Faculty Publications -- Chemistry Department   Published Research - Department of Chemistry

1-1-2009

Continuous and Smooth Potential Energy Surface for Conductorlike Screening Solvation Model Using Fixed Points with Variable Areas

Peifeng Su
University of Nebraska - Lincoln, psu2@unl.edu

Hui Li
University of Nebraska - Lincoln, hli4@unl.edu

Continuous and smooth potential energy surface for conductorlike screening solvation model using fixed points with variable areas

Peifeng Su and Hui Li[a)]
Department of Chemistry, University of Nebraska-Lincoln, Lincoln, Nebraska 68588, USA

Rigorously continuous and smooth potential energy surfaces, as well as exact analytic gradients, are obtained for a conductorlike screening solvation model (CPCM, a variant of the general COSMO) with Hartree–Fock (RHF, ROHF, UHF, and MCSCF) and density functional theory (R-DFT, RO-DFT, and U-DFT) methods using a new tessellation scheme, fixed points with variable areas (FIXPVA). In FIXPVA, spheres centered at atoms are used to define the molecular cavity and surface. The surface of each sphere is divided into 60, 240, or 960 tesserae, which have positions fixed relative to the sphere center and areas scaled by switching functions of their distances to neighboring spheres. Analytic derivatives of the positions and areas of the surface tesserae with respect to atomic coordinates can be obtained and used to evaluate the solvation energy gradients. Due to the accurate analytic gradients and smooth potential energy surface, geometry optimization processes using these methods are stable and convergent. © *2009 American Institute of Physics.* [DOI: 10.1063/1.3077917]

I. INTRODUCTION

Continuum solvation models are important computational models for theoretical studies of condensed phase chemistry.[1] In a continuum model the solvent is treated as a dielectric medium or a conductor, while the solute is represented by a distribution of charges. The polarizable continuum models [the earlier DPCM (Ref. 2) and the more recent IEF-PCM (Ref. 3)], the conductorlike screening models [COSMO,[4] GCOSMO,[5] and CPCM (Ref. 6)] and the SS-(V)PE (Ref. 7) models are representative continuum solvation models that use the surface elements method. The analytic gradients have been derived and implemented for DPCM,[8,9] IEF-PCM,[10,11] COSMO,[4] GCOSMO,[12] and CPCM,[6,13] and used for molecular geometry optimization. However, if an improper scheme is used for molecular surface tessellation, such analytic gradients are either inaccurate or numerically instable, or both, and geometry optimizations are often difficult to accomplish.

York and Karplus[14] showed that using a switching function to treat the appearance and disappearance of tesserae lead to a rigorously smooth potential energy surface for the COSMO solvation model; Senn *et al.*[15] implemented a similar approach for Car–Parrinello *ab initio* molecular dynamics. Their methods use atom-centered spheres (no additional spheres) to define the solvent accessible surface for solutes and use spherical Gaussian functions to describe the intertessera electrostatic interactions.

Based on the GEPOL (Ref. 16) scheme, which uses additional spheres to fill the crevice between neighboring spheres and define the solvent excluded surface (SES) for solutes, Pomelli[17] developed a similar switching function method for IEF-PCM and obtained smooth potential energy

surfaces. However, two issues may arise in the GEPOL scheme: (1) too many small spheres may be added and the number of surface tesserae may be significantly increased; (2) the additional spheres have small radii and tessera areas, which may lead to unsmooth potential energy surfaces.

Li and Jensen[18] developed a modified GEPOL scheme called GEPOL-AS (AS stands for area scaling), which, by scaling the areas of the tesserae that are too close to each other, can significantly increase the stability of IEF-PCM and CPCM geometry optimization procedures. A feature of the GEPOL-AS scheme is that for two separated spheres close to each other, the tesserae between the spheres are effectively scaled by a distance and angle switching function, as if they were excluded by the additional spheres used in the GEPOL scheme. However, the rigorously analytic gradients for the GEPOL-AS based IEF-PCM and CPCM methods are very difficult to obtain. In addition, the GEPOL-AS based PCM potential energy surfaces are not smooth.

Based on prior work described above, a new tessellation scheme, fixed points with variable area (FIXPVA), is developed for conductorlike screen solvation models (COSMO), and is implemented for CPCM, which is a specific variant of COSMO. FIXPVA can produce continuous and smooth potential energy surfaces, as well as exact analytic gradients for geometry optimization. In FIXPVA, atom-centered spheres (no additional spheres) are used, and the SESs for CPCM calculations are mimicked with switching functions.

This paper is organized as follows. In Sec. II, the general theory for COSMO/HF and COSMO/density functional theory (DFT) calculations is outlined, with a focus on the formulas for energy gradient calculation; then the FIXPVA scheme is described in detail, including the derivatives of the tessera coordinates and areas with respect to atomic coordinates. In Sec. III, the general methods used for COSMO calculations are briefly described. In Sec. IV, some numerical

---
a)Author to whom correspondence should be addressed. Electronic mail: hli4@unl.edu.

results are presented and discussed, with an emphasis on demonstrating the accuracy of the gradients and the smoothness of the potential energy surfaces. Finally, a conclusion is presented in Sec. V.

II. THEORY

A. COSMO

In COSMO, the induced surface charges on the tesserae, written as a vector $\mathbf{q}$, are obtained by solving a matrix equation,[4]

$$\mathbf{q} = -\lambda \mathbf{C}^{-1}\mathbf{V}, \tag{1}$$

where λ is a scaling factor that has a general expression $(\varepsilon - 1)/(\varepsilon + x)$, with ε being the dielectric constant of the solvent and x being a value between 0 and 2. There are evidences that $x = 1/2$ may produce better results over a large range of ε.[4,6,13] In this work only $x = 0$ and $\lambda = (\varepsilon - 1)/\varepsilon$ is discussed. The selection of the value for x does not affect the main results reported in this paper. In Eq. (1) vector $\mathbf{V}$ consists of the electrostatic potentials at the tesserae due to the solute nuclei and electrons; $\mathbf{C}$ is a geometric matrix, which has the following elements:

$$C_{ii} = 1.07\sqrt{\frac{4\pi}{a_i}}, \tag{2}$$

$$C_{ij} = \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \tag{3}$$

where $\mathbf{r}_i$, and a_i are, respectively, the center and area of tessera i. The $\mathbf{C}$ matrix is symmetric. The factor 1.07 in Eq. (2) was first obtained by Klamt and Schüürmann.[4]

In practice, the induced surface charges can be obtained by solving Eq. (1) separately for solute nuclear potential $\mathbf{V}_N$ and electronic potential $\mathbf{V}_e$, and the resultant charges are denoted as $\mathbf{q}_N$ and $\mathbf{q}_e$, respectively:

$$\mathbf{q}_N = -\frac{\varepsilon - 1}{\varepsilon}\mathbf{C}^{-1}\mathbf{V}_N, \tag{4}$$

$$\mathbf{q}_e = -\frac{\varepsilon - 1}{\varepsilon}\mathbf{C}^{-1}\mathbf{V}_e, \tag{5}$$

with

$$\mathbf{V}_N(i) = \sum_\alpha \frac{Z_\alpha}{|\mathbf{r}_\alpha - \mathbf{r}_i|}, \tag{6}$$

$$\mathbf{V}_e(i) = \sum_{\rho\sigma} P_{\rho\sigma}\mathbf{V}_{\rho\sigma}(i), \tag{7}$$

$$\mathbf{V}_{\rho\sigma}(i) = -\left\langle \rho \left| \frac{1}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \sigma \right\rangle, \tag{8}$$

where Z_α and $\mathbf{r}_\alpha$ are the nuclear charge and coordinates of atom α; ρ and σ, as well as μ and ν used in this paper, are Gaussian type basis functions; $P_{\rho\sigma}$ is the density matrix; $\mathbf{r}_e$ is the electronic coordinate.

Equation (5) can also be equivalently written as the basis set induced surface charges $\mathbf{q}_{\rho\sigma}$ contracted by the density matrix:

$$\mathbf{q}_e(i) = \sum_{\rho\sigma} P_{\rho\sigma}\mathbf{q}_{\rho\sigma}(i). \tag{9}$$

$\mathbf{q}_{\rho\sigma}$ are obtained by solving Eq. (1) for the basis set potentials $\mathbf{V}_{\rho\sigma}$ at the tesserae:

$$\mathbf{q}_{\rho\sigma} = -\frac{\varepsilon - 1}{\varepsilon}\mathbf{C}^{-1}\mathbf{V}_{\rho\sigma}, \tag{10}$$

with $\mathbf{V}_{\rho\sigma}$ given by Eq. (8). It is convenient to use $\mathbf{q}_{\rho\sigma}$ to form an operator, such as that used in Eq. (11).

