Renlong Ye, Xuemei Nie, Chung F. Wong, Xuedong Gong,
Yan A. Wang, Thomas Heine, and Baojing Zhou*

Variable van der Waals Radii Derived From a Hybrid Gaussian Charge Distribution Model for Continuum-Solvent Electrostatic Calculations

DOI 10.1515/zpch-2015-0746
Received December 9, 2015; accepted February 1, 2016

Abstract: We introduce a hybrid Gaussian charge distribution model (HGM) that partitions the molecular electron density into overlapping spherical atomic domains. The semi-empirical HGM consists of atom-centered spherical Gaussian functions and discrete point charges, which are optimized to reproduce the electrostatic potential on the molecular surface as well as the number of electrons in atom-centered and certain off-atom-centered spherical regions as closely as possible. In contrast, our previous Gaussian charge distribution model [J. Chem. Phys. **129**, 014509 (2008)] contained only spherical Gaussian functions and was not required to reproduce the number of electrons in off-atom-centered regions. Variable van der Waals (vdW) radii fluctuating around the Bondi radii are derived from the HGM based on the isodensity contour concept and further employed to define the molecular cavity in our quantum mechanical/Poisson–Boltzmann/surface area model as well as the polarizable continuum model. The variable vdW radii produce more accurate solvation free energies for 31 neutral molecules than the Bondi radii for both continuum solvent models (CSM) consistently. Moreover, for

***Corresponding author: Baojing Zhou,** Computational Institute for Molecules and Materials, School of Chemical Engineering, Nanjing University of Science and Technology, Xiaolingwei 200, 210094 Nanjing, China, e-mail: bzhou@mail.njust.edu.cn
Renlong Ye, Xuemei Nie, Xuedong Gong: Computational Institute for Molecules and Materials, School of Chemical Engineering, Nanjing University of Science and Technology, Xiaolingwei 200, 210094 Nanjing, China
Chung F. Wong: Department of Chemistry and Biochemistry and Center for Nanoscience, University of Missouri-Saint Louis, One University Boulevard, 63121 Saint Louis, USA
Yan A. Wang: Department of Chemistry, University of British Columbia, V6T 1Z1 Vancouver, Canada
Thomas Heine: School of Engineering and Science, Jacobs University Bremen Campus Bremen, Ring 1, 28759 Bremen, Germany

H atoms, the linear dependence of the atomic radii on the atomic partial charges is identified.

Keywords: Continuum Solvent Models, van-der-Waals Radii, Hybrid Gaussian Charge Distribution Model, Solvent Effects, Force Fields.

Dedicated to Professor Michael Springborg on the occasion of his 60th birthday

1 Introduction

When a molecule is transferred from vacuum to aqueous solution, its structure, properties, and reactivity may change significantly. One can choose either explicit-solvent (see, for example, Reference [1]) or implicit-solvent models (see, for example, References [2] and [3]) to account for this phenomenon. The latter approach is more appealing since its computational cost is much lower due to the elimination of the extensive sampling of solvent configurations required in the former [3]. Implicit-solvent models are also called continuum solvation models (CSM) since the solvent is treated as a dielectric continuum [2]. One popular CSM solves the following Poisson–Boltzmann (PB) equation:

$$\nabla \cdot [\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})] - \kappa^2(\mathbf{r})\phi(\mathbf{r}) = \rho(\mathbf{r}), \tag{1}$$

where $\phi(\mathbf{r})$ is the electrostatic potential, $\epsilon(\mathbf{r})$ is the dielectric constant at $\mathbf{r}$, $\kappa(\mathbf{r})$ is the Debye-Hückel parameter, and $\rho(\mathbf{r})$ is the charge distribution of the solute. Different numerical methods have been developed to solve the PB equation as reviewed in Reference [2].

In PB CSM calculations, one needs to define the dielectric boundary between the solute and the solvent appropriately [2]. A widely used scheme is to use a set of interlocking spheres to define the molecular cavity. Various schemes have been proposed to designate the radii of these spheres. They are often taken from the Bondi radii [4]. In the highly parameterized SMD model, the solute atoms of the same element adopt a fixed and optimized radius except for O atoms, whose sizes depend on the solvent but not on their chemical environment [5]. On the other hand, it is desirable to endow solute atoms with flexible radii since the shape of the electron cloud associated with each atom changes when a molecule is formed. The use of flexible atomic radii can lead to higher precisions [6–8] and can also to some extent compensate for the failure of the PB equation [Equation (1)] to describe nonlinear effects [9].

The simplest scheme for generating flexible atomic radii is probably to assume that atomic radii depend linearly on atomic partial charges [10–12]. It is also common to use atomic radii specially designed for PB calculations or developed based on molecular dynamics simulations [6, 7]. These models provide a different kind of flexibility in that different atom types are used to ascribe atomic radii. In the united atom model for Hartree-Fock, atomic radii for heavy atoms are formulated to rely on the detailed chemical environment, such as the number of H atoms attached to them, the nature of the chemical groups they are bonded to, and the hybridization state [8]. Beyond the interlocking sphere model, one can use the constant electron density contour obtained from quantum mechanical (QM) calculations to define the molecular cavity [13, 14]. In this way, additional flexibility can be gained because the variation in atomic radii is not limited by a finite number of atom types.

Our previous Gaussian charge distribution model (Gaussian model in short, GM) provides another variant for generating variable atomic radii [15]. The GM partitions the QM electron density of a molecule into overlapping spherical atomic domains. Then, the variable atomic radii can be naturally derived based on the isodensity contour concept. Note that the way to partition the molecular electron density is not unique. According to Bader's *atoms in molecules* (AIM) theory, the non-overlapping space enclosed within a zero flux surface surrounding the nucleus defines the shape of an atom unambiguously [16]. From a different perspective, Hirshfeld introduced auxiliary weight functions to divide the molecular electron density into overlapping non-spherical atomic domains [17]. Then, the product of the weight function and the molecular electron density yields the so-called stockholder's atomic domain.

As a semi-empirical model, the GM cannot closely resemble the QM electron density of a molecule everywhere in space. The deviation is particularly large in the core region, where the shapes of both the core and valence electron densities are very spiky. Since the core electrons are less likely to appear in the valence region, they exert little influence on the formation of the atomic domains in a molecule. Thus, we use only the QM valence electron density to build the GM [15]. In this way, the stress on the GM to mimic the QM electron density is lessened to some extent and the GM can better approximate the QM electron density in the bonding region. Variable van der Waals (vdW) radii were derived from the GM for the H, C, N, O, S atoms in 31 small neutral molecules [15]. Interestingly, only the sizes of the H atoms fluctuate significantly, while those of the other heavy atoms show much less variations. Although the variable vdW radii from the GM produce better solvation free energies than the Bondi radii in our quantum mechanical/Poisson–Boltzmann/surface area (QM/PB/SA) calculations, the improvement is small, with the RMS error reduced from 1.71 to 1.38 kcal/mol [15].

