Improvement of semiempirical response properties with charge-dependent response density

Timothy J. Giese and Darrin M. York

Citation: The Journal of Chemical Physics **123**, 164108 (2005); doi: 10.1063/1.2080007
View online: http://dx.doi.org/10.1063/1.2080007
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/123/16?ver=pdfcov
Published by the AIP Publishing

Articles you may be interested in

Water interactions with hydrophobic groups: Assessment and recalibration of semiempirical molecular orbital methods
J. Chem. Phys. **141**, 034106 (2014); 10.1063/1.4886655

Magnetic linear response properties calculations with the Gaussian and augmented-plane-wave method
J. Chem. Phys. **131**, 014106 (2009); 10.1063/1.3156803

LocalSCF method for semiempirical quantum-chemical calculation of ultralarge biomolecules
J. Chem. Phys. **121**, 1266 (2004); 10.1063/1.1764496

Resonance Raman spectra of uracil based on Kramers–Kronig relations using time-dependent density functional calculations and multireference perturbation theory
J. Chem. Phys. **120**, 11564 (2004); 10.1063/1.1697371

Calculating molecular electric and magnetic properties from time-dependent density functional response theory
J. Chem. Phys. **116**, 891 (2002); 10.1063/1.1420401

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 192.12.88.212 On: Fri, 12 Sep 2014 18:52:12

Improvement of semiempirical response properties with charge-dependent response density

Timothy J. Giese and Darrin M. York[a)]
Department of Chemistry, University of Minnesota, Minneapolis, Minnesota 55455

(Received 3 June 2005; accepted 29 August 2005; published online 26 October 2005)

The present work outlines a new method for treatment of charge-dependent polarizability in semiempirical quantum models for use in combined quantum-mechanical/molecular mechanical simulations of biological reactions. The method addresses a major shortcoming in the performance of conventional semiempirical models for these simulations that is tied to the use of a localized minimal atomic-orbital basis set. The present approach has the advantages that it uses a density basis that retains a set of linear-response equations, does not increase the atomic-orbital basis, and avoids the problem of artificial charge transfer and scaling of the polarizability seen in related models that allow atomic charges to fluctuate. The model introduces four new atom-based parameters and has been tested with the modified neglect of differential overlap d-orbital Hamiltonian against 1132 molecules and ions and shown to decrease the dipole moment and polarizability errors by factors of 2 and 10, respectively, with respect to density-functional results. The method performs impressively for a variety of charge states (from 2+ to 2−), and offers a potentially powerful extension in the design of next generation semiempirical quantum models for accurate simulations of highly charged biological reactions. © *2005 American Institute of Physics*.
[DOI: 10.1063/1.2080007]

I. INTRODUCTION

Semiempirical quantum models[1–4] have been and continue to be invaluable tools in the simulation of chemical reactions in large, complex systems.[5,6] The efficiency afforded by semiempirical quantum models allows fundamental aspects of chemical bonding to be captured while remaining tractable for linear scaling electronic structure calculation of very large systems[7,8] and hybrid quantum-mechanical/molecular-mechanical (QM/MM) simulations that span long times scales.[9,10] The accuracy and transferability of semiempirical methods are limited by the approximations inherent in their functional form and basis. Nonetheless, through careful parametrization, semiempirical methods can be made that rival the accuracy of high-level *ab initio* calculations for specific properties within a restricted range of chemistry.[4]

Semiempirical models continue to be actively developed,[1–4] including improvements in the treatment of dispersion interactions,[11,12] modified exchange-repulsion terms,[12] orthogonalization corrections,[11,13,14] improved hydrogen bonding models,[12,15,16] and reformulation from density functional-based perspectives.[17,18] Moreover, the use of these methods in combined QM/MM simulations[19] and the application to biological reactions[19–21] remain an area of intense interest and effort.

An important response property not well reproduced by conventional semiempirical methods is the electronic polarizability. Accurate modeling of polarization is likely to be important for highly polar and ionic interactions such as those encountered in the reactions of enzymes and especially ribozymes, and hence is a primary focus area in the development of next generation semiempirical quantum models for biological reactions. Polarization is sensitive to the size, angular momentum, and diffuseness of the basis functions used for the molecular-orbital degrees of freedom. The inability of semiempirical methods to reproduce reliable atomic and molecular polarizabilities[22] is largely due to the use of localized minimal valence basis sets.[23] This effect is emphasized in Table I, which compares the polarizability of isolated atoms and their ions computed with the AM1, PM3, and modified neglect of differential overlap (MNDO)/d semiempirical Hamiltonians to density-functional calcula-

TABLE I. Static dipole polarizabilities of atoms and ions.

		α (Å^3)				
Atom	Q	Expt.[a]	DFT[b]	PM3	AM1	MNDO/d
O	−2	3.880	3.546	0.000	0.000	0.000
	−1	⋯	1.626	0.023	0.029	0.030
	0	0.802	0.682	0.052	0.059	0.061
	1	⋯	0.382	0.081	0.128	0.128
S	−2	10.2	12.215	0.000	0.000	0.862
	−1	⋯	5.569	0.171	0.296	0.949
	0	2.90	2.703	0.396	0.578	1.041
	1	⋯	1.594	1.143	1.463	2.099
F	−1	1.040	0.987	0.000	0.000	0.000
	0	0.557	0.455	0.054	0.013	0.024
	1	⋯	0.259	0.084	0.026	0.047

[a]From Ref. 71.
[b]B3LYP/6-311++G(3*df*,2*p*). Note that the anionic DFT values contained positive orbital eigenvalues and are therefore suspect.

[a)]Author to whom correspondence should be addressed. Electronic mail: york@chem.umn.edu

 © 2005 American Institute of Physics

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 192.12.88.212 On: Fri, 12 Sep 2014 18:52:12

tions and available experimental estimates. The complete occupation of all spin orbitals, e.g., O^{2-}, is admittedly severe and is demonstrated for motivational purposes; however, the poor prediction of molecular polarizabilities of neutral molecules has also been shown to be underestimated by approximately 25% at the semiempirical level when compared to experiment.[24] Recently, neglect of diatomic differential overlap- (NDDO) based methods have been introduced[2,25,26] that make impressive progress in the direct prediction of molecular electronic polarizabilities. These predictive schemes are founded in variational perturbation theory,[27–29] which is used to derive analytic expressions for the polarizability in terms of *scaled* dipole integrals.

