



Symmetric versus asymmetric discretization of the integral equations in polarizable continuum solvation models

Adrian W. Lange, John M. Herbert*

Department of Chemistry, The Ohio State University, Columbus, OH 43210, United States

ARTICLE INFO

Article history:

Received 20 November 2010

In final form 28 April 2011

Available online 1 May 2011

ABSTRACT

Discretization of the integral equations that define widely-used ‘polarizable continuum’ solvation models fails to preserve certain properties of the integral operators. Consequently, the appropriate form of the finite-dimensional matrix equations is ambiguous, with two different asymmetric versions and also a symmetrized version as obvious possibilities. We demonstrate cases where solvation energies differ by as much as 24 kcal/mol amongst these variants. These differences are sometimes exacerbated by new discretization procedures that guarantee smooth potential energy surfaces. Formal and numerical arguments favor one particular formulation of the matrix equations.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Within the quantum chemistry community, the most widely-used class of dielectric continuum solvation models are ‘apparent surface charge’ (ASC) methods, better known as polarizable continuum models (PCMs) [1–3]. Given a definition for what constitutes the boundary between the solute and the dielectric continuum (*i.e.*, the ‘solute cavity’), these models afford a prescription for calculating an ‘apparent’ charge density on the cavity surface, whose electrostatic interaction with the solute approximates the solute/continuum electrostatic interaction. Because this interaction is represented using a two-dimensional surface charge density, ASC methods are far more computationally efficient than numerical solution of Poisson’s equation, which requires a three-dimensional volume integration.

The development of PCM theory originated thirty years ago with the work of Tomasi and co-workers [4], and has since been revised and elaborated manyfold into its modern form; see Refs. [1,3] for a history of this development. Nowadays, numerous levels of approximation are available within the PCM formalism, including the conductor-like screening model [5], which is typically known within the PCM framework as C-PCM [6] or GCOSMO [7], as well as the more sophisticated ‘integral equation formulation’ (IEF-PCM) [8–10]. The latter is formally equivalent [11], at the level of integral equations, to the ‘surface and simulation of volume polarization for electrostatics’ [SS(V)PE] method [12]. The IEF-PCM/SS(V)PE approach affords an exact solution for the solute continuum electrostatic interaction that arises from whatever part of the solute’s charge density exists within the cavity, and furthermore provides an approximate solution for the ‘volume

polarization’ that arises from any ‘escaped charge’, *i.e.*, that part of the solute’s charge density that penetrates beyond the cavity, into the dielectric medium.

The IEF-PCM/SS(V)PE approach, which is the focus of the present work, is derived in terms of certain integral operators that act on the surface charge and surface electric potential [8,12–14]. Numerical solution of the integral equations that define this model requires discretization of the cavity surface, resulting in finite-dimensional matrix equations. The matrix forms of the integral operators, however, fail to preserve certain properties of the corresponding integral operators. Consequently, the proper definition of the matrix equations is ambiguous, because the integral equations can be written in various equivalent forms that yield *inequivalent* matrix equations.

This issue has been noted in the past, but only very modest numerical analysis and discussion have been devoted to it. Two previous studies [15,16] report that solvation energies differ by no more than 0.01 kcal/mol amongst the variants of IEF-PCM/SS(V)PE that are considered here, although the data sets in these studies consisted of no more than four small molecules. (Lipparini et al. [17] also assert that different forms of the matrix equations yield solvation energies that differ by only ~ 1 kcal/mol, but their numerical tests are unpublished.) Other properties, including solute dipole moments, molecular orbital eigenvalues, and total surface charge are also reported to differ negligibly amongst these variants [15,18]. In light of these results, the importance of the particular form of the matrix equations has been largely ignored in the literature. As a result, no consensus actually exists in regard to which version of the matrix equations is best.

In contrast to previous studies, we present a thorough examination of various matrix formulations of IEF-PCM/SS(V)PE, from both formal and numerical points of view. We find cases in which these variants afford substantial differences in electrostatic solvation

* Corresponding author.

E-mail address: herbert@chemistry.ohio-state.edu (J.M. Herbert).

energies and total surface charge. This work also constitutes the first assessment of these differences for recently-developed ‘smooth PCMs’ [14,19], in which switching functions are used to guarantee that the solute’s potential energy surface is smooth. Our results indicate that these smooth PCMs are more sensitive to the particular form of the matrix equations than are previous PCM implementations, and our analysis favors one particular choice of these equations.

2. Theory

2.1. ASC PCM equations

The working equations for a variety of different ASC PCMs can be written in the general form [12,2,16,14]

$$\widehat{\mathcal{K}}\sigma(\vec{s}) = \widehat{\mathcal{R}}\phi_0(\vec{s}), \quad (1)$$

where $\sigma(\vec{s})$ is the ASC density and $\phi_0(\vec{s})$ is the solute’s electrostatic potential at the cavity surface. Different PCMs correspond to different choices for the integral operators $\widehat{\mathcal{K}}$ and $\widehat{\mathcal{R}}$. For the IEF-PCM/SS(V)PE model, which is the focus of the present work,

$$\widehat{\mathcal{K}} = \left(\widehat{\mathcal{F}} - \frac{f_\varepsilon}{2\pi} \widehat{\mathcal{D}} \right) \widehat{\mathcal{F}} \quad (2)$$

and

$$\widehat{\mathcal{R}} = -f_\varepsilon \widehat{\mathcal{Y}}, \quad (3)$$

where

$$f_\varepsilon = \frac{\varepsilon - 1}{\varepsilon + 1} \quad (4)$$

and

$$\widehat{\mathcal{Y}} = \widehat{\mathcal{F}} - \frac{1}{2\pi} \widehat{\mathcal{D}}. \quad (5)$$

In these equations, ε denotes the dielectric constant that characterizes the continuum and $\widehat{\mathcal{I}}$ is the identity operator. Definitions of the surface-Coulomb operator, $\widehat{\mathcal{F}}$, and surface-dipole operator, $\widehat{\mathcal{D}}$, can be found in the literature [8–10,12,14]. At present, it suffices to note that

$$\widehat{\mathcal{D}}\widehat{\mathcal{F}} = \widehat{\mathcal{F}}\widehat{\mathcal{D}}^\dagger. \quad (6)$$

Violation of this exact condition, within standard discretization schemes, leads to the ambiguity that is addressed here. This issue does not arise for the conductor-like models (COSMO, GCOSMO, and C-PCM), for which the operator $\widehat{\mathcal{D}}$ is absent in the definitions of $\widehat{\mathcal{K}}$ and $\widehat{\mathcal{R}}$ [12,14].

Upon discretization, the cavity surface is replaced by a finite set of surface grid points, $\{\vec{s}_i\}$, with surface areas $\{a_i\}$. Eq. (1) is thereby transformed into a finite-dimensional matrix equation,

$$\mathbf{K}\mathbf{q} = \mathbf{R}\mathbf{v}. \quad (7)$$

The electrostatic potential vector, \mathbf{v} , has elements $v_i = \phi_0(\vec{s}_i)$, and the vector \mathbf{q} consists of point charges q_i . If we define a diagonal matrix \mathbf{A} , with $A_{ij} = a_i\delta_{ij}$, then the discretized forms of $\widehat{\mathcal{K}}$ and $\widehat{\mathcal{R}}$ are

$$\mathbf{K} = \left(\mathbf{I} - \frac{f_\varepsilon}{2\pi} \mathbf{D}\mathbf{A} \right) \mathbf{S} \quad (8)$$

and

$$\mathbf{R} = -f_\varepsilon \mathbf{Y} = -f_\varepsilon \left(\mathbf{I} - \frac{1}{2\pi} \mathbf{D}\mathbf{A} \right).$$

Appropriate definitions for the matrix elements D_{ij} and S_{ij} have been discussed extensively in the literature [3,9,10,14,16,20,21]. (In particular, several choices have been suggested for the diagonal

elements, the definition of which is not entirely straightforward owing to singularities in certain integrands, when evaluated over a single surface element.) We use the definitions suggested in our previous work [14], which are appropriate for use with the smooth discretization schemes introduced below. We will revisit the topic of diagonal matrix elements in Section 2.2.

In general, discretization preserves the symmetry of $\widehat{\mathcal{F}}$, i.e., $\mathbf{S}^\dagger = \mathbf{S}$, and if the solute cavity consists of a single sphere then $\mathbf{D}^\dagger = \mathbf{D}$ as well [18,12]. However, realistic solute cavities are non-spherical, and for non-spherical cavities discretization fails to preserve Eq. (6). In other words,

$$\mathbf{D}\mathbf{A}\mathbf{S} \neq \mathbf{S}\mathbf{A}\mathbf{D}^\dagger$$

for realistic cavity shapes. As a consequence of this inequality, the step in going from Eq. (2) to Eq. (8) is not entirely justified, because we could just as well have written $\widehat{\mathcal{K}}$ in a form that involves $\widehat{\mathcal{F}}\widehat{\mathcal{D}}^\dagger$ rather than $\widehat{\mathcal{D}}\widehat{\mathcal{F}}$. Had we done so, then the factor of $\mathbf{D}\mathbf{A}\mathbf{S}$ in Eq. (8) would be replaced by $\mathbf{S}\mathbf{A}\mathbf{D}^\dagger$ upon discretization. In fact, one could use Eq. (6) to justify replacing $\mathbf{D}\mathbf{A}\mathbf{S}$ in Eq. (8) with any linear combination $c_1\mathbf{D}\mathbf{A}\mathbf{S} + c_2\mathbf{S}\mathbf{A}\mathbf{D}^\dagger$ such that $c_1 + c_2 = 1$.

In view of this ambiguity, let us replace the definition of \mathbf{K} in Eq. (8) with the more general form

$$\mathbf{K} = \mathbf{S} - \frac{f_\varepsilon}{2\pi} \mathbf{X}. \quad (11)$$

Eq. (8) is recovered if $\mathbf{X} = \mathbf{D}\mathbf{A}\mathbf{S}$. Taking $\mathbf{X} = \mathbf{S}\mathbf{A}\mathbf{D}^\dagger$ is also justified, as is the symmetric choice $\mathbf{X} = (\mathbf{D}\mathbf{A}\mathbf{S} + \mathbf{S}\mathbf{A}\mathbf{D}^\dagger)/2$. Other linear combinations are possible but are not considered here. Chipman uses the symmetric form consistently [2,12,22], in a method that he calls SS(V)PE; this form facilitates somewhat more efficient solution of Eq. (7). On the other hand, IEF-PCM calculations using all three of the aforementioned variants have been reported [15,17,18], along with other asymmetric variants that are not considered here [8,9].

