Alternative Approaches for the Calculation of Induction Energies: Characterization, Effectiveness, and Pitfalls

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY A · NOVEMBER 2001

Impact Factor: 2.69 · DOI: 10.1021/jp012393+

6 AUTHORS, INCLUDING:

Chris Chipot
French National Centre for Scientific Researc…
163 PUBLICATIONS **11,778** CITATIONS
SEE PROFILE

János G Ángyán
French National Centre for Scientific Research
160 PUBLICATIONS **5,103** CITATIONS
SEE PROFILE

Claude Millot
University of Lorraine
60 PUBLICATIONS **1,594** CITATIONS
SEE PROFILE

Alternative Approaches for the Calculation of Induction Energies: Characterization, Effectiveness, and Pitfalls

Christophe Chipot,*,† François Dehez,†,‡ János Ángyán,† Claude Millot,† M. Orozco,§ and F. Javier Luque||

Equipe de chimie et biochimie théoriques, Institut nancéien de chimie moléculaire, Unité mixte de recherche CNRS/UHP 7565, Université Henri Poincaré, BP 239, 54506 Vandœuvre-lès-Nancy Cedex, France, CEA Valrho, DEN/DTE/STME/LPCM, BP 111, 26702 Pierrelatte Cedex, France, Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Martíi Franquès 1, E−08028 Barcelona, Spain, and Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Avgda. Diagonal s/n, E−08028 Barcelona, Spain

Received: June 25, 2001; In Final Form: October 4, 2001

One of the practical difficulties precluding the generalized development of nonadditive, polarizable models for statistical simulations is rooted in the costly estimation of accurate induction energies, from which distributed polarizabilities can be derived. From a finite perturbation (FP) perspective, mapping the induction energy over a grid of points implies as many distinct quantum chemical calculations of the molecule interacting with a polarizing charge as the total number of points. Here, two alternative routes for computing accurate induction energies in a time-bound fashion are explored. The first one is based upon second-order perturbation theory and only involves a single quantum chemical calculation at the Hartree−Fock level of approximation to map the induction energy. The second one, less straightforward in its implementation, relies on a topological partitioning of the response charge density, also evaluated from a single quantum chemical calculation, yet at virtually any level of sophistication. Critical comparison with reference FP computations reveals that only appropriate scaling of the perturbative (PT) induction energies can warrant a faithful description of polarization phenomena. In the case of neutral molecules, a reasonable reproduction of molecular dipole polarizabilities is achieved when use is made of a simple scaling function that solely depends on the distance separating the points of the grid from the center of mass of the molecule. For anions, the marked anisotropy in the deviation of the PT induction energies from the target FP ones makes the definition of such a general scaling function virtually impossible. In sharp contrast, the approach based upon the topological partitioning of the response charge density does not require any adjustment or scaling, and, thus, constitutes a more robust and rigorous strategy for the computation of induction energies. Examination of distinct protocols for mapping the induction energy emphasizes the necessity to sample the space around the molecule far enough from the nuclei to reproduce molecular dipole polarizabilities accurately. Compared to the spatial extent of the grid, the density of points appears to be of lesser importance.

1. Introduction

The assumption that induction effects can be accounted for in an average fashion by means of a simple, appropriate parameterization is at the origin of the durable success of pairwise, additive potential energy functions for cost-effective statistical simulations of liquids. An important ingredient in the development of such nonpolarizable, effective force fields consists of increasing artificially the polarity of the participating molecules to compensate for missing, through-space intermolecular induction phenomena.[1] A popular approach for implicit polarization is based upon the observation that, compared to the experimental gas-phase quantities, molecular dipole moments computed at the Hartree−Fock (HF) level of theory, using a split-valence 6-31G(d) basis set, are systematically exaggerated.[2] In a number of instances where explicit nonadditivity effects can be neglected, intermolecular potentials parameterized with net atomic charges derived from the HF/6-31G(d) electrostatic potential provide a reasonable description of the properties of the liquid,[3] although explicit, mutual polarization of the constituent molecules and their environment is clearly absent. The advantage of representing the system using pairwise, additive potential energy functions lies in the cost-effectiveness of the statistical simulation, avoiding the computationally demanding estimation of induced moments at the expense of sampling of the phase space. Such an implicit polarization approach is, however, not equivalent to a rigorous, atomic-level description of the molecular response to a nonuniform, external electric field, thus constituting the inherent limitation of the approach. This is particularly relevant in statistical simulations of ionic species or highly polarizing systems interacting with polarizable ones. Furthermore, explicit polarization has proven to be pivotal in those cases where intramolecular induction effects are sizable.[4,5]

The massive increase of computational resources witnessed in recent years, pushing back the limitations of statistical

* Corresponding author: chipot@lctn.uhp-nancy.fr
† Université Henri Poincaré.
‡ CEA Valrho.
§ Departament de Bioquímica i Biologia Molecular, Universitat de Barcelona.
|| Departament de Fisicoquímica, Universitat de Barcelona.

10.1021/jp012393+ CCC: $20.00
Published on Web 11/21/2001

simulations, has allowed molecular assemblies of appreciable complexity to be tackled over realistic time-scales, using sophisticated potential energy functions. The generalized availability of such resources has inevitably contributed to the recent interest of nonadditive force fields, opening the way to the development of alternative approaches for modeling induction effects in chemical systems. To a large extent, the success of molecular simulations relying on polarizable potential energy functions is closely coupled to the ability of the modeler to design such functions. Arguably, the most popular scheme for constructing models of atomic polarizabilities is the one proposed by Applequist.[6] In essence, his method is based on a self-consistent determination of parameters characterizing the atomic polarizability that see each other through screened dipole−dipole interactions. Just like its refined version devised by Thole,[7] this heuristic approach implies a substantial component of arbitrariness in the parameterization of the model. In contrast, the partitioning scheme put forward by Stone,[8−9] combining an earlier formulation of the susceptibility function of the charge density[10] with the distributed multipole analysis (DMA) method,[11] is clearly more rigorous. Rational models of distributed polarizabilities are derived from high-quality quantum mechanical (QM) calculations of the response of an isolated molecule to an external perturbation, thereby providing one- and two-center charge−charge, charge−dipole, dipole−dipole, etc. parameters. A pivotal aspect of this approach concerns the partitioning of the response charge density into atomic and nonlocal contributions. Closely related, an alternative procedure supplying reasonably transferable, distributed polarizabilities[12] relies on a topological partitioning of the molecular space into atomic regions, according to the theory of "atoms in molecules" (AIM) devised by Bader.[13] The derived polarizability parameters have been shown to reproduce accurately, in a given region, the induced moments due to a local electrostatic potential and its successive derivatives experienced at another site. Inherently nonlocal,[8,9,13] the resulting models include a plethora of terms that rapidly become cumbersome to handle, especially in the context of statistical simulations. For the latter, the number of polarizability components should be reduced significantly to yield compact and tractable sets of parameters, limited to low-order terms, e.g., charge flow and one-center dipole polarizabilities, capable of describing induction phenomena at a minimal cost. As a result, the above methods are not necessarily compatible with such requirements, and the quality of the generated models strongly depends on their level of sophistication. For instance, limiting the description to atomic dipole polarizabilities, accurate reproduction of molecular quantities beyond the order of the dipole is hopeless.