Here, our goal is to improve the GM by introducing point charges and other physically meaningful constraints for its optimization. We show that the augmented GM can produce better variable vdW radii so that the accuracy of the PB CSM calculations can be further improved. The linear correlation between the variable atomic radii and the atomic partial charges for different elements is also investigated.

2 Methods

2.1 HGM for deriving variable vdW radii

The central approximation for the GM is a superposition of atom-centered Gaussian functions $\rho_i^{\mathrm{G}}(\mathbf{r})$ to represent the molecular electron density,

$$\begin{aligned}\rho(\mathbf{r}) &\approx \sum_i^N \rho_i^{\mathrm{G}}(\mathbf{r}) \\ &= \sum_i^N \frac{q_i^{\mathrm{G}}}{(\sqrt{2\pi}\sigma_i)^3}\exp\left\{-\frac{1}{2}\left(\frac{\mathbf{r}-\mathbf{R}_i}{\sigma_i}\right)^2\right\},\end{aligned} \tag{2}$$

where N is the total number of atoms and $\mathbf{R}_i$ is the position of the ith nuclei. The two parameters, q_i^{G} and σ_i give, respectively, the measures of the electronic charge associated with an atom and the extent to which the electron density is spread out. Note that when σ_i approaches zero, the Gaussian function reduces to the δ function. Thus, the parameters of the GM condense into a common point-charge representation. In the following, two strategies are employed to further improve the GM.

First, we add extra atom-centered point charges q_i to the above GM:

$$\rho(\mathbf{r}) \approx \sum_i^N \rho_i^{\mathrm{G}}(\mathbf{r}) + q_i(\mathbf{R}_i). \tag{3}$$

We refer to the modified GM as the hybrid Gaussian model (HGM) since it consists of both diffuse Gaussian functions and discrete point charges. The HGM can describe the electron density of heavy atoms better as the charges q_i can be used to represent their core electrons. We will show later that this is necessary for elements of the third row and beyond in the periodic table as they possess more than ten core electrons. Thus, the core electrons can be eliminated from the corresponding Gaussian functions. However, for the second-row elements, the two core electrons still need to be represented by the Gaussian function as in our previous work [15]. Otherwise, the Gaussian functions would be electron deficient and become less capable of describing the electron density in the bonding region. The

total number of variational parameters $\{q_i^G, \sigma_i, q_i\}$ is increased to $3N$, which renders the HGM more flexible than the GM.

For the above HGM, the electrostatic potential (ESP) at a point around the molecule is given by:

$$V_{el}^G(\mathbf{r}) = \sum_i^N \frac{Z_i - q_i}{\mathbf{r} - \mathbf{R}_i} - \sum_i^N \int \frac{\rho_i^G(\mathbf{r}')}{\mathbf{r} - \mathbf{r}'} d\mathbf{r}', \quad (4)$$

where Z_i is the nuclear charge of the ith atom and the integral is over all space. The deviation of $V_{el}^G(\mathbf{r})$ from the QM ESP, $V_{el}^Q(\mathbf{r})$, is evaluated in a least square manner:

$$R_1 = \frac{\sum_k \left[V_{el}^Q(\mathbf{r}_k) - V_{el}^G(\mathbf{r}_k)\right]^2}{\sum_k V_{el}^Q(\mathbf{r}_k)^2}, \quad (5)$$

where $\mathbf{r}_k$ is the position vector of a point between 1.4 times and 2.0 times the vdW surface of the molecule. R_1 needs to be minimized for the optimization of the HGM.

We use another constraint R_2 to provide extra control on the shape of the Gaussian functions [15]. We define the following R_2 to require the number of electrons produced by the HGM in the atom-centered spherical regions to approach a certain value,

$$R_2 = \sum_i^N \frac{[\int_{C_i} \rho_{val}^Q(\mathbf{r}) + q_{c,i} - \sum_j^N \int_{C_i} \rho_j^G(\mathbf{r}) - q_i]^2}{[\int_{C_i} \rho_{val}^Q(\mathbf{r})]^2}. \quad (6)$$

Here, $\rho_{val}^Q(\mathbf{r})$ is the valence electron density from the *ab initio* QM calculations and C_i represents the spherical region centered on the ith nuclei. We do not use the QM all-electron density in the above equation since this would generate too much stress on the HGM as explained in the previous section. For the third-row elements, we use $q_{c,i}$, which is set to 10, to make up for the number of missing core electrons. However, for the second-row elements, $q_{c,i}$ is set to 0 instead of 2. Otherwise, the resulting Gaussian functions would be too spiky and fail to mimic the QM valence electron density in the bonding region satisfyingly. Thus, the number of electrons in the core region would be underestimated by the Gaussian functions to some extent. However, the sacrifice of the accuracy in the core region would be compensated by the improved accuracy in the bonding region, which is much more important for defining the atomic domain. The sizes of the atom-centered spherical regions C_i need to be carefully chosen in order for the two constraints R_1 and R_2 to be ideally minimized. Based on numerical tests, for different elements, we choose the radius of C_i to be 40% of their Bondi radii.

Our second strategy for improving the HGM aims at the electron density in the bonding region. The HGM should appropriately capture the changes occurring

to the shapes of atomic domains due to the influences from different chemical environments. To achieve this, we further require the number of valence electrons in certain off-atom-centered regions to be reproduced by the HGM as closely as possible. We define the third constraint R_3 to measure the difference between the HGM and the QM valence electron density within the off-atom-centered regions:

$$R_3 = \sum_i^M \frac{[\int_{B_i} \rho_{\mathrm{val}}^{\mathrm{Q}}(\mathbf{r}) - \sum_j^N \int_{B_i} \rho_j^{\mathrm{G}}(\mathbf{r})]^2}{[\int_{B_i} \rho^{\mathrm{Q}}(\mathbf{r})]^2}, \tag{7}$$

where M runs over all the covalent bonds in a molecule. The integral is performed in a small sphere, B_i, somewhere in between the two nuclei. We first position the B_i at the centers of covalent bonds, i. e. the boundary points or bond critical points between bonded atoms [18–21]. However, this does not lead to substantial improvements. We try other options and find the minima of the valence electron densities along bond axes are much better choices for locating off-atom-centered regions, B_i. For the C–O and C–H bonds, such points are closer to the nuclei of the C atoms, only ca. 0.3 bohr from them (see Figure 4). Based on numerical tests, the optimal radius of B_i is set to 0.36 bohr, much smaller than the radii of atom-centered regions.