The electrostatic potentials created by the induced surface charges are then included into the Hartree–Fock or DFT Kohn–Sham equations to variationally determine the total molecular energy E_{total}:

$$\begin{aligned}
E_{\text{total}} = &\sum_{\mu\nu} P_{\mu\nu}\left(T_{\mu\nu} + V_{\mu\nu,N} - \frac{1}{2}\sum_i \left\langle \mu \left| \frac{\mathbf{q}_N(i)}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \nu \right\rangle\right) \\
&+ \frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}P_{\rho\sigma}\left(\langle \mu\nu|\rho\sigma\rangle\right) \\
&+ \sum_i \left\langle \mu \left| \frac{\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \nu \right\rangle\right) \\
&+ E_{xc}[\rho] + \left(E_{NN} + \frac{1}{2}\sum_\alpha\sum_i \frac{Z_\alpha \mathbf{q}_N(i)}{|\mathbf{r}_\alpha - \mathbf{r}_i|}\right. \\
&\left. + \frac{1}{2}\sum_{\rho\sigma} P_{\rho\sigma}\sum_\alpha\sum_i \frac{Z_\alpha \mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_\alpha - \mathbf{r}_i|}\right),
\end{aligned} \tag{11}$$

where $T_{\mu\nu}$ and $V_{\mu\nu}$ are the basis set kinetic energy and electron-nuclei potential energy integrals, respectively; $E_{xc}[\rho]$ is the exchange-correlation energy; E_{NN} is the nuclear repulsion energy.

In Eq. (11) the four terms involving $\mathbf{q}_N$ and $\mathbf{q}_{\rho\sigma}$ are, respectively, the electrostatic interactions between (1) solute electrons and induced surface charge due to the solute nuclei, (2) solute electrons and induced surface charge due to the solute electrons, (3) solute nuclei and induced surface charge due to the solute nuclei, and (4) solute nuclei and induced surface charge due to the solute electrons. These four terms can be combined to form a single term $\frac{1}{2}\mathbf{V}\mathbf{q}$, which is referred to as the solvation energy; the remaining terms have the same forms as those in the gas phase calculations, and can be written as E_{gas}:

$$E_{\text{total}} = \sum_{\mu\nu} P_{\mu\nu}(T_{\mu\nu} + V_{\mu\nu,N}) + \frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}P_{\rho\sigma}(\langle\boldsymbol{\mu}\boldsymbol{\nu}|\boldsymbol{\rho}\boldsymbol{\sigma}\rangle) + E_{\text{xc}}[\boldsymbol{\rho}] + E_{\text{NN}} - \frac{1}{2}\sum_{\mu\nu} P_{\mu\nu}\sum_{i}\left\langle\boldsymbol{\mu}\left|\frac{\mathbf{q}_N(i)}{|\mathbf{r}_e - \mathbf{r}_i|}\right|\boldsymbol{\nu}\right\rangle$$

$$+ \frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}P_{\rho\sigma}\sum_{i}\left\langle\boldsymbol{\mu}\left|\frac{\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_e - \mathbf{r}_i|}\right|\boldsymbol{\nu}\right\rangle + \frac{1}{2}\sum_{\alpha}\sum_{i}\frac{Z_\alpha\mathbf{q}_N(i)}{|\mathbf{r}_\alpha - \mathbf{r}_i|} + \frac{1}{2}\sum_{\rho\sigma} P_{\rho\sigma}\sum_{\alpha}\sum_{i}\frac{Z_\alpha\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_\alpha - \mathbf{r}_i|}$$

$$= E_{\text{gas}} + \frac{1}{2}\mathbf{V}_e\mathbf{q}_N + \frac{1}{2}\mathbf{V}_e\mathbf{q}_e + \frac{1}{2}\mathbf{V}_N\mathbf{q}_N + \frac{1}{2}\mathbf{V}_N\mathbf{q}_e = E_{\text{gas}} + \frac{1}{2}\mathbf{V}\mathbf{q}. \tag{12}$$

For Hartree–Fock methods, the exchange-correlation term $E_{\text{xc}}[\rho]$ is

$$E_{\text{xc}}[\rho] = -\frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}P_{\rho\sigma}\langle\mu\rho|\nu\sigma\rangle. \tag{13}$$

For DFT methods, $E_{\text{xc}}[\rho]$ has various forms, which are not the main concern of this paper.

B. Gradient

The general expressions of the nuclear gradients in continuum solvation models have been derived for DPCM,[8] IEF-PCM,[10,19] COSMO,[4] GCOSMO,[12] and CPCM.[6,13] In the following, for the purpose of establishing the necessary notations and describing the actual implementation, a short derivation of the COSMO gradients is presented.

In COSMO/HF and COSMO/DFT calculations, when the coordinate x (of atom X) changes by ∂x, (1) the nuclear coordinates of atom X will change by ∂x, (2) the basis set centered at atom X will change by ∂x and (3) the coordinates and areas of some COSMO tesserae will also change according to the surface tessellation scheme.

Following Eqs. (11) and (12), the first derivative of the total molecular energy with respect to x is

$$E_{\text{total}}^x = \sum_{\mu\nu} P_{\mu\nu}(T_{\mu\nu}^x + V_{\mu\nu,N}^x) + \frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}P_{\rho\sigma}\langle\boldsymbol{\mu}\boldsymbol{\nu}|\boldsymbol{\rho}\boldsymbol{\sigma}\rangle^x$$

$$+ E_{\text{xc}}^x[\boldsymbol{\rho}] + E_{\text{NN}}^x - \frac{1}{2}\sum_{\mu\nu} P_{\mu\nu}\sum_{i}\left\langle\boldsymbol{\mu}\left|\frac{\mathbf{q}_N(i)}{|\mathbf{r}_e - \mathbf{r}_i|}\right|\boldsymbol{\nu}\right\rangle^x$$

$$+ \frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}P_{\rho\sigma}\sum_{i}\left\langle\boldsymbol{\mu}\left|\frac{\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_e - \mathbf{r}_i|}\right|\boldsymbol{\nu}\right\rangle^x$$

$$+ \frac{1}{2}\sum_{\alpha}\sum_{i}\left[\frac{Z_\alpha\mathbf{q}_N(i)}{|\mathbf{r}_\alpha - \mathbf{r}_i|}\right]^x$$

$$+ \frac{1}{2}\sum_{\rho\sigma} P_{\rho\sigma}\sum_{\alpha}\sum_{i}\left[\frac{Z_\alpha\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_\alpha - \mathbf{r}_i|}\right]^x - \sum_{\mu\nu} W_{\mu\nu}S_{\mu\nu}^x, \tag{14}$$

where S is the basis set overlap integral matrix, W is the energy-weighted density matrix (note the orbital energy includes both the DFT exchange-correlation and COSMO contributions). The $W_{\mu\nu}S_{\mu\nu}^x$ term absorbs all the quantities containing the derivative of the density matrix [this is possible

because, and only if, the $\mathbf{C}$ matrix in Eq. (1) is symmetric]. Therefore, the exchange-correlation energy gradient $E_{\text{xc}}^x[\rho]$ in Eq. (14) only involves the derivatives of the exchange-correlation potential integrals and the density matrix (no derivatives of the density matrix). Equation (14) is an analogy of Pulay's original formula for the gas phase Hartree–Fock gradient, and can be derived using the same procedure.[20]

Equation (14) clearly shows that the total molecular energy gradient in COSMO/HF and COSMO/DFT methods consists of two parts: (1) one part, denoted as E_{gas}^x, has the same form as "the gas phase" gradient, but evaluated with the COSMO perturbed density matrix and the COSMO corrected energy-weighted density matrix; (2) the other part, denoted as E_{sol}^x, contains the derivatives of the COSMO solvation terms:

$$E_{\text{gas}}^x = \sum_{\mu\nu} P_{\mu\nu}(T_{\mu\nu}^x + V_{\mu\nu,N}^x) + \frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}P_{\rho\sigma}\langle\mu\nu|\rho\sigma\rangle^x$$

$$+ E_{\text{xc}}^x[\rho] + E_{\text{NN}}^x - \sum_{\mu\nu} W_{\mu\nu}S_{\mu\nu}^x, \tag{15}$$

$$E_{\text{sol}}^x = -\frac{1}{2}\sum_{\mu\nu} P_{\mu\nu}\sum_{i}\left\langle\boldsymbol{\mu}\left|\frac{\mathbf{q}_N(i)}{|\mathbf{r}_e - \mathbf{r}_i|}\right|\boldsymbol{\nu}\right\rangle^x$$

$$+ \frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}P_{\rho\sigma}\sum_{i}\left\langle\boldsymbol{\mu}\left|\frac{\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_e - \mathbf{r}_i|}\right|\boldsymbol{\nu}\right\rangle^x$$

$$+ \frac{1}{2}\sum_{\alpha}\sum_{i}\left[\frac{Z_\alpha\mathbf{q}_N(i)}{|\mathbf{r}_\alpha - \mathbf{r}_i|}\right]^x$$

$$+ \frac{1}{2}\sum_{\rho\sigma} P_{\rho\sigma}\sum_{\alpha}\sum_{i}\left[\frac{Z_\alpha\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_\alpha - \mathbf{r}_i|}\right]^x. \tag{16}$$

The first term in Eq. (16) can be expanded as

$$-\frac{1}{2}\sum_{\mu\nu} P_{\mu\nu}\sum_{i}\left\langle\boldsymbol{\mu}\left|\frac{\mathbf{q}_N(i)}{|\mathbf{r}_e - \mathbf{r}_i|}\right|\boldsymbol{\nu}\right\rangle^x$$

$$= -\frac{1}{2}\sum_{\mu\nu} P_{\mu\nu}\sum_{i}\left\langle\boldsymbol{\mu}\left|\frac{1}{|\mathbf{r}_e - \mathbf{r}_i|}\right|\boldsymbol{\nu}\right\rangle^x\mathbf{q}_N(i)$$

$$- \frac{1}{2}\sum_{\mu\nu} P_{\mu\nu}\sum_{i}\left\langle\boldsymbol{\mu}\left|\frac{1}{|\mathbf{r}_e - \mathbf{r}_i|}\right|\boldsymbol{\nu}\right\rangle\mathbf{q}_N^x(i). \tag{17}$$