Atomic point charges derived from the quantum mechanical electrostatic potential computed on a grid of points around the molecule[14−16] are largely utilized for parameterizing the Coulomb part of all-purpose force fields. In the spirit of this approach, a number of schemes targeted at the construction of models of distributed polarizabilities have been devised, based on a least-squares fitting procedure to the induction energy.[17−20] Their strength resides in the possibility to generate easily compact and flexible sets of polarizability parameters at any given order. Contrasting with the electrostatic potential, directly available from the wave function, the induction energy is somewhat more difficult to attain. In its most straightforward formulation, referred to as finite perturbation (FP),[19] it can be estimated by considering the interaction of the molecule with a nonpolarizable point charge. Whereas potential derived charges imply a single QM calculation to map the entire space around the molecule, discretized in the form of a grid of N_p points, evaluating induction energies for the same grid via FP would require N_p distinct QM calculations, hence constituting the computational bottleneck of the approach. Clearly, the significant computational effort involved in the determination of induction energies, including usually intramolecular electron correlation and employing large basis sets to warrant an appropriate reproduction of the molecular polarizabilities, limits the applicability of FP to sufficiently small, prototypical systems. In light of these conclusions, two alternatives were developed concurrently, aimed at a faster estimation of induction energies and relying essentially upon a single QM calculation − a *sine qua non* condition for the development of nonheuristic models of distributed polarizabilities. The first scheme, based on second-order perturbation theory, allows the generation of large grids of induction energies derived from one QM calculation at the HF level.[21,22] In essence, it relies upon an uncoupled form of the HF equations and provides only an approximation of the exact induction energy. Using adequately chosen scaling factors, this perturbative approach has been shown to yield induction energies in good agreement with the FP quantities.[22] The second, more elaborate scheme consists of computing induction energies from fully distributed models of polarizabilities obtained within the framework of the AIM theory.[12] The key feature of this approach lies in its ability to map grids of induction energies from a single QM calculation at a sophisticated level of approximation, followed by the topological partitioning of the electron density response into atomic regions.[23]

The main thrust of the present contribution is the search for an optimal, cost-effective approach for constructing models of distributed polarizabilities capable of describing accurately nonadditive phenomena in statistical simulations. A particular effort has been invested in the past decades in the optimization of basis sets of reasonable size, adapted to the reproduction of electric properties. Here, the molecular dipole polarizabilities for a series of small, prototypical molecules have been computed using the ELP,[24] the Sadlej,[25] and the Spackman[26] basis sets. It will be shown that the second probably constitutes the best compromise in terms of precision and computational effort for the evaluation of induction energies, and, hence, the development of models of distributed polarizabilities. Next, the merits and the limitations of two alternative approaches for calculating induction energies, namely the method based on second-order perturbation theory and that relying upon a topological partitioning of the response charge density, will be discussed. To analyze quantitatively the accuracy of these approaches, the computed induction energies of a series of molecules will be compared with those obtained via FP methodology. An important aspect in the derivation of distributed polarizabilities concerns the design of the grid of points over which the induction energy is mapped. The density of points, as well as the radial extent of the grid will be analyzed by considering how they affect the fitted polarizabilities. Last, models of distributed polarizabilities, including charge flow and isotropic dipole polarizabilities, will be built using the induction energies resulting from second-order perturbation theory and the topological partitioning of the response charge density. As a rigorous assessment of their respective quality, the molecular polarizabilities regenerated from these models will be compared with those obtained from second-order Møller−Plesset (MP2) computations and those determined experimentally.

2. Methods and Computational Details

2.1. Effects of Geometry and Basis Set. A critical issue in the construction of models of distributed polarizabilities is the

level at which the induction energy is computed to guarantee an accurate regeneration of molecular polarizabilities. It has been demonstrated that the faithful reproduction of these quantities via QM calculations requires a high level of sophistication.[27] In particular, to attain results in close agreement with the available experimental data, the introduction of intramolecular electron correlation, at least at the MP2 level of approximation, and the use of large, flexible basis sets are mandatory. Recent calculations of molecular polarizabilities employing the triple-ζ-like 6-311++G(2d,2p) basis set, have revealed that the latter was far from optimal for that purpose, always providing underestimated quantities.[23] Considerable effort has been devoted in recent years to the development of alternative basis sets aimed at a quantitatively accurate description of electric properties in molecular systems. In the present contribution, the merits of the ELP,[24] the Sadlej,[25] and the Spackman[26] basis sets for the computation of molecular dipole polarizabilities will be analyzed for H_2O, NH_3, H_2S, HCN, C_2H_2, C_2H_4, C_2H_6, C_6H_6, HCOH, HCOOH, and $HCONH_2$. These QM calculations were carried out at the MP2 level of approximation, utilizing two distinct geometries for each molecule, one optimized with the popular 6-31G(d) basis set, largely employed in the parameterization of macromolecular force fields, and the other with the 6-311++G(2d,2p) basis set. All geometry optimizations and estimations of molecular dipole polarizabilities were carried out using the Gaussian 98 suite of programs.[28] In terms of computational investment, the Spackman basis set probably constitutes the most attractive option for computing molecular polarizabilities, as it contains less than half as many basis functions as the ELP basis set. For instance, in the case of C_6H_6, the ELP, the Sadlej, and the Spackman basis sets involve respectively 294, 198, and 126 basis functions.

2.2. Alternative Schemes for Calculating Induction Energies. One of the fundamental issues that this article proposes to address concerns the comparison of approximate methods for the reliable computation of induction energies that can be employed routinely in the development of models of distributed polarizabilities. Within the reference method, referred to as FP,[19] the molecule interacts with a nonpolarizable charge, q_k, located at $\mathbf{r}_k$ on a grid of N_p points. The resulting induction energy can be expressed as

$$\mathcal{U}_{\mathrm{ind},k} = \mathcal{E}_{\mathrm{total},k} - \mathcal{E}^0 - q_k \mathcal{V}(\mathbf{r}_k) \tag{1}$$

$\mathcal{E}_{\mathrm{total},k}$ is the energy of the molecule in the presence of the point charge, requiring one individual QM calculation for each point k of the grid; N_p separate QM calculations are, therefore, needed to map the full grid of induction energies. $\mathcal{E}^0$ is the energy of the isolated molecule. $\mathcal{V}(\mathbf{r}_k)$ is the electrostatic potential at $\mathbf{r}_k$ generated by the isolated molecule.