We derive the variable vdW radii from the HGM with the same strategy as that in our previous work [15], i. e. using the isodensity contour concept. For atom i of the element I, the atomic boundary is reached when the Gaussian function $\rho_i^{\mathrm{G}}(\mathbf{r})$ drops to a fixed isodensity contour value, ρ_I^{iso}:

$$\rho_i^{\mathrm{G}}(\mathbf{r})|_{r=r_{i,I}} = \rho_I^{\mathrm{iso}}, \tag{8}$$

where $r_{i,I}$ is the variable vdW radius of atom i. Note that the average of the variable vdW radii derived this way depends on the value of ρ_I^{iso}, i. e. the former would increase with the decreasing of the latter. We adjust the isodensity contour value, ρ_I^{iso}, so that the average vdW radius for each element I is equal to its Bondi radius:

$$\frac{1}{n}\sum_i^n r_{i,I} = r_{I,\mathrm{Bondi}}. \tag{9}$$

Here n is the total number of atoms belonging to chemical species I in a test set of molecules.

Finally, we emphasize that the gist of our approach is to generate high quality spherical atomic domains from the QM electron density. Other spherical functions may be used to build the model to approximate the QM electron density and different constraints can be designed for its optimization. The optimal choices for the spherical functions and constraints depend on the nature of the molecular density. We summarize the computational protocol for the construction of the HGM in Figure 1.

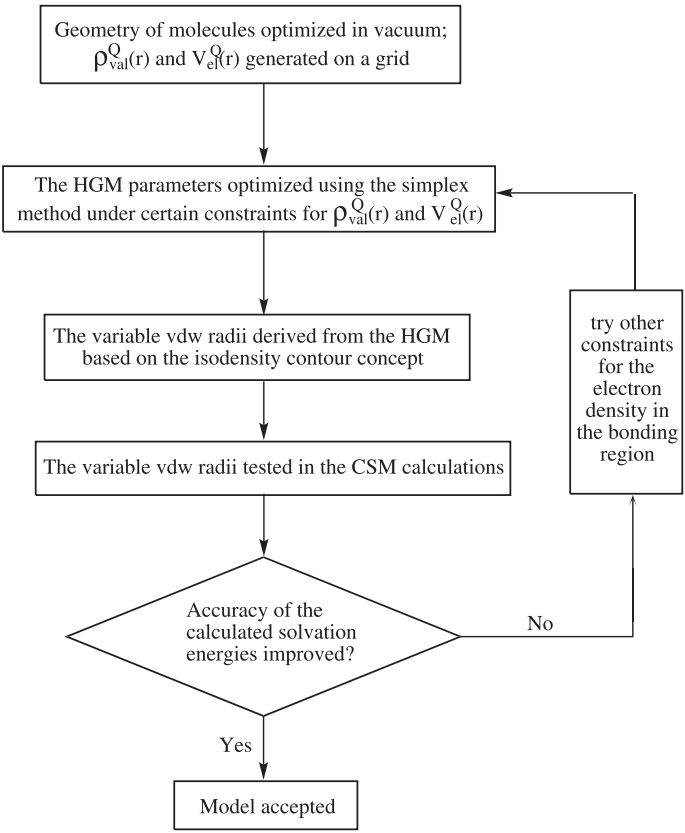

Figure 1: Flowchart indicating the computational protocol used for deriving the HGM.

2.2 Solvation energy calculation

The calculated solvation free energy includes the electrostatic and nonpolar contributions [22, 23],

$$G_{sol} = G_{el} + G_{np}. \tag{10}$$

In our QM/PB/SA model, the electrostatic energy, G_{el}, is given by:

$$G_{el} = G_{RF} + G_{wfd}, \tag{11}$$

where G_{RF} is the electrostatic interaction energy between the solute and the reaction field

$$G_{RF} = \frac{1}{2}\int \rho(\mathbf{r})\phi_{RF}(\mathbf{r})d\mathbf{r} + \frac{1}{2}\sum_i Z_i\phi_{RF}(\mathbf{r}). \tag{12}$$

The reaction field, $\phi_{\mathrm{RF}}(\mathbf{r})$, is calculated as:

$$\phi_{\mathrm{RF}}(\mathbf{r}) = \phi_{\mathrm{solvent}}(\mathbf{r}) - \phi_{\mathrm{vac}}(\mathbf{r}), \tag{13}$$

where $\phi_{\mathrm{solvent}}(\mathbf{r})$ and $\phi_{\mathrm{vac}}(\mathbf{r})$ are the ESP obtained by solving the PB equation [Equation (1)] in solvent and in vacuum. The second term, G_{wfd}, in Equation (10) is the energy associated with the distortion of the wave function in going from the vacuum to the solution [23]

$$G_{\mathrm{wfd}} = \left\langle \psi^{\mathrm{s}} \left| H^0 \right| \psi^{\mathrm{s}} \right\rangle - \left\langle \psi^{\mathrm{g}} \left| H^0 \right| \psi^{\mathrm{g}} \right\rangle. \tag{14}$$

Here, Ψ^g is the gas phase solute wave function, Ψ^s is the solvated solute wave function, and H^0 is the gas phase Hamiltonian. We solve the following Kohn-Sham (KS) equation (in Hartree a. u.) [24–26],

$$\left(-\frac{1}{2}\nabla^2 + \upsilon_{\mathrm{eff}}^{\mathrm{KS}}[\rho](\mathbf{r}) + \phi_{\mathrm{RF}}(\mathbf{r})\right)\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}), \tag{15}$$

where, ϵ_i is the eigenvalue of the ith orbital $\psi_i(\mathbf{r})$ and the KS effective potential $\upsilon_{\mathrm{eff}}^{\mathrm{KS}}[\rho](\mathbf{r})$ contains the Hartree, the exchange-correlation, and the ion-electron potentials. The electron density is generated via:

$$\rho(\mathbf{r}) = \sum_i f_i \left|\psi_i(\mathbf{r})\right|^2, \tag{16}$$

where f_i is the occupation number of $\psi_i(\mathbf{r})$. The outlying charges are not confined to the solute cavity and are treated as being immersed in a high dielectric media of solvent water. Since Eqs. (1), (13), and (15) are coupled, we solve them iteratively until self-consistency is reached [27].

The nonpolar energy, G_{np}, is evaluated as a surface area (SA) dependent term [22]

$$G_{\mathrm{np}} = \sigma A, \tag{17}$$

where A is the solvent accessible surface area [28] constructed by rolling a probe sphere of 1.4 Å along the vdW surface and σ is a surface tension coefficient set to 0.006 kcal/mol Å^{-2}. This crude nonpolar energy model is much less sensitive to the variation of the atomic radii than the electrostatic part. Since our goal is mainly to improve the accuracy of the electrostatic calculations, this model is adequate for our purpose.