The derivative of the induced nuclear charge $\mathbf{q}_N^x$ in Eq. (17) can be avoided by using the standard formula

$$(\mathbf{C}^{-1})^x = -\mathbf{C}^{-1}(\mathbf{C})^x\mathbf{C}^{-1} \tag{18}$$

to convert induced charges back to potentials:

$$-\frac{1}{2}\sum_{\mu\nu}P_{\mu\nu}\sum_i\left\langle\mu\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle\mathbf{q}_N^x(i) = \frac{1}{2}\sum_i\mathbf{V}_e(i)\mathbf{q}_N^x(i) = \frac{1}{2}\mathbf{V}_e\mathbf{q}_N^x$$

$$= \frac{1}{2}\mathbf{V}_e\left(-\frac{\varepsilon-1}{\varepsilon}\mathbf{C}^{-1}\mathbf{V}_N\right)^x$$

$$= -\frac{1}{2}\left(\frac{\varepsilon-1}{\varepsilon}\right)\mathbf{V}_e[(\mathbf{C}^{-1})^x\mathbf{V}_N + \mathbf{C}^{-1}\mathbf{V}_N^x]$$

$$= -\frac{1}{2}\left(\frac{\varepsilon-1}{\varepsilon}\right)\mathbf{V}_e[-\mathbf{C}^{-1}\mathbf{C}^x\mathbf{C}^{-1}\mathbf{V}_N] - \frac{1}{2}\left(\frac{\varepsilon-1}{\varepsilon}\right)\mathbf{V}_e\mathbf{C}^{-1}\mathbf{V}_N^x$$

$$= \frac{1}{2}\left(\frac{\varepsilon}{\varepsilon-1}\right)\mathbf{q}_e\mathbf{C}^x\mathbf{q}_N + \frac{1}{2}\mathbf{q}_e\mathbf{V}_N^x. \tag{19}$$

So, Eq. (17) becomes

$$-\frac{1}{2}\sum_{\mu\nu}P_{\mu\nu}\sum_i\left\langle\mu\left|\frac{\mathbf{q}_N(i)}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle^x = -\frac{1}{2}\sum_{\mu\nu}P_{\mu\nu}\sum_i\left\langle\mu\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle^x\mathbf{q}_N(i) + \frac{1}{2}\left(\frac{\varepsilon}{\varepsilon-1}\right)\mathbf{q}_e\mathbf{C}^x\mathbf{q}_N + \frac{1}{2}\mathbf{q}_e\mathbf{V}_N^x. \tag{20}$$

Similarly, the second, third and forth terms of Eq. (16) can be derived, so Eq. (16) becomes

$$E_{\text{sol}}^x = -\frac{1}{2}\sum_{\mu\nu}P_{\mu\nu}\sum_i\left\langle\mu\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle^x\mathbf{q}_N(i) + \frac{1}{2}\left(\frac{\varepsilon}{\varepsilon-1}\right)\mathbf{q}_e\mathbf{C}^x\mathbf{q}_N + \frac{1}{2}\mathbf{q}_e\mathbf{V}_N^x - \sum_{\mu\nu}P_{\mu\nu}\sum_i\left\langle\mu\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle^x\mathbf{q}_e(i)$$

$$+ \frac{1}{2}\left(\frac{\varepsilon}{\varepsilon-1}\right)\mathbf{q}_e\mathbf{C}^x\mathbf{q}_e + \frac{1}{2}\left(\frac{\varepsilon}{\varepsilon-1}\right)\mathbf{q}_N\mathbf{C}^x\mathbf{q}_N + \mathbf{q}_N\mathbf{V}_N^x + \frac{1}{2}\mathbf{q}_e\mathbf{V}_N^x + \frac{1}{2}\left(\frac{\varepsilon}{\varepsilon-1}\right)\mathbf{q}_N\mathbf{C}^x\mathbf{q}_e - \frac{1}{2}\sum_{\rho\sigma}P_{\rho\sigma}\sum_i\left\langle\rho\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\sigma\right\rangle^x\mathbf{q}_N(i). \tag{21}$$

For convenience, Eq. (21) can be written as

$$E_{\text{sol}}^x = -\sum_{\mu\nu}P_{\mu\nu}\sum_i\left\langle\mu\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle^x\mathbf{q}(i) + \mathbf{q}\mathbf{V}_N^x$$

$$+ \frac{1}{2}\left(\frac{\varepsilon}{\varepsilon-1}\right)\mathbf{q}\mathbf{C}^x\mathbf{q}. \tag{22}$$

Equation (22) shows that the derivatives of the solvation terms can be evaluated directly with the total induced surface charge $\mathbf{q}$.

The first term in Eq. (22) represents the forces between induced surface charges and solute electrons due to the changes of the basis set potential integrals at the tesserae when the x coordinate changes, and can be evaluated with the derivatives of the basis sets:

$$\left\langle\mu\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle^x = \left\langle\frac{\partial\mu}{\partial\mathbf{r}_\mu}\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle\frac{\partial\mathbf{r}_\mu}{\partial x}$$

$$+ \left\langle\mu\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\frac{\partial\nu}{\partial\mathbf{r}_\nu}\right\rangle\frac{\partial\mathbf{r}_\nu}{\partial x}$$

$$+ \left\langle\mu\left|\frac{\partial(|\mathbf{r}_e-\mathbf{r}_i|^{-1})}{\partial\mathbf{r}_i}\right|\nu\right\rangle\frac{\partial\mathbf{r}_i}{\partial x}, \tag{23}$$

where $\mathbf{r}_\mu$ and $\mathbf{r}_\nu$ are, respectively, the center points of basis functions μ and ν. If $\partial\mathbf{r}_\mu/\partial x = \partial\mathbf{r}_\nu/\partial x = \partial\mathbf{r}_i/\partial x$, which means

basis functions μ, ν and tessera i move together, Eq. (23) is exactly zero because the integral will not change (the relative positions of $\mathbf{r}_\mu$, $\mathbf{r}_\nu$, and $\mathbf{r}_i$ are not changing), and the last term in Eq. (23) can be evaluated with the derivatives of the basis set:

$$\left\langle\mu\left|\frac{\partial(|\mathbf{r}_e-\mathbf{r}_i|^{-1})}{\partial\mathbf{r}_i}\right|\nu\right\rangle = -\left\langle\frac{\partial\mu}{\partial\mathbf{r}_\mu}\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle$$

$$- \left\langle\mu\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\frac{\partial\nu}{\partial\mathbf{r}_\nu}\right\rangle. \tag{24}$$

Therefore, Eq. (23) becomes

$$\left\langle\mu\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle^x = \left\langle\frac{\partial\mu}{\partial\mathbf{r}_\mu}\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle\left(\frac{\partial\mathbf{r}_\mu}{\partial x} - \frac{\partial\mathbf{r}_i}{\partial x}\right)$$

$$+ \left\langle\mu\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\frac{\partial\nu}{\partial\mathbf{r}_\nu}\right\rangle\left(\frac{\partial\mathbf{r}_\nu}{\partial x} - \frac{\partial\mathbf{r}_i}{\partial x}\right), \tag{25}$$

where $\partial\mathbf{r}_\mu/\partial x$ and $\partial\mathbf{r}_\nu/\partial x$ are one if the basis functions μ and ν belong to atom X, and are zero otherwise; $\partial\mathbf{r}_i/\partial x$ depends on the surface tessellation scheme. The basis set of atom X will move and cause changes in the basis set potential integrals at all tesserae. In other words, the "electrons of atom X" (in the sense that they are described by the basis set of X)

feel the electrostatic forces from all surface charges. In the meantime, the coordinates of (some or all) tesserae will also change, and cause changes in the basis set potential integrals at themselves. In other words, the "tesserae of atom X" (in the sense that they move with atom X) feel the electrostatic forces from all electrons.

The second term in Eq. (22) represents the forces between induced surface charges and solute nuclei due to the changes of the nuclear potentials at each tessera when the x coordinate changes, and can be evaluated straightforwardly:

$$\mathbf{q}\mathbf{V}_N^x = \sum_\alpha \sum_i Z_\alpha \mathbf{q}(i)\left[|\mathbf{r}_\alpha - \mathbf{r}_i|^{-3} \cdot (\mathbf{r}_i - \mathbf{r}_\alpha) \right.$$
$$\left. \cdot \left(\frac{\partial \mathbf{r}_\alpha}{\partial x} - \frac{\partial \mathbf{r}_i}{\partial x} \right) \right]. \tag{26}$$

$\partial \mathbf{r}_\alpha / \partial x$ is one if atom $\mathbf{r}_\alpha$ is x, and is zero otherwise; $\partial \mathbf{r}_i / \partial x$ depends on the surface tessellation scheme. The nucleus of atom X will move by ∂x and cause changes in the potentials at all tesserae; in the meantime, the coordinates of (some or all) tesserae will also change and cause changes in the nuclear potentials at themselves. In other words, the nucleus of atom X feels the electrostatic forces from all surface charges, and the tesserae of atom X feel the electrostatic forces from all solute nuclei.