The first alternative to the variational definition of the induction energy highlighted by eq 1 stems from perturbation theory. At the first order, the interaction energy of a molecule interacting with the nonpolarizable point charge q_k located at $\mathbf{r}_k$ is given by the electrostatic potential

$$\mathcal{V}(\mathbf{r}_k) = \sum_A \frac{Z_A}{|\mathbf{r}_k - \mathbf{R}_A|} - \sum_a^{\mathrm{occ}} \sum_\mu \sum_\nu c_{\mu a}^* c_{\nu a} \left\langle \phi_\mu \left| \frac{1}{|\mathbf{r}_k - \mathbf{r}|} \right| \phi_\nu \right\rangle \tag{2}$$

where $|\phi_\mu\rangle$ stands for a basis function, $c_{\mu a}$ is the coefficient of atomic orbital μ in the occupied molecular orbital a, and Z_A is the nuclear charge of atom A located at $\mathbf{R}_A$. Induction phenomena appear at the second order of perturbation theory:

$$\mathcal{U}_{\mathrm{ind},k} = \left\langle \psi^{(0)} \left| \frac{q_k}{|\mathbf{r}_k - \mathbf{r}|} \right| \psi^{(1)} \right\rangle \tag{3}$$

Here, $|\psi^{(1)}\rangle$ is the first-order corrected molecular wave function. Equation 3 can be approximated within the framework of the HF theory, employing the following expression:[29]

$$\mathcal{U}_{\mathrm{ind},k} \simeq \sum_a^{\mathrm{occ}} \sum_r^{\mathrm{vir}} \frac{1}{\epsilon_a - \epsilon_r} \left[\sum_\mu \sum_\nu c_{\mu a}^* c_{\nu r} \left\langle \phi_\mu \left| \frac{q_k}{|\mathbf{r}_k - \mathbf{r}|} \right| \phi_\nu \right\rangle \right]^2 \tag{4}$$

ϵ_a and ϵ_r correspond, respectively, to the energy of occupied, a, and virtual, r, molecular orbitals of the isolated species. The strength of this PT approach lies in its reduced computational cost, as only one single QM calculation at the level is required to estimate the density matrix.[21,22] It also constitutes its inherent weakness. It should be clearly emphasized here that eq 4 provides only a coarse estimate of the actual induction energy. Furthermore, because it is based upon an uncoupled form of the HF equations, the PT scheme is limited to that level of theory, and the resulting induction energies should be scaled to compare with those derived from QM calculations, including electron correlation effects.

The second, more elaborate method for estimating induction energies equally relies upon a single QM calculation, carried out either at a time-dependent HF (TDHF), coupled perturbed HF (CPHF), or any higher level of theory. From this calculation, a topological analysis of the response charge density is performed to derive the components of distributed polarizabilities, $\alpha_{l\kappa,l'\kappa'}^{ss'} = \alpha_{l\kappa,l'\kappa'}(\mathbf{r}_s,\mathbf{r}_{s'})$, at a given rank l, $l' \leq L$, where L is the highest rank of the multipoles considered in what will be referred to as the model of topologically partitioned electric properties (TPEPs).[12] From the knowledge of these TPEPs, the induction energy resulting from the polarization of the molecule by the nonpolarizable charge q_k can be determined readily:

$$\mathcal{U}_{\mathrm{ind},k} = -\frac{1}{2} q_k^2 \sum_{s,l,m} \sum_{s',l',m'} T_{l\kappa,00}^{sk} \alpha_{l\kappa,l'\kappa'}^{ss'} T_{l'\kappa',00}^{s'k} \tag{5}$$

Here, s and s' denote two sites of the molecule, located at $\mathbf{r}_s$ and $\mathbf{r}_{s'}$, respectively. $T_{l\kappa,00}^{sk}$ is a matrix element of the electrostatic tensor that corresponds to multipole component $l\kappa$, giving at point s the electrostatic potential, or its successive derivatives, created by point charge q_k.[30]

In the present contribution, the induction energies computed using eqs 1, 4, and 5 will be compared for a variety of small, prototypical species, viz., C_2H_6, C_6H_6, $HCONH_2$, CH_3OH, HCOOH, and $HCOO^-$. For each compound, a grid of points consisting of seven concentrical van der Waals surfaces was generated, over which the induction energy was mapped. Use was made of a cutoff distance, r_{cut}, of 7.5 Å, limiting the spatial distribution of the points from any nucleus. Roughly speaking, points were distributed between 3 and 9 Å from the center of mass of each species, with a separation between the layers of $\Delta r = 0.75$ Å, making a total of 892, 1452, 962, 858, 878, and 790 points in the case of C_2H_6, C_6H_6, $HCONH_2$, CH_3OH, HCOOH, and $HCOO^-$, respectively. FP and TPEP induction energies were determined at the MP2/Sadlej level of approximation, whereas PT ones resulted from single HF/Sadlej calculations. All geometries were optimized at the MP2/6-311G++(2d,2p) level. The TPEP expansion was truncated at the

quadrupole and the hexadecapole polarizabilities for hydrogen and heavy atoms, respectively.

What makes the direct use of TPEPs for constructing nonadditive potential energy functions so cumbersome is the extraordinarily large number of components, which rapidly becomes difficult to handle when the size of the molecule increases, e.g., from 5790 for C_2H_6 to 18690 for C_6H_6. From the induction energy mapped over the grid of N_p points, it is, however, possible to derive far simpler, compact models of distributed polarizabilities. In practice, eq 5 will be employed to obtain a reduced set of N_c components $\alpha^{ss'}_{l\kappa,l'\kappa'}$ capable of reproducing the induction energies computed via eq 1, 4, or 5. In the latter case, the target induction energies have been calculated using the full TPEP model described in the previous paragraph. One straightforward route to constructing compact models of distributed polarizabilities consists of solving the normal equations of the least-squares problem.[19,31,32] Another, closely related way to build models of distributed polarizabilities is provided by the statistical analysis of distributed polarizabilities (SADP) method, which will be employed in what follows.[20] This method, offering a pictorial representation of the fitting procedure,[33] is based on a random selection of N_c points among the $N_p \gg N_c$ points of the grid and solving the $N_c \times N_c$ system of linear equations. Deriving a set of distributed polarizabilities following this protocol is reached by constructing for each component of the distributed polarizibilities a distribution function, $\mathcal{J}(\alpha^{ss'}_{l\kappa,l'\kappa'})$, from which the most probable value is inferred. An attractive feature of the SADP method is its ability to pinpoint pathological cases, often illustrated by flat distributions characterizing redundant or ill-defined components $\alpha^{ss'}_{l\kappa,l'\kappa'}$.[20,23]

