

## Efficient and accurate solvation energy calculation from polarizable continuum models

Shiang-Tai Lin<sup>a)</sup> and Chieh-Ming Hsieh

*Department of Chemical Engineering, National Taiwan University, Taipei 10617, Taiwan*

(Received 27 December 2005; accepted 22 August 2006; published online 25 September 2006)

A new approach is proposed to enhance the efficiency and accuracy for calculation of the long-range electrostatic interaction from implicit solvation models, i.e., the polarizable continuum model (PCM) and its variants, conductorlike PCM/conductorlike screening model and integral equation formalism PCM. In these methods the solvent electrostatics effects are represented by a set of 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 the inaccurate treatment of the apparent charge self-contribution to the potential and/or electric field. By taking into account 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 used. For molecules where gradient of apparent charge density is nonzero at the cavity surface, we propose a multiple-sampling technique which significantly lowers the calculated error compared to the original PCM methods, especially when very few numbers of tesserae are used. © 2006 American Institute of Physics.

[DOI: 10.1063/1.2354489]

### I. INTRODUCTION

The solvation (free) energy of a solute in solvent is an important quantity that can be used for determining various thermophysical properties of the solution,<sup>1–8</sup> phase,<sup>9–15</sup> and chemical<sup>16,17</sup> equilibrium of complex mixtures. Except for nonpolar systems, the long-range electrostatic interaction between the solute and solvent is, in general, the dominant component. It is found that the electrostatic response of solvent due to the presence of solute can be well estimated by treating the solvent surrounding the solute as a continuum with a macroscopic dielectric constant. As a consequence, such implicit solvation has been the focus of many theoretical developments on solvation calculations, also known as reaction field theory.<sup>18–21</sup>

In classical electrostatic theory, the solute (distributed charges)-solvent (dielectric continuum) interaction is described by the Poisson equation.<sup>21</sup> Tomasi and Persico<sup>22</sup> and more recently Tomasi *et al.*<sup>23</sup> provided an excellent review and summary on this various approaches of solving this partial differential equation. Among them, the boundary element approach, or the apparent surface charge (ASC) method,<sup>22</sup> gains great popularity in many applications, perhaps, due to its efficiency and simplicity when incorporated with molecular simulations.<sup>24–26</sup> The foundation of the ASC method lies on the fact that all aspects of electrostatic response of the dielectric solvent due to the presence of a solute can be represented in terms of hypothetical charges distributed on the boundary between the solute and solvent, that is, the surface

of the molecular cavity embedding the solute. In the case of infinite solvent dielectric constant, they are called screening charges. In practice, the cavity surface is tessellated into segments, called tesserae, each containing one point surface charge. With these approximations, the problem reduces to a set of linear equations whose solution gives the value of apparent charge on each tessera.

Both the calculation time and precision in calculated solvation energy depend on the number of tesserae ( $n$ ) used to describe the cavity surface. The precision can be improved by using more tesserae but at the cost of a significant increase in computational time (roughly proportional to  $n^3$  for matrix inverse<sup>27</sup>). This undesired behavior in the ASC method is a result of discretization of the apparent charges to tessera, where the knowledge of self-contribution to electrostatic potential and electric field at a tessera from the charge located on it is required. Due to the approximate treatment of size, curvature, and charge gradient effects on such self-contributions, the solvation energy converges slowly with  $n$  in commonly used ASC methods, such as polarizable continuum model (PCM),<sup>24</sup> conductorlike screening model (COSMO),<sup>26</sup> and integral equation formalism PCM (IEF-PCM).<sup>28</sup> Miertus *et al.* have addressed this issue and provided a first order correction to the self-contribution.<sup>24</sup> Here we investigate the importance of the tessera self-contribution in more detail and show that the calculated error can be minimized, and in some cases eliminated, and the efficiency can be enhanced with a careful treatment of the apparent charge self-contribution. We also show that the three methods are, in fact, identical for cases where the exact value of self-contribution can be determined.

<sup>a)</sup> Author to whom correspondence should be addressed. Electronic mail: stlin@ntu.edu.tw

## II. ELECTROSTATIC SOLVATION ENERGY FROM APPARENT SURFACE CHARGE (ASC) METHODS

In the ASC method, the potential  $\phi_D(\mathbf{r})$  due to the dielectric solvent is completely represented by hypothetical charges distributed on the surface of molecular cavity,

$$\phi_D(\mathbf{r}) = \int \frac{\sigma(\mathbf{r}_s)}{|\mathbf{r} - \mathbf{r}_s|} dr_s^2, \quad (1)$$

where  $\sigma(\mathbf{r}_s)$  is the apparent charge density located at the cavity surface  $\mathbf{r}_s$  and the integration is over the closed cavity surface. The electrostatic component of the solvation energy  $\Delta G^{*el}$  is easily evaluated from the following surface integration:

$$\Delta G^{*el} = \frac{1}{2} \int \sigma(\mathbf{r}_s) \phi_S(\mathbf{r}_s) dr_s^2, \quad (2)$$

where  $\phi_S(\mathbf{r}) = \int [\rho_S(\mathbf{r}') / |\mathbf{r} - \mathbf{r}'|] dr'^3$  is the electrostatic potential due to the solute. In practice the solute cavity surface is represented by segments, called tesserae, and Eq. (2) is more conveniently written in a matrix form

$$\Delta G^{*el} = \frac{1}{2} V_S^t q, \quad (3)$$

where  $V_S$  and  $q$  are column vectors (superscript  $t$  denotes matrix transpose) whose elements are the solute potential and apparent charge at each tessera  $k$ , i.e.,

$$V_{S,k} = \phi_S(\mathbf{r}_k) \quad \text{and} \quad q_k = \sigma(\mathbf{r}_k) s_k, \quad k = 1, 2, \dots, n, \quad (4)$$

with  $s_k$  being the surface area of tessera  $k$  and  $n$  being the total number of tessera on the solute cavity. In the following, we briefly review three most commonly used methods for evaluating  $q$ : PCM, conductorlike PCM (C-PCM)/COSMO, and IEF-PCM. All three methods can be cast into a general form  $gq = P$ , i.e., a set of  $n$  linear equations that gives a unique solution of  $q$  on each tessera for a given  $P$  and  $g$ . Matrix  $P$  contains the property (potential or electric field) at the molecular cavity due to the solute, and the product  $gq$  gives the corresponding property due to the apparent charges  $q$ . In the following we use lower case matrices for properties of the molecular cavity tesserae and upper case matrices for properties due to the solute.

### A. The polarizable continuum model (PCM)

The PCM method by Tomasi and co-workers<sup>22,24,29,30</sup> is one of the most robust ways of calculating the apparent charges at any molecular cavity. This method is based on the fact that  $\sigma$  is proportional to the normal component of the electric field at the cavity surface,<sup>22</sup>

$$\sigma_k = \frac{q_k}{s_k} = \frac{\epsilon - 1}{4\pi\epsilon} \bar{\nabla} \phi_k^{\text{in}} \cdot \bar{n}_k = -\frac{\epsilon - 1}{4\pi\epsilon} E_k, \quad (5)$$

where  $\mathbf{n}_k$  is the unit outward normal vector of tessera  $k$  and  $E_k = -\bar{\nabla} \phi_k^{\text{in}} \cdot \mathbf{n}_k$  is the total electric field in the direction of  $\mathbf{n}_k$  at the inner cavity surface which is the sum of contributions

from the solute  $E_{S,k}$  and the apparent charges  $E_{D,k}$  ( $E_k = E_{S,k} + E_{D,k}$ ). Writing Eq. (5) in a matrix form,

$$s^{-1} q = -\frac{\epsilon - 1}{4\pi\epsilon} I(E_S + eq) \quad \text{or} \quad \left( \frac{4\pi\epsilon}{\epsilon - 1} I + se \right) q = -sE_S, \quad (6)$$

where  $I$  is the identity matrix and the tessera matrix  $s$  is of dimension  $n \times n$  with its elements defined as

$$s_{kj} = \begin{cases} 0, & j \neq k \\ s_k, & j = k. \end{cases} \quad (7)$$

The  $n \times n$  matrix  $e$  (when multiplied by  $q$ ) produces the apparent charge contribution to the electric field normal to the cavity surface for each tessera,

$$e_{kj} = \begin{cases} \frac{(\mathbf{r}_k - \mathbf{r}_j) \cdot \mathbf{n}_k}{|\mathbf{r}_k - \mathbf{r}_j|^3}, & j \neq k \\ e_{kk}, & j = k, \end{cases} \quad (8)$$

where  $\mathbf{r}_k$  and  $e_{kk}q_k$  are the position vector and the self-contribution to the electric field at tessera  $k$ , respectively. The calculation of  $e_{kk}$  will be discussed in detail in the next section. For a given solute (thus  $E_S$  and  $V_S$  are known) and solvent dielectric constant, Eq. (6) can be used to calculate  $q$  and thus  $\Delta G^{*el}$  from Eq. (3).