3 Computational details

We build the HGM for the 31 small neutral molecules (Figure 2), which were studied in our previous work [15, 27]. These molecules are selected because their sol-

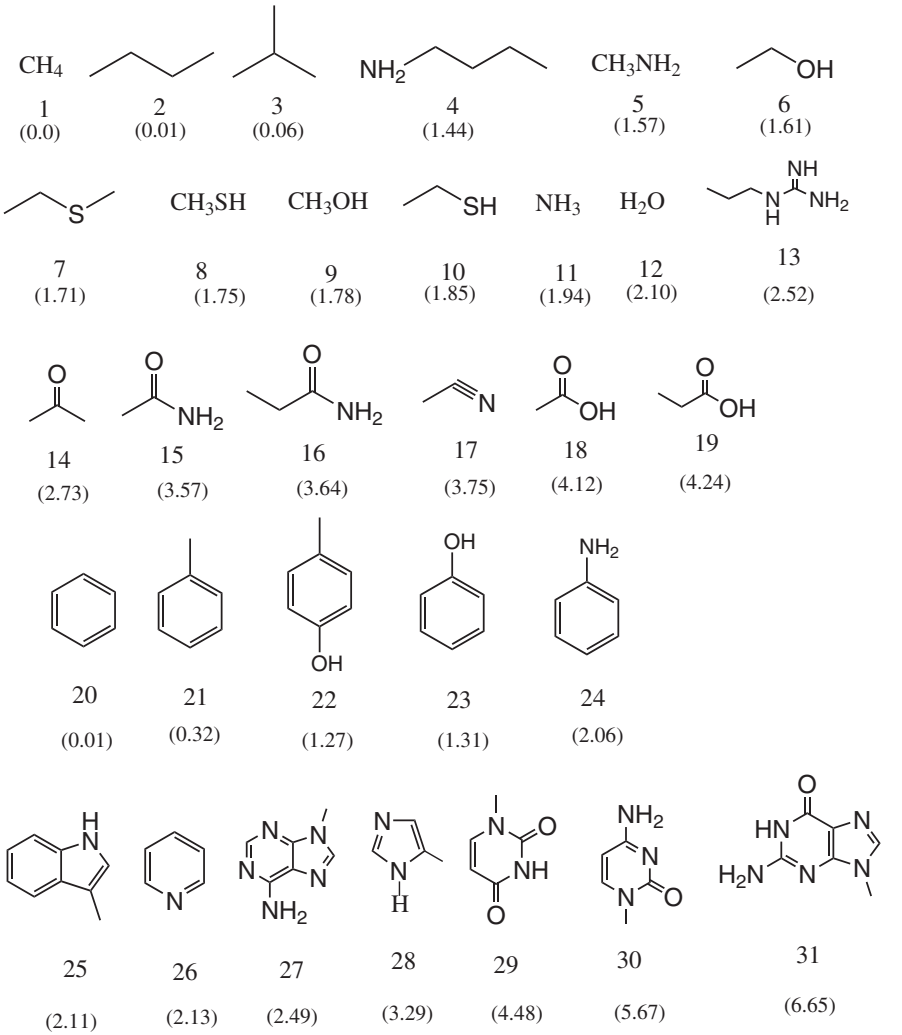

Figure 2: The 31 small neutral molecules studied in this work. The number in parenthesis is the gas-phase dipole moment (debye).

vation free energies have been well measured and they are frequently encountered in organic, biological, and pharmaceutical chemistry research. Following Reference [15], the 31 molecules are divided into three groups: acyclic compounds (molecule 1–19), carbocyclic compounds (molecule 20–24), and heterocyclic compounds (molecule 25–31). Within each group, the molecules are ordered by their dipole moments.

Table 1: The initial guesses of Gaussian parameters, $\{q_i^G, \sigma_i, q_i\}$, for five elements. Z_i and q^{MK} are nuclear charges and Merz–Kollman partial atomic charges.

	H	C	N	O	S
$q_i^G(e^-)$			$Z_i - q_i - q^{MK}$		
σ_i(a.u.)	0.7	1.1	1.0	0.8	1.7
$q_i(e^-)$	0	0	0	0	10

All geometries were optimized by density functionary theory (DFT) calculations in gas phase with the 6-31G(d) basis set using the Gaussian09 package [29]. The Becke3LYP functional [30–33] is used for the electron exchange and correlation. The QM charge density and ESP were generated on fine grids of 6–8 points/bohr using the 6-31G(d) basis set and "*cubgen*" utility of the Gaussian09 suite of programs.

We use the simplex method [34] to minimize the following objective function in order to optimize the HGM,

$$R = \alpha R_1 + \beta R_2 + \gamma R_3, \tag{18}$$

where the weights associated with the three constraints α, β, and γ are set to 0.6, 0.4, and 0.5, respectively. The initial values for $\{q_i^G, \sigma_i, q_i\}$ for the five elements are listed in Table 1. During the parameter optimization, q_i would fluctuate around 0 and 10 for the second-row and third-row elements, respectively. We find the total number of electrons in a molecule is well conserved in the optimized HGM.

The variable vdW radii are derived form the HGM using the isodensity contour ρ^{iso} deduced from Equation (9). Then, they are used to define the molecular cavity in the QM/PB/SA model we developed previously [27]. We coupled a QM code SIESTA [35] with the PB solver in the University of Houston Brownian dynamics (UHBD) program [36, 37]. The coupled KS and PB equations are solved iteratively until the solute electron density and the reaction field are self-consistent [27, 38]. The default double-ζ plus single polarization basis set in SIESTA was used and the grid cutoff was set to 100 Rydberg. The nonpolar energy is calculated using the UHBD program [36, 37]. These computational details are the same as those in Reference [15].

We also test the variable vdW radii using the polarizable continuum model (PCM) as implemented in the Gaussian09 package [29]. We used the 6-31G(d) and 6-311++G(d,p) basis sets for the electrostatic PCM calculations. The scaling factor α for the Bondi radii is set to its default value of 1.1 for the two basis sets. For the variable vdW radii, α is set to 1.1 and 1.14 for the 6-31G(d) and 6-311++G(d,p) basis sets, respectively. In our calculations, the nonpolar part is still calculated

from the SA based formula [Equation (17)]. We find the SA dependent nonpolar energy model works quite well for the 31 neutral molecules studied in this work. We further use the SMD model in the Gaussian09 package [29] for comparison purpose.

4 Results and discussions

4.1 ESP and electron density: GM vs. HGM

The QM ESP on the molecular surface is well reproduced by the two GMs and the Merz–Kollman point-charge model [39] as compared in Figure 3 (a). For carbocyclic compounds and heterocyclic compounds, $\sqrt{R_1}$ is less than 0.2. Larger errors are seen in several acyclic compounds with small dipole moments, such as butane. The total number of QM valence electrons in the atom-centered regions is also well reproduced by the two GMs as shown in Figure 3 (b). The largest relative root mean square (RRMS) errors of the GM and the HGM are ca. 0.27 and 0.25, respectively. For the former, $\sqrt{R_2}$ contains large fluctuations, while for the latter $\sqrt{R_2}$ is more stable and its average is ca. 0.1.