As a pertinent test of the intrinsic quality of the fitted models of distributed polarizabilities, molecular polarizabilities, $\alpha_{l\kappa,l'\kappa'}$, were regenerated at origin $\mathbf{r}_0$ by translating the distributed components at sites s and s', $\alpha^{ss'}_{l\kappa,l'\kappa'}(\mathbf{r}_s,\mathbf{r}_{s'})$, according to

$$\alpha_{l\kappa,l'\kappa'} = \sum_{s,s'}\sum_{l_s,\kappa_s l_{s'},\kappa_{s'}} \mathcal{W}_{l\kappa,l'\kappa'}(\mathbf{r}_0 - \mathbf{r}_s)\, \alpha_{l_s\kappa_s,l_s'\kappa_s'}(\mathbf{r}_s, \mathbf{r}_{s'}) \mathcal{W}_{l'\kappa',l\kappa}(\mathbf{r}_0 - \mathbf{r}_{s'}) \quad (6)$$

where, $\mathcal{W}_{l\kappa,l'\kappa'}(\mathbf{r})$ is a translation function defined by Stone.[34,30]

Grids of points were constructed with the program GRID 3.1.[35] Evaluation of induction energies by means of the PT scheme was done with the MOPETE suite of programs.[36] QM calculations prior to the topological analysis of the response charge density were performed using MOLCAS.[37] Determination of the TPEPs was carried out using the FOURIER program.[38]

2.3. Effects of the Grid. A critical issue not discussed hitherto concerns the definition of the grid over which the induction energy is mapped. For the most part, the influence of its spatial extent and the separation of its constituent points on the derived distributed polarizabilities remains unclear. Careful examination of the grid in the case of point charges fitted to the electrostatic potential led to the conclusion that insofar as the space surrounding the molecule of interest is sampled appropriately, addition of points in the periphery is superfluous.[32,39] In fact, since the number of points lying close to the van der Waals envelope is small compared to the total number of points forming the grid, any additional point far from the nuclei will contribute to the reproduction of low-order moments, modulating the short-range effects of higher order contributions that are best described in the vicinity of the molecule. For this reason, points found at distances from the nucleus exceeding three times the van der Waals radius of the latter are discarded arbitrarily.[35,40] It has also been noted that an envelope separated by 0.1 Å for concentrical layers and a grid step of 0.5 Å for Cartesian grids[41] were appropriate, smaller separations of the points leading to comparable derived point charges.[39] Moreover, the envelope beyond which the electrostatic potential is computed has been oversized by doubling each van der Waals radius to ensure that contamination of $\mathcal{V}(\mathbf{r}_k)$due to the penetration of the electron clouds is negligible.[40]

These observations, mutatis mutandis, may be applied for the derivation of fitted models of distributed polarizabilities. In addition to penetration effects that alter the pure multipolar part of the electrostatic potential, it is important to ascertain that the polarizing charge q_k is not too close to the nuclei, where it is likely to induce strong fields that entail nonnegligible hyperpolarization contributions. Experience has shown that doubling van der Waals radii, as was done for fitting point charges to the electrostatic potential, safely circumvents this problem.[20] The issue of spatial extent and grid step requires, however, more attention because the influence of a distant polarizing charge on a highly polarizable molecule can be appreciable. The strength of the PT method and that based on the use of the TPEPs resides in their ability to generate rapidly large grids of induction energies, the number of points being virtually independent from the computational effort, in contrast with the FP approach. Here, four different grids were constructed for C_6H_6, employing distinct cutoff distances, r_{cut}, beyond which points are discarded and separations of the concentrical van der Waals envelopes, Δr, viz. (i) $r_{cut} = 5.4$ Å, i.e., approximately three times the van der Waals radius of carbon, and $\Delta r = 0.15$ Å, (ii) $r_{cut} = 5.4$ Å and $\Delta r = 0.75$ Å, (iii) $r_{cut} = 9.0$ Å and $\Delta r = 0.15$ Å, and (iv) $r_{cut} = 9.0$ Å and $\Delta r = 0.75$ Å, yielding, respectively, 4368, 1044, 7224, and 1452 points. Furthermore, two additional Cartesian grids of regularly spaced points along the x, y, and z directions were built, using a grid step of 0.5 Å and a cutoff distance corresponding to (i) three and (ii) five times the van der Waals radius of the participating atoms, thus, yielding, respectively, 4785 and 25999 points. For each grid, induction energies were calculated from the TPEPs, although the PT method could have been equally employed for that purpose.

For all models of distributed polarizabilities reported herein, the root-mean-square deviation (RMSD) between the reference, QM induction energy, $\mathcal{U}_{ind,k}$, and that regenerated from the models, $\tilde{\mathcal{U}}_{ind,k}$, will be calculated by means of the following expression:

$$\mathrm{RMSD} = \left\{ \frac{1}{N_p} \sum_{k=1}^{N_p} [\mathcal{U}_{ind,k} - \tilde{\mathcal{U}}_{ind,k}]^2 \right\}^{1/2} \quad (7)$$

together with the relative error, defined as

$$\Delta\epsilon = 100 \times \frac{1}{N_p} \sum_{k=1}^{N_p} \left| \frac{\mathcal{U}_{ind,k} - \tilde{\mathcal{U}}_{ind,k}}{\mathcal{U}_{ind,k}} \right| \quad (8)$$

3. Results and Discussion

3.1. Effects of Geometry and Basis Set. As can be seen in Table 1, the largest number of basis functions does not necessarily ensure an optimal accord between the quantum chemically calculated molecular dipole polarizabilities and the experimentally determined ones. In fact, in the light of the RMSD and the mean errors between the latter quantities, the best results are obtained with the Sadlej basis set, which broadly

TABLE 1: Molecular Dipole Polarizabilities[a] of Selected Prototypical Molecules Using the ELP, the Sadlej, and the Spackman Basis Sets in Conjunction with MP2/6-31G(d) and MP2/6-311++G(2d,2p) Geometries