### B. The conductorlike screening model (COSMO)

In PCM, the apparent charges are determined with the fact that  $q$  at any point on the molecular cavity surface is proportional to the net electric field at the same location [Eq. (7)] with the proportionality constant defined by the dielectric constant of the solvent. In the limit of infinite dielectric, Klamt and Schuurmann<sup>26</sup> propose to use the fact that the net potential at the cavity surface should be zero in the calculation of  $q$ . This leads to a simpler formulation, called COSMO,

$$V_S + vq = 0. \quad (9)$$

The  $n \times n$  matrix  $v$  (when multiplied by  $q$ ) produces the potential due to the screening charges for each tessera,

$$v_{kj} = \begin{cases} \frac{1}{|\mathbf{r}_k - \mathbf{r}_j|}, & j \neq k \\ v_{kk}, & j = k, \end{cases} \quad (10)$$

where  $\mathbf{r}_k$  and  $v_{kk}q_k$  are the position vector and the self-contribution to the electrostatic potential at tessera  $k$ , respectively. In COSMO, the screening charges  $q$  can be obtained from solving Eq. (9) once the solute is specified ( $V_S$  is known).

### C. The integral equation formalism polarizable continuum model (IEF-PCM)

Recently, Tomasi and co-workers<sup>28,31</sup> proposed an integral equation formalism of the PCM which allows the consideration of dielectric anisotropy of the solvent. For the special case of having an isotropic dielectric solvent, IEF-PCM can be easily derived by, first, considering the solute dissolved in a solvent of infinite dielectric constant,

$$V_S + vq^* = 0 \quad \text{or} \quad q^* = -v^{-1}V_S$$

and

$$(4\pi I + se)q^* = -sE_S \quad \text{or} \quad q^* = -(4\pi I + se)^{-1}sE_S.$$

Thus the potential and electric field (normal to the tessera) from the solute on the solute cavity surface are related,

$$(4\pi I + se)^{-1}sE_S = v^{-1}V_S \quad \text{or} \quad E_S = s^{-1}(4\pi I + se)v^{-1}V_S. \quad (11)$$

Replacing  $E_S$  in Eq. (6) with that in Eq. (11), we obtain the IEF-PCM equation for  $q$  for a solvent of finite dielectric constant  $\epsilon$ ,<sup>23,32,33</sup>

$$\left(\frac{4\pi\epsilon}{\epsilon-1}I + se\right)q = -(4\pi I + se)v^{-1}V_S \quad \text{or} \\ v(4\pi I + se)^{-1}\left(\frac{4\pi\epsilon}{\epsilon-1}I + se\right)q = -V_S. \quad (12)$$

Thus instead of the use of  $E_S$  in PCM, IEF-PCM needs the input of  $V_S$  and reduces to the COSMO form when  $\epsilon$  approaches infinity.

To completely determine the matrix  $g$ , we still need the tessera self-contribution ( $v_{kk}, e_{kk}$ ) to the potential (COSMO, IEF-PCM) or the electric field (PCM, IEF-PCM).

### III. TESSERA SELF-CONTRIBUTION TO POTENTIAL AND ELECTRIC FIELDS

Here we wish to determine the 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. In principle,  $v_{kk}$  and  $e_{kk}$  should be functions of the area ( $s_k$ ), the curvature ( $\kappa_k$ ), and the spatial gradient of the apparent charge ( $\nabla q_k$ ). If the exact functional dependence of  $v_{kk}$  and  $e_{kk}$  on these parameters was known and used, exactly the same apparent charges  $q$  for any given solute would be obtained from PCM, IEF-PCM, and (as the solvent dielectric constant approaches infinity) COSMO models. (Note that this is strictly true only for classical solutes. For quantum mechanical solutes, there is a finite amount of solute electron leaking outside the solute cavity whose effects are not considered here.) Unfortunately this is not yet known for general cases and different approximations are currently employed in the three models. Nevertheless,  $v_{kk}$  and  $e_{kk}$  must satisfy the boundary condition as the segment size approaches zero,

$$\left. \begin{aligned} e_{kk}s_k &= -2\pi \\ v_{kk}s_k &= 0 \\ q_k/s_k &= \sigma_k \end{aligned} \right\} \quad \text{as} \quad s_k \rightarrow 0. \quad (13)$$

#### A. Tessera area and curvature corrections

Klamt and Schuurmann<sup>26</sup> evaluated  $v_{kk}$  by considering the solvation of a point charge located at the center of a spherical cavity,

$$v_{kk} = a_k \sqrt{\frac{4\pi}{s_k}}, \quad (14)$$

with the coefficient  $a_k$  being set to a constant of 1.07. Wang and Ford<sup>34</sup> showed that  $e_{kk}$  can be expressed as the following based on Gauss's theorem:

$$e_{kk} = -\frac{2\pi}{s_k} \left(1 - \kappa_k a_k \sqrt{\frac{s_k}{4\pi}}\right), \quad (15)$$

with the coefficient  $a_k$  being set to a constant of 1.0. (Note that the curvature  $\kappa$  is positive for convex tesserae and negative for concave ones.) Interestingly, both constants are set to 1.07 in the IEF-PCM method. Both Eqs. (14) and (15) satisfy the limiting condition in Eq. (13). However, the use of a constant value of  $a$  in Eqs. (14) and (15) leads to different values of  $G^{*el}$  from the three methods, all of which deviate from the theoretical value for the simplest case of a point charge located at the center of a spherical cavity (solvation of ions) except for the limit when  $n \rightarrow \infty$ . Such discrepancy can be resolved when the exact values for  $a$  are used.

In the Appendix, we show that coefficient  $a_k$  in Eqs. (14) and (15) should be the same when considering the solvation of a point charge located at the center of a spherical cavity,

$$a_k = \left( \sqrt{\frac{4\pi}{\kappa_k^2 s_k}} - \sqrt{\frac{1}{4\pi s_k} \sum_{i=1, i \neq k}^n \frac{s_i}{|\mathbf{r}_i - \mathbf{r}_k|}} \right), \quad (16)$$

and the limiting value of  $a$  as  $s$  approaches zero is unity,

$$\lim_{s_k \rightarrow 0} a_k = 1. \quad (17)$$

If the tesserae are evenly distributed on the unit sphere, Eq. (16) can be well approximated with a simpler equation (Appendix)

$$a_k^{ED} = 1 + \frac{\alpha x + \beta x^2}{1 + \delta x + \gamma x^2 + \eta x^3}, \quad (18)$$

with  $x = \ln(1 + \kappa_k^2 s_k / e)$  and the coefficients  $\alpha=2430$ ,  $\beta=500$ ,  $\delta=23\,700$ ,  $\gamma=3000$ , and  $\eta=16\,000$ . Note that Eq. (18) leads to the following limiting values:

$$a_k^{ED}(\kappa_k^2 s_k = 0) = a_k^{ED}(\kappa_k^2 s_k = \infty) = 1.$$

Although Eq. (16) is valid only for spherical cavities, Eq. (18) provides a "mean-field-like" value for coefficient  $a$  and thus is applicable to cavities of arbitrary shapes. The effect from size differences of nearby segments can be approximated with the following equation:

$$a_k = a_k^{ED} \sqrt{\frac{s'_k}{s_k}} - \sqrt{\frac{1}{4\pi s_k} \sum_{i=1}^{m(nr)} \frac{s_i}{|\mathbf{r}_i - \mathbf{r}_k|}}, \quad (19)$$

where the summation is over the  $m$  neighboring segments nearby segment  $k$ ,  $s'_k = \sum_{i=1}^m s_i$ , and  $a'_k(m, s'_k, \kappa_k) = [a_k^{ED}(s'_k, \kappa_k) - 1] / e^{(m-1)b} + 1$ . It is noteworthy that Eq. (19) reduces to Eq. (16) in the limit  $m \rightarrow \infty$  and to Eq. (18) for  $m=0$ . The value of  $m$  depends on how many nearest neighboring segments are included. [In the computer program, we create a list of the first nearest neighbors ( $nr=1$ ) for each segment. The second, third, and so forth nearest neighbors can be determined from this list. The value of  $nr$  is set to zero (thus  $m=0$ ) for