The purpose of this work is to characterize the degree to which current NDDO-based semiempirical methods are able to properly polarize in response to long-range electrostatics, and to propose and test corrections to the formalism that afford an avenue for considerable improvement without a significant increase in computational cost. The approach taken in the present work is to expand the energy in terms of the electron-density response to second order, using the self-consistent semiempirical density as a reference in the expansion. The density response functions are attenuated by the semiempirical atomic charge, and become more polarizable as electron density builds up around atoms—an important consideration for anionic systems such as biological phosphates and phosphoranes. The corrections to the polarizability enter as a post-self-consistent-field (SCF) correction and do not expand the atomic-orbital basis. Hence, the method is quite inexpensive in that it does not need to be calculated at every step of the SCF procedure, and does not increase the dimensions of the Hamiltonian or Fock matrices.

The outline of the paper is as follows. Section II describes the chemical-potential equalization (CPE) implementation within the semiempirical method and the reference data used for parametrization. Section III compares the polarizability of molecules predicted from *ab initio*, semiempirical, and the semiempirical calculations supplemented with the CPE correction. The paper concludes in Sec. IV with a summary of the key results and outline of directions for future research.

II. METHODS

Section II A describes the CPE correction to the semiempirical energy and Sec. II B discusses the reference data used in the fitting procedure.

A. Chemical-potential equalization correction to the semiempirical energy

The purpose of this subsection is to formulate a CPE (Ref. 30) correction to semiempirical models in order to improve the response properties. In density-functional theory, the total electronic energy, $E[\rho]$, can be expressed in terms of the electron density ρ as

$$E[\rho] = F[\rho] + \int \rho(\mathbf{r}) v(\mathbf{r}) d^3r, \tag{1}$$

where $F[\rho]$ is a universal functional of the density ρ that contains the electron kinetic energy plus the classical and nonclassical electron-electron interaction energies, and $v(\mathbf{r})$ is the external potential that arises from the nuclei plus any additional applied fields.

Consider the expansion of the electronic energy about some reference density, ρ_{ref}, that is close to the equilibrated ground-state density ρ_0,

$$E[\rho] = E[\rho_{\mathrm{ref}}] + \Delta E^{(1,2,\cdots)}_{\rho_{\mathrm{ref}}}[\delta\rho], \tag{2}$$

where $\delta\rho = \rho - \rho_{\mathrm{ref}}$, and

$$\begin{aligned}\Delta E^{(1,2,\cdots)}_{\rho_{\mathrm{ref}}}[\delta\rho] &= E[\rho] - E[\rho_{\mathrm{ref}}] \\ &= \int \left[\frac{\delta E[\rho]}{\delta\rho(\mathbf{r})}\right]_{\rho_{\mathrm{ref}}} \delta\rho(\mathbf{r}) d^3r + \frac{1}{2}\int\int \delta\rho(\mathbf{r}) \\ &\quad \times \left[\frac{\delta^2 E[\rho]}{\delta\rho(\mathbf{r})\,\delta\rho(\mathbf{r}')}\right]_{\rho_{\mathrm{ref}}} \delta\rho(\mathbf{r}') d^3r d^3r' \\ &\quad + \cdots . \end{aligned} \tag{3}$$

The ground-state energy, $E[\rho_0]$, and electron density, ρ_0, are derived from the solution of the constrained variational condition

$$\delta\left\{E[\rho] - \mu\left(\int \rho(\mathbf{r}) d^3r - N\right)\right\} = 0, \tag{4}$$

or equivalently

$$\delta\left\{\Delta E^{(1,2,\cdots)}_{\rho_{\mathrm{ref}}}[\delta\rho] - \Delta\mu\left(\int \delta\rho(\mathbf{r}) d^3r - \Delta N\right)\right\} = 0, \tag{5}$$

where $N = \int \rho(\mathbf{r}) d^3r$ and $\Delta N = \int \delta\rho(\mathbf{r}) d^3r$.

The approach taken in the present work is to use the SCF semiempirical density as the reference density in the Taylor-series expansion. Under the assumption that the semiempirical density is reasonably close to the true ground-state density, the Taylor expansion can be truncated at second order to provide a post-SCF linear-response correction that considerably improves the charge-dependent molecular response properties. The method has the advantages that it is variational, requires only a solution of a set of linear equations, and does not increase the atomic-orbital basis or the dimensions of the Hamiltonian, Fock, or single-particle density matrices.

In order to cast the variational procedure into a linear algebraic form, the response density is expanded in an auxiliary density basis as

$$\delta\rho(\mathbf{r}) = \sum_i c_i \varphi_i(\mathbf{r}), \tag{6}$$

where the functions $\varphi_i(\mathbf{r})$ are *not* themselves expanded in the atomic-orbital basis. If the energy expression in Eq. (2) is truncated at second order, i.e.,

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 192.12.88.212 On: Fri, 12 Sep 2014 18:52:12

$$E[\rho] \approx E[\rho_{\mathrm{ref}}] + \Delta E^{(1,2)}_{\rho_{\mathrm{ref}}}[\delta\rho], \tag{7}$$

then the the equation for the response energy, Eq. (3), can be cast into a linear algebraic form as[30]

$$\Delta E^{(1,2)}_{\rho_{\mathrm{ref}}}[\delta\rho] = \mathbf{c}^T \cdot \mathbf{m} + \tfrac{1}{2}\mathbf{c}^T \cdot \boldsymbol{\eta} \cdot \mathbf{c}, \tag{8}$$

where, formally, the matrix elements are given by

$$m_i = \int \left[\frac{\delta E[\rho]}{\delta\rho(\mathbf{r})}\right]_{\rho_{\mathrm{ref}}} \varphi_i(\mathbf{r}) d^3r, \tag{9}$$

$$\eta_{ij} = \int\int \varphi_i(\mathbf{r}) \left[\frac{\delta^2 E[\rho]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\right]_{\rho_{\mathrm{ref}}} \varphi_j(\mathbf{r}') d^3r d^3r'. \tag{10}$$

The constrained variational solution for the CPE coefficient vector, $\mathbf{c}$, is given by

$$\mathbf{c} = -\boldsymbol{\eta}^{-1} \cdot (\mathbf{m} - \Delta\mu\mathbf{d}), \tag{11}$$

where

$$\Delta\mu = \frac{\Delta N + \mathbf{d}^T \cdot \boldsymbol{\eta} \cdot \mathbf{m}}{\mathbf{d}^T \cdot \boldsymbol{\eta} \cdot \mathbf{d}}, \tag{12}$$

and the elements of the matrix $\mathbf{d}$ are given by

$$d_i = \int \varphi_i(\mathbf{r}) d^3r. \tag{13}$$

In the present work, the model considered shall assume $\Delta N = 0$; i.e., the number of electrons is set at the semiempirical level, and not altered by the post-SCF CPE correction.