The third term in Eq. (22) represents the forces between induced surface charges and can be written separately for the diagonal and off-diagonal parts of the $\mathbf{C}$ matrix:

$$\frac{1}{2}\left(\frac{\varepsilon}{\varepsilon - 1} \right)\mathbf{q}(i)\mathbf{C}_{ii}^x \mathbf{q}(i) = -\frac{1}{2}\left(\frac{\varepsilon}{\varepsilon - 1} \right)\frac{1.07\sqrt{4\pi}}{2(a_i)^{3/2}}\frac{\partial a_i}{\partial x}\mathbf{q}(i)\mathbf{q}(i), \tag{27}$$
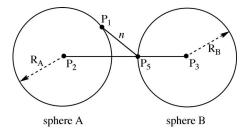
$$\frac{1}{2}\left(\frac{\varepsilon}{\varepsilon - 1} \right)\mathbf{q}(i)\mathbf{C}_{ij}^x \mathbf{q}(j) = -\frac{1}{2}\left(\frac{\varepsilon}{\varepsilon - 1} \right)|\mathbf{r}_i - \mathbf{r}_j|^{-3}(\mathbf{r}_i - \mathbf{r}_j)$$
$$\cdot \left(\frac{\partial \mathbf{r}_i}{\partial x} - \frac{\partial \mathbf{r}_j}{\partial x} \right)\mathbf{q}(i)\mathbf{q}(j). \tag{28}$$
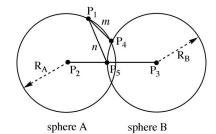
Equation (27) represents the self-repulsion force of the charges within tessera i due to its area changes; Eq. (28) represents the forces between the charges on tesserae i and j: the forces are seemingly halved by the 1/2 factor, but are actually not, because the $\mathbf{C}$ matrix is symmetric (i.e., $\mathbf{C}_{ij} = \mathbf{C}_{ji}$). The values of $\partial a_i / \partial x$ and $\partial \mathbf{r}_i / \partial x$ depend on the surface tessellation scheme.

Clearly, the key for evaluating Eq. (22) is to obtain the derivatives of the areas (a_i) and coordinates ($\mathbf{r}_i$) of the tesserae with respect to the atomic coordinate x.

C. Fixed point with variable area

A simple and efficient surface tessellation scheme, FIXPVA, is described here. FIXPVA is able to provide the analytic derivatives of the areas (a_i) and coordinates ($\mathbf{r}_i$) of the tesserae, thus enabling the evaluation of analytic gradients. In addition, FIXPVA can generate rigorously continuous and smooth potential energy surfaces, which are crucial for geometry optimizations.
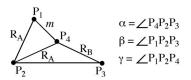

FIG. 1. Two distances (m and n) are used for two switching functions to scale the area of a tessera at P_1.

The main features of the FIXPVA scheme can be summarized as

(a)   Each atom is assigned with a sphere and no additional spheres are used;

(b)   each sphere is divided into a certain number of tesserae (such as 60, 240, or 960);

(c)   the position of a tessera is always fixed relative to the center of the associated sphere;

(d)   the area of a tessera is a smooth function of its distances to other spheres.

In the following, more detailed descriptions are presented.

The position of a tessera is always fixed relative to the center of its sphere (i.e., the atom). As a result, the derivatives of the tessera coordinate $\mathbf{r}_i$ with respect to the atomic coordinate x (of atom X) can be evaluated efficiently:

$$\frac{\partial \mathbf{r}_i}{\partial x} = \begin{cases} 1, & \text{if } i \text{ is on sphere } X, \\ 0, & \text{if } i \text{ is not on sphere } X. \end{cases} \tag{29}$$

Equation (29) can be used to evaluate Eqs. (25), (26), and (28).

Two spheres, A and B, are used to illustrate the FIXPVA scheme (Fig. 1). The area of a tessera i at point P_1 on the surface of sphere A (radius is R_A) is initially assigned as $a_0 = 4\pi R_A^2 / 60$ (assume the sphere is divided into 60 tesserae). Then, according to two distances to another sphere B, m and n, the area of tessera i is scaled by two functions, f_1 and f_2:

$$a_i = f_1 \cdot f_2 \cdot a_0, \tag{30}$$

where f_1 and f_2 are the well-known fifth-order polynomial functions:[21]

$$f_1 = \begin{cases} 1, & m > m_2, \\ 10\left(\dfrac{m^2 - m_1^2}{m_2^2 - m_1^2}\right)^3 - 15\left(\dfrac{m^2 - m_1^2}{m_2^2 - m_1^2}\right)^4 + 6\left(\dfrac{m^2 - m_1^2}{m_2^2 - m_1^2}\right)^5, & m_2 \geq m > m_1, \\ 0, & m_1 \geq m, \end{cases} \tag{31}$$

$$f_2 = \begin{cases} 1, & n > n_2, \\ 10\left(\dfrac{n^2 - n_1^2}{n_2^2 - n_1^2}\right)^3 - 15\left(\dfrac{n^2 - n_1^2}{n_2^2 - n_1^2}\right)^4 + 6\left(\dfrac{n^2 - n_1^2}{n_2^2 - n_1^2}\right)^5, & n_2 \geq n > n_1, \\ 0, & n_1 \geq n. \end{cases} \tag{32}$$

It is noted that the same type of switching function was used in the method of York and Karplus,[14] but the variables are different.

As shown in Fig. 1, m is the distance from the tessera P_1 to the surface intersecting point P_4 of a neighboring sphere (P_1, P_2, P_3, and P_4 are on the same plane); n is the distance from the tessera P_1 to the surface point P_5 of a neighboring sphere (P_2, P_5, and P_3 are on the same line):

$$m^2 = |\mathbf{r}_1 - \mathbf{r}_4|^2, \tag{33}$$

$$n^2 = |\mathbf{r}_1 - \mathbf{r}_5|^2. \tag{34}$$

The values of 0.02 and 0.3 Å, respectively, are used for m_1 and m_2. In principle, a value of zero can be used for m_1. However, this may lead to cases in which two tesserae on two spheres have zero distance. Although their areas are zero when the distance is zero, the numerical instability in dealing with these "zero-divided-by-zero" problems is better avoided. The value for m_2 cannot be too small or too large—otherwise the switching function f_1 will be either too sharp (sudden scaling) or too wide (unnecessary scaling). Numerical tests show that 0.3 Å is likely the minimum value that can always lead to satisfactory smooth potential energy surfaces.

The values of 1.0 and 1.5 Å, respectively, are used for n_1 and n_2. These values are determined according to the fact that there should be no solvent molecules between two non-bonded heavy atoms (such as C, N, O, S, and P in both solute and solvent molecules) that are 5–6 Å from each other. In continuum solvation model calculations spheres with radii of ~ 2.0 Å are used for heavy atoms so values of 1.0–1.5 Å are reasonable for n_1 and n_2. In addition, f_2 is necessary for two separated spheres because in this case m does not exist so two tesserae on two separated spheres can have full areas and arbitrarily small distances. In principle, it is possible to use only one switching function of one distance, such as the distance of a tessera to the surface of another sphere, for all the tesserae, as used by York and Karplus[14] However, it is difficult to use only one switching function to achieve a similar result produced by the f_1 and f_2 used in the present work.

Figure 2 visualizes how the surface areas are affected by switching functions for two spheres with radii =2 Å. Points I, J, K, and L and the center points of the two spheres are on the x-y plane, and the center point of the left sphere has a

coordinate (0, 0), and the coordinates of I, J, K, and L are (0, 2), (1, 1.732), (1,732,1), and (2,0), respectively (all in Å). The coordinate of the right sphere is (x, 0) with x changes from 0 to 6 Å.

Tesserae at points I, J, K, and L on the left sphere originally have full areas ($1.0a_0$). As the right sphere approaches and merges in, the area of tessera L starts to be scaled when the distance between the two sphere centers is 5.5 Å ($R_{\text{left}} = 2.0$ Å, $R_{\text{right}} = 2.0$ Å, $n_2 = 1.5$ Å), and is scaled to zero when the distance is 5.0 Å. For tessera L, only the switching function f_2 is meaningful because it is already zero before the spheres contact. Tessera K is almost solely controlled by f_2, but f_1 also contribute after the spheres merge into each other. As a result, the area of tessera K starts to be scaled at a distance of 4.85 Å (where $n < 1.5$ Å) and is scaled to almost zero at a distance of 4.0 Å (where $n = 1.035$ Å). Tessera J is controlled solely by f_1 (since n always > 1.732 Å$> n_2$), and is scaled at distances 2.5 and 2.1 Å. Similar to J, tessera I is also controlled solely by f_1, and is scaled at distances 0.6 and 0.1 Å.

In general, the area of a tessera is scaled by the switching functions associated with all the other spheres. In an actual calculation, of course, it is only effectively scaled by the switching functions associated with a few neighboring spheres since all distant spheres have $f_1 = f_2 = 1$. The derivative of the area of a tessera with respect to an atomic coordinate x (of atom X) depends on the derivatives of the
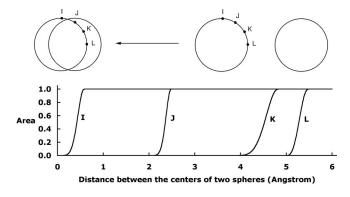

FIG. 2. The areas of tesserae I, J, K, and L on the left sphere are smoothly switched to zero as the right sphere approaches and cuts in. The radii of the two spheres are both of 2 Å.

switching functions associated with the sphere centered at atom X and the switching functions associated with spheres centered at other atoms.