Whatever the definition of \mathbf{K} , ASC PCM calculations consist of solving Eq. (7) for the surface charge vector, \mathbf{q} , given the solute’s electrostatic potential. The surface charges can then be used to evaluate the polarization energy [14],

$$E_{pol} = \frac{1}{2} \mathbf{q}^\dagger \mathbf{v} = \frac{1}{2} \mathbf{v}^\dagger \mathbf{Q}\mathbf{v}, \quad (12)$$

which is the electrostatic contribution to the solvation energy. In Eq. (12), we have introduced the *solvent response matrix*,

$$\mathbf{Q} = \mathbf{K}^{-1} \mathbf{R}. \quad (13)$$

The total energy is

$$W = E_0 + E_{pol}, \quad (14)$$

where E_0 represents the internal energy of the solute, computed in the field of the ASC. When the solute is described using quantum mechanics, the solute’s wavefunction must be converged in the presence of the surface charges, so the system of linear equations in Eq. (7) must be solved at each self-consistent field cycle.

2.2. Smooth discretization

We have recently developed a general discretization procedure (applicable to C-PCM, IEF-PCM, and related PCMs) that we call the ‘switching/Gaussian’ (SWIG) procedure [21,14]. This approach rectifies a long-standing problem with ASC PCMs, namely, that straightforward pointwise discretization or tessellation (as in the widely-used GEPOL algorithm [23–25]) often leads to discontinuities in the solute’s potential energy surface. These may arise due to tessellation algorithms that fail to treat nuclear perturbations in a symmetric fashion [26,27], or else due to the appearance or disappearance of grid points as the solute geometry changes [14,19,21,27–29], which is equivalent to a discontinuous change

in cavity surface area and in the dimension of the system of equations in Eq. (7). Discontinuities of the latter type can hinder the convergence of geometry optimizations (or prevent them from converging at all), and can lead to other artifacts including spurious vibrational frequencies, non-variational solvation energies, and catastrophic failure to conserve energy in molecular dynamics simulations [14,21].

A variety of ‘smooth’ discretization methods have been proposed in recent years, in attempt to alleviate these problems [14,19–21,26–28,30]. We have documented numerical problems with several of these smooth discretization procedures, and shown that these problems are absent in the SWIG approach and a related ‘ISWIG’ procedure that is described below [14,21]. A detailed description of these PCM implementations, including formulas for the matrix elements D_{ij} and S_{ij} , can be found in Ref. [14]. The essential aspects are summarized here.

The SWIG discretization algorithm has two key features: a switching function to attenuate the point \vec{s}_i ’s contribution to Eq. (7), as \vec{s}_i passes through a buffer region surrounding the cavity surface; and Gaussian blurring of the surface charges q_i , in order to eliminate numerical problems associated with the singular Coulomb potential. Such problems are exacerbated when a switching function is introduced to ensure continuity, because this function allows surface grid points to approach one another more closely than would be possible if all interior grid points were simply discarded. The close approach of point charges can lead to unwanted fluctuations in the solvation energy or its gradient, as a function of the nuclear coordinates [14,21].

This sort of numerical instability is more often associated with the \mathbf{D} matrix rather than with the \mathbf{S} matrix, because $D_{ij} \propto r_{ij}^{-3}$ whereas $S_{ij} \propto r_{ij}^{-1}$, where $r_{ij} = |\vec{s}_i - \vec{s}_j|$ [14]. Thus, D_{ij} fluctuates more rapidly than S_{ij} when r_{ij} is small. As such, IEF-PCM and related methods that utilize \mathbf{D} may be more susceptible to numerical instabilities than is C-PCM, where \mathbf{D} is absent. This fact was previously noted in Ref. [8], where it was suggested that a modification of the GEPOL tessellation algorithm is required in the context of IEF-PCM. Alternatively, we will show that instabilities can be avoided using Gaussian blurring of the point charges, as this technique modifies the \mathbf{D} matrix elements such that $D_{ij} \rightarrow 0$ as $r_{ij} \rightarrow 0$ [14].

To avoid numerical instabilities and unwanted energy fluctuations, the SWIG algorithm uses the charges q_i as amplitudes for spherical Gaussian functions [14,28],

$$g_i(\vec{r}) = q_i (\zeta_i^2 / \pi)^{3/2} \exp(-\zeta_i^2 |\vec{r} - \vec{s}_i|^2). \quad (15)$$

The functions g_i are centered at the surface grid points, \vec{s}_i , and we use atom-centered Lebedev grids [31] to generate these points. Lebedev grids are already widely used in quantum chemistry codes, and this form of discretization does not increase the complexity of analytic gradient expressions, in the way that surface tessellation schemes sometimes do [32]. The Gaussian exponents, ζ_i , are taken from Ref. [33], where they were optimized to reproduce solvation energies for the Born ion model. The use of surface Gaussians instead of point charges ensures that the representation of $\sigma(\vec{s})$ is continuous even though a discrete set of grid points is used.

The switching function is the second key feature of SWIG. In Ref. [14], we presented two different switching functions, the first of which was originally proposed by York and Karplus [28], who developed a smooth version of COSMO. We use the term ‘SWIG’ to refer to this particular switching function, used in conjunction with Gaussian blurring. An alternative switching function was introduced in Ref. [14], which uses both the radii of the atomic spheres and the Gaussian exponents to determine how rapidly a surface point is attenuated. Consistent with the terminology

introduced in Ref. [14], we call this the ‘improved SWIG’ (ISWIG) method. Both SWIG and ISWIG discretization are examined here.

2.3. Alternative discretization schemes

We also examine several other discretization schemes that will aid in understanding the origin of certain discrepancies amongst the \mathbf{K} -matrix variants. A simple point-charge (PC) discretization is obtained by placing point charges q_i at the Lebedev grid points and discarding any grid points that reside within the cavity. Alternatively, we might modify these point charges using Eq. (15), and we refer to this discretization procedure as ‘Gaussian blurring only’ (GBO). Neither the PC nor the GBO procedure employs a switching function, hence both approaches are apt to produce discontinuities in the potential surface. As such, these procedures are *not* recommended for chemical applications, but they do shed light on the role of the switching function and the Gaussian blurring procedure.

In contrast to the Lebedev grids that we prefer, most contemporary PCM calculations utilize the GEPOL tessellation procedure [23–25]. To ascertain whether any of the anomalies documented here originate in the use of Lebedev grids, we will also report ‘PC-GEPOL’ and ‘GBO-GEPOL’ calculations, where GEPOL grid points and surface areas are used in place of Lebedev grid points and quadrature weights. Although the GEPOL algorithm has the ability to add extra spheres to approximate solvent-excluded surfaces, we do not utilize these extra spheres, so that the GEPOL calculations should be directly comparable to calculations using Lebedev grids.

Another discretization scheme that we shall examine is the fixed-point, variable area (FIXPVA) algorithm [19], which uses a switching function in conjunction with a point-charge representation of $\sigma(\vec{s})$. Unlike the SWIG and ISWIG approach to attenuating grid points, the FIXPVA procedure does not allow grid points to penetrate into the interior of the cavity. Instead, FIXPVA uses the switching function to scale the individual surface areas a_i . An unfortunate side effect of this approach is that it sometimes completely scales away important surface elements, thus significantly underestimating the cavity surface area [19,34,21] and—even worse—leaving holes in the cavity surface [21]. This matter is analyzed in more detail herein.

Finally, we investigate several different choices for defining the diagonal elements of the \mathbf{D} matrix. Two definitions of D_{ii} are common in the PCM literature. [3,27]. One of these, originally derived by Purisima and Nilar [35], amounts to a sum rule for each row of \mathbf{D} :

$$D_{ii} = -\frac{1}{a_i} \left(2\pi + \sum_{j \neq i} D_{ij} a_j \right). \quad (16)$$

The other common definition, from Mennucci et al. [9], is based on an exact relationship for spherical cavities. For a surface grid point, \vec{s}_i , residing on a spherical cavity surface of radius R_i , one can show that

$$D_{ii} = -\frac{S_{ii}}{2R_i}, \quad (17)$$

which is then taken to define D_{ii} for arbitrary cavity shapes. Previous comparisons of symmetric and asymmetric \mathbf{K} matrices [15,16,18] have used Eq. (16) exclusively, but in the context of SWIG/ISWIG discretization, we have shown that the use of Eq. (16) often violates the negative-definiteness condition for \mathbf{Q} . As such, our previous SWIG/ISWIG calculations have used Eq. (17) to define D_{ii} [14,36]. This definition preserves the negative-definiteness of \mathbf{Q} [14].

Here, we will show that the choice of Eq. (16) versus Eq. (17) affects the extent to which solvation energies for $\mathbf{X} = \mathbf{DAS}$ differ

from those predicted when $\mathbf{X} = \mathbf{SAD}^\dagger$. For consistency with our previous work [14], the SWIG/ISWIG notation will imply the use of Eq. (17); alternatively, we will refer to the *substitution* of Eq. (16), in place of Eq. (17), as ‘subSWIG/subISWIG’ discretization. The PC, GBO, and FIXPVA calculations use Eq. (17), except where it is explicitly stated that the sum rule in Eq. (16) is used instead.

3. Computational details

Previous assessments of symmetric versus asymmetric \mathbf{K} matrices have been limited to data sets consisting of H_2O , $\text{C}_2\text{H}_5\text{OH}$, $(\text{CH}_3)_2\text{NH}$, $\text{CH}_3\text{C}(\text{O})\text{NH}_2$, NO^+ , and CN^- [15,18,16]. Here, we take the 20 amino acids that constitute the universal genetic code as a representative test set of molecules, since the side chains in these molecules exhibit a variety of chemical properties (polar versus non-polar, charged versus neutral, aliphatic versus aromatic, etc.). The geometries of the amino acids were generated, in their zwitterionic forms, using the TINKER program (version 4.2), and were not further optimized. IEF-PCM calculations were carried out in water, $\epsilon = 78.39$.