The response density is expanded in a set of atom-centered dipolar Gaussian functions of the form

$$\varphi(\mathbf{r}) = -2\zeta_i(q_i)^2 \left(\frac{\zeta_i(q_i)^2}{\pi}\right)^{3/2} (u - U_i) \exp^{-\zeta_i(q_i)^2 |\mathbf{r} - \mathbf{R}_i|^2}, \tag{14}$$

where u is either x, y, or z (a component of the spatial coordinate), U_i is either X_i, Y_i, or Z_i (a component of the atomic position of atom i), and i is an atom index. Note that it is implied that the Gaussian exponents, $\zeta_i(q_i)$, are explicit functions of the atomic charges, q_i, and will be discussed in more detail below. In principle, the basis used for the response density can be made more general, for example, by inclusion of monopolar functions, higher-order multipolar functions, and functions at auxiliary positions other than the atomic centers. In the present work, monopolar functions were neglected in order to avoid artificial charge-transfer phenomena inherent in the conventional linear-response framework of CPE,[30] electronegativity equalization,[31,32] charge equilibration,[33] fluctuating charge,[34] and related models.[35,36] Charge-transfer artifacts have been studied extensively and discussed for these linear-response models.[37–40] Nonetheless, this topic remains an active area of research.[36,38,39] The proposed correction scheme can be described as being a hybrid SCF/linear-response approach: atomic charge transfer is not treated with a linear-response model, but is instead handled at the semiempirical level in the conventional fashion of populating occupied molecular orbitals during the nonlinear SCF procedure. The use of atom-centered dipolar Gaussian functions removes the possibility of interatomic charge transfer, in the traditional sense, and circumvents the need to apply explicit Lagrange multiplier constraint on the normalization in Eq. (12), i.e., $\Delta\mu = 0$.

Consider the linear term in Eq. (3) that gives rise to the m_i matrix elements of Eq. (9). The functional derivative can be written as

$$\left[\frac{\delta E[\rho]}{\delta\rho(\mathbf{r})}\right]_{\rho_{\mathrm{ref}}} = \left[\frac{\delta F[\rho]}{\delta\rho(\mathbf{r})}\right]_{\rho_{\mathrm{ref}}} + \upsilon(\mathbf{r}) = \left[\frac{\delta T_S[\rho]}{\delta\rho(\mathbf{r})}\right]_{\rho_{\mathrm{ref}}} + \left[\frac{\delta E_{\mathrm{XC}}[\rho]}{\delta\rho(\mathbf{r})}\right]_{\rho_{\mathrm{ref}}} + \phi_{\mathrm{ref}}(\mathbf{r}) + \upsilon(\mathbf{r}), \tag{15}$$

where the terms $T_S[\rho]$ and $E_{\mathrm{XC}}[\rho]$ are the noninteracting kinetic energy and exchange-correlation energy, respectively, and $\phi_{\mathrm{ref}}(\mathbf{r})$ is the electrostatic potential due to the reference electron density $\rho_{\mathrm{ref}}(\mathbf{r})$. Note that if the reference electron density is the exact ground-state electron density (i.e., $\rho_{\mathrm{ref}} = \rho_0$), this term is a constant everywhere in space equal to the electronic chemical potential.

In the present work, a model is considered where the electrostatic potential of the semiempirical reference density is damped in the short range by a smooth switching function. The matrix and vector elements take the form

$$m_i = \int \varphi_i(\mathbf{r}) \nu_f(\mathbf{r}) d^3r + \sum_{j=1}^{N} D_{ij}(r_{ij}) \int \varphi_i(\mathbf{r}) \phi_j(\mathbf{r}) d^3r, \tag{16}$$

$$\eta_{ij} = D_{ij}(r_{ij}) \int\int \frac{\varphi_i(\mathbf{r})\varphi_j(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3r\, d^3r', \tag{17}$$

where D_{ij} is a short-ranged pairwise damping function between atoms i and j (see below), $\nu_f(\mathbf{r})$ is the electrostatic potential of applied electric fields, and $\phi_j(\mathbf{r})$ is the electrostatic potential due to a second-order multipole expansion of the Mulliken partitioned atomic density (i.e., the Mulliken atomic charge and dipole) of atom j. The pairwise damping function takes the form

$$D_{ij}(r_{ij}) = D(r_{ij}; C_i, W_i) \cdot D(r_{ij}; C_j, W_j), \tag{18}$$

where

$$D(r_{ij}; C_i, W_i) = \begin{cases} 0 & \text{if } S < 0 \\ S^3(10 - 15S + 6S^2) & \text{if } 0 \leqslant S \leqslant 1 \\ 1 & \text{if } S > 1, \end{cases} \tag{19}$$

and $S \equiv S(r_{ij}; C_i, W_i)$ is a switching function defined by

$$S(r_{ij}; C_i, W_i) = \left(r_{ij} - \left(C_i - \tfrac{1}{2}W_i\right)\right)/W_i. \tag{20}$$

The switching function $S(r_{ij}; C_i, W_i)$ for atom i is a function of the interatomic separation, r_{ij}, and controlled by the switching parameters C_i (the center of the switching region) and W_i (the width of the switching region), as illustrated in Fig. 1. This model was found to considerably improve the response properties, although alternate models have not been exhaustively explored.

The dipole moment of the system can be expressed as

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 192.12.88.212 On: Fri, 12 Sep 2014 18:52:12

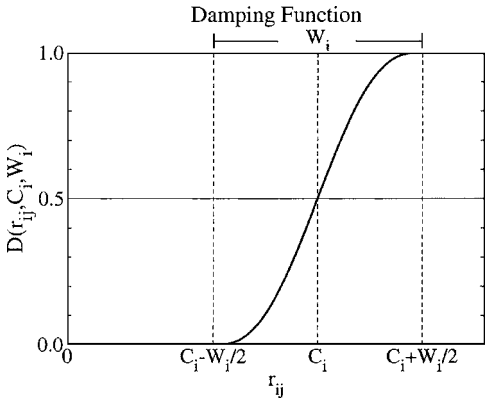

FIG. 1. Illustration of the damping function in Eq. (19) in relation to the parameters C_i and W_i.

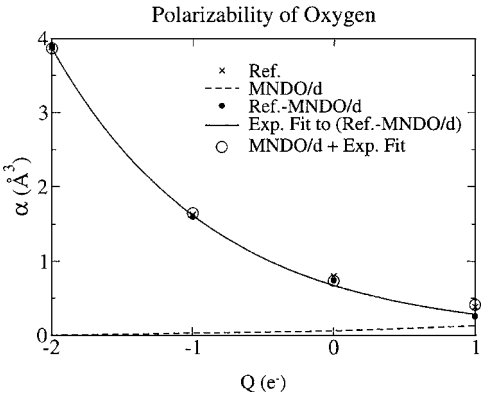

FIG. 2. Polarizability of oxygen as a function of charge. Ref. refers to the experimental results listed in Table I or the DFT values where experimental estimates are unavailable. MNDO/d are the MNDO/d values listed in Table I. Ref.-MNDO/d are the differences between the reference and MNDO/d polarizabilities, i.e., they are the desired CPE polarizability corrections. Exp. Fit to (Ref.-MNDO/d) is a least-squares fit of an exponential function to the desired CPE correction [see Eq. (23)]. MNDO/d+Exp. Fit are the MNDO/d polarizabilities supplemented with a CPE correction resulting from the exponential fit.

$$\mu_u = \mu_u^{\mathrm{SE}} + \sum_{i \in u}^{N} c_i, \tag{21}$$

where μ_u is either the x, y, or z component of the dipole moment, μ_u^{SE} is the semiempirical dipole moment component u, and the summation runs over response basis functions corresponding to component u. The polarizability is computed from finite field differentiation of the dipole moment.