For example, if tessera i (on sphere A) is close to three other spheres (B, C, and D), three sets of switching functions are involved:

$$a_i = (f_1 \cdot f_2)(g_1 \cdot g_2)(h_1 \cdot h_2) \cdot a_0, \tag{35}$$

where f, g, and h are switching functions for sphere B, C, and D, respectively.

Using sphere B (Fig. 1) to represent a neighboring sphere, the derivative of a_i with respect to the x_3 coordinate of the center point, P_3 (x_3, y_3, z_3), of sphere B is

$$\frac{\partial a_i}{\partial x_3} = (g_1 \cdot g_2)(h_1 \cdot h_2) \cdot a_0 \cdot \left(f_1 \frac{\partial f_2}{\partial x_3} + f_2 \frac{\partial f_1}{\partial x_3} \right). \tag{36}$$

Clearly, the derivatives of m^2 and n^2 are required to obtain the derivatives of f_1 and f_2.

$$\frac{\partial f_1}{\partial x_3} = \left[30 \left(\frac{m^2 - m_1^2}{m_2^2 - m_1^2} \right)^2 - 60 \left(\frac{m^2 - m_1^2}{m_2^2 - m_1^2} \right)^3 \right.$$
$$\left. + 30 \left(\frac{m^2 - m_1^2}{m_2^2 - m_1^2} \right)^4 \right] \frac{1}{m_2^2 - m_1^2} \frac{\partial (m^2)}{\partial x_3}, \tag{37}$$

$$\frac{\partial f_2}{\partial x_3} = \left[30 \left(\frac{n^2 - n_1^2}{n_2^2 - n_1^2} \right)^2 - 60 \left(\frac{n^2 - n_1^2}{n_2^2 - n_1^2} \right)^3 \right.$$
$$\left. + 30 \left(\frac{n^2 - n_1^2}{n_2^2 - n_1^2} \right)^4 \right] \frac{1}{n_2^2 - n_1^2} \frac{\partial (n^2)}{\partial x_3}. \tag{38}$$

Here the derivatives of m^2 and n^2 with respect to x_3 are derived. As shown in Fig. 1, c is the distance between the center points, P_2 (x_2, y_2, z_2) and P_3 (x_3, y_3, z_3), of spheres A and B; d is the distance between tesserae P_1 (x_1, y_1, z_1) and the center point of sphere B:

$$c = \sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2 + (z_2 - z_3)^2}, \tag{39}$$

$$d = \sqrt{(x_1 - x_3)^2 + (y_1 - y_3)^2 + (z_1 - z_3)^2}. \tag{40}$$

Using the cosine formula, m^2 can be expressed as (angles shown in Fig. 1):

$$m^2 = 2R_A^2(1 - \cos \gamma)$$
$$= 2R_A^2[1 - \cos(\alpha - \beta)]$$
$$= 2R_A^2(1 - \cos \beta \cos \alpha - \sin \beta \sin \alpha), \tag{41}$$

with

$$\cos \beta = \frac{R_A^2 + c^2 - d^2}{2R_A c}; \tag{42}$$

$$\sin \beta = \frac{\sqrt{(2R_A c)^2 - (R_A^2 + c^2 - d^2)^2}}{2R_A c},$$

$$\cos \alpha = \frac{R_A^2 + c^2 - R_B^2}{2R_A c}; \tag{43}$$

$$\sin \alpha = \frac{\sqrt{(2R_A c)^2 - (R_A^2 + c^2 - R_B^2)^2}}{2R_A c}.$$

Without showing the details, the derivative of m^2 with respect to the coordinate x_3 of the center point of sphere B can be obtained:

$$\frac{\partial (m^2)}{\partial x_3} = \frac{(x_3 - x_2)}{c^4} [(R_A^2 + c^2 - d^2) \cdot (R_A^2 + c^2 - R_B^2) + F_{ab}]$$
$$- \frac{1}{c^2} [(x_1 - x_2)(R_A^2 + c^2 - R_B^2) + (x_3 - x_2)$$
$$\times (R_A^2 + c^2 - d^2)] - \frac{1}{2c^2} \frac{\partial F_{ab}}{\partial x_3}, \tag{44}$$

where

$$F_{ab}$$
$$= \sqrt{(R_A + c + d) \cdot (R_A + c - d) \cdot (R_A - c + d) \cdot (-R_A + c + d) \cdot (R_A + c + R_B) \cdot (R_A + c - R_B) \cdot (R_A - c + R_B) \cdot (-R_A + c + R_B)}, \tag{45}$$

$$\frac{\partial F_{ab}}{\partial x_3} = \frac{1}{2} F_{ab} \cdot \left[\frac{\left(\frac{\partial c}{\partial x_3} + \frac{\partial d}{\partial x_3} \right)}{(R_A + c + d)} + \frac{\left(\frac{\partial c}{\partial x_3} - \frac{\partial d}{\partial x_3} \right)}{(R_A + c - d)} + \frac{\left(-\frac{\partial c}{\partial x_3} + \frac{\partial d}{\partial x_3} \right)}{(R_A - c + d)} + \frac{\left(\frac{\partial c}{\partial x_3} + \frac{\partial d}{\partial x_3} \right)}{(-R_A + c + d)} \right] + \frac{1}{2} F_{ab} \cdot \left(\frac{\partial c}{\partial x_3} \right) \left[\frac{1}{(R_A + c + R_B)} \right.$$
$$\left. + \frac{1}{(R_A + c - R_B)} + \frac{-1}{(R_A - c + R_B)} + \frac{1}{(-R_A + c + R_B)} \right], \tag{46}$$

with [see Eqs. (39) and (40)]

$$\frac{\partial c}{\partial x_3} = \frac{x_3 - x_2}{c}; \quad \frac{\partial d}{\partial x_3} = \frac{x_3 - x_1}{d}. \quad (47)$$

It is trivial to show that

$$\begin{aligned}
\frac{\partial(n^2)}{\partial x_3} = &- 2(x_1 - x_5)\left[1 - \frac{R_B}{c} + \frac{(x_2 - x_3)^2}{c^3}R_B\right] \\
&- 2(y_1 - y_5)\left[\frac{(x_2 - x_3)(y_2 - y_3)}{c^3}R_B\right] \\
&- 2(z_1 - z_5)\left[\frac{(x_2 - x_3)(z_2 - z_3)}{c^3}R_B\right],
\end{aligned} \quad (48)$$

with

$$n^2 = (x_1 - x_5)^2 + (y_1 - y_5)^2 + (z_1 - z_5)^2, \quad (49)$$

$$x_5 = x_3 + (x_2 - x_3)R_B/c;$$

$$y_5 = y_3 + (y_2 - y_3)R_B/c; \quad (50)$$

$$z_5 = z_3 + (z_2 - z_3)R_B/c.$$

In short, the analytic derivatives of the switching functions for the tesserae can be readily and efficiently evaluated with the Cartesian coordinates of the tesserae, the sphere center points and the radii of the spheres, and used to calculate the derivatives of the tessera areas, which are required by Eq. (27).

In Fig. 1, sphere A (of atom A) can also move while other spheres (atoms) remain still. Since the distances from tessera i to other spheres will change as A moves, a_i will change. The derivatives of a_i with respect to the coordinates of P_2 (x_2, y_2, z_2), the center point of sphere A, can be evaluated by summing up the negative derivatives of a_i with respect to the coordinates of the center points of all the rest spheres. For example, if only spheres A and B are considered (Fig. 1),

$$\frac{\partial a_i}{\partial x_2} = -\frac{\partial a_i}{\partial x_3}. \quad (51)$$

III. IMPLEMENTAION AND COMPUTATIONAL METHODOLOGY

The FIXPVA tessellation scheme was implemented in the GAMESS (Ref. 22) program for the CPCM/HF and CPCM/DFT methods, with which all the calculations were performed. The CPCM program in GAMESS was implemented by Li and Jensen[18] in a previous work on the basis of the IEF-PCM program originally implemented by Tomasi et al.[23] and Li et al.[24] The GEPOL and GEPOL-AS schemes are also available in GAMESS.

In FIXPVA, the partition of each sphere surface into 60 initial tesserae is based on pentakis dodecahedron (60 nonequilateral but equal area triangles on a sphere), the same as in the GEPOL scheme. The projections of the 60 triangle center points on the sphere are defined as the center points of the tesserae, and 1/60 of the sphere area is assigned to each tessera as the initial area, which is scaled by the switching functions.