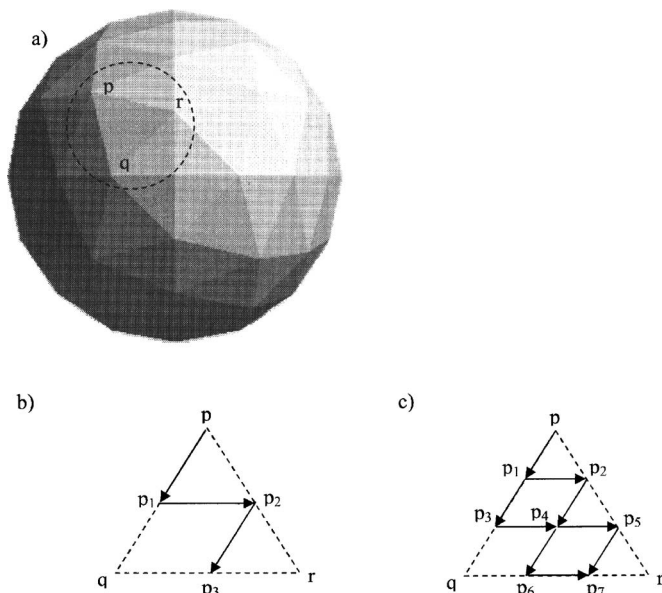


FIG. 1. Illustration of the generation of correlated samples for charge gradient corrections. (a) Three arbitrarily chosen segment centers  $p$ ,  $q$ , and  $r$  on a sphere. (b) The generation of three additional samples by moving point  $p$  to  $p_1$ ,  $p_2$ , and  $p_3$ . (c) The extension of the method to generate more samples (seven additional samples in this case).

spherical shape cavities and to 1 for others.] To denote the usage of Eq. (19) for  $a_k$  in Eqs. (14) and (15), we will refer to the methods as PCM-acc, COSMO-acc, and IEF-PCM-acc (acc stands for area and curvature corrected).

## B. Charge gradient correction

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 functional dependence of tessera self-contribution on the gradient of  $q$  (spatial nonuniformity); however, he/she would end up with the need of solving a set of much more complicated nonlinear equations [ $g(\nabla q)q=P$ ].

As an alternative, we propose an indirect approach in which  $q$  and  $G^{*el}$  are evaluated using several slightly different tessellations while keeping the total number of tesserae ( $n$ ) constant. In this technique, we first find (arbitrarily) the center of three segments that are in contact with one another on the surface of a sphere [for example, points  $p$ ,  $q$ , and  $r$  in Fig. 1(a)]. We then rotate point  $p$  to the midpoints of arcs  $pq$  [ $p_1$  in Fig. 1(b)],  $pr$  [ $p_2$  in Fig. 1(b)], and  $qr$  [ $p_3$  in Fig. 1(b)] while keeping the relative positions of all the tesserae unchanged (solid body rotation). In this way, we generate four correlated tessellation samples for the sphere. This method can be extended to generate more samples by using the 1/3 and 2/3 points of the lateral arcs of the triangle  $pqr$  [Fig. 1(c)]. The average value of  $\Delta G^{*el}$  from four correlated tessellation samples [Fig. 1(b)] provides a better (unbiased) measure for  $G^{*el}$  for a certain value of  $n$  when there is gradient of apparent charges on the cavity surface. We refer to the use of correlated sampling technique as cs4.

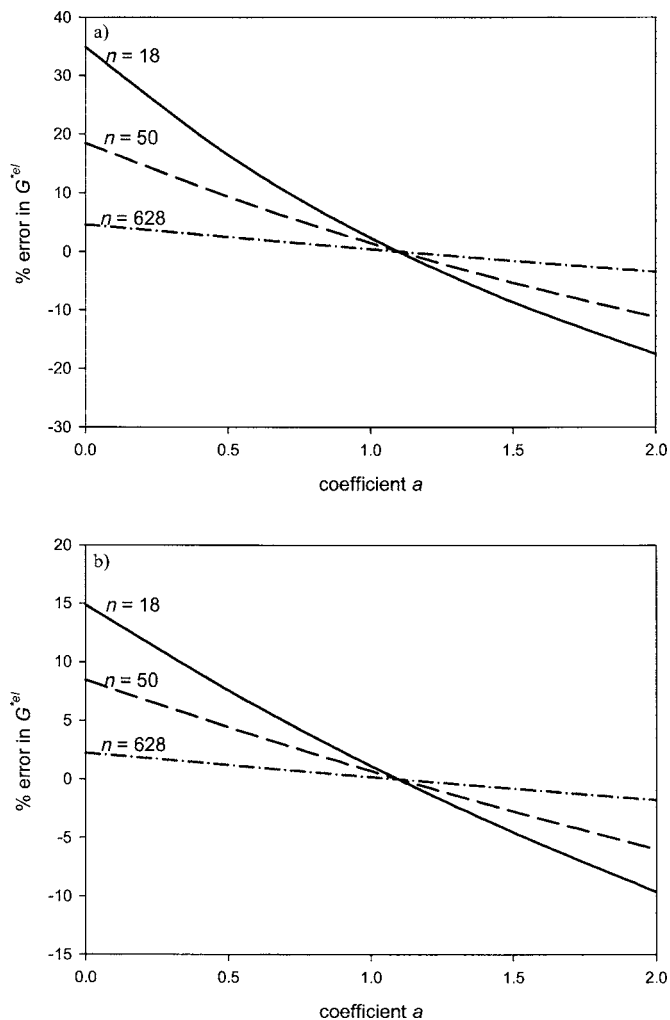


FIG. 2. Effect of the variation of the tessera self-contribution coefficient  $a$  on  $\Delta G^{*el}$  from COSMO (a) and PCM (b) for a point charge at the center of a sphere in a solvent of infinite dielectric constant.

## IV. RESULTS AND DISCUSSION

To demonstrate the importance of the tessera area, curvature, and charge gradient corrections, we consider three types of systems: (1) a point charge located at the center of a spherical cavity, (2) a point charge located at a point away from the center of a spherical cavity, and (3) realistic molecules having point charges on the nuclei. There are two reasons for the choice of the first two systems. First, analytical solutions to the Poisson equation are available.<sup>21</sup> Second, the combination of these two represents the major ingredients in most modern implicit solvation calculation for real molecules, where a molecular cavity is represented as overlapping van der Waals spheres centered at the atoms of the molecule.<sup>35,36</sup> It is noteworthy that although a point source charge is used in this study, there is no fundamental difficulty to implement the proposed method to any quantum mechanical packages.