Polarizability is strongly a function of charge state (see Table I). In the CPE polarizability correction model, for a particular atom i with Mulliken charge q_i, the Gaussian exponent $\zeta_i(q_i)$ determines the atomic polarizability correction $\Delta\alpha_i(q_i)$ through the following relationship:

$$\zeta_i(q_i) = \left(\frac{3}{\Delta\alpha_i(q_i)} \sqrt{\frac{\pi}{2}} \right)^{1/3}. \tag{22}$$

Given a series of reference and computed semiempirical polarizabilities corresponding to different atomic charge states (see Table I), the difference of which is the desired polarizability correction, it was observed that $\Delta\alpha_i(q_i)$ can be well represented (see Fig. 2) by a parametrized exponential function of the form

$$\Delta\alpha_i(q_i) \approx \Delta\alpha_i(0)\exp^{-3B_i q_i}, \tag{23}$$

which leads to the relationship

$$\zeta_i(q_i) = \left(\frac{3}{\Delta\alpha_i(0)} \sqrt{\frac{\pi}{2}} \right)^{1/3} \exp^{B_i q_i}, \tag{24}$$

where $\Delta\alpha_i(0)$ and B_i are the independent parameters in the model. As mentioned in Sec. I, the polarizability of isolated atoms represents a rather extreme breakdown of the use of localized minimal basis sets. For this reason, the parameters $\Delta\alpha_i(0)$ and B_i should be adjusted to improve molecular polarizabilities as opposed to atomic polarizabilities. The examination of atomic polarizabilities in Table I and Fig. 2 is useful, however, in providing insight and motivation for the chosen approximate functional form of the charge-dependent behavior of the CPE correction.

In summary, the present work formulates a model whereby the electron density from a semiempirical calculation is used as the reference density that provides the basis for a post-SCF CPE correction term. This term uses a simple damped Coulomb model based on a Mulliken point charge and dipole representation from the semiempirical reference density, along with charge-dependent atomic polarizabilities, to create a linear-response correction that significantly improves atomic and molecular polarizabilities. There are four new independent atomic parameters introduced in this model: $\Delta\alpha_i(0)$ and B_i, which adjust the Gaussian exponent with respect to partial charge and C_i and W_i, which screen the Coulomb interactions at short range. The next section describes the construction of a systematic set of reference data and subsequent model parametrization procedure.

B. Reference data, semiempirical calculations, and parametrization

Molecular dipoles and polarizabilities for 1132 molecules were obtained from the QCRNA online density-functional theory (DFT) database.[41] The complete protocol used to construct the database has been described in detail elsewhere[42] and applied extensively to biological phosphate systems[42,43] and reactions important in RNA catalysis.[44–46] A brief description of the protocol's essential features, as it relates to this work, is now summarized. Density-functional calculations were performed using the B3LYP exchange-correlation functional[47,48] with the 6-31++G(d,p) basis set for geometry optimizations followed by single point energy refinement using the 6-311++G$(3df,2p)$ basis set. Polarizabilities were calculated at B3LYP/6-31++G(d,p) and dipole moments were obtained at B3LYP/6-311++G$(3df,2p)$ level of theory.

At the semiempirical level, treatment of reactions involving biological phosphates requires an accurate d-orbital representation.[49,50] These methods have been applied to phosphate hydrolysis reactions in solution using hybrid quantum-mechanical/molecular mechanical (QM/MM) methods[9,10] and implicit solvation models.[51,52] All semiempirical calculations were performed at the DFT-optimized geometries using the MNDO/d Hamiltonian[50] implemented

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 192.12.88.212 On: Fri, 12 Sep 2014 18:52:12

TABLE II. MNDO/d+CPE parameters.

Atom	$\Delta\alpha(0)$ (Å^3)	B (a.u.)	C (Å)	W (Å)
H	0.153 952 332 6	3.000 000 000 0	3.219 124 687 0	0.500 383 232 3
C	0.353 947 441 2	1.612 172 447 4	1.228 391 575 1	0.500 000 000 0
N	0.522 941 855 9	0.453 729 824 4	1.194 685 057 9	0.500 000 000 0
O	0.256 531 909 1	0.599 999 976 2	1.698 055 077 3	1.153 712 830 1
Na[a]	0.000 000 000 0	0.000 000 000 0	3.600 000 000 0	1.700 000 000 0
P	1.156 634 671 6	0.300 000 000 0	0.429 556 901 9	0.676 132 367 2
S	1.199 009 815 3	0.520 249 266 4	1.832 954 240 4	0.666 491 012 9
Cl	1.202 270 615 2	0.300 454 686 9	2.319 485 909 8	0.503 013 377 8

[a]Na was not given a response dipole function because the cation polarizability should be negligible; however, it does still participate in the damping of the Coulomb response of other atoms.

in a modified version of MNDO97.[53] Note that atoms in the first and second periods do not utilize d orbitals, and for these atoms the MNDO/d Hamiltonian is equivalent to MNDO. The convergence criterion for the SCF energy was 10^{-6} eV and 0.001 a.u. electric fields were used in the finite field determination of the polarizability.

Of the 1132 molecules selected from the QCRNA database,[41] 133 were randomly selected for parametrization. To avoid ambiguity in the calculation of the dipole moment, only the neutral molecule dipole moments were included in the parametrization (66 molecules). The charge states represented by the other randomly selected molecules are 3(2+), 17(1+), and 47(1−).