Options with 240 and 960 initial tesserae per sphere have also been implemented for the GEPOL, GEPOL-AS and FIXPVA scheme. To obtain 240 initial tesserae for a sphere, each of the 60 nonequilateral triangles is subdivided into four identical triangles and projected onto the sphere. Although the four subdivided triangles have identical areas, their projections on the sphere have different surface areas that can differ by ~7%. The situation is similar for 960 initial tesserae. The projections of the subdivided triangle center points are taken as the initial tessera center points; and the projections of the subdivided triangle areas are calculated using the Gauss–Bonnet formula and taken as the initial tessera areas (the actual implementation uses precalculated scaling factors to obtain the correct areas). The center points and areas of the boundary tesserae are recalculated in GE-POL and GEPOL-AS; while in FIXPVA, the initial areas will be scaled. It is found that the continuity and smoothness of the FIXPVA-CPCM potential energy surfaces are insensitive to the density of tesserae, and 60 per sphere is as good as 240 and 960, and is recommended for general use for its efficiency (see discussion on the rotational variance). In all the calculations reported in this paper, 60 tesserae per sphere were used.

In the CPCM calculations, the solvent was water with $\varepsilon = 78.39$. Spheres with radii of 0, 2.124, 2.016, 1.908, 2.52, and 2.76 Å were used for H, C, N, O, S, and Cu atoms, respectively, to define the molecular cavity; no additional spheres were used. Using zero radii for H atoms means that they do not contribute to the formation of the surface. The induced surface charges were determined by a semi-iterative DIIS procedure[24,25] with no charge renormalization. Geometry optimization is performed in internal coordinates generated by the automatic delocalized coordinate algorithm.[26] The 6-31G* basis set was used for all the HF and B3LYP calculations in this study.

IV. RESULTS AND DISCUSSION

A. Gradients

Tables I–III present both the analytic and numerical gradients obtained with CPCM (using FIXPVA tessellation scheme) and RHF, MCSCF, ROHF, UHF, R-B3LYP, RO-B3LYP, and U-B3LYP methods. The numerical gradients were computed with double displacements (forward and backward) using a step size of 0.001 au for each step. Other DFT methods were also tested and similar results were obtained.

Table I presents the total energy gradients for acetate (Fig. 3) in water computed with CPCM/RHF/6-31G* and CPCM/R-B3LYP/6-31G* at the geometry optimized in the gas phase with RHF/6-31G*. The agreement between the analytic and numerical gradients is perfect. Because the analytic gradients are "exact", the errors are in the numerical values due to the finite displacements. For the RHF case, the maximum error in the numerical gradients is 5.4 $\times 10^{-6}$ a.u., and the root-mean-square error is 2.5

TABLE I. Gradients (a.u.) calculated for CH_3COO^- at the geometry optimized in the gas phase with RHF/6-31G*.

Coordinates	CPCM/RHF/6-31G*		CPCM/R-B3LYP/6-31G*	
	Analytic	Numerical	Analytic	Numerical
C1X	−0.000 931 9	−0.000 930 7	−0.000 705 6	−0.000 705 6
C1Y	0.000 179 7	0.000 179 5	0.000 308 3	0.000 308 2
C1Z	−0.005 375 7	−0.005 375 9	−0.002 582 3	−0.002 581 9
C2X	0.003 836 9	0.003 837 5	0.006 360 2	0.006 360 3
C2Y	−0.000 237 0	−0.000 242 2	−0.000 481 5	−0.000 481 8
C2Z	0.028 739 6	0.028 735 1	0.039 188 0	0.039 184 8
O3X	−0.002 723 3	−0.002 719 6	−0.037 725 1	−0.037 726 8
O3Y	0.000 252 2	0.000 252 2	0.000 401 7	0.000 401 6
O3Z	−0.012 572 5	−0.012 572 3	−0.019 407 8	−0.019 406 4
O4X	0.000 321 1	0.000 318 3	0.030 514 9	0.030 514 1
O4Y	−0.000 185 7	−0.000 185 6	−0.000 110 3	−0.000 110 4
O4Z	−0.009 952 9	−0.009 952 3	−0.026 786 7	−0.026 786 2
H5X	0.000 656 7	0.000 656 4	−0.005 741 2	−0.005 740 1
H5Y	0.000 000 5	0.000 000 5	−0.000 159 6	−0.000 159 8
H5Z	−0.000 441 5	−0.000 441 1	0.004 406 5	0.004 406 7
H6X	−0.000 602 7	−0.000 608 1	0.003 743 6	0.003 742 1
H6Y	−0.000 076 4	−0.000 076 4	−0.005 681 1	−0.005 680 5
H6Z	−0.000 188 0	−0.000 188 0	0.002 523 7	0.002 523 4
H7X	−0.000 556 7	−0.000 562 0	0.003 553 3	0.003 551 4
H7Y	0.000 066 7	0.000 066 6	0.005 722 4	0.005 721 9
H7Z	−0.000 209 0	−0.000 208 9	0.002 658 5	0.002 658 4
Max error		0.000 005 4		0.000 003 2
rms error		0.000 002 5		0.000 001 1

$\times 10^{-6}$ a.u.. For the R-B3LYP case, the maximum error in the numerical gradients is 3.2×10^{-6} a.u., and the root-mean-square error is 1.1×10^{-6} a.u.

Table II shows the CPCM/MCSCF (Ref. 27) gradients for the CHCOO⁻ radical (Fig. 3) in its first excited state, which is singlet ($S=0$). The geometry was optimized in the gas phase with MCSCF/6-31G* method. In the MCSCF calculation, 28 electrons in 14 orbitals are frozen, and 2 elec-

TABLE II. Gradients (a.u.) calculated with CPCM/MCSCF/6-31G* for CHCOO⁻ radical in its first excited state (singlet) at the geometry optimized in the gas phase with the MCSCF/6-31G* method.

Coordinates	Analytic	Numerical
C1X	0.003 631 6	0.003 632 0
C1Y	0.000 113 5	0.000 102 6
C1Z	−0.003 764 1	−0.003 765 1
C2X	0.004 891 8	0.004 892 8
C2Y	−0.001 298 7	−0.001 289 8
C2Z	0.015 969 1	0.015 970 4
O3X	−0.002 034 8	−0.002 036 0
O3Y	0.001 448 5	0.001 448 0
O3Z	−0.008 472 0	−0.008 472 0
O4X	−0.002 882 6	−0.002 882 4
O4Y	−0.000 197 9	−0.000 200 2
O4Z	−0.005 604 8	−0.005 605 2
H5X	−0.003 606 1	−0.003 606 7
H5Y	−0.000 065 4	−0.000 058 8
H5Z	0.001 871 8	0.001 872 6
Max error		0.000 010 9
rms error		0.000 004 1

trons are in 2 active orbitals. Although there are four possible determinants, this is actually a three-determinant MCSCF calculation due to the symmetry constraint (singlet). The agreement between the analytic and numerical gradients is also perfect: the maximum error in the numerical gradients is 10.9×10^{-6} a.u., and the root-mean-square error is 4.1×10^{-6} a.u.

Table III lists the CPCM/UHF, CPCM/ROHF, CPCM/U-B3LYP, and CPCM/RO-B3LYP gradients for the CHCOO⁻ radical (Fig. 3) in its ground state, which is triplet ($S=1$). The geometry was optimized in the gas phase with UHF/6-31G* method. Again, the agreement between the analytic and numerical gradients is perfect. The maximum errors in the numerical gradients vary from 2.2×10^{-6} to 5.8×10^{-6} a.u., and the root-mean-square errors vary from 0.8×10^{-6} to 2.1×10^{-6} a.u..

These data evidently show that using FIXPVA, very accurate analytic gradients can be obtained for both CPCM/HF and CPCM/DFT methods. The accuracy of the analytic gradients must be at the same level ($\sim 10^{-6}$ a.u.) or better. The errors in the analytic gradients are mainly caused by the errors in solving the SCF equations and the induced surface charges, and can be, in principle, systematically reduced by tightening the convergence criteria for the SCF and CPCM iterations. It is worth noting that *ab initio* or DFT calculations use various approximations or cutoffs in basis set integral evaluation, SCF convergence, and especially derivative integral evaluation, so default settings in GAMESS produce the gas phase gradients good to about 1×10^{-6} a.u.. It is expected that the FIXPVA-CPCM HF and DFT gradients are accurate to the same limitations as the gas phase code, and so, in this sense, the exact gradients have been obtained for FIXPVA-CPCM HF and DFT methods.

B. Surface area and induced surface charge

If the same set of atomic radii is used for a molecule with no additional spheres, the tesserae in the boundary region of two intercepting spheres are downscaled or zeroed by switch functions in FIXPVA, leading to a much smaller total surface area as compared to GEPOL and GEPOL-AS. A concern is that the induced surface charge in the FIXPVA scheme may significantly deviate from the expected theoretical value, which is $-(\varepsilon-1)/\varepsilon$ of the net molecular charge if all the electron density is included in the cavity.

Table IV lists the total number and total area of the tesserae generated for acetate (structure optimized in the gas phase with RHF/6-31G*) by GEPOL, GEPOL-AS, and FIXPVA schemes. The actual surface area of the molecular cavity is 97.83 Å² as calculated by GEPOL, and the corresponding FIXPVA value is 88.77 Å², approximately 10%

FIG. 3. (Color online) Structures of acetate and two CHCOO⁻ radicals (singlet and triplet).

TABLE III. Gradients (a.u.) calculated for CHCOO⁻ radical in its ground state (triplet) at the geometry optimized in the gas phase with the UHF/6-31G* method.