### A. A point charge at the center of a spherical cavity

Figure 2 illustrates the sensitivity of the error in calculated  $\Delta G^{*el}$  ( $[(\Delta G_{calc}^{*el} - \Delta G_{theory}^{*el}) / \Delta G_{theory}^{*el}] \times 100\%$ ) on the variation of coefficient  $a$  for  $\epsilon = \infty$  (note that in this

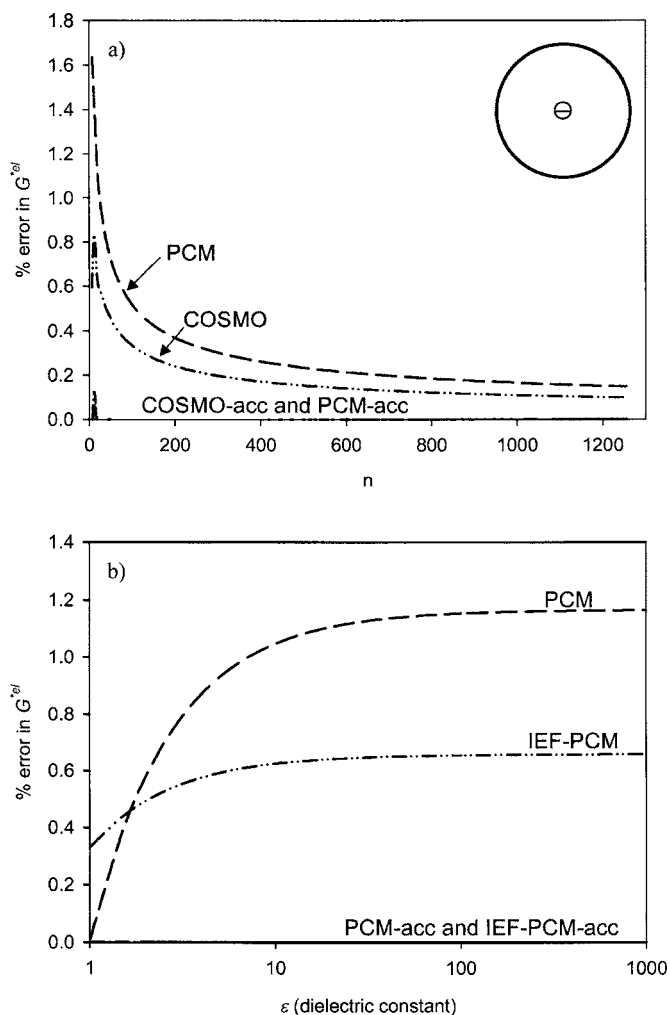


FIG. 3. Comparison of error in calculated  $\Delta G^{*el}$  with (acc) and without tessera area and curvature corrections for a point charge at the center of a spherical cavity. (a) Variation of error with the number of tessellation ( $\epsilon = \infty$ ). (b) Variation of error with the solvent dielectric constant ( $n = 18$ ).

limit IEF-PCM is equivalent to COSMO). It is seen that the error varies from 35% to -20% for COSMO as the value of  $a$  changes from 0 to 2. The error is smaller (15% to -10%) in the case of PCM. The error also varies with the number of tesserae used to represent the cavity. The error is reduced as the spherical cavity is partitioned into more tesserae. It is interesting to note that the error approaches minimum for all cases around  $a = 1.10$ , similar to that found by others.<sup>37,38</sup>

The effectiveness of inclusion of tessera curvature and area corrections [Eq. (19) for  $a_k$  in Eqs. (14) and (15)] is shown in Fig. 3(a). In these calculations, the neighboring parameter  $m$  is set to zero [thus Eq. (19) reduces to Eq. (18)]. It can be seen that the error in calculated  $\Delta G^{*el}$  from the original methods (PCM and COSMO) converges slowly with  $n$ . In contrast, with the inclusion of correction (PCM-acc and COSMO-acc), the calculated error is essentially zero, regardless of the number of tessellations  $n$  used.

Figure 3(b) compares the accuracy of PCM and IEF-PCM at various values of solvent dielectric constant ( $\epsilon$ ) when the number of tesserae ( $n$ ) is set to 18. It is found that the error in both methods increases with the value of  $\epsilon$ . However, applying area and curvature corrections to the segment

self-contributions, the error reduces to essentially zero for all values of  $\epsilon$ . In addition, it can be seen [Fig. 3(b)] that the error for PCM and IEF-PCM is largest in the limit of infinite dielectric constant. Thus we put the focus on this limit in subsequent analysis.

There are two important observations here. First, when the exact value of coefficient  $a$  is used, all three ASC models are equally accurate. Second, when the same but approximate values are used, the three models produces values of  $\Delta G^{*el}$  that are different.

It is also important to note that although the percentage errors shown in Fig. 3 are quite small (in the range of 0.2%–2%), they could have a significant impact on the phase equilibrium predictions. The value of solvation free energy of ions in aqueous solution is on the order of 100 kcal/mol (e.g.,  $\Delta G^{*sol}$  of  $\text{Na}^+$  in water is -90 kcal/mol at room temperature). An error of 0.2 kcal/mol in  $\Delta G^{*sol}$  would result in an error of about 40% in the calculated activity coefficient.<sup>4</sup>

### B. A point charge located away from the center of a spherical cavity

When the point charge is moved away from the center of the spherical cavity, there is a nonuniform distribution of apparent charges on the cavity surface. For such cases, the computed value of  $\Delta G^{*el}$  depends strongly on the relative position of the centers of tessera and the source charge. Slightly rotating the positions of the surface segments would lead to a very different value of  $\Delta G^{*el}$ . Figure 4 shows the maximum and minimum errors (solid curves) in the calculated value of  $\Delta G^{*el}$  for several different degrees of tessellation ( $n$ ) for a unit point charge located at a distance of  $0.8R$  away from the center of a spherical cavity ( $R$  is the radius). In all cases the  $n$  tesserae are uniformly distributed on the cavity surface. The only difference is their relative positions to the point charge. Since we do not know the exact dependence of  $v_{kk}$  and  $e_{kk}$  on  $\nabla q$ , i.e., we do not have the exact value of coefficient  $a$ , we do not expect to get the same results from the various ASC methods.

The absolute errors from the original COSMO [Fig. 4(a)] and PCM [Fig. 4(b)] may give significant errors unless a large amount of segments is used in tessellation (e.g., more than 100% for COSMO and 300% for PCM when  $n < 20$  and  $\epsilon = \infty$ ). However, the maximum possible errors quickly reduce to less than 50% (COSMO) and 100% (PCM) when the  $\Delta G^{*el}$  is determined from the average of four correlated samples (described above) for  $n = 20$ . Thus the use of correlated sampling is very important for obtaining reliable solvation energies when there is a gradient of apparent charges on the cavity surface.

There are two additional observations to make from Fig. 4. First, the uncertainty (maximum-error–minimum-error) in PCM is greater than that from COSMO. Although we do not know the exact functional dependence of  $e_{kk}$  and  $v_{kk}$  on  $\nabla q$ , these results indicate that  $e_{kk}$  is much more sensitive to  $\nabla q$  than  $v_{kk}$  is. Thus the ignorance of  $\nabla q$  in  $e_{kk}$  leads to a greater error. The second observation is that the maximum and minimum errors are not symmetric around zero. Both PCM and COSMO tend to overestimate the value of  $\Delta G^{*el}$  (too nega-