The four atom-based parameters $\Delta\alpha_i(0)$, B_i, C_i, and W_i were adjusted for the atoms H, C, N, O, Na, P, S, and Cl to minimize the χ^2 function

$$\chi^2 = \sum_{i}^{N=133} w_\alpha(\alpha_i - \alpha_{\mathrm{ref},i})^2 + \sum_{i}^{N_0=66} w_\mu(\mu_i - \mu_{\mathrm{ref},i})^2, \qquad (25)$$

where the first term sums over the 133 polarizabilities weighted by $w_\alpha = 11.0$ Å^{-6} and the second term sums over the 66 neutral molecule dipole moments with weights $w_\mu = 100.0$ D^{-2}. The larger weight for the dipoles is chosen due to the smaller number of sample points and the relatively smaller reference values compared to the polarizabilities. Parameter space was explored with a combination of genetic algorithm[54,55] and direction set minimization.[56] Genetic algorithm optimization procedures have been shown to be very useful in semiempirical parameter optimizations.[57–60] The implementation of the genetic algorithm was based in part on the description by Goldberg.[54]

III. RESULTS AND DISCUSSION

The parameters resulting from the optimization to the 133 randomly selected molecules are presented in Table II. The polarizabilities of the 1132 molecules from the QCRNA database[41] were computed with MNDO/d and MNDO/d supplemented with the CPE correction to the electron density, denoted as MNDO/d+CPE. The neutral molecule dipole moments and molecular polarizabilities for the entire set of data are tabulated and presented as supplementary material.[61]

A. Polarizability of atoms and monatomic ions

The development of the MNDO/d+CPE model with charge-dependent atomic polarizabilities requires a more quantitative understanding of the polarizability of atoms and their positively and negatively charged ions. Table I compares experimentally estimated and *ab initio* computed polarizabilities of several atoms and their ions to values obtained from several conventional semiempirical methods, including PM3,[62] AM1,[63] and MNDO/d (Refs. 50 and 64) from finite field differentiation of the dipole moment. As the number of electrons on an atom increases, the polarization should increase; however, semiempirical models often exhibit the opposite behavior. In the limit that the valence spin-orbitals become occupied, no polarization can be achieved. The MNDO/d sulfur dianion does not reach this severe limit due to the mixing of auxiliary d-basis functions with the valence p-basis functions. In the less severe case of neutral atoms, the polarizability is underestimated by approximately a factor of 10. Previous studies have found semiempirical methods to underestimate molecular polarizabilities by approximately 25% when compared to experiment.[24]

B. Dipole moments of neutral molecules

Table III compares the MNDO/d and MNDO/d+CPE dipole moments with the reference DFT values for 491 neutral molecules. The relative errors for the MNDO/d dipole moments are two to three times smaller than the corresponding relative errors in the polarizability (discussed in the next section), suggesting that the semiempirical methods are considerably more accurate in determination of the reference

TABLE III. Dipole error statistics. The columns in respective order indicate molecular charge (Q), number of molecules (N), and errors statistics on the dipole moments in Debye: average reference value (AVG), mean signed error (MSE), mean unsigned error (MUE), maximum error (MAX), and root mean square error (rms). The numbers in parentheses are percent relative error values (kEL). MNDO/d+CPE error statistics are also in normal font, and MNDO/d error statistics are shown in italics on the line immediately below. The eight molecules whose symmetry prohibit formation of a new dipole moment were excluded from the statistics.

Q	N	AVG	MUE (REL)	MSE (REL)	MAX (REL)	rms
0	491	2.490	0.250(11.8)	0.027(2.6)	2.476(176.0)	0.376
			0.517(25.0)	*0.253(13.4)*	*3.295(484.7)*	*0.708*

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 192.12.88.212 On: Fri, 12 Sep 2014 18:52:12

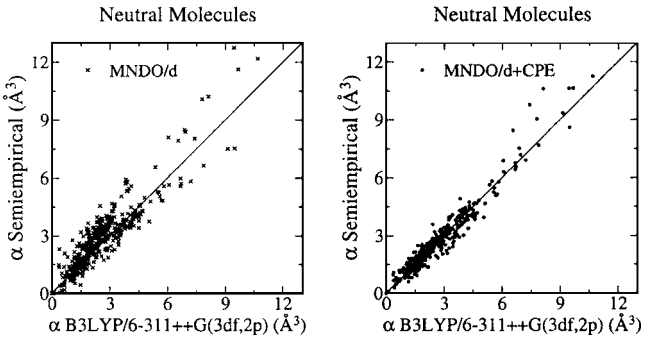

FIG. 3. Comparison of the semiempirical dipole moments with MNDO/d +CPE and without MNDO/d the CPE correction to B3LYP/ 6-311++G(3*df*,2*p*).

density (related to the relative atomic electronegativity values) than their ability to respond to applied fields and long-range electrostatics. Nonetheless, MNDO/d+CPE further improves the mean unsigned error (MUE) in the dipole moments by a factor of 2 and the mean signed error (MSE) by an order of magnitude. This improvement is illustrated in Fig. 3. It has been demonstrated previously that the NDDO-based semiempirical methods such as MNDO/d provide fairly accurate electron densities, from which accurate one-electron properties such as electrostatic potentials can be derived.[65] This aspect is exploited in the MNDO/d+CPE model to obtain improved charge-dependent semiempirical response properties.

C. Polarizabilities of neutral molecules, cations, and anions

The results of the comparison with the B3LYP/ 6-31++G(*d*,*p*) (DFT) polarizabilities are presented in Table IV and are broken down by charge state. It is interesting to note that the MUE and MSE of MNDO/d (italicized) are consistently equal in magnitude and opposite in sign, indicating that MNDO/d systematically underestimates the polarizability by 30%–40%. When considering all molecules of every charge state or any particular charge state, the MNDO/d+CPE MUE and MSE values are reduced by an order of magnitude relative to the corresponding MNDO/d MUE and MSE values. The systematic improvement of the MNDO/d+CPE polarizabilities relative to the MNDO/d values is illustrated in Fig. 4, which plots the regression of the MNDO/d and MNDO/d+CPE polarizabilities with respect to the reference DFT values for each charge state. The ability to make such drastic improvement over all charge states is directly related to the use of charge-dependent polarizability correction. Note that the parametrization set contained only three molecules with a 2+ charge and *no* molecules with a 2− charge, yet an order of magnitude improvement in the errors, are observed, suggesting that the exponential behavior chosen for the charge-dependent polarizability is quite reasonable.