Coordinates	CPCM/UHF/6-31G*		CPCM/ROHF/6-31G*		CPCM/U-B3LYP/6-31G*		CPCM/RO-B3LYP/6-31G*	
	Analytic	Numerical	Analytic	Numerical	Analytic	Numerical	Analytic	Numerical
C1X	0.000 186 8	0.000 188 7	−0.003 260 6	−0.003 261 7	−0.007 369 3	−0.007 368 7	−0.008 238 9	−0.008 238 9
C1Y	0.000 658 6	0.000 660 2	0.000 614 9	0.000 614 7	0.001 302 5	0.001 302 4	0.001 284 0	0.001 283 9
C1Z	−0.007 143 3	−0.007 145 5	−0.004 461 5	−0.004 455 7	−0.012 007 2	−0.012 007 8	−0.011 212 7	−0.011 213 6
C2X	0.005 537 7	0.005 541 8	0.007 873 3	0.007 873 9	−0.000 601 3	−0.000 601 5	0.000 397 5	0.000 398 6
C2Y	−0.002 206 3	−0.002 206 5	−0.002 192 5	−0.002 192 5	−0.003 493 0	−0.003 492 2	0.003 467 1	−0.003 466 3
C2Z	0.024 248 8	0.024 250 3	0.022 948 7	0.022 948 8	0.049 283 6	0.049 281 4	0.048 313 3	0.048 312 3
O3X	−0.003 036 6	−0.003 040 1	−0.002 511 1	−0.002 512 2	−0.036 847 7	−0.036 849 3	−0.036 759 0	−0.036 758 7
O3Y	0.001 665 0	0.001 664 9	0.001 605 1	0.001 605 0	0.002 920 5	0.002 920 1	0.002 900 1	0.002 899 8
O3Z	−0.009 419 9	−0.009 420 4	−0.008 409 8	−0.008 409 6	−0.019 145 1	−0.019 145 4	−0.018 693 8	−0.018 693 7
O4X	−0.000 735 8	−0.000 736 4	−0.001 052 4	−0.001 052 0	0.034 774 0	0.034 774 0	0.034 374 9	0.034 373 6
O4Y	−0.000 079 4	−0.000 079 4	−0.000 049 5	−0.000 049 4	−0.000 199 5	−0.000 199 9	−0.000 197 1	−0.000 197 6
O4Z	−0.008 170 8	−0.008 173 7	−0.008 778 4	−0.008 778 8	−0.023 232 0	−0.023 231 5	−0.023 175 7	−0.023 174 7
H5X	−0.001 952 1	−0.001 952 5	−0.001 049 2	−0.001 051 1	0.010 044 3	0.010 043 5	0.010 225 6	0.010 225 7
H5Y	−0.000 037 9	−0.000 039 8	0.000 022 0	0.000 021 9	−0.000 530 5	−0.000 530 3	−0.000 520 0	−0.000 520 1
H5Z	0.000 485 1	0.000 488 4	−0.001 298 9	−0.001 301 4	0.005 100 7	0.005 101 1	0.004 768 9	0.004 772 0
Max error		0.000 004 1		0.000 005 8		0.000 002 2		0.000 003 1
rms error		0.000 002 1		0.000 001 8		0.000 000 8		0.000 001 0

less. The total number of tesserae in GEPOL and GEPOL-AS is 164, but is 114 in the FIXPVA case due to the exclusion of the boundary tesserae that have their center points lying in neighboring spheres. In GEPOL and GEPOL-AS, all boundary tesserae are kept and assigned with new center coordinates and areas.

Although the total area in the FIXPVA case is ∼10% less, the total induced surface charge is almost the same as those in the GEPOL and GEPOL-AS cases. If all the electron density is contained in the cavity, the theoretical total induced surface charge should be $(\varepsilon-1)/\varepsilon = 77.39/78.39 = 0.987\,243e$ for acetate anion. Due to the incomplete inclusion of the electron density, the theoretical total induced surface charge should be slightly less than $0.987\,243e$ in this case. The values obtained in the GEPOL and GEPOL-AS cases are ∼$0.976e$, ∼1% less than $0.987\,243e$; the values in the FIXPVA cases are ∼$0.970e$, 1.7% less than 0.987 243. Therefore, all these induced surface charges are close to the expected theoretical value.

For acetate, the solvation energy is around −68 kcal/mol, a typical value for univalent ions of that size. The value obtained with the FIXPVA scheme is 1.1 kcal/mol less negative than the GEPOL values (Table IV). For acetic acid, the solvation energy is around −10 kcal/mol, and the FIXPVA

value is 0.3 kcal/mol less negative than the GEPOL values. For many purposes, these differences are insignificant, and the same set of radii can be interchangeably used for CPCM regardless which tessellation scheme, FIXPVA or GEPOL, is used.

Finally, it is noted that FIXPVA is a different tessellation scheme from GEPOL, and it is not anticipated that the same set of radii parametrized for GEPOL is perfect for FIXPVA. If a better agreement to experiments is desired, a new set of radii shall be parameterized for FIXPVA-CPCM as done for other tessellation schemes.

C. Smooth potential energy surface

The energy profiles of NaCl dissociation in aqueous solution are computed by CPCM/RHF/6-31G* using GEPOL, GEPOL-AS, and FIXPVA schemes with a 0.1 Å increment in the Na–Cl distance (Fig. 4). Radii of 1.8 and 2.76 Å are used for Na⁺ and Cl⁻, respectively, to define the solute surface. It is noted that these calculations are not intended to reproduce any experimental or simulated dissociation curve, but merely to examine the smoothness of the potential energy surfaces in the CPCM.

Clearly, the GEPOL curve fluctuates at the Na–Cl dis-

TABLE IV. The number of tesserae (NTS), surface area, total surface charges, and molecular energies (E) computed for acetate in water.

Methods	NTS	Area (Å)	CPCM/RHF/6-31G*		CPCM/B3LYP/6-31G*	
			Surface charge (e)	E (a.u.)	Surface charge (e)	E (a.u.)
GEPOL	164	97.83	0.976 489	−227.333 067	0.975 904	−228.481 667
GEPOL-AS	164	97.30	0.976 504	−227.333 059	0.975 912	−228.481 654
FIXPVA	114	88.77	0.970 468	−227.331 347	0.969 694	−228.479 868
The gas phase			0	−227.225 068	0	−228.377 000

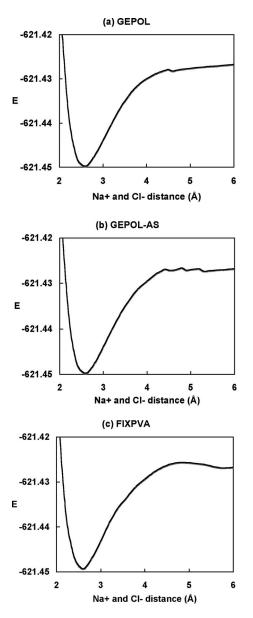FIG. 4. Energy profiles calculated with CPCM/RHF/6-31G* for NaCl dissociation.

tance of 4.56 Å, which is the sum of R_{Na} and R_{Cl}, corresponding to the sudden separation of the Na and Cl spheres [Fig. 4(a)]. Due to the area scaling, the GEPOL-AS curve start to fluctuate at a shorter distance, ∼4.2 Å, and continues to fluctuate until the distance is over 5.3 Å [Fig. 4(b)]. Although in general GEPOL-AS performs better than GEPOL, for this particular case, it is not better.

As expected, the FIXPVA curve is smooth over the entire range [Fig. 4(c)]. The smoothness of the FIXPVA-CPCM potential energy surfaces depends on the parameters m_1, m_2, n_1, and n_2 for the switching functions [see Eqs. (31) and (32)]. Apparently, these parameters have been given reasonable values and the smoothness is satisfactory.

There is a noticeable barrier of 0.0121 hartree (or 0.76 kcal/mol) in the FIXPVA curve [Fig. 4(c)]. As the Na$^+$ and Cl$^-$ ions approach each other, the total energy increases from a minimum of −621.426 89 hartree at 5.8 Å to a maximum of −621.425 68 hartree at 4.9 Å. The scaling would start at

6.06 Å (from 1.8+2.76+1.5) if a tessera happens to be on the line P_2P_3 in Fig. 1. An inspection on the change of the cavity surface area clearly shows that this barrier is a result of the scaling of a tessera starting at 5.8 Å (which means the tessera is not on the line P_2P_3 in Fig. 1]. Although the GEPOL-AS curve does not show a barrier, its level is noticeably raised in the 4.3–5.3 Å region, corresponding to the area scaling. No barrier is found in the GEPOL curve.

Why such a barrier exists in the FIXPVA calculation can be readily explained by the fact that when the solvent molecules start to be removed between the ions, the ions become desolvated so the solvation energy of the system becomes less negative, i.e., the beginning of the scaling of the tesserae reduces the screening of the ions by the solvent and reduces the solvation energy. This reduction in the solvation energy will likely produce a barrier on the energy curves such as in the NaCl case. Whether such a barrier exists in the real physical process can hardly be answered by merely using continuum solvation models; molecular dynamics or Monte Carlo simulations are likely to provide more insights into this interesting issue. For example, Smith and Dang's[28] simulation clearly indicates that the Potential of Mean Force, which is defined as the free energy of a pair of ions as the function of distance, has a ∼1.6 kcal/mol peak at ∼3.7 Å for Na$^+$Cl$^-$. Therefore, the FIXPVA scheme is not physically wrong in predicting such a barrier. However, as mentioned at the beginning of this section, these calculations are not intended to reproduce any experimental or simulation results. To do so, a careful parameterization of the switching functions and probably the atomic radii as functions of distance, is required, and is beyond the scope of this paper.