		MP2/ELP		MP2/Sadlej		MP2/Spackman	
		DZP[b]	TZP[b]	DZP	TZP	DZP	TZP
C_2H_2	$\alpha_{10,10}$	30.909	30.556	31.186	30.830	30.888	30.535
	$\alpha_{11c,11c}$	18.938	18.814	18.708	18.572	16.925	16.837
	$\alpha_{11s,11s}$	18.938	18.814	18.708	18.572	16.925	16.837
	$\bar{\alpha}$	22.928	22.728	22.867	22.658	21.579	21.403
C_2H_4	$\alpha_{10,10}$	34.359	33.999	34.494	34.134	34.188	33.827
	$\alpha_{11c,11c}$	22.631	22.537	22.559	22.464	21.468	21.396
	$\alpha_{11s,11s}$	25.178	24.947	25.317	25.080	25.039	24.809
	$\bar{\alpha}$	27.389	27.161	27.457	27.226	26.898	26.677
C_2H_6	$\alpha_{10,10}$	30.816	30.873	31.069	31.111	30.955	29.496
	$\alpha_{11c,11c}$	27.046	27.062	27.247	27.252	27.067	26.124
	$\alpha_{11s,11s}$	27.046	27.062	27.247	27.252	27.067	26.124
	$\bar{\alpha}$	28.579	28.332	28.521	28.538	28.363	27.248
H_2O	$\alpha_{10,10}$	9.696	9.546	9.900	9.751	9.327	9.188
	$\alpha_{11c,11c}$	9.654	9.574	9.622	9.542	9.104	9.042
	$\alpha_{11s,11s}$	10.131	9.925	10.266	10.063	10.196	9.980
	$\bar{\alpha}$	9.827	9.682	9.929	9.785	9.542	9.403
NH_3	$\alpha_{10,10}$	13.687	13.508	13.886	13.710	13.638	13.445
	$\alpha_{11c,11c}$	13.687	13.508	13.886	13.710	13.638	13.445
	$\alpha_{11s,11s}$	15.574	15.519	15.757	15.695	14.983	14.912
	$\bar{\alpha}$	14.316	14.178	14.510	14.372	14.086	13.934
C_6H_6	$\alpha_{10,10}$	45.342	45.234	45.287	45.171	43.015	42.922
	$\alpha_{11c,11c}$	81.630	81.113	81.919	81.401	80.793	80.283
	$\alpha_{11s,11s}$	81.630	81.113	81.919	81.401	80.793	80.283
	$\bar{\alpha}$	69.534	69.153	69.708	69.324	68.230	67.829
$HCOO^-$	$\alpha_{10,10}$	37.614	36.922	37.245	36.551	33.735	33.019
	$\alpha_{11c,11c}$	25.136	25.020	24.802	24.684	19.985	19.858
	$\alpha_{11s,11s}$	42.719	42.613	42.199	42.080	37.203	36.988
	$\bar{\alpha}$	35.156	34.852	34.749	34.438	30.308	29.955
H_2S	$\alpha_{10,10}$	23.201	23.085	24.399	24.260	23.800	23.668
	$\alpha_{11c,11c}$	23.261	23.214	25.229	25.150	24.263	24.198
	$\alpha_{11s,11s}$	23.733	23.435	24.576	24.274	24.766	24.448
	$\bar{\alpha}$	23.398	23.245	24.735	24.561	24.276	24.105
HCN	$\alpha_{10,10}$	22.488	22.144	22.700	22.359	21.863	21.537
	$\alpha_{11c,11c}$	14.112	13.983	13.943	13.811	13.281	13.181
	$\alpha_{11s,11s}$	14.112	13.983	13.943	13.811	13.281	13.181
	$\bar{\alpha}$	16.904	16.703	16.862	16.660	16.142	15.966
HCOOH	$\alpha_{10,10}$	16.506	16.391	16.503	16.388	30.888	15.477
	$\alpha_{11c,11c}$	28.078	27.747	28.329	28.004	26.828	26.517
	$\alpha_{11s,11s}$	23.989	23.687	24.024	23.725	23.725	23.426
	$\alpha_{11s,11c}$	0.723	0.780	0.696	0.753	0.890	0.948
	$\bar{\alpha}$	22.040	22.608	22.952	22.706	22.040	21.807
H_2CO	$\alpha_{10,10}$	22.408	22.106	22.583	22.283	21.323	21.069
	$\alpha_{11c,11c}$	12.985	12.885	12.910	12.810	12.445	12.367
	$\alpha_{11s,11s}$	17.923	17.787	17.926	17.789	17.206	17.072
	$\bar{\alpha}$	17.772	17.593	17.806	17.593	17.001	16.833
H_2CONH_2	$\alpha_{10,10}$	20.745	20.641	20.791	20.686	19.934	19.854
	$\alpha_{11c,11c}$	36.879	36.422	37.134	36.683	36.137	35.697
	$\alpha_{11s,11s}$	27.878	27.584	27.987	27.696	27.345	27.053
	$\alpha_{11s,11c}$	-0.881	-0.859	-0.929	-0.911	-0.774	-0.756
	$\alpha_{10,11c}$	0.166	0.169	0.175	0.178	0.220	0.221
	$\bar{\alpha}$	28.501	28.355	28.637	28.355	27.805	27.535
RMSD[c] (a.u.)		1.27	1.13	1.12	1.05	1.12	1.40
$\bar{\Delta}$[d]		0.24	0.38	0.01	0.20	0.72	1.01

[a] All molecular dipole polarizabilities in atomic units (a.u.). [b] Geometries optimized at the MP2/6-31G(d) (DZP) and at the MP2/6-311++G(2d,2p) (TZP) levels of approximations. [c] Root-mean-square deviation between the experimental[45] and the QM estimates. [d] Mean error between the experimental[45] and the QM estimates.

corresponds to an intermediate between the costly ELP and the less expensive Spackman basis sets.[23] While the ELP and the Sadlej basis sets give comparable molecular polarizabilities, the latter should be preferred over the former on account of its lesser cost. The Spackman basis set yields systematically underestimated quantities, ca. 0.4 a.u., thus constituting a less appealing choice for the computation of induction energies. Interestingly enough, in a number of instances, geometries optimized using the smaller 6-31G(d) basis set led to a better agreement between the calculated and the experimental dipole polarizabilities.

TABLE 2: Molecular Dipole Polarizabilities[a] of Selected Molecules at the MP2/Sadlej//MP2/6-311++G(2d,2p) Level of Approximation