The RRMS errors of the two GMs for the number of electrons in the off-atom-centered regions are compared in Figure 3 (c). For the GM, abnormally large $\sqrt{R_3}$ occurs to molecules 7, 8 and 10, which contain S atoms. This is mainly caused by the mismatch between the number of electrons represented by the GM (ca. 16) and the number of QM valence electrons (ca. 6) for S atoms. For other molecules, $\sqrt{R_3}$ of the GM are much smaller, with an average ca. 0.5. On the other hand, $\sqrt{R_3}$ of the HGM are reduced to ca. 0.25 for all 31 molecules. These results suggest that the HGM approximates the QM valence electron density better than the GM.

The electron densities generated by the two GMs along the axes of the O–C bond in acetic acid and the N_2C=O bond in 1-methyl-uracil are illustrated in Figure 4. The QM valence electron densities are also depicted for comparison. We can see that the QM electron density along the double C=O bond is bulkier than that along the single C–O bond. This is not surprising since there is more build-up of the electron density in the bonding region for the double C=O bond than for the single C–O bond. Consequently, the O atom in the C–O single bond is smaller than that in the N_2C=O double bond. However, this difference is clearly underestimated by the GM. In contrast, the HGM adequately differentiate the two situations, and therefore the difference in the sizes of the two types of O atoms increases sufficiently.

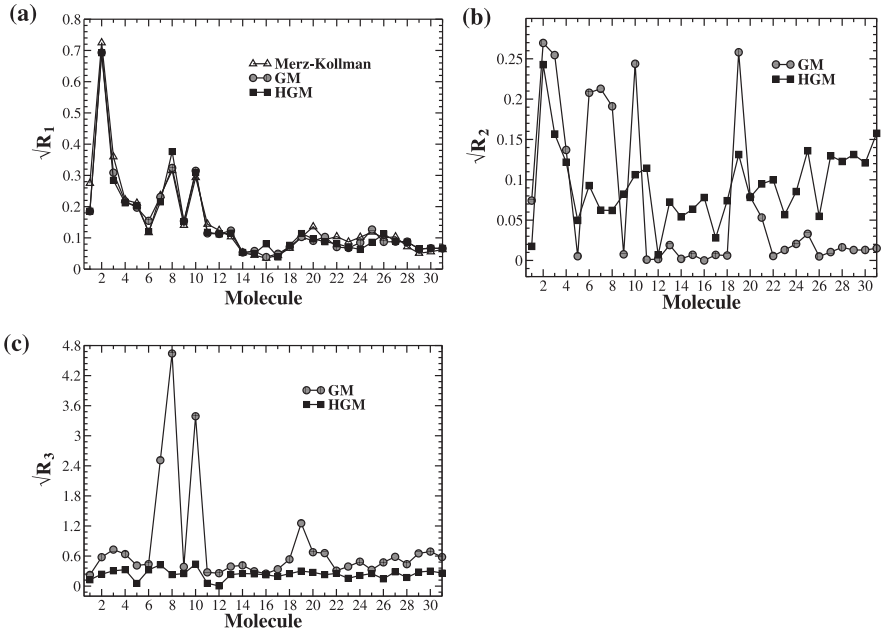

Figure 3: RRMS errors of (a) the electrostatic potential, (b) the total number of valence electrons in atom-centered regions, and (c) the total number of valence electrons in off-atom-centered regions produced by the Merz–Kollman model (open up triangle), the GM (dark circle), and the HGM (solid square).

4.2 Variable vdW radii for QM/PB/SA calculations

We extract the variable vdW radii from the HGM and the number of H, C, N, O atoms in 31 molecules are 197, 114, 29, and 16, respectively. As shown in Table 2, different isodensity contours need to be used for the four elements in order for their average radii to approach the Bondi radii. The full width at half maximum w_i of the Gaussian function $\rho_i^{\mathrm{G}}(\mathbf{r})$ is proportional to the Gaussian parameter σ_i [Equation (2)], i. e. $w_i = 2.3548\sigma_i$. For the four elements, the average width of $\rho_i^{\mathrm{G}}(\mathbf{r})$ increase in the following order: $\bar{w}_{\mathrm{H}} < \bar{w}_{\mathrm{O}} < \bar{w}_{\mathrm{N}} < \bar{w}_{\mathrm{C}}$. For each element, $\bar{w}$ and ρ^{iso} from the HGM are slightly lower than those from the GM.

The standard deviation σ of the variable vdW radii reveals their fluctuation intensities. For the GM, these variations decrease in the following order: H ≫ C > O ≃ N. Thus, only the sizes of the H atoms contain significant fluctuations. For the HGM, the largest fluctuation in the atomic radii still occurs to the H atoms, albeit it decreases by ca. 10%. On the other hand, the fluctuations in the sizes of both the N and O atoms increase noticeably, almost comparable to that of the H

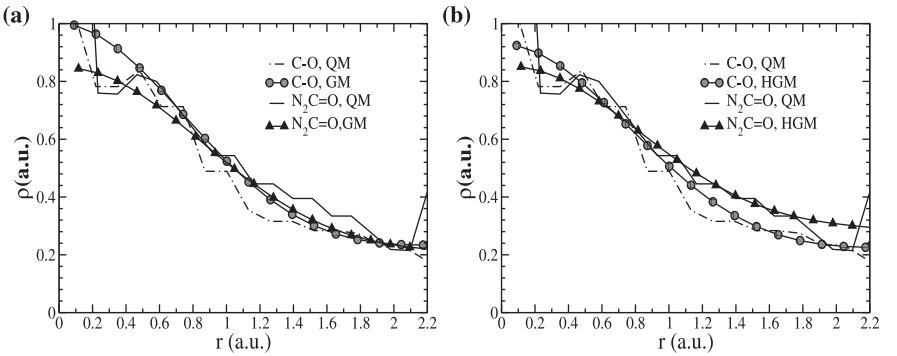

Figure 4: Electron densities along the O–C bond axis of acetic acid and O=CN_2 bond axis of 1-methyl-uracil produced from (a) the QM model and GM and (b) the QM model and HGM.

Table 2: Bondi radii r_{Bondi} and a comparison of key parameters for the GM and HGM: the average full width at half maximum $\bar{w}$ of the Gaussian functions, the isodensity contour ρ^{iso}, and the standard deviations σ of the variable vdW radii.

	H	C	N	O
r_{Bondi} (Å)	1.20	1.70	1.55	1.52
GM				
$\bar{w}$ (a.u.)	1.797	2.809	2.312	1.990
ρ^{iso} (10^{-3} a.u.)	1.3	5.7	5.9	2.8
σ (a.u.)	0.210	0.061	0.025	0.028
HGM				
$\bar{w}$ (a.u.)	1.686	2.656	2.141	1.947
ρ^{iso} (10^{-3} a.u.)	0.98	4.6	3.5	2.3
σ (a.u.)	0.188	0.065	0.144	0.142

atoms. Little change occurs to the sizes of the C atoms, whose fluctuation remains small. In the following, we use vdW-r-GM, and vdW-r-HGM to refer to the variable vdW radii derived from the GM and HGM, respectively.