The solute molecules in our calculations are described using either the Hartree–Fock/6-31+G* method, or else using the AMBER99 force field [37]. In the former case, the solute cavity is constructed from a union of atom-centered spheres, using van der Waals radii taken from Bondi [38] (except for hydrogen [39]). These radii are scaled by a factor of 1.2 for use in cavity construction [1]. For AMBER99 solutes, we construct a solvent-accessible surface [40] by adding a solvent probe radius of 1.4 Å to the unscaled Lennard-Jones radius of each atom. All atomic radii and nuclear coordinates are provided in the Supplementary data.

Calculations were performed using a locally-modified version of Q-CHEM [41], in which we have recently implemented hybrid quantum mechanics/molecular mechanics methods [42] and also PCM methods [14,21]. In addition to SWIG and ISWIG discretization [14], we have implemented the FIXPVA discretization algorithm as described in Ref. [19], except that our version uses Lebedev grids to discretize the atomic spheres, rather than the GEPOL algorithm. (We also use somewhat different atomic radii than those used in Ref. [19].) We present a comparison of our implementation of FIXPVA to the implementation in the most recent release of GAMESS [43] (version 1 Oct., 2010) for C-PCM calculations in Supplementary data.

For comparisons of Lebedev grids versus GEPOL grids in IEF-PCM calculations, we generated GEPOL grids with GAMESS [43] and read them into Q-Chem. In order to use GBO on the GEPOL grids, we needed to determine an optimal value of the parameter ζ that controls the width of the Gaussian surface charges, via [14,20,21,28]

$$\zeta_i = \frac{\zeta}{\sqrt{w_i R_i}}. \quad (18)$$

Here, w_i is the quadrature weight of the i th grid point and R_i is the radius of the sphere on which the i th point resides. Scalmani and Frisch [20] assert that optimal parameters for GEPOL grids lie in the range of $4.5 \leq \zeta \leq 4.8$, but they do not explicitly report their values. By numerical optimization, we arrived at the values of $\zeta = 4.7477177485$ and $\zeta = 4.7396506415$ for GEPOL grids of 60 and 240 points per atom, respectively. In C-PCM calculations, these values reproduce Born ion energies within $\sim 10^{-6}$ kcal/mol across a variety of dielectric constants, ion charges, and cavity radii. Note that these ζ values fall within the range specified by Scalmani and Frisch.

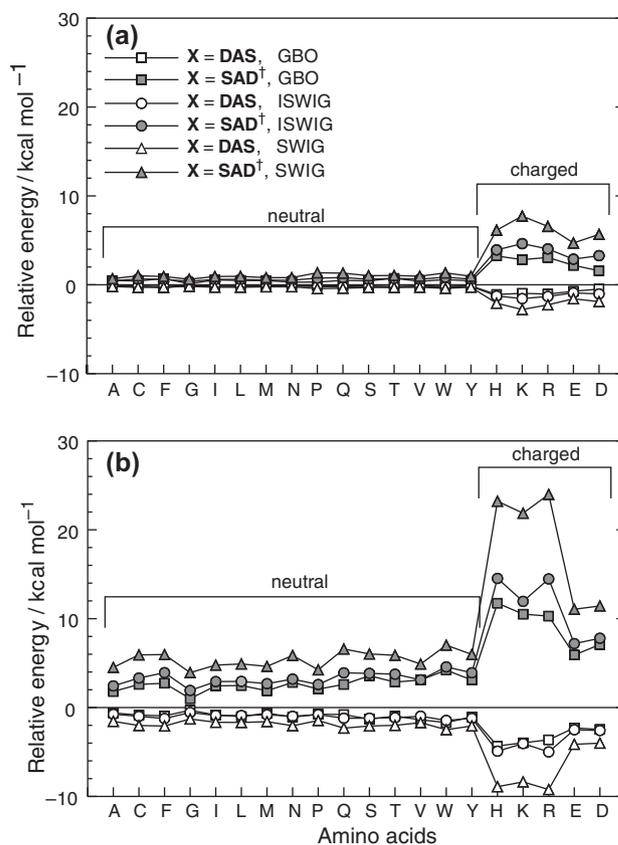


Figure 1. Relative energies of the amino acids in water, obtained at (a) the AMBER99 level and (b) the HF/6-31+G* level. The cavity surface is discretized using 590 Lebedev points per atomic sphere, and solution-phase energies obtained using $\mathbf{X} = \mathbf{DAS}$ or $\mathbf{X} = \mathbf{SAD}^\dagger$ in Eq. (11) are reported relative to the energy obtained using a symmetrized form of \mathbf{X} . The GBO discretization uses Gaussian blurring only, whereas SWIG and ISWIG use a switching function in conjunction with Gaussian blurring.

4. Results

4.1. Solvation energies with Lebedev grids

Figure 1 shows the differences in the total energy, W , among the three different forms of \mathbf{K} that are considered here. We report the energies obtained using the asymmetric $\mathbf{X} = \mathbf{SAD}^\dagger$ and $\mathbf{X} = \mathbf{DAS}$ forms of \mathbf{K} , relative to those obtained using the symmetric form, for the GBO, SWIG, and ISWIG discretization schemes. In these calculations, the solute molecules are described using either the AMBER99 force field (Figure 1a) or else at the Hartree–Fock (HF) level (Figure 1b), and $N = 590$ Lebedev grid points are used per atomic sphere.

For the neutral amino acids described at the AMBER99 level, the energy differences amongst the various forms of \mathbf{K} are $\lesssim 1$ kcal/mol, which is smaller than the intrinsic error in solvation energies computed at the IEF-PCM level [44,45]. At the HF/6-31+G* level, however, the discrepancies for the neutral amino acids approach 7 kcal/mol in several cases. For the charged amino acids, which exhibit the largest polarization energies, the discrepancies between different forms of \mathbf{K} approach 8 kcal/mol for AMBER99 solutes and 24 kcal/mol for HF solutes. The effects seen here for neutral versus charged solutes are further examined in Section 5.2.

The SWIG and ISWIG methods, which are compared in Figure 1, differ only in the choice of switching function. Compared to SWIG, ISWIG does not allow the surface charges to penetrate as deeply into the interior of the cavity [14], and we observe that ISWIG

affords much smaller (although still significant) differences between various forms of \mathbf{K} . The GBO method, wherein Gaussian blurring is used but where grid points cannot penetrate into the cavity at all, affords slightly smaller variations than are observed using ISWIG. This suggests a trend in which differences amongst the \mathbf{K} matrices are magnified as the extent to which the grid points can penetrate into the cavity becomes larger. These results appear to favor ISWIG over SWIG, as the former affords energies in reasonably good agreement with GBO results. Nevertheless, substantial variations among the different \mathbf{K} matrices still exist for GBO discretization, which demonstrates that the switching function is not entirely to blame for these variations.

Consistent with results reported in Ref. [2], the non-symmetric form of \mathbf{K} with $\mathbf{X} = \mathbf{DAS}$ tends to agree best with the symmetric form. However, previous comparisons of these alternate \mathbf{K} matrices [2,18,15], using a point-charge discretization and no switching function, have found that the differences in solvation energies amongst them are extremely small ($\lesssim 0.01$ kcal/mol [18]), which is clearly not the case here. One possible reason for the larger differences observed here could be due to the more complicated topography of the solute cavities in our calculations. (Recall that $\mathbf{DAS} = \mathbf{SAD}^\dagger$ for a spherical cavity.) Previous numerical comparisons of different \mathbf{K} matrices have employed either spherical cavities [2] or else ‘united atom’ cavities [46] that, due to the fairly small size of the molecules that were considered, consist of only a few spheres [15,18]. The definition of D_{ij} also contributes to the lack of asymmetry amongst the \mathbf{K} -matrix variants, as we shall show in Section 4.5.

The results in Figure 1 demonstrate that the switching function alone cannot explain large energy differences among the different forms of \mathbf{K} . Therefore, it is interesting to compare results obtained using FIXPVA discretization, where a switching function is used but Gaussian blurring is not. Using FIXPVA, we find that the various choices for \mathbf{K} are in excellent agreement with one another, using either Eq. (16) or Eq. (17) to define D_{ij} . Variations in solvation energies amongst the two asymmetric forms of \mathbf{K} are no larger than 0.2 kcal/mol for AMBER99 solutes and 0.5 kcal/mol for HF solutes, even for the charged amino acids. However, the fact that the dependence on \mathbf{K} is small does not necessarily imply that solvation energies or surface charge distributions are accurate.

For AMBER99 solutes, where there is no escaped charge, one can compute the *exact* electrostatic solvation energy (for a given solute cavity) by numerical solution of Poisson’s equation, which we accomplish using the adaptive Poisson–Boltzmann solver (APBS) [47]. We have carefully converged the APBS results with respect to the three-dimensional integration grid¹. However, because the energy computed with APBS (and also the PCMs considered in this work) is only the *electrostatic* part of the solvation energy, one should not expect these results to reproduce experimental solvation energies, which include non-electrostatic contributions as well. The more important consideration, for the purpose of this work, is whether the APBS and PCM results agree with one another. We find that converged APBS energies lie within 0.13 kcal/mol of the energy that is obtained using IEF-PCM with $\mathbf{X} = \mathbf{DAS}$ and GBO discretization, for each of the amino acids. (See the [Supplementary data](#).) In the absence of escaped charge, exact IEF-PCM calculations *should* afford an exact solution for the electrostatic solvation energy, E_{pol} , so these results serve to demonstrate that Gaussian blurring does not change this fact, at least not for $\mathbf{X} = \mathbf{DAS}$. As such, we take GBO