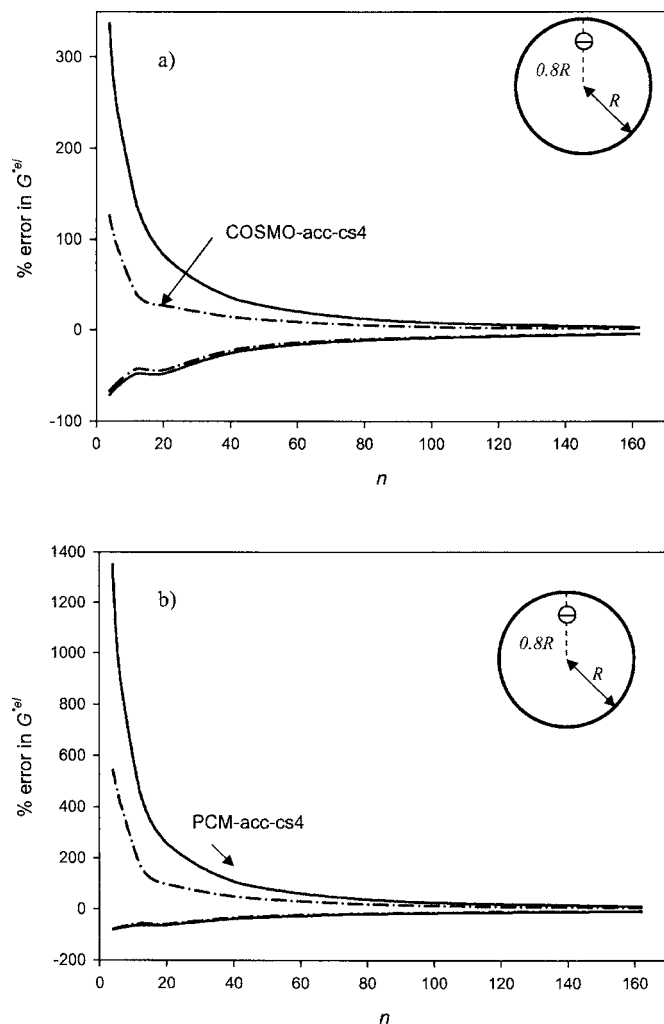


FIG. 4. Error in the calculated  $\Delta G^{*el}$  for a unit point charge located  $0.8R$  away from the center of a spherical cavity from the COSMO (a) and PCM (b) models. The dashed lines are the maximum and minimum errors from one arbitrary tessellation. The solid curves are those from averages over four tessellation samples (acc-cs4).

tive). The largest (most negative) value of  $\Delta G^{*el}$  is found when one of the tessera is located right on the line of the center of the sphere and the point charge. In such a case, the apparent charge on that particular tessera is especially large, resulting in the large value of  $\Delta G^{*el}$ . Such asymmetric behavior of error in  $\Delta G^{*el}$  is less prominent when the charge correction is applied using four correlated tessellation samples. This is another indication that such correction provides a nonbiased (thus more reliable) value of  $\Delta G^{*el}$ .

It is useful to address on the efficiency of the proposed method. Although the charge gradient correction requires the calculation of four tessellation samples, the total amount of time used is still less than that from the original method because of a significant reduction in the need of the amount of tesserae ( $n$ ) to achieve a certain level of accuracy. To illustrate, the relative time used is  $t_c/t_o = (4n_c^3)/n_o^3$ , where  $n_c$  and  $n_o$  are the number of surface segments ( $n$ ) needed from the original and the method proposed here (cs4) (the power of 3 is a result of the fact that the computational time in ACS calculation is dominated by the matrix inverse). To achieve a maximum possible error of 50% from PCM [Fig. 4(b)],

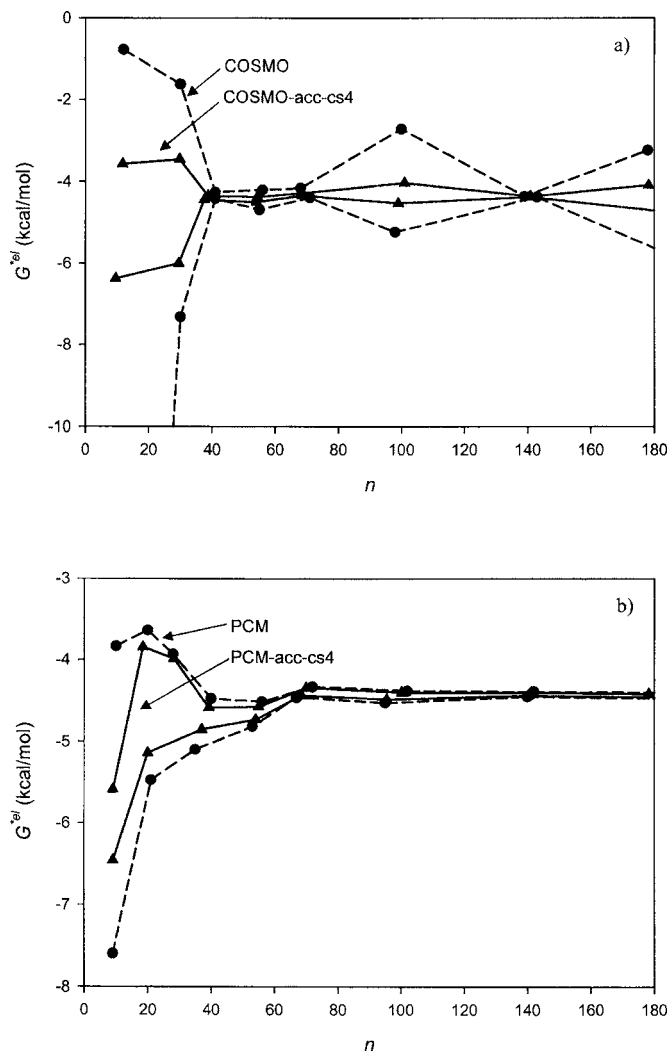


FIG. 5. Comparison of the calculated  $\Delta G^{*el}$  for water in a solvent of infinite dielectric constant with and without tessera self-contribution corrections using (a) COSMO and (b) PCM models. Closed circles and solid lines show the maximum and minimum values of  $\Delta G^{*el}$  from original ASC models using 100–5000 tessellation samples for a certain value of  $n$ . The triangle and solid curves show the maximum and minimum errors from the average of four correlated tessellation samples (acc-cs4).

about 70 ( $n_o$ ) points are need using the original method and the required segments reduces to 40 ( $n_c$ ) when correlated samplings are applied. Thus, there is a reduction of about 25% of the computational time ( $t_c/t_o=0.75$ ) using the new method. In addition, the calculation of the four correlated tessellation samples can be done simultaneously on different CPUs and thus is quite suitable for parallel computation.

### C. Solvation of ionic and neutral molecules

Figures 5–8 compare the calculated  $\Delta G^{*el}$  for two neutral [hydrogen cyanide (HCN) and water ( $H_2O$ )] and two ionic [acetic acid ( $CH_3COO^-$ ) and octadecanoic acid ( $C_{17}H_{35}COO^-$ )] molecular species with and without the area, curvature, and charge gradient corrections. The solvent dielectric constant is set to infinity. The solute molecules are represented as having point charges located at the nuclei (atomic charges). The atomic charges are calculated using the charge equilibration method<sup>39</sup> implemented in CERIU2.<sup>40</sup>

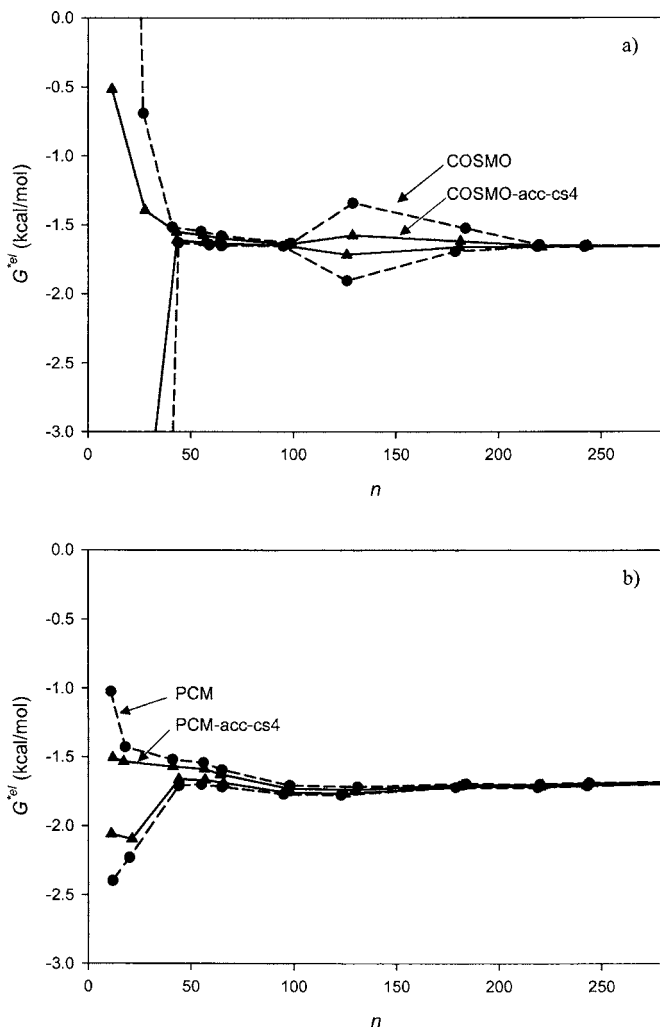


FIG. 6. Comparison of calculated  $\Delta G^{*el}$  for HCN in a solvent of infinite dielectric constant with and without tessera self-contribution corrections using (a) COSMO and (b) PCM models. The legends are the same as in Fig. 5.