D. $Na^+\cdots H_2O$ dipole and polarizability curves

To better demonstrate the improvement of the response properties to applied fields and long-range electrostatic interactions, the induced dipole moment of $Na^+\cdots H_2O$ was examined with B3LYP/6-311++G(3*df*,2*p*) and compared to MNDO/d and MNDO/d+CPE (Fig. 5). The dipole moment of a charged system depends on the origin, and in order to make comparison, the $Na^+\cdots H_2O$ system was for all calculations constructed such that the sodium ion was located at the origin and the oxygen of the water was located at some distance R_{Na-O} away. The internal geometry of the water was held fixed at the experimental gas-phase water geometry and orientated such that the sodium ion lies along the C_2 axis of the water, with the oxygen closer to the ion than the hydrogens. The dipole polarizability of Na^+ is negligibly small and the choice of origin allows the induced dipole moment to be interpreted as contributions arising from the dipole response of the water and the charge transfer between the water and the sodium ion. In the long-range limit ($R_{Na-O} > 4$ Å), MNDO/d underestimates the induced dipole moment by approximately 20%–30%, whereas MNDO/d+CPE agrees reasonably well with B3LYP/6-311++G(3*df*,2*p*). These observations are in direct agreement with the polarizability error statistics discussed earlier. At small distances (R_{Na-O}

TABLE IV. Polarizability error statistics. The columns in respective order indicate molecular charge (Q), number of molecules (N), and errors statistics on the polarizability values in Å^3: average reference value (AVG), mean signed error (MSE), mean unsigned error (MUE), maximum error (MAX), and root mean square error (rms). The numbers in parentheses are percent relative error values (REL). MNDO/d+CPE error statistics are shown in normal font, and MNDO/d error statistics are shown in italics on the line immediately below.

Q	N	AVG	MUE (REL)	MSE (REL)	MAX (REL)	rms
ALL	1132	14.312	0.471(3.6)	−0.051(0.1)	−6.492(50.6)	0.747
			5.413(38.0)	*−5.411(−37.9)*	*−18.603(−88.1)*	*5.832*
0	499	13.075	0.352(3.1)	0.271(2.2)	1.603(38.2)	0.413
			4.093(32.2)	*−4.089(−32.2)*	*−9.091(−58.1)*	*4.325*
−1	488	15.428	0.467(3.3)	−0.290(−1.9)	−4.963(50.6)	0.745
			6.398(42.3)	*−6.398(−42.3)*	*−10.197(−88.1)*	*6.530*
+1	74	12.662	0.412(4.1)	0.348(3.6)	1.222(20.8)	0.497
			4.410(34.0)	*−4.410(−34.0)*	*−9.791(−49.4)*	*4.932*
−2	56	18.623	1.683(8.8)	−1.460(−7.4)	−6.492(−24.1)	2.130
			10.228(55.4)	*−10.228(−55.4)*	*−18.603(−73.8)*	*10.532*
+2	15	11.166	0.351(3.6)	0.294(3.2)	0.823(7.4)	0.419
			4.252(37.9)	*−4.252(−37.9)*	*−8.662(−43.3)*	4.829

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 192.12.88.212 On: Fri, 12 Sep 2014 18:52:12

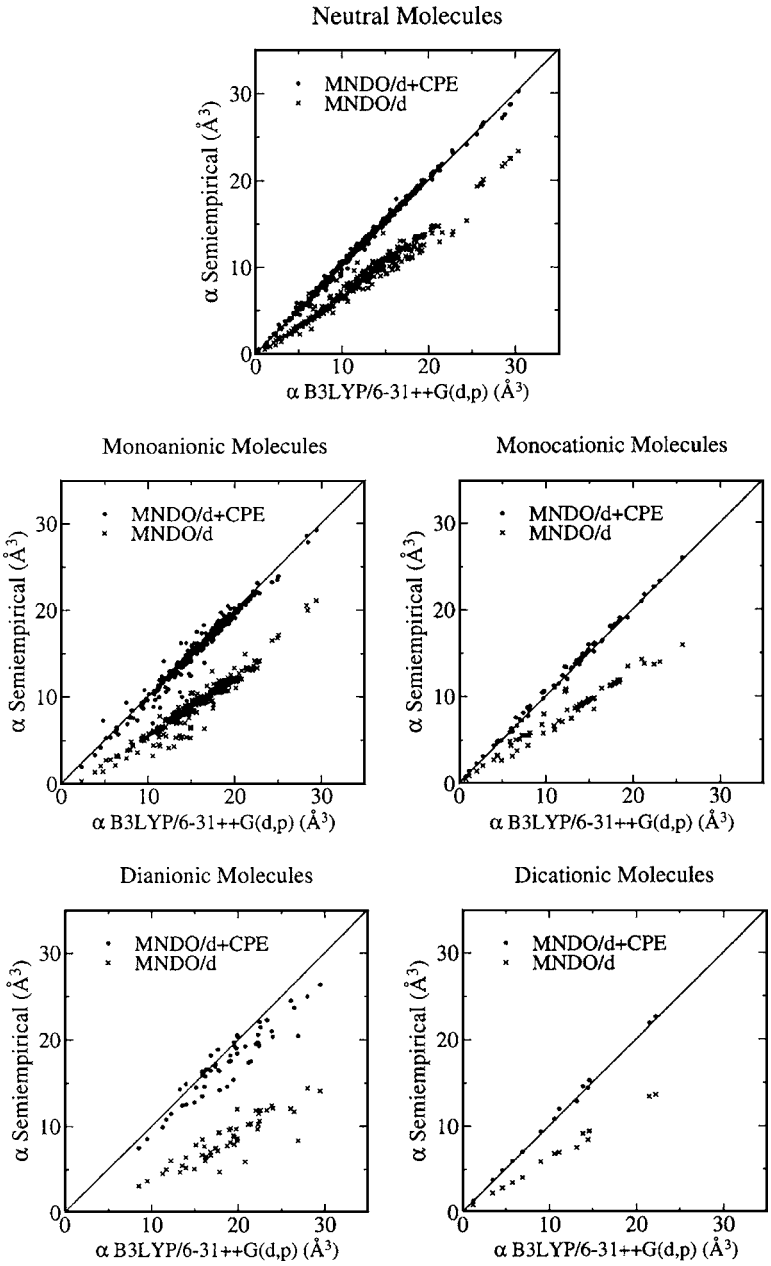

FIG. 4. Comparison of the semiempirical polarizabilities with MNDO/d+CPE and without MNDO/d the CPE correction to B3LYP/6-31++G(d,p).

<3 Å), MNDO/d+CPE converges to the MNDO/d result due to the short-range damping of the Coulomb interactions.