	$\alpha_{10,10}$	$\alpha_{11c,11c}$	$\alpha_{11s,11s}$	$\bar{\alpha}$	$\bar{\alpha}_{exp}$[45]
C_2H_2	30.830	18.572	18.572	22.658	22.47
C_2H_4	34.134	22.464	25.080	27.226	28.70
C_2H_6	31.111	27.252	27.252	28.538	30.17
H_2O	9.751	9.542	10.063	9.785	9.79
NH_3	13.710	13.710	15.695	14.372	15.25
C_6H_6	45.171	81.401	81.401	69.324	67.49
$HCOO^-$	36.551	24.684	42.080	34.438	–
H_2S	24.260	25.150	24.274	24.561	25.51
HCN	22.359	13.811	13.811	16.660	17.48
HCOOH	16.388	28.133	23.596	22.706	22.95
H_2CO	22.283	12.810	17.789	17.593	16.53
H_2CONH_2	20.686	36.776	27.605	28.355	27.53
$CH_3CH_2NO_2$	33.685	45.715	53.376	44.259	–
CH_3CN	38.985	24.042	24.042	29.023	30.23
CH_3F	17.798	16.226	16.226	16.750	20.04
CH_3OH	19.775	20.258	23.095	21.043	22.41
CH_3SH	33.819	33.523	42.641	36.661	–
CO_2	28.287	12.967	12.967	18.074	19.65
CH_3COCH_3	33.700	45.123	46.374	41.732	43.12
$HCOCH_3$	23.737	36.130	30.254	30.040	30.98
HF	6.401	5.295	5.295	5.664	5.40
CH_3COOCH_3	36.421	46.938	54.319	45.893	46.34
CH_3OCH_3	30.546	37.811	31.400	33.252	35.70
CH_3NH_2	23.953	24.796	28.325	25.691	27.06
C_6H_5OH	47.403	84.246	93.073	74.907	74.91
C_5H_5N	41.144	73.197	77.425	63.922	64.10
RMSD[b] (a.u.)					1.32
$\bar{\Delta}$[c]					0.04

[a] All molecular dipole polarizabilities in atomic units (a.u.). These quantities correspond to the eigenvalues of the dipole polarizability tensor. [b] Root-mean-square deviation between the experimental[45] and the QM estimates. [c] Mean error between the experimental[45] and the QM estimates.

The symptomatic case of anion $HCOO^-$ provides a glaring example of the difficulties involved in the determination of the polarizabilities of negatively charged species. The results reported in Table 1 can, indeed, vary by more than 4 a.u., depending on the basis set employed. This example illustrates the well-known necessity to include diffuse functions for the treatment of anions or neutral molecules bearing lone pairs extending far from the nuclei.[42] On that point, the basis set of Spackman is clearly inadequate and should be replaced by a larger, more complete one.

As an additional test, the basis set of Sadlej was employed subsequently to estimate the molecular dipole polarizabilities of a larger set of molecules, viz. $CH_3CH_2NO_2$, CH_3CN, CH_3F, CH_3OH, CH_3SH, CO_2, CH_3COCH_3, $HCOCH_3$, HF, CH_3COOCH_3, CH_3OCH_3, CH_3NH_2, C_6H_5OH, and C_5H_5N. As can be observed in Table 2, the general agreement between the quantum chemically calculated quantities and the experimentally determined ones (the RMSD between the quantum chemically calculated and the experimentally determined molecular dipole polarizabilities are equal to 1.16 and 0.61 a.u., respectively) confirms that MP2/Sadlej constitutes an appropriate strategy for modeling induction phenomena.

3.2. Alternative Schemes for Calculating Induction Energies. As mentioned previously, the bottleneck in the development of full QM models of distributed polarizabilities fitted to the induction energy is rooted in the tremendous cost involved in the evaluation of the latter on large grids of points.[19,20] In this spirit, only the conception of a fast and reliable approach for mapping the induction energy can guarantee the success of such fitting procedures, readily usable in the design of nonadditive potential energy functions, as was the case for potential

derived charges in pairwise, additive force fields. This explains why alternative routes to the expensive FP have been sought lately. In this section, the induction energies resulting from the interaction of C_2H_6, C_6H_6, $HCONH_2$, CH_3OH, HCOOH, and $HCOO^-$ with a nonpolarizable, polarizing charge, computed using FP, the PT approach, and that based on the use of TPEPs, are compared critically. Direct comparison of the induction energies obtained from these three schemes is not possible because of the difference in the levels of theory at which the QM calculations were carried out. In particular, PT induction energies result from an HF/Sadlej computation, whereas the quantities derived from FP and from the TPEPs were evaluated at the MP2/Sadlej level of approximation, albeit TPEPs could have been obtained just as well from CPHF calculations. To assess the performances of the PT approach, the induction energies computed from eq 4 for neutral molecules should, therefore, be scaled by an appropriate factor compensating for the systematic underestimation of $\mathcal{U}_{\mathrm{ind},k}$. The case of anions is far more problematic, as will be discussed below. The choice of an adapted scaling factor is not obvious for two major reasons. First, as has been commented on, scaling of the PT induction energies depends radially on the location of the polarizing charge, q_k, from the nuclei. It has been observed that the scaling factor decreases rapidly from the points near the molecule to those lying far from it. In fact, the distribution of $\mathcal{U}_{\mathrm{ind},k}(\mathrm{FP})/\mathcal{U}_{\mathrm{ind},k}(\mathrm{PT})$ as a function of r, the distance separating the center of mass of the molecule from point k of the grid, not too unexpectedly exhibits a plateau when r is sufficiently large. A function, $f(r)$, which consists of powers of $1/r$, is proposed to model the evolution of the scaling factor; other functional forms were investigated but did not improve the description. The second reason lies in the PT approach itself. Conceptually, it is based upon an uncoupled form of the HF equations, thus, providing always a lower bound of the reference, MP2/Sadlej FP, induction energies. Interestingly enough, the Gaussian 98 suite of programs[28] provides two distinct estimates of molecular dipole polarizabilities: an "exact" one, α_{exact}, relying on the derivative of the energy with respect to the electric field, and an "approximate" one, α_{UCHF}, obtained by means of the uncoupled HF equations (UCHF). Consequently, it would be natural to model $\mathcal{U}_{\mathrm{ind},k}(\mathrm{FP})/\mathcal{U}_{\mathrm{ind},k}(\mathrm{PT})$ by the ratio $\bar{\alpha}_{\mathrm{exact}}/\bar{\alpha}_{\mathrm{UCHF}}$, in which $\bar{\alpha}$ denotes the trace of the molecular polarizability. PT induction energies were, therefore, scaled by an r-dependent factor, $\zeta(r)$, defined by

$$\zeta(r) = \frac{\bar{\alpha}_{\mathrm{exact}}}{\bar{\alpha}_{\mathrm{UCHF}}} f(r) = \frac{\bar{\alpha}_{\mathrm{exact}}}{\bar{\alpha}_{\mathrm{UCHF}}}\left(a_0 + \frac{a_1}{r} + \frac{a_2}{r^2}\right) \qquad (9)$$

where $a_0 = 0.82455$, $a_1 = 1.54964$, and $a_2 = -0.94522$ are parameters fitted to reproduce the evolution of $f(r)$ in a series of small, prototypical molecules.