The geometry of the molecules is optimized in vacuum using the Dreiding force field<sup>41</sup> and the Connolly surface<sup>35</sup> is used as the solvation cavity.<sup>7</sup> For these calculations the number of neighboring segments  $m$  is set to include the first nearest neighbors ( $nr=1$ ). The maximum and minimum values of calculated  $\Delta G^{*el}$  in Figs. 5–8 are determined from 100 to 5000 tessellation samples (applying the method described in Sec. III B for each van der Waals sphere on the atoms) for a certain value of  $n$ .

Because the Connolly surface of molecules is roughly the exposed surface of spheres of atomic van der Waals radii centered on each atom, the general features we found here are quite similar to the those of the previous two model systems using a single sphere. The original methods (COSMO or PCM) are quite sensitive to the tessellation of the cavity surface especially for small values of  $n$ . This effect is more prominent for charged (Figs. 8 and 9) and large (Fig. 9) species. This is an indication that the accuracy in the calculated  $\Delta G^{*el}$  strongly depends on the quality of tessellation of the surface of molecular species. The proposed correction method (COSMO-acc-cs4 and PCM-acc-cs4) takes

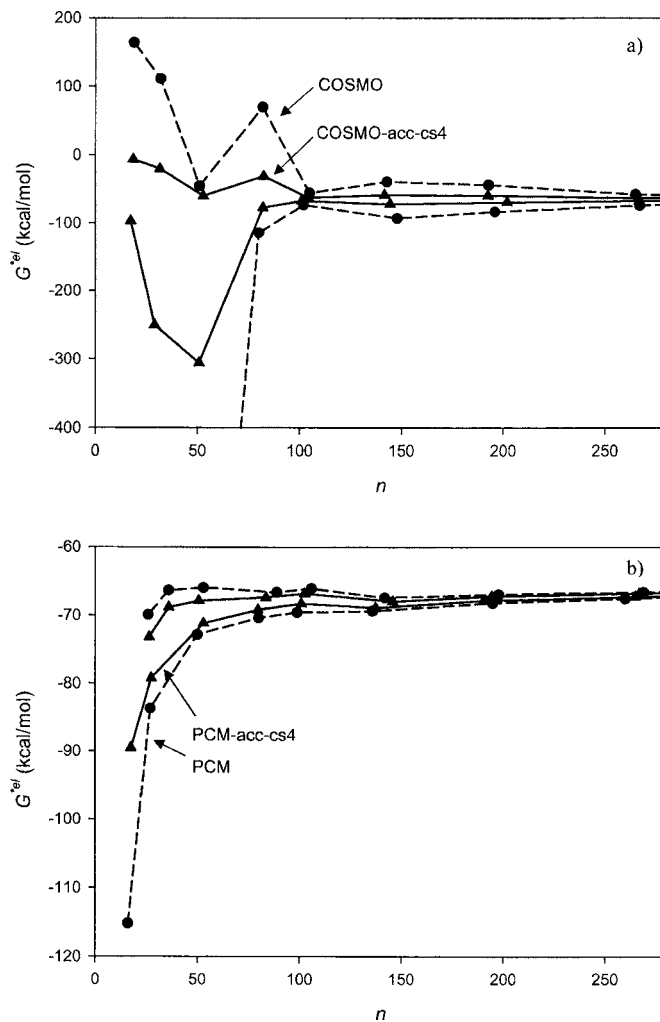


FIG. 7. Comparison of calculated  $\Delta G^{*el}$  for  $\text{CH}_3\text{COO}^-$  in a solvent of infinite dielectric constant with and without tessera self-contribution corrections using (a) COSMO and (b) PCM models. The legends are the same as in Fig. 5.

into account the effects of tesserae area, curvature, and charge gradient and thus produces more accurate values for  $\Delta G^{*el}$  at much lower values of  $n$ .

## V. CONCLUSIONS

In this work, we investigated the source of error in common solvation calculations and developed a new method that leads to highly accurate and reliable solvation free energy with much less computational time. We show that the tessera self-contribution to the potential and electric field plays a critical role in the calculation of the electrostatic contribution to the solvation free energy ( $\Delta G^{*el}$ ). Once this is correctly handled, all the ASC methods produce identical values of  $\Delta G^{*el}$ . The segment and area corrections are important for solvation of ions, whose apparent charge distribution is uniform on the cavity surface. The charge gradient correction becomes more important for neutral and ionic molecules, where the apparent charge distributions are nonuniform. The use of the average value from four correlated tessellation samples may greatly enhance the accuracy of the calculated  $\Delta G^{*el}$ . The proposed method allows for the use of a lower

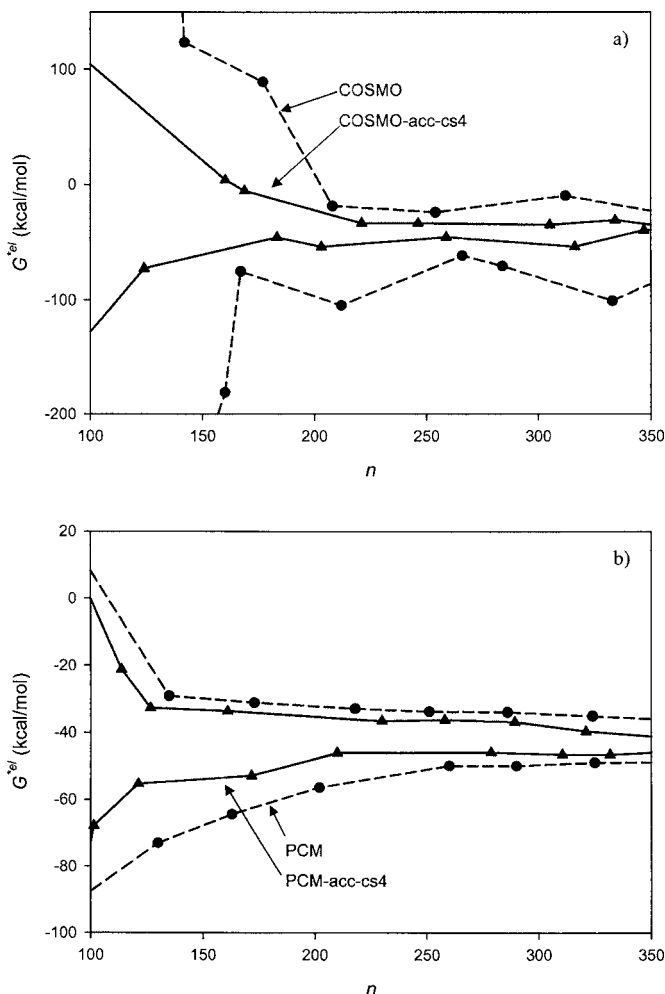


FIG. 8. Comparison of calculated  $\Delta G^{el}$  for  $C_{17}H_{35}COO^-$  in a solvent of infinite dielectric constant with and without tessera self-contribution corrections using (a) COSMO and (b) PCM models. The legends are the same as in Fig. 5.

degree of tessellation for the molecular surface. This would be critical for solvation calculations for large molecules (protein, DNA, polyelectrolyte, etc.) where a large number of tesserae are generally needed.