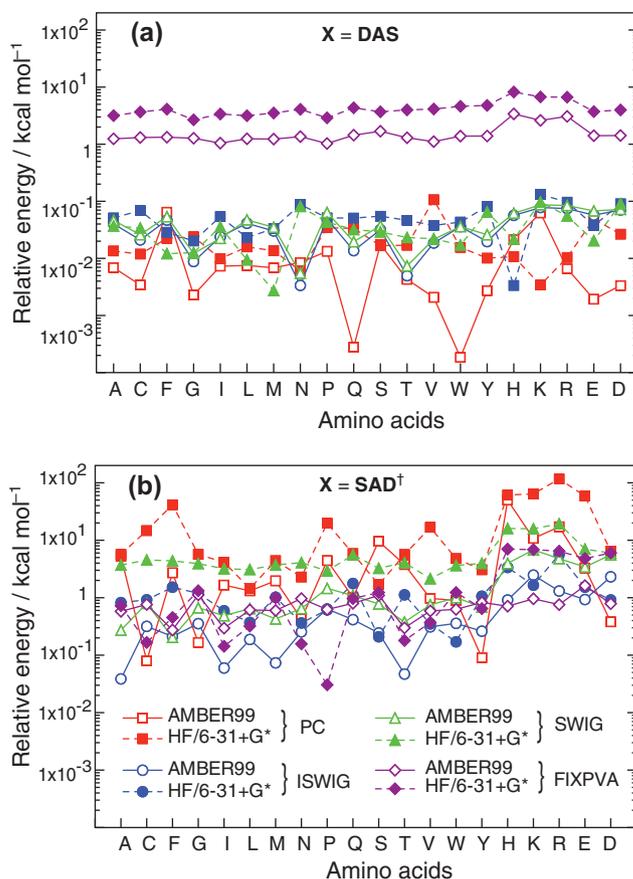


Figure 2. Energies of the amino acids in water, relative to results obtained using GBO discretization, for (a) the $\mathbf{X} = \mathbf{DAS}$ and (b) the $\mathbf{X} = \mathbf{SAD}^\dagger$ form of \mathbf{K} . Cavity surfaces were discretized using 590 Lebedev points per atomic sphere.

discretization as the benchmark result, for comparison to other discretization schemes.

Figure 2 compares these GBO benchmarks to energies computed using alternative discretization schemes. For the $\mathbf{X} = \mathbf{DAS}$ form of \mathbf{K} (Figure 2a), the PC, SWIG, and ISWIG procedures all agree with GBO results to within ~ 0.1 kcal/mol for AMBER99 solutes and to within ~ 1 kcal/mol for HF solutes. FIXPVA results are the outliers and differ by > 1 kcal/mol, both for AMBER99 and HF solutes.

For the $\mathbf{X} = \mathbf{SAD}^\dagger$ variant of \mathbf{K} (Figure 2b), there is far more scatter amongst the data, with ISWIG and FIXPVA in best agreement with the GBO results. (Note, however, that for $\mathbf{X} = \mathbf{SAD}^\dagger$ the GBO results differ by as much as 4.5 kcal/mol from converged APBS energies, as shown in the [Supplementary data](#).) The SWIG method is in poor agreement with GBO results for the charged amino acids, and the PC method is extremely irregular.

All together, ISWIG exhibits the best agreement with GBO results. These results seem to favor the use of $\mathbf{X} = \mathbf{DAS}$, for which the ISWIG, SWIG, PC, and GBO results are all in excellent agreement. The data also suggest that FIXPVA does not provide solvation energies in good agreement with other discretization methods, despite the fact that this approach is largely free of differences amongst the different choices for \mathbf{K} .

4.2. Convergence with respect to the surface grid

In a previous study [14], we found that $N = 110$ affords negligible violations of Gauss’ Law, and solvation energies that are converged within ~ 0.1 kcal/mol of the $N \rightarrow \infty$ limit, yet discrepancies among the various forms of \mathbf{K} persist even at $N = 590$. In fact, these discrepancies increase with N : the

¹ The APBS grid parameters that we employ reproduce Born ion energies to within ~ 0.2 kcal/mol for several different solvent dielectric constants and ion charges. In addition, our grid parameters are comparable to those used published APBS convergence tests [48–50]. These studies conclude that the accuracy of APBS, as compared to analytical or other highly-accurate models, systematically increases with increasing grid density.

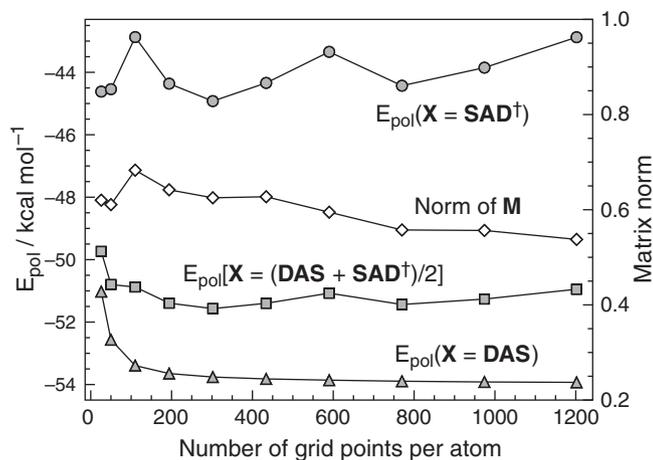


Figure 3. Convergence of the polarization energy (axis at left), as a function of the number of Lebedev grid points per atomic sphere, for histidine described at the AMBER99 level with SWIG discretization. Results for three alternative \mathbf{K} matrices are shown, along with the norm of the matrix $\mathbf{M} = \mathbf{DAS} - \mathbf{SAD}^\dagger$ (axis at right).

maximum variation observed at the HF/6-31+G* level is 15 kcal/mol for $N = 50$, 16 kcal/mol for $N = 110$, and 24 kcal/mol for $N = 590$. (Results for $N = 50$ and $N = 110$ can be found in the Supplementary data.)

In Figure 3, we test the convergence of the discretization by computing E_{pol} as a function of N . For this test, we use the histidine molecule described at the AMBER99 level, since histidine exhibits some of the largest discrepancies amongst the various choices for \mathbf{K} . For the $\mathbf{X} = \mathbf{DAS}$ form of \mathbf{K} , E_{pol} converges smoothly (and fairly rapidly) as a function of N , but the same cannot be said for $\mathbf{X} = \mathbf{SAD}^\dagger$, where E_{pol} does not appear to have converged even for $N = 1202$. In addition, the norm of the matrix $\mathbf{M} = \mathbf{DAS} - \mathbf{SAD}^\dagger$ is not significantly different at $N = 1202$ than it is at $N = 50$. So long as $\|\mathbf{M}\|$ remains large, the various forms of \mathbf{K} will continue to afford significantly different results.

Results for the symmetric form of \mathbf{K} are also shown in Figure 3. This is the form used by Chipman in SS(V)PE calculations [12,2,22], and we note that it inherits the oscillations in E_{pol} that arise from $\mathbf{X} = \mathbf{SAD}^\dagger$, although these oscillations are somewhat damped by the smooth convergence of the $\mathbf{X} = \mathbf{DAS}$ form.

4.3. Cavity surface area

We next examine the total cavity surface area predicted by each method. The total surface area is a well-defined quantity in PCMs, and it plays an important role in the ASC PCMs that are considered here, because individual surface areas appear in definitions of the \mathbf{S} and \mathbf{D} matrices [3,27]. The PC and GBO surface area is defined by the Lebedev weights $\{w_i\}$ and the atomic radii $\{R_i\}$,

$$\sum_i a_i = \sum_I \sum_{i \in I} R_I^2 w_i, \quad (19)$$

and is exact in the limit $N \rightarrow \infty$. For smooth PCMs, w_i in Eq. (19) is replaced by $w_i F_i$, where F_i is the switching function [21,14].

In Table 1, we report errors in the total surface area, relative to PC results, for each discretization method. The PC areas are comparable to GEPOL areas, having <1% error. Both SWIG and ISWIG also exhibit errors of <1%, but FIXPVA exhibits alarmingly large errors, and severely underestimates the surface area. The fact that FIXPVA underestimates the surface area has been noted previously [19,34,21], but the errors that we observe are significantly larger

Table 1

Errors in the total cavity surface area, for the amino acid data set using $N = 590$ Lebedev points per atom (SWIG, ISWIG, and FIXPVA) as well as using $N = 60$ GEPOL points per atom. The mean signed error (MSE), root mean square error (RMSE), and maximum signed error (Max) are listed, in percent, taking PC results with Lebedev grids as the benchmark.

Method	% error, AMBER99 solutes		
	MSE	RMSE	Max
GEPOL	0.1	0.3	0.7
SWIG	-0.8	0.4	-1.5
ISWIG	-0.3	0.3	-1.0
FIXPVA	-20.4	2.4	-24.5
Method	% error, HF/6-31+G* solutes		
	MSE	RMSE	Max
GEPOL	0.0	0.4	0.7
SWIG	-0.2	0.3	-0.7
ISWIG	-0.1	0.3	-0.6
FIXPVA	-40.0	4.2	-47.6

than in previous studies. An explanation for this observation is provided in Section 5.1.

4.4. Induced surface charge and Gauss' Law

Gauss' Law provides a diagnostic for gauging the accuracy of the discretization. For a cavity with unit dielectric inside, and dielectric constant ϵ outside, Gauss' Law states that the total induced charge on the cavity surface, \mathcal{Q}_{surf} , is proportional to the total charge contained within the cavity, \mathcal{Q}_{in} , according to

$$\mathcal{Q}_{surf} = -\left(\frac{\epsilon - 1}{\epsilon}\right)\mathcal{Q}_{in}. \quad (20)$$

For solutes described by a force field, \mathcal{Q}_{in} is simply the overall solute charge. This is not the case for HF solutes, owing to the presence of escaped charge, but we can still apply Eq. (20) easily if we take \mathcal{Q}_{in} to be the total nuclear charge. In this case, \mathcal{Q}_{surf} is that part of the ASC that is induced by the electrostatic potential arising from the nuclei.

Deviations from Eq. (20) for the aqueous amino acids are summarized in Table 2. Previously [14], we showed that $N = 590$ Lebedev points per atom is sufficient to reduce the Gauss' Law error below $0.001e$ in C-PCM calculations on this same set of molecules. The results presented here show that this same level of accuracy is achievable in IEF-PCM calculations, provided that one chooses $\mathbf{X} = \mathbf{DAS}$. Although the $\mathbf{X} = \mathbf{SAD}^\dagger$ variant is only slightly inferior for AMBER99 solutes, for HF solutes it affords errors as large as $55e$ in the nuclear part of \mathcal{Q}_{surf} .

