 2027 (1994).

24   A. Klamt and G. Schuurmann, J. Chem. Soc.-Perkin Trans. 2 (5), 799 (1993).

25   E. Cances, B. Mennucci, and J. Tomasi, J. Chem. Phys. **107** (8), 3032 (1997).

26   J. Tomasi, B. Mennucci, and E. Cances, J. Mol. Struct. **464** (1-3), 211 (1999).

27   W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes - The Art of Scientific Computing*. (Cambridge, New York, 1986).

28   B. Z. Wang and G. P. Ford, J. Chem. Phys. **97** (6), 4162 (1992).