The MNDO/d-induced dipole moment is larger than the DFT values at small separation. Examination of the Mulliken charge of the ion indicates that the semiempirical charge transfer between the ion and water is exaggerated. One mechanism for correcting this error is to include monopole response functions into the CPE correction; however, as

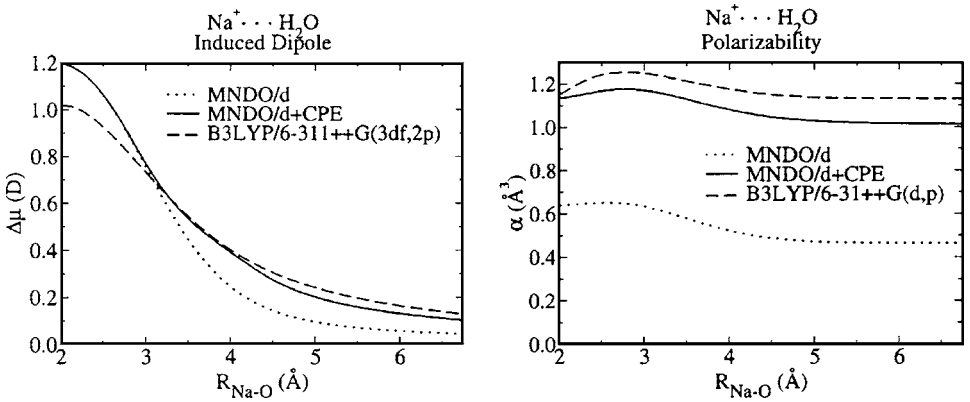

FIG. 5. The induced dipole moment and polarizability of the $Na^+\cdots H_2O$ complex as a function of Na–O separation. Na is located at the origin and the water geometry is fixed at the experimental values and oriented such that the oxygen is closer to the sodium than the hydrogens, located along the z axis, and the hydrogens are located in the x-y plane. For this system, none of the atoms utilize d orbitals, and thus the MNDO/d model is equivalent to MNDO.

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 192.12.88.212 On: Fri, 12 Sep 2014 18:52:12

already discussed, proper treatment of charge transfer within the CPE framework is still an area of active development.[36,38,39] A more tractable approach in improving the short-range interactions is through reparametrization of the semiempirical parameters. It is likely that the semiempirical parameters can be adjusted to improve short-range interactions without sacrificing the ability to properly polarize since the CPE correction, by nature of the short-range screening, improves the density response to moderate-to-long-range interactions. Thus, the CPE correction is an attractive approach relative to previous work[66,67] which adjusts the semiempirical parameters specifically to reproduce nonlinear optical properties, since such a specific reparametrization may otherwise reduce the ability to accurately describe bond strengths. To accommodate response properties while still having the parametric freedom to accurately describe covalent bonding, the use of extended basis sets has been proposed;[68–70] however, the increased size of the resulting Fock matrix and consequent increase in computational time required to obtain converged energies and forces quickly become prohibitively expensive to apply in QM/MM simulations. Therefore, the proposed CPE correction for the improvement of long-range response interactions is an attractive alternative.

IV. CONCLUSION

The present work addresses a major shortcoming in the performance of conventional semiempirical quantum models most commonly used in linear scaling electronic structure and combined QM/MM calculations of biological molecules and reactions: inadequate treatment of charge-dependent polarizability. This response property is of fundamental importance to reliably model highly charged systems such as metalloenzymes, ribozymes, and many other biological systems. The deficiency in conventional semiempirical models is primarily related to the exclusive use of a localized minimal atomic-orbital basis set for description of the electron-density response. It has been shown that with these models, polarizability tends to decrease with an increasing anionic charge, especially for atoms and small molecules, in stark contrast with the behavior observed experimentally and with high-level density-functional calculations. As anionic charge builds up, low-lying virtual orbitals become populated and are not available to elicit a second-order polarization response. The proposed MNDO/d+CPE model corrects this behavior in a manner that does not increase the atomic-orbital basis, retains a set of linear-response equations, and avoids the problem of artificial charge transfer and scaling of the polarizability seen in related models that allow atomic charges to fluctuate.

The proposed MNDO/d+CPE model introduces four new atom-based parameters. These parameters were fitted against polarizabilities and dipole moments for a set of 133 randomly selected molecules out of a database of 1132 molecules and ions. The resulting MNDO/d+CPE model had considerably reduced errors in the polarizability by an order of magnitude relative to MNDO/d across the entire set and performed reliably for neutral, cationic, and anionic systems. Errors in the dipole moment were decreased by a factor of 2. The drastic decrease in polarizability error is observed for a wide range of molecular charge states (from 2+ to 2−) and even for charge states that were not involved in the training set (2− charge state). The ability of the MNDO/d+CPE method to reproduce polarizabilities over such a wide range of molecules and charge states is tied to the use of charge-dependent polarizability corrections. The neglect of monopolar response functions at the CPE level circumvents well-known problems of artificial charge transfer and scaling of the polarizability inherent in conventional many-body linear-response models. The present work demonstrates a considerable step forward in the design of next generation semiempirical quantum models for accurate simulations of highly charged biological reactions such as phosphoryl transfer in kinases and phosphatases, and RNA catalysis.

ACKNOWLEDGMENTS

One of the authors (D.M.Y.) is grateful for financial support provided by the National Institute of Health (Grant No. GM62248), and the Army High Performance Computing Research Center (AHPCRC) under the auspices of the Department of the Army, Army Research Laboratory (ARL) under Cooperative Agreement No. DAAD19-01-2-0014. Computational resources were provided by the Minnesota Supercomputing Institute.

[1] W. Thiel, in *Advances in Chemical Physics*, edited by I. Prigogine and S. A. Rice (Wiley, New York, 1996), Vol. 93, pp. 703–757.
[2] T. Clark, J. Mol. Struct.: THEOCHEM **530**, 1 (2000).
[3] W. Thiel, in *Handbook of Molecular Physics and Quantum Chemistry*, edited by S. Wilson (Wiley, Chichester, 2003), Vol. 2, pp. 487–502.
[4] T. Bredow and K. Jug, Theor. Chem. Acc. **113**, 1 (2005).
[5] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[6] M. Garcia-Viloca, D. G. Truhlar, and J. Gao, J. Mol. Biol. **327**, 549 (2003).
[7] J. Khandogin, K. Musier-Forsyth, and D. M. York, J. Mol. Biol. **330**, 993 (2003).
[8] J. Khandogin and D. M. York, Proteins **56**, 724 (2004).
[9] B. A. Gregersen, X. Lopez, and D. M. York, J. Am. Chem. Soc. **125**, 7178 (2003).
[10] B. A. Gregersen, X. Lopez, and D. M. York, J. Am. Chem. Soc. **126**, 7504 (2004).
[11] P. Winget, C. Selçuki, A. Horn, B. Martin, and T. Clark, Theor. Chem. Acc. **110**, 254 (2003).
[12] M. I. Bernal-Uruchurtu, M. T. C. Martins-Costa, and M. F. R.-L. C. Millot, J. Comput. Chem. **21**, 572 (2000).
[13] M. Kolb and W. Thiel, J. Comput. Chem. **14**, 775 (1993).
[14] W. Weber and W. Thiel, Theor. Chem. Acc. **103**, 495 (2000).
[15] M. Bernal-Uruchurtu and M. Ruiz-López, Chem. Phys. Lett. **330**, 118 (2000).
[16] T. J. Giese, E. C. Sherer, C. J. Cramer, and D. M. York, J. Chem. Theory Comput. (in press).
[17] M. Elstner, T. Frauenheim, E. Kaxiras, G. Seifert, and S. Suhai, Phys. Status Solidi B **217**, 357 (2000).
[18] M. Elstner, Q. Cui, P. Munih, E. Kaxiras, T. Frauenheim, and M. Karplus, J. Comput. Chem. **24**, 565 (2003).
[19] A. Warshel, *Computer Modeling of Chemical Reactions in Enzymes and Solutions* (Wiley, New York, 1991).
[20] A. Warshel, Acc. Chem. Res. **35**, 385 (2002).
[21] M. Garcia-Viloca, J. Gao, M. Karplus, and D. G. Truhlar, Science **303**, 186 (2004).
[22] W. A. Parkinson and M. C. Zerner, J. Chem. Phys. **94**, 478 (1991).
[23] M. Dewar, Y. Yamaguchi, and S. Suck, Chem. Phys. Lett. **59**, 541 (1974).