Despite these tremendous errors in \mathcal{Q}_{surf} for $\mathbf{X} = \mathbf{SAD}^\dagger$, this method does not appear to produce any individual charges q_i that are anomalously large, at least not for any of the Gaussian-blurred discretization schemes considered here. In all cases, $|q_i| \lesssim 0.1e$. (We have previously shown that artificially large surface charges can sometimes appear in the absence of Gaussian blurring [21].) On the other hand, Table 2 shows that the surface charge induced with FIXPVA discretization seems to be less sensitive to the \mathbf{X} variants, yet it is less accurate, as compared to the other discretization methods, for $\mathbf{X} = \mathbf{DAS}$.

4.5. Solvation energies with GEPOL grids

All of the results presented above use Lebedev quadrature grids, but we have verified that large discrepancies between the $\mathbf{X} = \mathbf{DAS}$ and $\mathbf{X} = \mathbf{SAD}^\dagger$ forms of \mathbf{K} persist when GEPOL grids are used instead. Restricting our attention to HF/6-31+G* solutes (as AMBER99 results are quite similar), Figure 4 shows PC-GEPOL and GBO-GE-

Table 2

Gauss' Law error statistics for IEF-PCM. Listed are the mean absolute error (MAE), the root mean square error (RMSE), and the maximum error (Max), evaluated over the amino acid data set, with $\epsilon = 78.39$ and $N = 590$. All values are given in atomic charge units.

Method	Total charge errors (AMBER99)					
	$\mathbf{X} = \text{DAS}$			$\mathbf{X} = \text{SAD}^\dagger$		
	MAE	RMSE	Max	MAE	RMSE	Max
GBO	7.0×10^{-5}	6.5×10^{-5}	2.1×10^{-4}	2.9×10^{-2}	3.5×10^{-2}	1.1×10^{-1}
SWIG	2.9×10^{-4}	3.7×10^{-4}	1.3×10^{-3}	6.6×10^{-2}	8.3×10^{-2}	2.6×10^{-1}
ISWIG	2.0×10^{-4}	2.6×10^{-4}	9.1×10^{-4}	4.0×10^{-2}	5.0×10^{-2}	1.6×10^{-1}
FIXPVA	1.5×10^{-2}	7.6×10^{-3}	3.4×10^{-2}	1.8×10^{-2}	9.7×10^{-3}	4.1×10^{-2}
	Nuclear charge errors (HF/6-31+G*)					
GBO	7.9×10^{-3}	5.6×10^{-3}	2.2×10^{-2}	1.6×10^1	6.4×10^0	3.2×10^1
SWIG	1.8×10^{-2}	3.7×10^{-3}	2.7×10^{-2}	3.0×10^1	1.1×10^1	5.5×10^1
ISWIG	1.3×10^{-2}	3.1×10^{-3}	2.1×10^{-2}	1.8×10^1	6.9×10^0	3.3×10^1
FIXPVA	9.6×10^{-1}	2.4×10^{-1}	1.4×10^0	1.6×10^0	4.4×10^{-1}	2.4×10^0

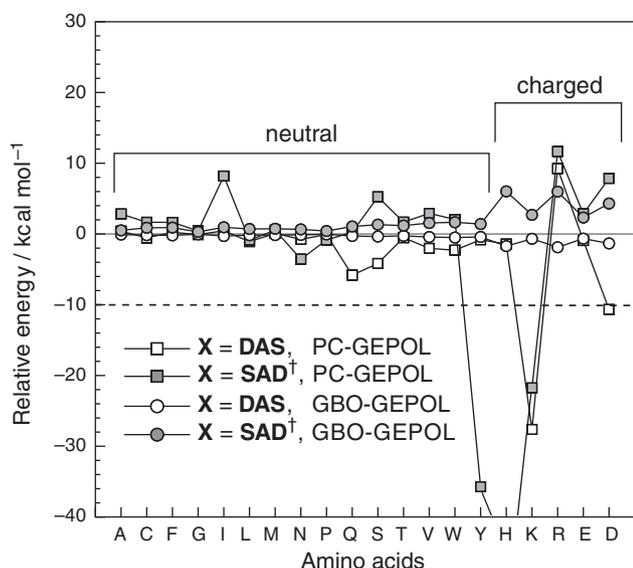


Figure 4. Relative energies of the amino acids in water, obtained at the HF/6-31+G* level. The cavity surface is discretized using 60 GEPOL points per atomic sphere, and solution-phase energies obtained using $\mathbf{X} = \text{DAS}$ or $\mathbf{X} = \text{SAD}^\dagger$ in Eq. (11) are reported relative to the energy obtained using a symmetrized form of \mathbf{X} . For comparison, the dashed line at -10 kcal/mol indicates the lower bound in Figure 1. For histidine ('H'), the PC-GEPOL data point for $\mathbf{X} = \text{SAD}^\dagger$ appears at -2544 kcal/mol, far outside the range of this figure.

POL solvation energies for the amino acid data set. As in the case of Lebedev grids, large variations are observed between the two asymmetric forms of \mathbf{K} , especially in the case of PC-GEPOL discretization. Given that Gaussian blurring greatly reduces the magnitude of these variations, they are likely caused at the GC-GEPOL level by point charges in close proximity.

These results show that both Lebedev and GEPOL grids are susceptible to \mathbf{K} -matrix asymmetries to similar extents. For GBO discretization, the differences observed in Figure 4 (GEPOL grids with 60 points per sphere) are not quite as large as those in Figure 1 (Lebedev grids with 590 points per sphere), which is due partly to the difference in grid densities. As shown in Supplementary data, differences in predicted solvation energies between different forms of \mathbf{K} tend to increase with increasing grid density. This trend is born out by calculations using GEPOL grids with 240 points per atom, which show even larger variations than those documented in Figure 4. Clearly, these variations do not originate in the use of Lebedev grids.

4.6. Definition of D_{ii}

Previous investigations of discrepancies between different forms of \mathbf{K} have exclusively used the sum rule in Eq. (16) to define D_{ii} , whereas all of the calculations presented above employ Eq. (17). In Table 3, we report results of PC-GEPOL and GBO-GEPOL calculations using the sum rule of Eq. (16) to define D_{ii} . Remarkably, the variations among the different forms of \mathbf{K} disappear, and we obtain results that are consistent with previous studies finding only small discrepancies between different forms of \mathbf{K} [15,16]. It is especially surprising that the sum rule manages to fix the wild disagreement for the PC-GEPOL calculations (cf. Figure 4).

In light of these results, we are compelled to see if the asymmetry in SWIG disappears in subSWIG. Table 3 summarizes solvation energies computed using the subSWIG, subISWIG, GBO, and PC discretization schemes, when the sum rule in Eq. (16) is used to define D_{ii} , rather than Eq. (17). Discrepancies between the different forms of \mathbf{K} disappear almost completely for the GBO and PC calculations reported in Table 3, whereas some differences remain in the subSWIG and subISWIG results, probably due to the presence of grid points inside of the cavity. Nevertheless, these differences are greatly reduced as compared to those reported in Figure 1, where Eq. (17) was used to define D_{ii} .

5. Analysis

5.1. FIXPVA

In Section 4.3 we noted that the FIXPVA algorithm significantly underestimates cavity surface areas. Wang and Li [34] suggest an *ad hoc* re-parameterization of the FIXPVA switching function, for the purpose of surface area calculations. Here, we show that the problem is intrinsic to the FIXPVA switching procedure itself.

The FIXPVA algorithm ensures a smooth potential surface for the solute by means of two different switching functions, f_1 and f_2 . Consider two intersecting spheres, A and B, with centers at points P_A and P_B as depicted in Figure 5a. Then, for a discretization point P_1 on the surface of sphere A, $f_1 = f_1(d_{12})$ where d_{12} is the distance from P_1 to the point P_2 , which is defined by the intersection of spheres A and B and the (P_1, P_A, P_B) plane. The other switching function is $f_2 = f_2(d_{13})$, where d_{13} represents the distance from P_1 to the point P_3 where sphere B intersects the line that connects P_A and P_B . As d_{12} and/or d_{13} becomes small, the product $f_1(d_{12}) f_2(d_{13})$ is used to scale down the area associated with the grid point P_1 .

The FIXPVA method appears to have been designed with an eye toward intersections of the type depicted in Figure 5a, where P_3 lies between P_A and P_B . Other types of intersections are possible,

Table 3
Statistics for relative energies of the amino acids in water, obtained at the HF/6-31+G* level, using the sum rule of Eq. (16) to define D_i . The mean absolute difference (MAD), root mean square difference (RMSD), and maximum absolute difference (Max) in the solvation energies (in kcal/mol) are tabulated for the two asymmetric \mathbf{X} matrices, relative to the symmetric version.

Method	GEPOL, 60 points per sphere					
	$\mathbf{X} = \mathbf{DAS}$			$\mathbf{X} = \mathbf{SAD}^j$		
	MAD	RMSD	Max	MAD	RMSD	Max
PC-GEPOL	3.3×10^{-2}	4.8×10^{-2}	2.2×10^{-1}	1.3×10^{-1}	1.2×10^{-1}	4.8×10^{-1}
GBO-GEPOL	1.3×10^{-2}	9.4×10^{-3}	3.3×10^{-2}	4.6×10^{-2}	2.3×10^{-2}	1.2×10^{-1}
	Lebedev, 590 points per sphere					
PC	1.8×10^{-2}	1.3×10^{-2}	4.3×10^{-2}	5.0×10^{-2}	1.1×10^{-1}	4.3×10^{-2}
GBO	1.8×10^{-2}	1.3×10^{-2}	4.4×10^{-2}	2.1×10^{-2}	1.4×10^{-2}	5.3×10^{-2}
subSWIG	5.2×10^{-1}	1.6×10^0	7.3×10^0	7.8×10^{-1}	1.7×10^0	6.2×10^0
subISWIG	1.5×10^{-1}	1.1×10^{-1}	4.8×10^{-1}	1.9×10^{-1}	3.5×10^{-1}	1.5×10^0

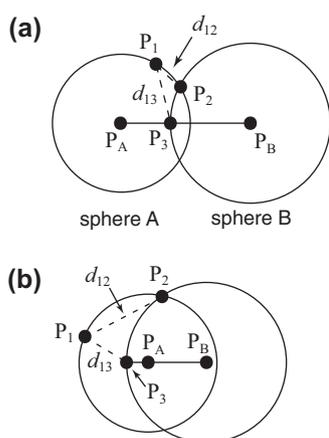


Figure 5. Geometrical definitions used to define the FIXPVA switching functions.

however, including the one depicted in Figure 5b. In this case, the distance d_{AB} between P_A and P_B is smaller than the radius of sphere B, a scenario that occurs readily if sphere A represents a hydrogen atom. Whereas in Figure 5a, it seems reasonable that P_1 might be somewhat attenuated, since this point would soon pass into the interior of the cavity if d_{AB} were to decrease, in Figure 5(b) the point P_1 clearly lies on the exterior of the cavity, and is necessary for the description of sphere A. Because d_{12} is small, however, this point (and every other point on the surface of sphere A) may nevertheless be substantially attenuated by the switching function f_1 .