## ACKNOWLEDGMENTS

We thank the financial support from Grant No. NSC 93-2218-E-002-136 by the National Science Council of Taiwan and Grant No. 94-S-A09 by the Tzong Jwo Jang Educational Foundation and the National Center for High-Performance Computing in Taiwan for the computational facility.

## APPENDIX: TESSERA SELF-CONTRIBUTION TO POTENTIAL AND ELECTRIC FIELDS

Consider a point charge  $Q$  located at the center of a spherical cavity of radius  $R$  surrounded by a continuum of dielectric constant  $\epsilon$ . The distribution of potential and electric field normal to the cavity surface due to the point charge ( $\phi_Q, \mathbf{E}_Q$ ) and apparent charge ( $\phi_q, \mathbf{E}_q$ ) can be obtained exactly by solving the Poisson equation,<sup>21</sup>

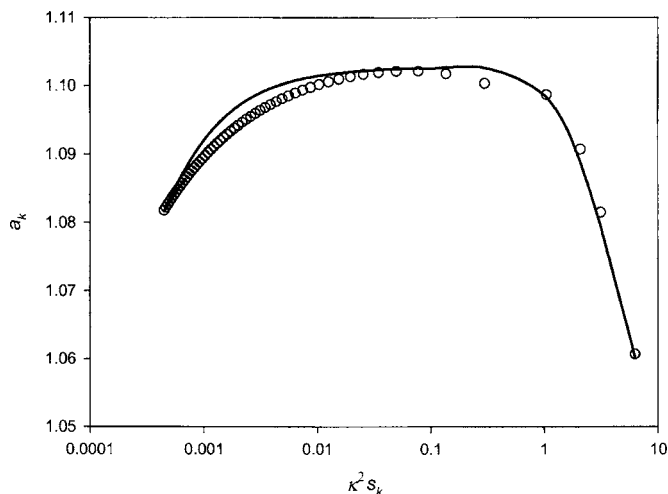


FIG. 9. The variation of coefficient  $a$  as a function of the size and curvature of evenly distributed tesserae on a sphere. The open circles are the exact values of  $a$  calculated from Eq. (A4) [or (16)] and the solid line is the best fit via Eq. (A11) [or (18)].

$$\phi_Q = \frac{Q}{r},$$

$$\mathbf{E}_Q = -\frac{\partial \phi_Q}{\partial r} \hat{\delta}_r = \frac{Q}{r^2} \hat{\delta}_r,$$

$$\phi_q = \begin{cases} -\frac{\epsilon-1}{\epsilon} \frac{Q}{R} & \text{for } r \leq R \\ -\frac{\epsilon-1}{\epsilon} \frac{Q}{r} & \text{for } r \geq R, \end{cases}$$

$$\mathbf{E}_q = \begin{cases} 0 & \text{for } r < R \\ -\frac{\epsilon-1}{\epsilon} \frac{Q}{r^2} & \text{for } r > R. \end{cases}$$
(A1)

Furthermore, the apparent charge density is

$$\sigma = -\frac{1}{4\pi} \frac{\epsilon-1}{\epsilon} (E_Q + E_q)|_{r=R} = -\frac{1}{4\pi} \frac{\epsilon-1}{\epsilon} \frac{Q}{R^2}.$$
(A2)

Assuming that the cavity surface is composed of  $n$  tesserae, each of area  $s_k$  ( $\sum_{k=1}^n s_k = 4\pi R^2$ ) and apparent charge  $q_k = s_k \sigma$ , the potential at any tessera  $k$  can be written as

$$\frac{Q}{R} + \sum_{i=1}^n v_{ik} q_i = \phi_Q(\mathbf{r}_k) + \phi_q(\mathbf{r}_k) = \frac{Q}{R} - \frac{\epsilon-1}{\epsilon} \frac{Q}{R}.$$

Thus

$$\sum_{i=1}^n v_{ik} s_i = 4\pi R = 4\pi/\kappa$$

$$\Rightarrow v_{kk} = \frac{4\pi}{\kappa s_k} - \sum_{i=1, i \neq k}^n \frac{s_i}{s_k} \frac{1}{|\mathbf{r}_i - \mathbf{r}_k|} = a_k \sqrt{\frac{4\pi}{s_k}},$$
(A3)

with



$$a_k = \left( \sqrt{\frac{4\pi}{\kappa_k^2 s_k}} - \sqrt{\frac{1}{4\pi s_k} \sum_{i=1, i \neq k}^n \frac{s_i}{|\mathbf{r}_i - \mathbf{r}_k|}} \right). \quad (\text{A4})$$

Note that Eq. (A3) is valid for all values of  $\epsilon$ , although the original derivation given by Klamt and Schuurmann<sup>26</sup> only considered the case of  $\epsilon = \infty$ .

Similarly, the electric field normal to the tessera at the inner surface of any tessera  $k$  is

$$\frac{Q}{R^2} + \sum_{i=1}^n e_{ik} q_i = \frac{Q}{R^2} + 0.$$

Therefore,

$$e_{kk} = - \sum_{i=1, i \neq k}^n e_{ik} = - \sum_{i=1, i \neq k}^n \frac{(\mathbf{r}_i - \mathbf{r}_k) \cdot \mathbf{r}_k}{|\mathbf{r}_i - \mathbf{r}_k|^3 |\mathbf{r}_k|}$$

$$= - \sum_{i=1, i \neq k}^n \frac{1}{|\mathbf{r}_i - \mathbf{r}_k|} \frac{1}{2R} = - \frac{2\pi}{s_k} \left( 1 - \frac{a_k}{R} \sqrt{\frac{s_k}{4\pi}} \right), \quad (\text{A5})$$

$$e_{kk} = - \frac{2\pi}{s_k} + \frac{v_{kk}}{2R},$$

where the coefficient  $a$  is the same as in Eq. (A4). To obtain Eq. (A5) we have also used the identities  $(\mathbf{r}_i - \mathbf{r}_k) \cdot \mathbf{r}_k = |\mathbf{r}_i - \mathbf{r}_k| |\mathbf{r}_k| \cos \alpha$  with the cosine of the included angle  $\alpha$ ,  $\cos \alpha = |\mathbf{r}_i - \mathbf{r}_k| / 2R$ . Now consider the electric field normal to the tessera at the outer surface of any tessera  $k$ ,

$$\frac{Q}{R^2} + \sum_{i=1}^n e_{ik} q_i = \frac{Q}{R^2} - \frac{\epsilon - 1}{\epsilon} \frac{Q}{R^2}$$

and

$$e_{kk} = 4\pi - \sum_{i=1, i \neq k}^n e_{ik} = \frac{2\pi}{s_k} \left( 1 + \frac{a}{R} \sqrt{\frac{s_k}{4\pi}} \right). \quad (\text{A6})$$

Combining Eqs. (A5) and (A6) we have the final result for  $e_{kk}$ ,

$$e_{kk} = - \frac{2\pi}{s_k} \left( 1 - \kappa_k a \sqrt{\frac{s_k}{4\pi}} \right), \quad (\text{A7})$$

with curvature  $\kappa (=1/R)$  being positive for convex tesserae and negative for concave ones. Thus we prove that both PCM, COSMO, and IEF-PCM should share the same value of  $a$  [Eq. (A4)] for uniformly distributed apparent charges on a spherical cavity.