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 192.12.88.212 On: Fri, 12 Sep 2014 18:52:12

[24] N. Matsuzawa and D. A. Dixon, J. Phys. Chem. **96**, 6232 (1992).
[25] G. Schürer, P. Gedeck, M. Gottschalk, and T. Clark, Int. J. Quantum Chem. **75**, 17 (1999).
[26] B. Martin, P. Gedeck, and T. Clark, Int. J. Quantum Chem. **77**, 473 (2000).
[27] D. Rinaldi and J. Rivail, Theor. Chim. Acta **32**, 57 (1974).
[28] J. L. Rivail and A. Cartier, Mol. Phys. **36**, 1085 (1978).
[29] D. Rinaldi and J. Rivail, Theor. Chim. Acta **32**, 243 (1974).
[30] D. M. York and W. Yang, J. Chem. Phys. **104**, 159 (1996).
[31] W. J. Mortier, K. Van Genechten, and J. Gasteiger, J. Am. Chem. Soc. **107**, 829 (1985).
[32] R. F. Nalewajski, J. Phys. Chem. **89**, 2831 (1985).
[33] A. K. Rappe and W. A. Goddard III, J. Phys. Chem. **95**, 3358 (1991).
[34] S. W. Rick, S. J. Stuart, and B. J. Berne, J. Chem. Phys. **101**, 6141 (1994).
[35] J. Morales and T. J. Martínez, J. Phys. Chem. A **105**, 2842 (2001).
[36] R. Chelli, M. Pagliai, P. Procacci, G. Cardini, and V. Schettino, J. Chem. Phys. **122**, 074504 (2005).
[37] J. Morales and T. J. Martínez, J. Phys. Chem. A **108**, 3076 (2004).
[38] S. M. Valone and S. R. Atlas, J. Chem. Phys. **120**, 7262 (2004).
[39] J. Cioslowski and B. B. Stefanov, J. Chem. Phys. **99**, 5151 (1993).
[40] T. J. Giese and D. M. York, J. Chem. Phys. **120**, 7939 (2004).
[41] T. J. Giese, B. Gregersen, J. Khandogin, Y. Liu, E. Mayaan, A. Moser, K. Nam, K. Range, and D. M. York, http://riesling.chem.umn.edu/QCRNA
[42] K. Range, M. J. McGrath, X. Lopez, and D. M. York, J. Am. Chem. Soc. **126**, 1654 (2004).
[43] E. Mayaan, K. Range, and D. M. York, JBIC, J. Biol. Inorg. Chem. **9**, 807 (2004).
[44] C. S. López, O. N. Faza, B. A. Gregersen, X. Lopez, A. R. de Lera, and D. M. York, ChemPhysChem **5**, 1045 (2004).
[45] C. S. López, O. N. Faza, A. R. de Lera, and D. M. York, Chem.-Eur. J. **11**, 2081 (2005).
[46] Y. Liu, X. Lopez, and D. M. York, Chem. Commun. (Cambridge) **31**, 3909 (2005).
[47] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[48] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[49] W. Thiel and A. A. Voityuk, Theor. Chim. Acta **81**, 391 (1992).
[50] W. Thiel and A. A. Voityuk, J. Phys. Chem. **100**, 616 (1996).
[51] J. Khandogin, B. A. Gregersen, W. Thiel, and D. M. York, J. Phys. Chem. B **109**, 9799 (2005).
[52] B. A. Gregersen, J. Khandogin, W. Thiel, and D. M. York, J. Phys. Chem. B **109**, 9810 (2005).
[53] W. Thiel, Program MNDO97 (University of Zurich, Zurich, Switzerland, 1998).
[54] D. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning* (Addison-Wesley, Reading, MA, 1989).
[55] D. A. Coley, *An Introduction to Genetic Algorithms for Scientists and Engineers* (World Scientific, Singapore, 1999).
[56] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and W. P. Flannery, *Numerical Recipes in Fortran*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[57] I. Rossi and D. G. Truhlar, Chem. Phys. Lett. **233**, 231 (1995).
[58] T. R. Cundari, J. Deng, and W. Fu, Int. J. Quantum Chem. **77**, 421 (2000).
[59] A. A. Voityuk and N. Rösch, J. Phys. Chem. A **104**, 4089 (2000).
[60] M. C. Hutter, J. R. Reimers, and N. S. Hush, J. Phys. Chem. B **102**, 8080 (1998).
[61] See EPAPS Document No. E-JCPSA6-123-316538 for further discussion. This document can be reached via a direct link in the online article's HTML reference section or via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html).
[62] J. J. P. Stewart, J. Comput. Chem. **10**, 209 (1989).
[63] M. J. S. Dewar, E. Zoebisch, E. F. Healy, and J. J. P. Stewart, J. Am. Chem. Soc. **107**, 3902 (1985).
[64] M. J. Dewar and W. Thiel, J. Am. Chem. Soc. **99**, 4899 (1977).
[65] D. Bakowies and W. Thiel, J. Comput. Chem. **17**, 87 (1996).
[66] V. J. Docherty, D. Pugh, and J. O. Morley, J. Chem. Soc., Faraday Trans. 2 **81**, 1179 (1985).
[67] J. O. Morley, V. J. Docherty, and D. Pugh, J. Chem. Soc., Perkin Trans. 2 **1987**, 1361.
[68] M. G. Papadopoulos, J. Waite, and C. A. Nicolaides, J. Chem. Phys. **77**, 2527 (1982).
[69] J. Waite, M. G. Papadopoulos, and C. A. Nicolaides, J. Chem. Phys. **77**, 2536 (1982).
[70] J. Waite and M. G. Papadopoulos, J. Chem. Phys. **82**, 1427 (1985).
[71] *CRC Handbook of Chemistry and Physics*, 83rd ed., edited by D. R. Lide (CRC, Boca Raton, FL, 2003).

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 192.12.88.212 On: Fri, 12 Sep 2014 18:52:12