However, a close examination of the FIXPVA code as implemented in GAMESS [43] reveals that the situation depicted in Figure 5b is avoided by arbitrarily setting $f_1 = 1$ when $d_{AB} < R_A$. This aspect of FIXPVA is not discussed in Ref. [19], and as a result, our previously-reported implementation of FIXPVA [21,14] lacked this feature; we have since modified our implementation to reflect this undisclosed feature of the algorithm. (In the Supplementary data, we demonstrate that energies, surface areas, and induced surface charges obtained with our current implementation of FIXPVA reproduce those obtained using GAMESS, when identical GEPOL grids are used, and furthermore differ negligibly between Lebedev and GEPOL grids of similar grid density.)

This undisclosed aspect of the FIXPVA algorithm is in fact a significant flaw that limits its applicability, because it introduces a discontinuity in the switching function f_1 whenever R_A is less than the parameter $n_2 = 1.5 \text{ \AA}$ from Ref. [19]. This parameter controls the upper bound of the switching region for f_1 . Referring to Figure 5, the discontinuity occurs for atom A (with $R_A < n_2$) at the point when P_A and P_3 coincide because $n_1 < d_{13} < n_2$ (i.e. it is within

the switching region for f_1). Once P_A is inside sphere B, however, f_1 will be immediately set to unity. An example of this discontinuity, using the GAMESS implementation of FIXPVA, is provided in the Supplementary data. This discontinuity is reproducible in our own implementation of FIXPVA.

In Ref. [19], where the FIXPVA algorithm was introduced, this artifact was avoided by setting the hydrogen atom radii to zero and setting all other radii to be larger than n_2 . More generally, one could re-parameterize the FIXPVA switching functions in an effort to avoid this problem, but n_2 will always set a lower bound on the atomic radii that can be safely used in FIXPVA calculations. As such, the FIXPVA approach is potentially useful in the context of united-atom cavities [46], but is less useful for all-atom approaches or for any sort of cavity with radii below n_2 . Our FIXPVA calculations with the HF solutes use hydrogen radii of 1.32 \AA and are therefore subject to the f_1 discontinuity, which may contribute to the observed errors in energy, surface area, and induced surface charge.

Regardless of the value of the FIXPVA parameters n_1 and n_2 , the underestimation of cavity surface areas that is documented in Table 1 results primarily from the fact that the FIXPVA algorithm scales away the areas associated with grid points that lie near the seams of intersecting spheres, in order to prevent close approach of surface charges. This scaling, however, results in ‘holes’ in the cavity surface [21], at places where the spheres intersect. The lack of grid points to gather surface charge in these regions leads to large violations of Gauss’ Law (Table 2), and suggests that the FIXPVA solvation energies should not be taken seriously, despite their apparent lack of dependence upon the form of \mathbf{K} .

5.2. Dependence of the solvation energy on the electrostatic potential

That the neutral amino acids afford better agreement amongst the various alternative \mathbf{K} matrices, as compared to the charged amino acids, is ultimately a consequence of the quadratic dependence of E_{pol} on the electrostatic potential; see Eq. (12). Suppose that we were to scale the electrostatic potential by a factor λ , at a fixed cavity geometry and therefore a fixed response matrix, \mathbf{Q} . Then we could write

$$E_{pol}(\lambda) = \frac{1}{2} \lambda^2 \mathbf{v}^t \mathbf{Q} \mathbf{v}. \quad (21)$$

Suppose next that we compute E_{pol} for the two asymmetric \mathbf{K} matrices, using the same solute and cavity geometry in each case. Then the difference between the two solvation energies would be

$$\Delta E_{pol}(\lambda) = \frac{1}{2} \lambda^2 \mathbf{v}^t (\mathbf{Q}^{\mathbf{DAS}} - \mathbf{Q}^{\mathbf{SAD}^j}) \mathbf{v}. \quad (22)$$

As such, any difference between the \mathbf{Q} matrices is magnified by λ^2 . For AMBER99 solutes, one can verify numerically that ΔE_{pol} increases by a factor of λ^2 if the atomic point charges are scaled by λ .

For HF/6-31+G* solutes, it is not so straightforward to scale \mathbf{v} by a factor, so we instead consider a sequence of ionized histidine solutes, His $^{n+}$ ($n = 0, 1, 2, 3$). We find that the energy difference $\Delta W = W_{\text{DAS}} - W_{\text{SAD}^\dagger}$ is fit very well by a quadratic function of n . (See the [Supplementary data](#).) For neutral histidine, $\Delta W = 6.5$ kcal/mol using SWIG discretization, which is not too much different from the discrepancies observed for neutral amino acids using the same PCM parameters (*cf.* Figure 1b), but ΔW increases rapidly as the solute charge increases, with $\Delta W = 191.7$ kcal/mol obtained for His $^{3+}$.

In view of these results, one should expect molecules that elicit a smaller electrostatic potential to afford better agreement among the various forms of \mathbf{K} . This explains the relatively large discrepancies that are observed for the charged amino acids.

5.3. Dependence of the surface charge on the electrostatic potential

As noted in Section 4.4, the $\mathbf{X} = \text{SAD}^\dagger$ form of \mathbf{K} exhibits large violations of Gauss' Law when Gaussian blurring is used. Since the induced surface charges depend linearly on the surface electrostatic potential, $\mathbf{q} = \mathbf{Q}\mathbf{v}$, one might expect larger deviations from Gauss' Law as the magnitude of \mathbf{v} increases for a fixed cavity geometry.

We investigate this trend by means of HF/6-31+G* calculations on a series of homonuclear diatomic molecules. The solute cavity consists of just two spheres, which represents the simplest non-trivial case (since $\text{DAS} = \text{SAD}^\dagger$ for a spherical cavity [18,12]). We fix the internuclear distance at 1.0 Å and use a radius of 1.4 Å for both atomic spheres, with $N = 590$ Lebedev points per sphere. As such, the surface grid (and therefore the matrix \mathbf{Q}) is held fixed as the atomic number, Z , is varied. The magnitude of the nuclear contribution to \mathbf{v} increases linearly with Z , hence the nuclear contribution to $\mathcal{Q}_{\text{surf}}$ should increase linearly with Z as well. The nuclear charge is entirely contained within the cavity, so Eq. (20) should hold for this part of $\mathcal{Q}_{\text{surf}}$.

Deviations from Gauss' Law, as a function of Z , are shown in Figure 6. For the nuclear contribution to $\mathcal{Q}_{\text{surf}}$, numerical errors in Gauss' Law depend linearly on Z (*i.e.*, on the magnitude of the electrostatic potential), for both asymmetric variants of \mathbf{K} , which

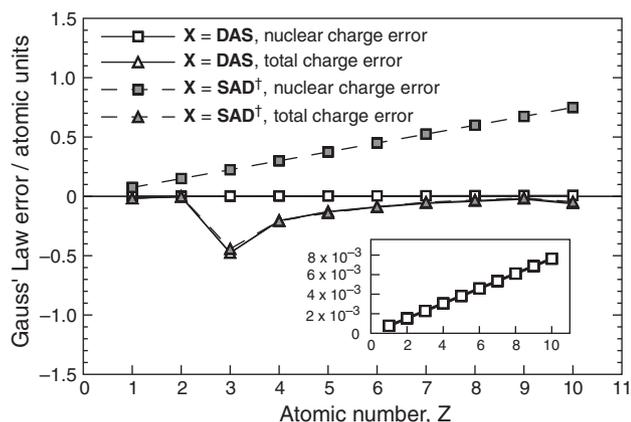


Figure 6. Deviations from Gauss' Law for the nuclear contribution to the induced surface charge, and for the total (nuclear + electronic) induced surface charge, for two different versions of IEF-PCM. (The inset shows an enlarged view of the nuclear charge error for $\mathbf{X} = \text{DAS}$.) The solutes are a series of homonuclear diatomic molecules described at the HF/6-31+G* level. The bond lengths, solute cavities, and discretization points are the same for each molecule. SWIG, ISWIG, and GBO discretization produce essentially identical results, so only SWIG results are shown here.

seems reasonable in view of the fact that the total nuclear contribution to $\mathcal{Q}_{\text{surf}}$ also increases linearly with Z . The magnitude of the error is quite small ($< 0.01e$) for $\mathbf{X} = \text{DAS}$, whereas for $\mathbf{X} = \text{SAD}^\dagger$ the error reaches $0.75e$ at $Z = 10$.

Interestingly, deviations from Gauss' Law for the total induced surface charge (considering both the nuclear and electronic contributions to the electrostatic potential) are nearly equivalent for the two asymmetric versions of IEF-PCM. These deviations are not necessarily errors, since for HF solutes the total surface charge need not obey Eq. (20), owing to the presence of escaped charge. (In fact, the Z -dependent trend in the total deviation from Gauss' Law correlates with atomic electronegativities; the electronegativity is smallest for $Z = 3$, leading to the largest escaped charge and also the largest deviation from Eq. (20).) In the $\mathbf{X} = \text{SAD}^\dagger$ case, this escaped charge somehow cancels the nuclear charge error, and the total deviation from Eq. (20) is the same as it is for $\mathbf{X} = \text{DAS}$.