The vdW-r-HGM of atoms in the four selected molecules are provided in Table S1 of the Supporting Information. For O atoms, their radii are sensitive to their local chemical environments, i. e. the coordination number and the bonding atoms. We can see that the sizes of double-bonded O atoms are usually larger than the Bondi radius of 1.52 Å, in contrast the radii of single-bonded O atoms are often less than it. However, the O atom in water molecule is a bit larger than the Bondi radius. This is mainly because the two H atoms in a water molecule only have one electron each. Thus, the repulsion between O and H atoms is relatively week so that the electron cloud of the former can spread out more in space.

Table 3: RMS error (in kcal/mol) in the calculated hydration energies of the 31 molecules divided into three groups. The Bondi radii, the variable vdW radii from the GM, and that from the HGM are used to define the dielectric boundaries for the QM/PB/SA calculations.

	acyclic	carbocyclic	heterocyclic	total
Bondi	1.39	1.69	2.38	1.71
GM	1.38	1.69	1.14	1.38
HGM	1.01	1.52	0.80	1.07

We calculate the solvation free energies for 31 neutral molecules with our QM/PB/SA model to test the Bondi radii, vdW-r-GM and vdW-r-HGM. The experimental solvation free energies are used to benchmark the QM/PB/SA calculations using the three sets of atomic radii. The use of the Bondi radii leads to a total RMS error of 1.71 kcal/mol (Table 3). The largest RMS error of 2.38 kcal/mol occurs to the heterocyclic compounds. The use of the vdW-r-GM reduces the total RMS error by 0.33 kcal/mol. However, the vdW-r-GM produces improvement only for the heterocyclic compounds. It is encouraging that the use of the vdW-r-HGM doubles this improvement produced by the vdW-r-GM. With a reduction of 0.64 kcal/mol, the total RMS error is now only ca. 1.0 kcal/mol. Moreover, the vdW-r-HGM produces improvements for all three types of molecules, which increase in the following order: carbocyclic < acyclic < heterocyclic. For heterocyclic compounds, the reduction in the RMS error reaches 1.58 kcal/mol.

The calculated solvation free energy consists of the electrostatic and nonpolar contributions [Equation (10)]. These two contributions calculated from using the three different radii are compared for the four selected molecules in Table 4. We see that the electrostatic part dominates the solvation free energy, although the small positive nonpolar contribution is not negligible. Further inspection discloses that the nonpolar part is only slightly altered by the use of different sets of atomic radii and the change is within 0.1 kcal/mol. Thus, the improved accuracy from using the vdW-r-HGM is mainly due to the electrostatic part.

The correlation between the experimental and calculated hydration energies using the vdW-r-GM and vdW-r-HGM for these 31 molecules is plotted in Figure 5. When the vdW-r-GM is used, solvation free energies of acyclic compounds are systematically overestimated (too negative), while those of heterocyclic compounds are systematically underestimated (too positive). Those systematic errors are significantly reduced when the vdW-r-HGM is used. For carbocyclic compounds, the experimental hydration energies are also better reproduced. The correlation coefficients R between experimental and calculated hydration energies are 0.971 and 0.985 for the vdW-r-GM and vdW-r-HGM, respectively.

Table 4: Experimental and calculated hydration free energies from the QM/PB/SA model for the four selected molecules (in kcal/mol).

molecule	model	G_{el}	G_{np}	G_{sol}	exptl
H_2O	Bondi	−5.45	0.74	−4.71	
	GM	−5.46	0.74	−4.72	−6.30
	HGM	−6.22	0.72	−5.50	
acetic	Bondi	−8.94	1.19	−7.75	
acid	GM	−9.24	1.19	−8.05	−6.70
	HGM	−7.64	1.20	−6.44	
4-cresol	Bondi	−5.81	1.65	−4.16	
	GM	−6.34	1.69	−4.65	−6.13
	HGM	−8.01	1.63	−6.38	
9-methyl-	Bondi	−21.21	1.97	−19.24	
guanine	GM	−23.06	1.98	−21.08	−22.40
	HGM	−25.34	1.96	−23.38	

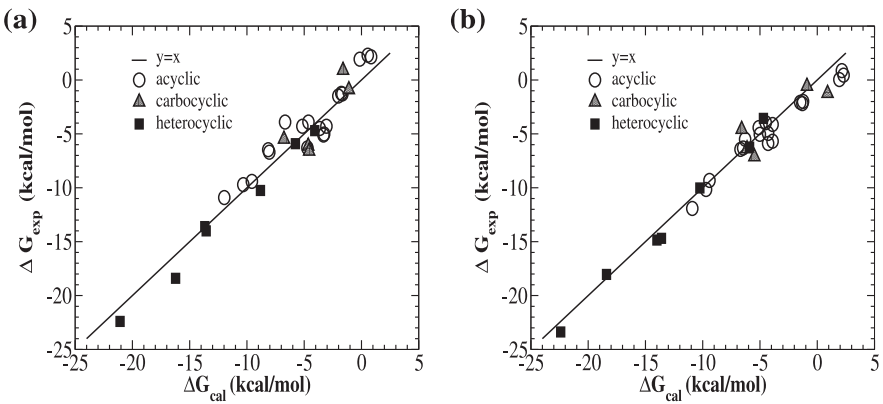

Figure 5: Correlation plots between the experimental and calculated hydration free energies for 31 molecules. The variable vdW radii generated from the GM (a) and HGM (b) are employed to define the molecular cavity in the QM/PB/SA model.