In Figures 1 and 2, the reference FP induction energies are plotted against the quantities estimated from eqs 4 and 5. Considering that raw PT induction energies depart nonuniformly from the FP ones by up to ca. 40%, they were scaled using eq 9 for comparison purposes. It can be observed that, insofar as neutral compounds are concerned, scaled PT induction energies and those derived from TPEPs behave similarly. The error relative to the reference FP quantities, i.e., $|\mathcal{U}_{\mathrm{ind},k}(\mathrm{FP}) - \mathcal{U}_{\mathrm{ind},k}(\mathrm{PT/TPEP})|/\mathcal{U}_{\mathrm{ind},k}(\mathrm{FP})$, never exceeds 7.3% for the former and 6.2% for the latter. These errors clearly arise from distinct origins. In the case of PT induction energies, deviation from the reference FP quantities is rooted in the level of theory, viz. uncoupled HF equations, in conjunction with an ad hoc, empirical scaling. In contrast, the accuracy of the induction

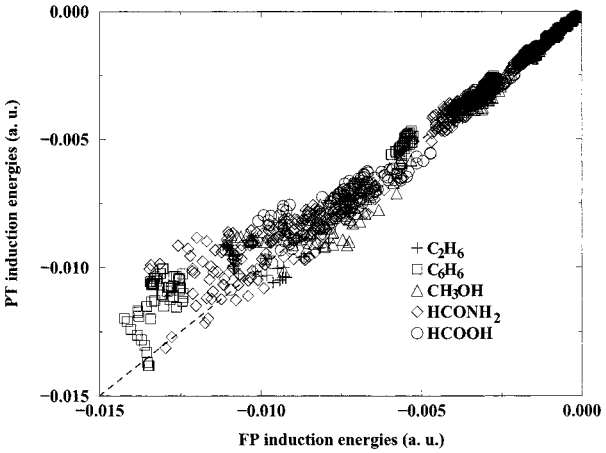

Figure 1. FP induction energies versus PT induction energies for all the neutral chemical species.

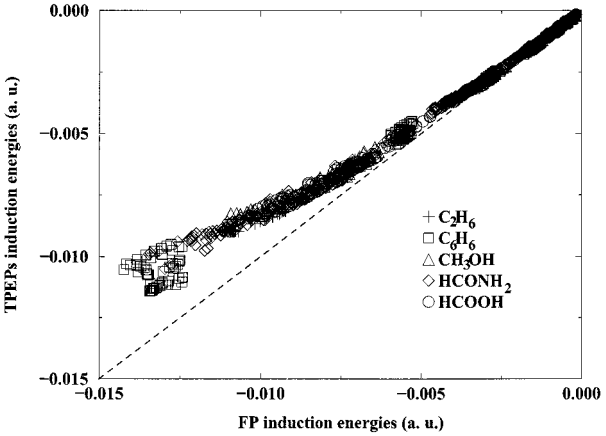

Figure 2. FP induction energies versus induction energies derived from TPEPs for all the neutral species.

energies derived from TPEPs is necessarily dependent on the rank at which the expansion of eq 5 is truncated. It is worth noting that scaling factor $\zeta(r)$ defined in eq 9 provides only an average correction at a given value of r. In reality, the dispersion of $\mathcal{U}_{\mathrm{ind},k}(\mathrm{FP})/\mathcal{U}_{\mathrm{ind},k}(\mathrm{PT})$ is the largest when r is small, i.e., for the few grid points close to the nuclei. This explains why the error between scaled PT and FP induction energies, arising mostly from these few points, remains, nonetheless, moderate.

The case of anion $HCOO^-$ is significantly more problematic, and illustrates the difficulties of the PT approach to handle such species. Here, an appropriate scaling factor is far more difficult to seek due to the erratic behavior of $\mathcal{U}_{\mathrm{ind},k}(\mathrm{FP})/\mathcal{U}_{\mathrm{ind},k}(\mathrm{PT})$ as a function of r, oscillating roughly between 0.5 and 3.5 (see Figure 3). The previously witnessed systematic underestimation of the induction energy does not hold anymore for negatively charged compounds. It is, therefore, extremely difficult to define an average scaling factor in the spirit of eq 9, capable of correcting the PT results on the sole basis of a radial dependence. Using this equation, the error relative to the FP induction energies, initially equal to 38.1%, is even increased. A closer examination of $\mathcal{U}_{\mathrm{ind},k}(\mathrm{FP})/\mathcal{U}_{\mathrm{ind},k}(\mathrm{PT})$ around $HCOO^-$ reveals a strong angular dependence of this quantity, distributed symmetrically with respect to the plane of the anion. It can be observed that $\mathcal{U}_{\mathrm{ind},k}(\mathrm{FP})/\mathcal{U}_{\mathrm{ind},k}(\mathrm{PT}) \geq 1.0$ only in a small solid angle about the C_2 axis of the anion. Clearly, scaling of the PT induction energies is now a function of both the radial, φ, and the azimuthal, θ, angles, in addition to the distance, r, separating the center of mass of the anion from the points of the grid. θ,φ-dependence can be introduced conveniently using the real

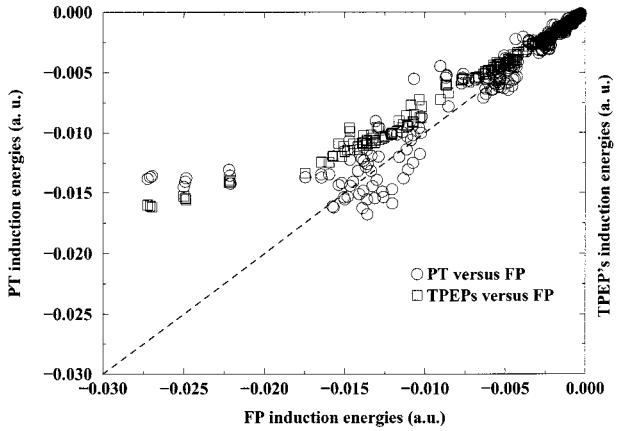

Figure 3. FP induction energies versus PT/TPEPs induction energies for $HCOO^-$.

spherical harmonics, $y_l^m(\theta, \varphi)$, the order of which is inherently related to the symmetry of the anion, thereby making the definition of a general scaling factor, $\zeta(r, \theta, \varphi) = \bar{\alpha}_{\text{exact}}/\bar{\alpha}_{\text{UCHF}}\, f(r) \sum_{l,m} c_l^m\, y_l^m(\theta, \varphi)$ a cumbersome task, probably impossible to accomplish. Here, in the particular case of anion $HCOO^-$, it can be shown that $l = 2$ real spherical harmonics should be taken into account, while first-order ones, $l = 1$, are extraneous. It can be further seen that $y_2^{\pm 1}(\theta, \varphi)$ and $y_2^{-2}(\theta, \varphi)$ do not play any role in the distribution of $\mathcal{V}_{\text{ind},k}(\text{FP})/\mathcal{V}_{\text{ind},k}(\text{PT})$ around $HCOO^-$. As a result, the scaling factor for this anion can be expressed as