This analysis explains the results in Table 2. The Gauss' Law error for the amino acids at the AMBER99 level is small because the atomic partial charges are small, and thus the magnitude of \mathbf{v} is relatively small. The total nuclear charge is much larger, and the $\mathbf{X} = \text{SAD}^\dagger$ version of IEF-PCM exhibits large deviations from Gauss' Law, for the nuclear part of $\mathcal{Q}_{\text{surf}}$.

5.4. Definition of D_{ii}

The results in Section 4.6 demonstrate that enforcing the sum rule in Eq. (16) removes the discrepancies in solvation energies obtained using different forms of \mathbf{K} . This is a perplexing result, given that it only modifies the diagonal elements of \mathbf{D} ; the off-diagonal elements are the same whether using Eq. (16) or Eq. (17). On the other hand, the sum rule in Eq. (16) causes D_{ii} to depend on off-diagonal elements of \mathbf{D} , whereas the definition in Eq. (17) does not. While we do not yet have a completely satisfactory explanation for how the sum rule eliminates the discrepancies between different forms of \mathbf{K} , we suggest that it is probably related to an increase in the diagonal dominance of the matrix \mathbf{DA} , as documented below.

In Figure 7, we plot every value of $D_{ii}a_i$ across the entire amino acid data set. One can clearly see that the sum rule inflates many of the magnitudes of $D_{ii}a_i$ relative to the definition in Eq. (17). Moreover, given the definition in Eq. (17) all $D_{ii}a_i$ values are negative and are sharply peaked within the range of $-0.4 \leq D_{ii}a_i \leq 0$. The sum-rule definition, in contrast, exhibits many significantly larger values, over a wide range that even spans some positive values.

We find that inflation of the diagonal element tends to be most pronounced for rows of \mathbf{DA} that contain large off-diagonal elements, such that $|\sum_{j \neq i} D_{ij}a_j| \gg 2\pi$. In such a case, $|D_{ii}a_i| \approx |\sum_{j \neq i} D_{ij}a_j|$, and therefore the i th row of \mathbf{DA} is diagonally dominant. Furthermore, the off-diagonal elements of \mathbf{DA} decay rapidly away from the diagonal, since $(\mathbf{DA})_{ij} \propto r_{ij}^{-3}$. In the context of the sum-rule definition of D_{ii} , these two facts mean that, in practice, $\mathbf{DA} \approx \mathbf{AD}^\dagger$, and therefore $\mathbf{K} \approx \mathbf{K}^\dagger$ regardless of the choice of \mathbf{X} . In contrast, the definition of D_{ii} in Eq. (17) does not allow \mathbf{DA} to become so overwhelmingly diagonally dominant, and large off-diagonal elements of \mathbf{DA} sometimes lead to large differences in the solvation energy, depending upon the choice for \mathbf{X} .

Although we cannot definitively state that inflated diagonal elements D_{ii} are incorrect, we have pointed out in previous work [14] that the D_{ii} sum rule can compromise the negative-definiteness of the response matrix, \mathbf{Q} (thus making PCM energies non-variational) when switching functions are employed. From another point of view, it also seems somewhat unphysical that D_{ii} , which is related to the *self-field* interaction of the i th surface element [20], should depend on any surface element other than the i th one. That is, Eq. (16) has the strange property that the self-field interaction over a given surface element is a function of the

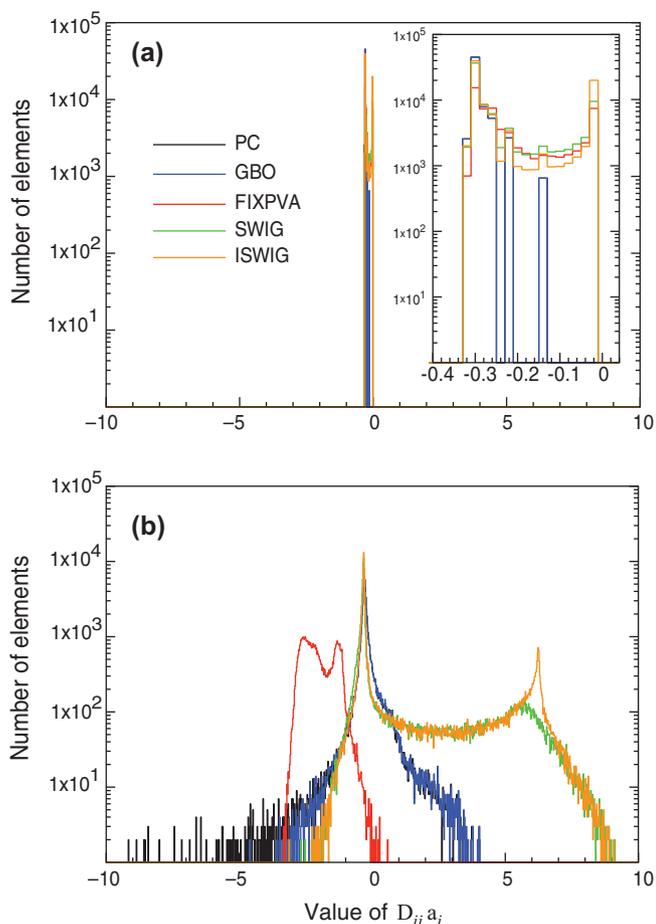


Figure 7. Histogram of dimensionless $D_{ii}a_i$ values, collected from the Lebedev grids (590 points per atom) across all 20 of the amino acids in the HF/6-31+G* calculations. In panel (a), we show the results when Eq. (17) is used to define D_{ii} , and the inset shows a magnified view of the narrow range of $D_{ii}a_i$ values. In panel (b), we use Eq. (16) to define D_{ii} .

positions of all other surface elements. Another odd property is that the sum rule produces positive values for $D_{ii}a_i$ (see Figure 7), which is problematic because \mathbf{DA} is supposed to be the matrix analog of the negative-definite operator $\hat{\mathcal{D}}$, and a negative-definite matrix must necessarily have all negative entries along its diagonal. Values of $D_{ii}a_i > 0$ correspond to the unphysical situation of an attractive self-field interaction over the i th surface element [14]. Thus, even though the sum rule eliminates differences in E_{pol} amongst the different forms of \mathbf{K} , there is no guarantee that the energy and/or surface charges are physically reasonable. Therefore, we will not appeal to the sum rule to eliminate these differences in SWIG/ISWIG discretization.

5.5. Conductor limit

If there is no escaped charge, then the exact relation

$$\hat{\mathcal{F}}\sigma(\vec{s}) = -\phi_0(\vec{s}) \quad (23)$$

is valid in limit $\varepsilon \rightarrow \infty$ [12]. On the other hand, Eq. (1) reduces to Eq. (23) in this limit, regardless of whether there is any escaped charge. As such, the reaction field is inexact *outside* of the cavity, if there is escaped charge [12].

Although the limiting form of the C-PCM and IEF-PCM equations has been derived previously [9,12,14,17], less attention has been given to the discretized matrix equations in this limit. The discretized form of the PCM working equations can be written as

$-\mathbf{Q}^{-1}\mathbf{q} = -\mathbf{v}$. For C-PCM, $\mathbf{Q} = -[(\varepsilon - 1)/\varepsilon]\mathbf{S}^{-1}$ [14] and this equation becomes

$$\left(\frac{\varepsilon}{\varepsilon - 1}\right)\mathbf{S}\mathbf{q} = -\mathbf{v}. \quad (24)$$

For the $\mathbf{X} = \mathbf{DAS}$ form of IEF-PCM, the corresponding equation is

$$\mathbf{Y}^{-1}\left(\frac{1}{\varepsilon}\mathbf{I} - \frac{1}{2\pi}\mathbf{DA}\right)\mathbf{S}\mathbf{q} = -\mathbf{v}, \quad (25)$$

whereas for the $\mathbf{X} = \mathbf{SAD}^\dagger$ form one has

$$\mathbf{Y}^{-1}\mathbf{S}\left(\frac{1}{\varepsilon}\mathbf{I} - \frac{1}{2\pi}\mathbf{AD}^\dagger\right)\mathbf{q} = -\mathbf{v}. \quad (26)$$

For a conductor, we simply have $\mathbf{S}\mathbf{q} = -\mathbf{v}$, and it is immediately clear that Eqs. (24) and (25) reduce to this form in the limit $\varepsilon \rightarrow \infty$. Furthermore, the correct limit is obtained from Eq. (25) regardless of whether $\mathbf{DAS} = \mathbf{SAD}^\dagger$ or not. This feature of $\mathbf{X} = \mathbf{DAS}$ has also been pointed out elsewhere [17,18].

In the limit $\varepsilon \rightarrow \infty$, Eq. (26) can be rearranged to yield

$$\mathbf{S}\mathbf{q} - \frac{1}{2\pi}\mathbf{SAD}^\dagger\mathbf{q} = -\mathbf{v} + \frac{1}{2\pi}\mathbf{DA}\mathbf{v}. \quad (27)$$

To obtain the conductor limit ($\mathbf{S}\mathbf{q} = -\mathbf{v}$) from this equation, it must be the case that

$$-\mathbf{SAD}^\dagger\mathbf{q} = \mathbf{DA}\mathbf{v} = -\mathbf{DAS}\mathbf{q}, \quad (28)$$

or in other words, $\mathbf{SAD}^\dagger = \mathbf{DAS}$. As such, we expect that both the $\mathbf{X} = \mathbf{SAD}^\dagger$ and the symmetrized version of IEF-PCM will afford incorrect limits as $\varepsilon \rightarrow \infty$.

To test this presumption, we investigated the conductor limit numerically, using histidine described at the AMBER99 level. Figure 8 shows the solvation energy as a function of ε , computed using C-PCM and also both asymmetric forms of IEF-PCM. The conductor-like model tracks the $\mathbf{X} = \mathbf{DAS}$ model fairly closely (with small discrepancies when ε is small, as expected), and by the time $\varepsilon \approx 80$, the two methods are indistinguishable. In contrast, the \mathbf{SAD}^\dagger variant of IEF-PCM is in fair agreement with the other two models when $\varepsilon = 2$, but already exhibits a large discrepancy by the time $\varepsilon = 10$, which only increases as $\varepsilon \rightarrow \infty$. This discrepancy is entirely the result of Eq. (27) failing to yield the correct conductor limit when $\mathbf{DAS} \neq \mathbf{SAD}^\dagger$.