4.3 Charge dependent radii for H atoms

Previous CSM calculations by Tawa et al. suggested that the calculated solvation free energies of molecules are particularly sensitive to the value of the hydrogen radius [40]. The variable vdW radii for the H atoms derived from the two GMs are plotted against the Mulliken atomic partial charges in Figure 6. The linear correlation is identified between the two parameters, i. e. the vdW radii decrease almost linearly with the increasing positive charges. We fit the variable vdW radii $r_{i,\alpha}$ to

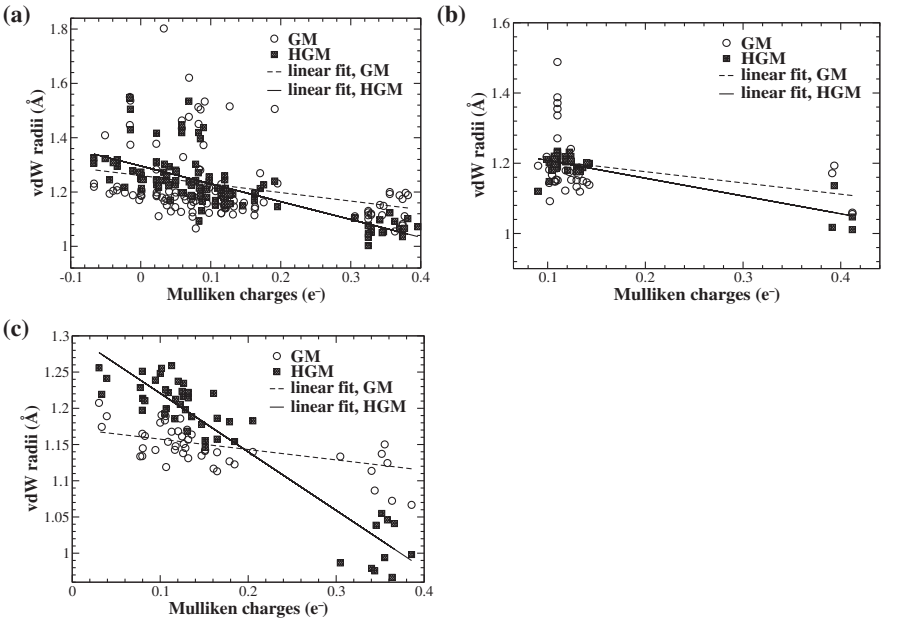

Figure 6: Correlation plot between the Mulliken partial charges and variable vdW radii derived from the GM and HGM for H atoms in (a) acyclic, (b) carbocyclic, and (c) heterocyclic compounds.

the Mulliken charges $q_{i,\alpha}^{\mathrm{M}}$ with the linear regression model,

$$r_{i,\alpha} = A_\alpha + B_\alpha q_{i,\alpha}^{\mathrm{M}}, \tag{19}$$

where α refers to the three types of compounds (Table 5). The HGM produces steeper slopes (more negative B_α) than the GM for all three types of compounds. The discrepancy between the two GMs is particularly large for the heterocyclic compounds. The correlation coefficients from the HGM are greater than 0.7, meanwhile those from the GM are less than 0.5.

We further divide the H atoms into three groups according to the elements they are bounded to: oxygen, nitrogen, and carbon. The average radii r_{ave} of the three types of H atoms and the standard deviations from the GM and the HGM are compared in Table 6. For both the GM and HGM, H atoms bonded to the less electronegative C are slightly larger than the Bondi radius, whereas those bonded to the more electronegative N are smaller than the Bondi radius. However, the sizes of hydroxyl hydrogens are overestimated by the GM. In our previous work, we had to rectify this defect by artificially setting the radii of hydroxyl hydrogens to 1.08 Å so that the computed solvation energies of alcohol molecules can

Table 5: The linear regression parameters (Equation 19) for the acyclic, carbocyclic, and heterocyclic compounds: A and B are the fitting coefficients and R stands for the correlation coefficient.

	GM			HGM		
	A	B	R	A	B	R
acyclic	1.261	−0.313	0.28	1.296	−0.657	0.74
carbocyclic	1.240	−0.320	0.32	1.259	−0.507	0.85
heterocyclic	1.172	−0.143	0.48	1.302	−0.808	0.94

Table 6: Comparison of the GM and the HGM: average atomic radii (r_{ave}) of H atoms that are bounded to O (H^a), N (H^b), and C atoms (H^c) and the standard deviations (s.d.) from the mean, which is the Bondi radius.

	H^a	H^b	H^c
GM			
r_{ave} (Å)	1.20	1.14	1.21
s.d. (a.u.)	0.0082	0.21	0.22
HGM			
r_{ave} (Å)	1.08	1.04	1.23
s.d. (a.u.)	0.04	0.04	0.08

be more accurate [15]. It is encouraging that this optimal size is naturally produced by the HGM. Thus, the too large sizes of hydroxyl hydrogens from the GM are mainly caused by its inadequate accuracy in the bonding region. We also see that the sizes of hydroxyl hydrogens derived from the GM exhibit little fluctuation, while the variations in the sizes of all three types of H atoms from the HGM are similar.

4.4 Transferability of variable vdW radii

The development of CSM has a long history and various implementations are available for solving the PB equation. We further test the transferability of our variable vdW radii in PCM calculations. However, the nonpolar part of the solvation energy is still evaluated as the SA dependent term [Equation (17)]. The highly parameterized SMD model is also employed for comparison purpose.

The RMS errors in the solvation energies for the 31 molecules produced by the four schemes are compared in Table 7. For the Bondi radii, the total RMS error from the PCM calculations using the 6-31G(d) basis set is a bit

Table 7: RMS errors (in kcal/mol) in the hydration energies of the 31 molecules obtained from the PCM and SMD calculations. The Bondi radii, the variable vdW radii from the GM, HGM, and HGM-*l*H are used to define the molecular cavity in the PCM.

	acyclic	carbocyclic	heterocyclic	total
6-31G(d)				
Bondi	1.01	1.19	2.03	1.33
GM	1.32	1.31	1.03	1.26
HGM	1.01	1.33	0.65	1.00
HGM-*l*H	0.93	1.07	0.98	1.04
SMD	0.84	0.88	1.93	1.18
6-311++G(d,p)				
Bondi	1.06	1.18	0.94	1.06
GM	1.05	1.30	1.21	1.13
HGM	0.87	1.20	0.84	0.93
HGM-*l*H	0.88	0.96	0.83	0.88
SMD	0.63	0.81	0.91	0.73

lower than that from our QM/PB/SA calculations using the double-ζ plus single polarization basis set (Table 3). The use of the vdW-r-GM produces little improvement, i. e. less than 0.1 kcal/mol. In contrast, the vdW-r-HGM significantly reduces the total RMS error by more than 0.3 kcal/mol. The improvement mainly comes from the heterocyclic compounds. This is consistent with the results from the QM/PB/SA calculations. The SMD model is very accurate for the acyclic and carbocyclic molecules. However, the RMS error is a bit large for the heterocyclic molecules. With the use of the 6-31G++(d,p) basis set, the RMS errors from using the Bondi radii are significantly lowered. The vdW-r-GM becomes slightly worse than the Bondi radii, however the vdW-r-HGM remains superior to the Bondi radii, especially for the acyclic compounds. The SMD model produces the most accurate results for all three types of compounds.