The limiting value of  $a$  as  $n$  approaches infinity can be determined as follows. Consider one of the  $n$  segments located at the  $z$  axis. The surface area is related to half of the solid angle  $\theta_k$  as

$$s_k = 2\pi R^2 (1 - \cos \theta_k) \Rightarrow \cos \theta_k = 1 - \frac{s_k}{2\pi R^2}. \quad (\text{A8})$$

In the limit of infinitely small segment sizes, the summation in Eq. (A4) can be replaced with integration,

$$\lim_{s_i \rightarrow 0} \sum_{i=1, i \neq k}^n \frac{s_i}{|\mathbf{r}_i - \mathbf{r}_k|} = \int_s \frac{ds}{|\mathbf{r}_i - \mathbf{r}_k|} = \int_{\theta_k}^{\pi} \frac{2\pi R^2 \sin \theta d\theta}{\sqrt{2R} \sqrt{(1 - \cos \theta)}}$$

$$= 4\pi R - 2\sqrt{\pi s_k}. \quad (\text{A9})$$

Therefore,

$$\lim_{s \rightarrow 0} a_k = \lim_{s \rightarrow 0} \sqrt{\frac{1}{4\pi s_k}} \left( 4\pi R - \sum_{i=1, i \neq k}^n \frac{s_i}{|\mathbf{r}_i - \mathbf{r}_k|} \right)$$

$$= \sqrt{\frac{1}{4\pi s_k}} (4\pi R - 4\pi R + \sqrt{4\pi s_k}) = 1. \quad (\text{A10})$$

Figure 9 (open circles) shows the exact values of coefficient  $a$  calculated from Eq. (A4) for a sphere tessellated to equal-sized segments. (Note that distributing points evenly on a sphere is not trivial and the variation of  $a$  with  $n$  depends on how the tessellation is done. In this work, the center point of each tessera is taken as the vertex of a corresponding polyhedron for  $n=4, 6, 8, 12$ , and  $20$ . For  $n > 20$ , the vertices are obtained from triangulation of the faces of an icosahedron.) It can be seen that the coefficient  $a$  has a maximum value of 1.102 at  $\kappa_k s_k = 0.0777$  ( $n=162$  for a unit sphere) and tends to the theoretical value of 1 as  $\kappa_k s_k$  decreases. To describe the dependence of  $a$  as a function of  $\kappa s$  for evenly distributed points, we use the following equation (solid line in Fig. 9):

$$a_k^{\text{ED}} = 1 + \frac{\alpha x + \beta x^2}{1 + \delta x + \gamma x^2 + \eta x^3}, \quad (\text{A11})$$

with  $x = \ln(1 + \kappa_k^2 s_k / e)$  and the coefficients  $\alpha=2430$ ,  $\beta=500$ ,  $\delta=23\,700$ ,  $\gamma=3000$ , and  $\eta=16\,000$ . Equation (A11) guarantees that  $a$  takes the value of unity for any infinitely small ( $s=0$ ) and/or flat ( $\kappa=0$ ) tessera.

<sup>1</sup>R. W. Taft, J. L. M. Abboud, M. J. Kamlet, and M. H. Abraham, *J. Solution Chem.* **14**, 153 (1985).

<sup>2</sup>M. J. Kamlet, R. M. Doherty, M. H. Abraham, Y. Marcus, and R. W. Taft, *J. Phys. Chem.* **92**, 5244 (1988).

<sup>3</sup>J. Schulte, J. Durr, S. Ritter, W. H. Hauthal, K. Quitzsch, and G. Maurer, *J. Chem. Eng. Data* **43**, 69 (1998).

<sup>4</sup>S. T. Lin and S. I. Sandler, *AIChE J.* **45**, 2606 (1999).

<sup>5</sup>A. Klamt, F. Eckert, and M. Hornig, *J. Comput.-Aided Mol. Des.* **15**, 355 (2001).

<sup>6</sup>A. Klamt, F. Eckert, M. Hornig, M. E. Beck, and T. Burger, *J. Comput. Chem.* **23**, 275 (2002).

<sup>7</sup>S. T. Lin, J. Chang, S. Wang, W. A. Goddard, and S. I. Sandler, *J. Phys. Chem. A* **108**, 7429 (2004).

<sup>8</sup>P. Kollman, *Chem. Rev. (Washington, D.C.)* **93**, 2395 (1993).

<sup>9</sup>A. Ben-Naim, *Solvation Thermodynamics*, 1st ed. (Plenum, New York, 1987).

<sup>10</sup>C. J. Cramer and D. G. Truhlar, *Chem. Rev. (Washington, D.C.)* **99**, 2161 (1999).

<sup>11</sup>A. Klamt, *J. Phys. Chem.* **99**, 2224 (1995).

<sup>12</sup>A. Klamt and F. Eckert, *Fluid Phase Equilib.* **172**, 43 (2000).

<sup>13</sup>A. Klamt, V. Jonas, T. Burger, and J. C. W. Lohrenz, *J. Phys. Chem. A* **102**, 5074 (1998).

<sup>14</sup>A. Klamt, G. J. P. Krooshof, and R. Taylor, *AIChE J.* **48**, 2332 (2002).

<sup>15</sup>S. T. Lin and S. I. Sandler, *Ind. Eng. Chem. Res.* **41**, 899 (2002).

<sup>16</sup>J. R. Pliego, *Chem. Phys. Lett.* **367**, 145 (2003).

<sup>17</sup>C. J. Cramer and D. G. Truhlar, *J. Am. Chem. Soc.* **113**, 8552 (1991).

<sup>18</sup>V. M. Born, *Z. Phys.* **1**, 45 (1920).

<sup>19</sup>J. G. Kirkwood, *J. Chem. Phys.* **2**, 351 (1934).

<sup>20</sup>L. Onsager, *J. Am. Chem. Soc.* **58**, 1486 (1936).

<sup>21</sup>C. J. F. Bötcher, *Theory of Electric Polarisation*, 1st ed. (Elsevier, Amsterdam, 1952).

- <sup>22</sup>J. Tomasi and M. Persico, Chem. Rev. (Washington, D.C.) **94**, 2027 (1994).
- <sup>23</sup>J. Tomasi, B. Mennucci, and R. Cammi, Chem. Rev. (Washington, D.C.) **105**, 2999 (2005).
- <sup>24</sup>S. Miertuš, E. Scrocco, and J. Tomasi, Chem. Phys. **55**, 117 (1981).
- <sup>25</sup>D. J. Tannor, B. Marten, R. Murphy, R. A. Friesner, D. Sitkoff, A. Nicholls, M. Ringnalda, W. A. Goddard, and B. Honig, J. Am. Chem. Soc. **116**, 11875 (1994).
- <sup>26</sup>A. Klamt and G. Schuurmann, J. Chem. Soc., Perkin Trans. 2 **1993**, 799.
- <sup>27</sup>W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, New York, 1986).
- <sup>28</sup>E. Cancès, B. Mennucci, and J. Tomasi, J. Chem. Phys. **107**, 3032 (1997).
- <sup>29</sup>S. Miertus and J. Tomasi, Chem. Phys. **65**, 239 (1982).
- <sup>30</sup>J. L. Pascualahir, E. Silla, J. Tomasi, and R. Bonaccorsi, J. Comput. Chem. **8**, 778 (1987).
- <sup>31</sup>J. Tomasi, B. Mennucci, and E. Cancès, J. Mol. Struct. **464**, 211 (1999).
- <sup>32</sup>M. Cossi and V. Barone, J. Chem. Phys. **109**, 6246 (1998).
- <sup>33</sup>H. Li and J. H. Jensen, J. Comput. Chem. **25**, 1449 (2004).
- <sup>34</sup>B. Z. Wang and G. P. Ford, J. Chem. Phys. **97**, 4162 (1992).
- <sup>35</sup>M. L. Connolly, J. Appl. Crystallogr. **16**, 548 (1983).
- <sup>36</sup>J. L. Pascualahir, E. Silla, and I. Tunon, J. Comput. Chem. **15**, 1127 (1994).
- <sup>37</sup>D. M. Chipman, J. Chem. Phys. **110**, 8012 (1999).
- <sup>38</sup>F. W. Chen and D. M. Chipman, J. Chem. Phys. **119**, 10289 (2003).
- <sup>39</sup>A. K. Rappe and W. A. Goddard, J. Phys. Chem. **95**, 3358 (1991).
- <sup>40</sup>CERIUS2, Molecular Simulations Inc., San Diego, CA, 1999.
- <sup>41</sup>S. L. Mayo, B. D. Olafson, and W. A. Goddard, J. Phys. Chem. **94**, 8897 (1990).