D. Rotational variance

Due to the asymmetric tessellation in the current implementation of the GEPOL, GEPOL-AS, and FIXPVA scheme, the CPCM/HF and CPCM/DFT energies are not rotationally invariant: molecules may have a preference in their orientations to minimize the energy.

Acetate in 20 random orientations was used to test the rotational variances of the GEPOL, GEPOL-AS, and FIXPVA scheme. When 60 initial tesserae per sphere were used it is found that the GEPOL and GEPOL-AS schemes exhibit small rotational variance (both are 0.1 kcal/mol) while the FIXPVA exhibits a relatively large rotational variance (0.9 kcal/mol). When 240 initial tesserae per sphere were used the FIXPVA rotational variance reduces to 0.2 kcal/mol, which is small. Continuum solvation models rely on parameterization of the atomic radii and the errors in the calculated solvation energy for ions are usually larger than 1 kcal/mol. Therefore, 60 tesserae per sphere is still a good option for its efficiency. In future studies, it is possible to devise a better tessellation scheme that can generate smooth potential energy surfaces while remain rotationally invariant.

E. Geometry optimization

Due to the rotational variance, CPCM geometry optimizations should, in principle, be performed in Cartesian coordinates so the molecules are free to rotate in order to find the
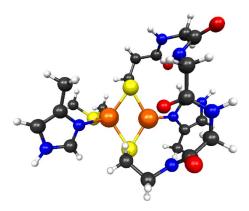
FIG. 5. (Color online) A 74-atom model of the Cu_A center in Cu_A azurin extracted from the PDB file 1CC3. The formula of the model is $Cu_2S_3O_4N_8C_{21}H_{36}$. It uses 702 Gaussian basis functions in the CPCM/U-B3LYP/6-31G* calculation.

true minima. However, it is well known that geometry optimization performed in Cartesian coordinates is much slower than that in internal coordinates. CPCM geometry optimizations performed in Cartesian coordinates are even slower due to the flatter potential energy surfaces. Tests show that even for very small molecules such as acetic acid, FIXPVA-CPCM geometry optimization in Cartesian coordinates cannot converge in hundreds of steps, and is impractical. In general, internal coordinates are recommended. Using internal coordinates, the orientation of a molecule will be largely fixed, and the internal geometry of the molecule is optimized to minimize the energy.
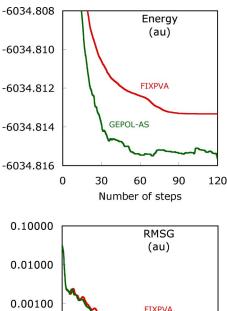
A 74-atom model molecule (Fig. 5) of the Cu_A center of the Cu_A azurin was extracted from the x-ray structure 1CC3 in PDB,[29] and then optimized using the CPCM/U-B3LYP/6-31G* method. This model molecule has a net charge of $+1e$, and a spin multiplicity=2. For comparison, both the GEPOL-AS (Ref. 18) and FIXPVA schemes were used. The automatically generated delocalized internal coordinates were used in the geometry optimization.[26] The energy and root-mean-square gradient (RMSG) profiles in the optimization processes are plotted in Fig. 6.

The FIXPVA optimization went smoothly: the energy monotonically decreases and the default convergence criteria (10^{-4} a.u. for maximum gradient and 0.333×10^{-4} a.u. for the RMSG) are met at the 110th step. The GEPOL-AS energy curve is not smooth, and the same convergence criteria cannot be met in 124 steps. Compared to GEPOL-AS, the FIXPVA optimized energy is ~1.2 kcal/mol higher, which is expected due to its reduced surface area.

Other tests for large and small molecules using various wave functions show that FIXPVA-CPCM geometry optimizations can always be performed robustly. The largest molecule tested so far contains 145 atoms and 1286 Gaussian-type basis functions.

V. CONCLUSION

A new molecular surface tessellation method, FIXPVAs are described in this paper for COSMO and its variant CPCM. The main features of FIXPVA are that the positions of the surface tesserae are fixed relative to their center atoms,
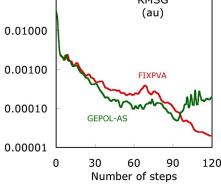

FIG. 6. (Color online) The energy and rms-gradient profiles in CPCM/U-B3LYP/6-31G* geometry optimizations of the 74-atom Cu_A model extracted from 1CC3 (Fig. 5).

and the tessera areas are smooth functions of their distances to neighboring spheres. For each tessera, two switching functions are used for each neighboring sphere. The analytic derivatives of the tessera positions and areas with respect to atomic coordinates can be obtained and used for evaluating the nuclear gradients. Numerical tests show that (1) the accuracy of the FIXPVA gradients is ~10^{-6} a.u., the same as that of the gas phase HF and DFT calculations, (2) the potential energy surfaces obtained with FIXPVA are smooth, and (3) FIXPVA geometry optimization processes are stable and convergent.

ACKNOWLEDGMENTS

[1] J. Tomasi, B. Mennucci, and R. Cammi, Chem. Rev. (Washington, D.C.) **105**, 2999 (2005).
[2] S. Miertus, E. Scrocco, and J. Tomasi, Chem. Phys. **55**, 117 (1981).
[3] E. Cances, B. Mennucci, and J. Tomasi, J. Chem. Phys. **107**, 3032 (1997).
[4] A. Klamt and G. Schuurmann, J. Chem. Soc., Perkin Trans. 2 **1993**, 799.
[5] T. N. Truong and E. V. Stefanovich, Chem. Phys. Lett. **240**, 253 (1995).
[6] V. Barone and M. Cossi, J. Phys. Chem. A **102**, 1995 (1998).
[7] D. M. Chipman, Theor. Chem. Acc. **107**, 80 (2002); D. M. Chipman and M. Dupuis, ibid. **107**, 90 (2002).
[8] R. Cammi and J. Tomasi, J. Chem. Phys. **101**, 3888 (1994).
[9] V. Barone, M. Cossi, and J. Tomasi, J. Comput. Chem. **19**, 404 (1998).
[10] E. Cances and B. Mennucci, J. Chem. Phys. **109**, 249 (1998).

[11] E. Cances, B. Mennucci, and J. Tomasi, J. Chem. Phys. **109**, 260 (1998); M. Cossi and V. Barone, *ibid.* **109**, 6246 (1998); B. Mennucci, R. Cammi, and J. Tomasi, *ibid.* **110**, 6858 (1999).

[12] T. N. Truong and E. V. Stefanovich, J. Chem. Phys. **103**, 3709 (1995).

[13] M. Cossi, N. Rega, G. Scalmani, and V. Barone, J. Comput. Chem. **24**, 669 (2003).

[14] D. M. York and M. Karplus, J. Phys. Chem. A **103**, 11060 (1999); J. Khandogin, B. A. Gregersen, W. Thiel, and D. M. York, J. Phys. Chem. B **109**, 9799 (2005).

[15] H. M. Senn, P. M. Margl, R. Schmid, T. Ziegler, and P. E. Blochl, J. Chem. Phys. **118**, 1089 (2003).

[16] J. L. Pascualahuir, E. Silla, and I. Tunon, J. Comput. Chem. **15**, 1127 (1994); E. Silla, I. Tunon, and J. L. Pascualahuir, *ibid.* **12**, 1077 (1991); E. Silla, F. Villar, O. Nilsson, J. L. Pascualahuir, and O. Tapia, J. Mol. Graphics **8**, 168 (1990); J. L. Pascualahuir and E. Silla, J. Comput. Chem. **11**, 1047 (1990).

[17] C. S. Pomelli, J. Comput. Chem. **25**, 1532 (2004).

[18] H. Li and J. H. Jensen, J. Comput. Chem. **25**, 1449 (2004).

[19] M. Cossi, G. Scalmani, N. Rega, and V. Barone, J. Chem. Phys. **117**, 43 (2002).

[20] P. Pulay, Mol. Phys. **17**, 197 (1969).

[21] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, J. Chem. Phys. **76**, 637 (1982).

[22] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, J. Comput. Chem. **14**, 1347 (1993); M. S. Gordon and M. W. Schmidt, in *Theory and Applications of Computational Chemistry*, edited by C. E. Dykstra, G. Frenking, K. S. Kim, and G. E. Scuseria (Elsevier, New York, 2005).

[23] J. Tomasi, B. Mennucci, and E. Cances, J. Mol. Struct.: THEOCHEM **464**, 211 (1999).

[24] H. Li, C. S. Pomelli, and J. H. Jensen, Theor. Chem. Acc. **109**, 71 (2003).

[25] C. S. Pomelli, J. Tomasi, and V. Barone, Theor. Chem. Acc. **105**, 446 (2001).

[26] J. Baker, A. Kessi, and B. Delley, J. Chem. Phys. **105**, 192 (1996).

[27] J. Ivanic and K. Ruedenberg, Theor. Chem. Acc. **106**, 339 (2001).

[28] D. E. Smith and L. X. Dang, J. Chem. Phys. **100**, 3757 (1994).

[29] H. Robinson, M. C. Ang, Y. G. Gao, M. T. Hay, Y. Lu, and A. H. J. Wang, Biochemistry **38**, 5677 (1999).