There are two ways to apply the variable vdW radii in the CSM calculations. For the direct use of our approach, one needs to construct the HGM for the target molecules following the protocol designed in this work. Alternatively, one can avoid the efforts of building the HGM and simply use the resulting variable vdW radii. This will be highly desirable for large molecules, for which the construction of the HGM can be very expensive. Then, it is necessary to relate the variation of atomic radii to some chemical descriptor. For H atoms, the atomic partial charge turns out to be a good chemical descriptor. We can take advantage of the linear relationship between the atomic radii and atomic partial charges for H atoms. We assign the charge-dependent radii to H atoms based on Equation (19). For other

heavy atoms, we still use the Bondi radii to define their sizes since a good chemical descriptor linked to the variation of atomic sizes is yet to be found. We refer to this scheme as HGM-*l*H to emphasize the fact that only the H atomic radii are flexible. Based on our numerical tests, the linear regression model from the GM is better for acyclic molecules, while for carbocyclic and heterocyclic compounds, the linear regression model from the HGM is superior (Table 5). The final HGM-*l*H model is also examined in the PCM calculations. As shown in Table 7, for the 31 molecules studied here, the atomic radii from the HGM-*l*H are of similar quality as those from the HGM when the small 6-31G(d) basis set is used. For the large 6-311++G(d,p) basis set, the former is slightly better than the latter.

5 Conclusions

We developed a semi-empirical model, HGM, to incorporate flexibility into the Bondi radii. This approach involves partitioning the molecular density into spherical atomic domains and taking advantage of the iso-density contour concept. The improvement of the HGM over the GM mainly stems from its more accurate description of the electron density in the bonding region. The variable vdW radii from the HGM are superior to those from our previous GM as demonstrated by their applications in our QM/PB/SA calculations as well as in the PCM calculations. The HGM produces variable vdW radii for not only H atoms but also N and O heavy atoms. However, a linear correlation between the variable vdW radii and the atomic partial charges is only identified for the H atoms.

Acknowledgement: BZ gratefully acknowledges the financial support from "the Fundamental Research Funds for the Central Universities", No. 30915011314. CFW acknowledges the support from a University of Missouri Research Board Award. We appreciate the useful comments from Dr. Norman Hamer to improve the manuscript.

References

1. C. L. Brooks III, M. Karplus, and B. M. Pettitt, in *Advances in Chemical Physics*, edited by I. Prigogine and S. A. Rice (Wiley, New York, 1988), Vol. LXXI.
2. J. Tomasi, B. Mennucci, and R. Cammi, Chem. Rev. **105** (2005) 2999.
3. C. J. Cramer and D. G. Truhlar, Chem. Rev. **99** (1999) 2161.

4. A. Bondi, J. Phy. Chem. **68** (1964) 441.
5. A. V. Marenich, C. J. Cramer, and D. G. Truhlar, J. Phys. Chem. B **113** (2009) 6378.
6. E. V. Stefanovich and T. N. Truong, Chem. Phys. Lett. **244** (1995) 65.
7. M. Nina, D. Beglov, and B. Roux, J. Phys. Chem. B **101** (1997) 5239.
8. V. Barone, M. Cossi, and J. Tomasi, J. Chem. Phys. **107**, 3210 (1997).
9. S. W. Rick and B. J. Berne, J. Am. Chem. Soc. **116** (1994) 3949.
10. D. Horvath, D. Van Belle, G. Lippens, and S. J. Wodak, J. Chem. Phys. **104** (1996) 6679.
11. B. Ginovska, D. M. Camaioni, M. Dupuis, C. A. Schwerdtfeger, and Q. Gil, J. Phys. Chem. A **112** (2008) 10604.
12. G. Hou, X. Zhu, and Q. Cui, J. Chem. Theory Comput. **6** (2010) 2303.
13. B. M. Deb, R. Singh, and N. Sukumar, J. Mol. Struct. (THEOCHEM) **259** (1992) 121.
14. J. B. Foresman, T. A. Keith, K. B. Wiberg, J. Snoonian, and M. J. Frisch, J. Phys. Chem. **100** (1996) 16098.
15. B. Zhou, M. Agarwal, and C. F. Wong, J. Chem. Phys. **129** (2008) 014509.
16. R. F. W. Bader, *Atoms in Molecules: A Quantum Theory* (Clarendon, Oxford, UK, 1990).
17. F. L. Hirshfeld, Theor. Chim. Acta **44** (1977) 129.
18. D. Zhao and Z. Yang, J. Comput. Chem. **35** (2014) 965.
19. J. S. Murray and P. Politzer, WIREs Comput. Mol. Sci. **1** (2011) 153.
20. P. Politzer and J. S. Murray, Theor. Chem. Acc. **108** (2002) 134.
21. P. Politzer, J. S. Murray, and P. Lane, J. Comput. Chem. **24** (2003) 505.
22. D. Sitkoff, K. A. Sharp, and B. Honig, J. Phys. Chem. **98** (1994) 1978.
23. D. J. Tannor, B. Marten, R. Murphy, R. A. Friesner, D. Sitkoff, A. Nicholls, M. Ringnalda, W. A. Goddard III, and B. Honig, J. Am. Chem. Soc. **116** (1994) 11875.
24. W. Kohn and L. J. Sham, Phys. Rev. **140** (1965) A1133.
25. R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Clarendon, New York, 1989).
26. R. M. Dreizler, E. K. U. Gross, *Density Functional Theory: An Approach to the Quantum Many-Body Problem* (Springer-Verlag, Berlin, 1990).
27. M. Wang and C. F. Wong, J. Phys. Chem. A **110** (2006) 4873.
28. F. A. Richards, Annu. Rev. Biophys. Bioeng. **6** (1977) 151.
29. M. J. Frisch, G. W. Trucks, H. B. Schlegel et al., Gaussian 09, Revision, A.02, Gaussian, Inc., Wallingford CT, 2009.
30. A. D. Becke, Phys. Rev. A **38** (1988) 3098.
31. A. D. Becke, J. Chem. Phys. **98** (1993) 5648.
32. C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37** (1988) 785.
33. S. H. Vosko, L. Wilik, and M. Nusair, Can. J. Phys. **58** (1980) 1200.
34. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in Fortran* (Cambridge University Press, Cambridge, 1992).
35. J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, J. Phys.: Condens. Matter **14** (2002) 2745.
36. M. E. Davis, J. D. Madura, B. A. Luty, and J. A. McCammon, Comput. Phys. Commun. **62** (1991) 187.
37. J. D. Madura, J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. R. Scott, and J. A. McCammon, Comput. Phys. Commun. **91** (1995) 57.
38. M. Wang and C. F. Wong, J. Chem. Phys. **126** (2007) 026101.

39. B. H. Besler, K. M. Merz, and P. A. Kollman, J. Comput. Chem. **11** (1990) 431.
40. G. J. Tawa, R. L. Martin, L. R. Pratt, and T. V. Russo, J. Phys. Chem. **100** (1996) 1515.
41. D. Jacquemin, A. Planchat, C. Adamo, and B. Mennucci, J. Chem. Theory Comput. **8** (2012) 2359.

Supplementary material: The online version of this article
(DOI: 10.1515/zpch-2015-0746) provides supplementary material for authorized users.