To understand the agreement between C-PCM and the $\mathbf{X} = \mathbf{DAS}$ form of IEF-PCM, we rearrange Eq. (1) for IEF-PCM to obtain

$$\left[\hat{\mathcal{F}} + \frac{1}{\varepsilon}\left(2\hat{\mathcal{D}}^{-1} - \hat{\mathcal{F}}\right)\right]\hat{\mathcal{F}}\sigma = -\left(\frac{\varepsilon - 1}{\varepsilon}\right)\phi_0. \quad (29)$$

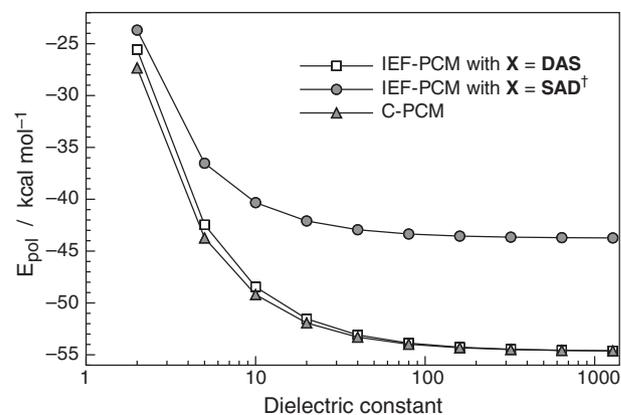


Figure 8. Approach to the conductor limit, for histidine described at the AMBER99 level using SWIG discretization. The same cavity is used in each case and is discretized using 110 Lebedev points per atomic sphere.

The only difference between this equation and the corresponding C-PCM equation is the term on the left that includes the factor of $1/\varepsilon$ (cf. Eq. (24)). In other words, C-PCM introduces an error on the order of ε^{-1} in the calculation of the polarization energy, and is therefore increasingly accurate as ε increases [5,7].

6. Conclusions

For IEF-PCM/SS(V)PE calculations using SWIG or ISWIG discretization [14], we recommend the use of the asymmetric ‘DAS’ form of the \mathbf{K} matrix,

$$\mathbf{K} = \mathbf{S} - \frac{1}{2\pi} \left(\frac{\varepsilon - 1}{\varepsilon + 1} \right) \mathbf{DAS}. \quad (30)$$

With this choice for \mathbf{K} , the IEF-PCM method exhibits the following desirable properties.

- (1) The electrostatic solvation energy, E_{pol} , converges smoothly and rapidly as the surface grid density increases.
- (2) In the absence of escaped charge, numerical errors in E_{pol} are < 0.2 kcal/mol and deviations from Gauss’ Law are $\lesssim 0.02\%$ (based on the amino acid data set).
- (3) The method has the correct limiting behavior as $\varepsilon \rightarrow \infty$.
- (4) All of the aforementioned features are preserved when IEF-PCM is implemented using the SWIG or ISWIG smooth discretization schemes.

The last point serves to underline our previous conclusion [14] that the ISWIG method represents a stable, accurate, and robust approach to cavity discretization.

Although previous studies report little difference between IEF-PCM results obtained using Eq. (30) versus its transpose [15–18], we observe numerous instances where the transposed (‘SAD[†]’) form of the equations leads to large errors and/or erratic behavior. The symmetrized version of the IEF-PCM equations, which has been employed in some previous work [2,12,22], is also subjected to these errors, since this form inherits some of the ill behavior of the SAD[†] form, including an incorrect limit as $\varepsilon \rightarrow \infty$.

Previous studies report [15–18] that E_{pol} is nearly independent of the choice of \mathbf{K} may be attributed to the use of a sum rule (Eq. (16)) to define the diagonal elements of \mathbf{D} . Here, we have demonstrated that this sum rule removes these variations by inflating the diagonal matrix elements of \mathbf{DA} , to the point that \mathbf{DA} is strongly diagonally-dominant and therefore $\mathbf{K} \approx \mathbf{K}^\dagger$. However, this compensatory effect of the sum rule can lead to unphysical solvation energies and/or surface charges, and we do not recommend it as a means to alleviate discrepancies amongst the various forms of \mathbf{K} .

Acknowledgments

This work was supported by an NSF CAREER award (CHE-0748448). Calculations were performed at the Ohio Supercom-

puter Center (Project No. PAS-0291). J.M.H. is an Alfred P. Sloan Foundation fellow.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cplett.2011.04.092.

References

- [1] J. Tomasi, M. Persico, Chem. Rev. 94 (1994) 2027.
- [2] D.M. Chipman, Theor. Chem. Acc. 107 (2002) 80.
- [3] J. Tomasi, B. Mennucci, R. Cammi, Chem. Rev. 105 (2005) 2999.
- [4] S. Miertuš, E. Scrocco, J. Tomasi, Chem. Phys. 55 (1981) 117.
- [5] A. Klamt, G. Schüürmann, J. Chem. Soc. Perkin Trans. 2 (1993) 799.
- [6] V. Barone, M. Cossi, J. Phys. Chem. A 102 (1998) 1995.
- [7] T.N. Truong, U.N. Nguyen, E.V. Stefanovich, Int. J. Quantum Chem. Symp. 30 (1996) 1615.
- [8] E. Cancès, B. Mennucci, J. Tomasi, J. Chem. Phys. 107 (1997) 3032.
- [9] B. Mennucci, E. Cancès, J. Tomasi, J. Phys. Chem. B 101 (1997) 10506.
- [10] B. Mennucci, R. Cammi, J. Tomasi, J. Chem. Phys. 109 (1998) 2798.
- [11] E. Cancès, B. Mennucci, J. Chem. Phys. 114 (2001) 4744.
- [12] D.M. Chipman, J. Chem. Phys. 112 (2000) 5558.
- [13] E. Cancès, B. Mennucci, J. Math. Chem. 23 (1998) 309.
- [14] A.W. Lange, J.M. Herbert, J. Chem. Phys. 133 (2010) 244111.
- [15] M. Cossi, G. Scalmani, N. Rega, V. Barone, J. Chem. Phys. 117 (2002) 43.
- [16] D.M. Chipman, M. Dupuis, Theor. Chem. Acc. 107 (2002) 90.
- [17] F. Lipparini, G. Scalmani, B. Mennucci, E. Cancès, M. Caricato, M.J. Frisch, J. Chem. Phys. 133 (2010) 014106.
- [18] M. Cossi, N. Rega, G. Scalmani, V. Barone, J. Chem. Phys. 114 (2001) 5691.
- [19] P. Su, H. Li, J. Chem. Phys. 130 (2009) 074109.
- [20] G. Scalmani, M.J. Frisch, J. Chem. Phys. 132 (2010) 114110.
- [21] A.W. Lange, J.M. Herbert, J. Phys. Chem. Lett. 1 (2010) 556.
- [22] D.M. Chipman, J. Chem. Phys. 124 (2006) 224111.
- [23] J.L. Pascual-Ahuir, E. Silla, J. Comput. Chem. 11 (1990) 1047.
- [24] E. Silla, I. Tuñón, J.L. Pascual-Ahuir, J. Comput. Chem. 12 (1991) 1077.
- [25] J.L. Pascual-Ahuir, E. Silla, I. Tuñón, J. Comput. Chem. 15 (1994) 1127.
- [26] C.S. Pomelli, J. Comput. Chem. 25 (2004) 1532.
- [27] C.S. Pomelli, in: B. Mennucci, R. Cammi (Eds.), Continuum Solvation Models in Chemical Physics, Wiley, Chichester, UK, 2007, pp. 49–63.
- [28] D.M. York, M. Karplus, J. Phys. Chem. A 103 (1999) 11060.
- [29] H. Li, J.H. Jensen, J. Comput. Chem. 25 (2004) 1449.
- [30] H.M. Senn, P.M. Margl, R. Schmid, T. Ziegler, P.E. Blöchl, J. Chem. Phys. 118 (2003) 1089.
- [31] V.I. Lebedev, USSR Comp. Math. Math+ 15 (1975) 44.
- [32] M. Cossi, B. Mennucci, R. Cammi, J. Comput. Chem. 17 (1996) 57.
- [33] B.A. Gregersen, D.M. York, J. Chem. Phys. 122 (2005) 194110.
- [34] Y. Wang, H. Li, J. Chem. Phys. 131 (2009) 206101.
- [35] E.O. Purisima, S.H. Nilar, J. Comput. Chem. 16 (1995) 681.
- [36] A.W. Lange, J.M. Herbert, J. Chem. Phys. 134 (2011) 117102.
- [37] J. Wang, P. Cieplak, P.A. Kollman, J. Comput. Chem. 21 (2000) 1049.
- [38] A. Bondi, J. Phys. Chem. 68 (1964) 441.
- [39] R.S. Rowland, R. Taylor, J. Phys. Chem. 100 (1996) 7384.
- [40] B. Lee, F.M. Richards, J. Mol. Biol. 55 (1971) 379.
- [41] Y. Shao et al., Phys. Chem. Chem. Phys. 8 (2006) 3172.
- [42] A.W. Lange, J.M. Herbert, J. Am. Chem. Soc. 131 (2009) 3913.
- [43] M.W. Schmidt et al., J. Comput. Chem. 14 (1993) 1347.
- [44] A. Klamt, B. Mennucci, J. Tomasi, V. Barone, C. Curutchet, M. Orozco, F.J. Luque, Acc. Chem. Res. 42 (2009) 489.
- [45] J.P. Guthrie, I. Povar, Can. J. Chem. 87 (2009) 1154.
- [46] V. Barone, M. Cossi, J. Tomasi, J. Chem. Phys. 107 (1997) 3210.
- [47] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Proc. Natl. Acad. Sci. USA 98 (2001) 10037.
- [48] S.Y. Yu, W. Geng, G.W. Wei, J. Chem. Phys. 126 (2007).
- [49] Y.C. Zhou, M. Feig, G.W. Wei, J. Comput. Chem. 29 (2008) 87.
- [50] E. Yap, T. Head-Gordon, J. Chem. Theory Comput. 6 (2010) 2214.