$$\zeta(r, \theta, \varphi) = \frac{\bar{\alpha}_{\text{exact}}}{\bar{\alpha}_{\text{UCHF}}}\left(a_0 + \frac{a_1}{r} + \frac{a_2}{r^2}\right) \times [c_{00} + c_{20}(3\cos^2\theta - 1) + c_{22}\sin^2\theta(\cos^2\varphi - \sin^2\varphi)] \quad (10)$$

where $a_0 = 0.66143$, $a_1 = 6.79301$, $a_2 = -5.91835$, $c_{00} = 0.29592$, $c_{20} = -0.11348$, and $c_{22} = -0.19254$ are parameters optimized using a Levenberg−Marquardt fitting procedure. Using eq 10, the error between the scaled PT and the reference FP induction energies drops from 38.1 to 15.6%. Whereas the improvement is significant, the inclusion of second-order real spherical harmonics is evidently insufficient to ensure optimal scaling. In sharp contrast, the induction energies evaluated using the TPEPs turn out to be much closer to the target FP values, viz. the error is ca. 9.7%, highlighting the obvious superiority of the approach based on a topological analysis of the response charge density for dealing with anions.

3.3. Effects of the Grid. Models of distributed polarizabilities consisting of charge flow plus isotropic dipole polarizabilities were determined subsequently by means of the SADP scheme.[20] The choice of isotropic dipole polarizabilities is dictated by the desire to construct compact models suitable for statistical simulations and to avoid possible numerical instabilities[19,20] that generally translate into flat or poorly defined distribution functions, $\mathcal{J}(\alpha_{l\kappa,l'\kappa'}^{ss'})$. From these cost-effective models, molecular dipole polarizabilities were regenerated using eq 6 and confronted critically to the corresponding target MP2/Sadlej values.

As can be observed in Table 3, the sensitivity of both the distributed polarizabilities and the reproduced molecular quantities to the spatial extent of the grid points, i.e., r_{cut}, is appreciable. Limiting the distribution of the points to approximately three times the van der Waals radii of the participating atoms is clearly not enough to guarantee a faithful description of the prevailing molecular dipole component. The reason lies in the exaggerated weight of the short-range, higher-order contributions that modulate the reproduction of molecular dipole polarizabilities. Depending upon the grid, discrepancies in the regenerated molecular polarizabilities can attain ca. 10 a.u., with respect to the reference MP2/Sadlej values. Such is the case, for instance, of $\alpha_{1m,1m}$ components, using the orthogonal, Cartesian grid of 4785 points. Whereas the latter would prove appropriate to derive net atomic charges from the electrostatic potential and reproduce not only the molecular dipole moment but higher-order components as well, accurate representation of induction phenomena requires that the space around the molecule be sampled thoroughly over a sufficiently large radial extent from the nuclei. This result is not totally surprising considering the long-range nature of the interaction of a polarizing charge with the molecule of interest.

Table 3 also suggests that the role played by the density of grid points, Δr, on both the distributed polarizabilities and the regenerated molecular quantities is very moderate. To ascertain this observation, seven orthogonal Cartesian grids were constructed, using distinct densities, viz. $\Delta r = 0.50$, 0.55, 0.65, 0.75, 1.00, 1.50, and 2.00 Å, corresponding to 25999, 19495, 11819, 7710, 3244, 966, and 407 points, respectively. As Table 3 confirms, the influence of Δr on the fitted polarizabilities is only marginal. It can be noted that as the density of points diminishes, so does the carbon−carbon charge flow polarizability, $\alpha_{00,00}^{\text{CC}}$. At the same time, the $\alpha_{1m,1m}$ components of the dipole molecular polarizability decrease moderately from 86.246 to 85.790 a.u., thereby indicating that the choice of Δr is most likely not as crucial as that of r_{cut}. Whereas in the case of the FP methodology, the role of Δr governs the computational cost involved in the mapping of the induction energy around the molecule, this aspect is clearly of lesser concern for either the PT approach or that based on the TPEPs.

3.4. Models of Distributed Polarizabilities. As has been seen so far, the scaled PT approach and that relying upon the TPEPs apparently provide comparable induction energies, at least for neutral chemical species. This result stems from the statistical analysis of the deviation, $|\mathcal{V}_{\text{ind},k}(\text{FP}) - \mathcal{V}_{\text{ind},k}(\text{PT/TPEP})|/\mathcal{V}_{\text{ind},k}(\text{FP})$, which can be artificially small, considering that the main source of error lies near the nuclei, and is essentially concealed by the vast majority of points in the periphery, for which the accord is much better (see Figure 1). Determination of models of distributed polarizabilities and reconstruction of molecular polarizabilities, directly comparable with the reference MP2/Sadlej quantities, should prove the true worth of the two alternative schemes for the computation of induction energies. In this section, models of distributed polarizabilities consisting of charge flow plus isotropic dipole polarizabilities were derived for C_2H_6, C_6H_6, $HCONH_2$, CH_3OH, HCOOH, and $HCOO^-$, using large Cartesian grids of points for which r_{cut} corresponds roughly to five times the van der Waals radii of the participating atoms and $\Delta r = 0.5$ Å, 21581, 25999, 19362, 19269, 19093, and 19099 points, respectively.

The distributed polarizabilities computed from the SADP procedure[20] are reported in Table 4, together with the corresponding molecular dipole polarizabilities derived using translation eq 6. From the onset, it can be observed that models of atomic polarizabilities constructed from induction energies based on TPEPs always yield smaller RMSDs and mean errors than those obtained from the PT approach. Whereas establishing the superiority of one scheme over the other cannot be done in the sole light of this trend (both the RMSD and the mean error reflecting the quality of the models and not the deviation from hypothetically "exact" induction energies), it may, nonetheless,

TABLE 3: Effects of the Grid Definition on the Distributed Polarizabilities[a,b] and the Regenerated Molecular Dipole Polarizabilities[b] of C_6H_6 at the MP2/Sadlej//MP2/6-311++G(2d,2p) Level of Approximation

	vdW envelopes				
r_{cut}	5.5 Å	5.5 Å	7.5 Å	7.5 Å	
Δr	0.15 Å	0.75 Å	0.15 Å	0.75 Å	MP2/Sadlej
$\alpha^{CC}_{00,00}$	−1.793	−1.765	−1.675	−1.679	
$\alpha^{CH}_{00,00}$	−0.511	−0.573	−0.564	−0.607	
$\alpha^{CC}_{1m,1m}$	7.551	7.542	7.484	7.487	
RMSD (10^{-3} a.u.)	0.374	0.523	0.338	0.490	
$\Delta\epsilon$ (%)	3.16	3.73	2.94	3.50	
$\alpha_{10,10}$	45.308	45.253	44.906	44.920	45.171
$\alpha_{11c,11c}$	89.154	89.279	89.934	87.576	81.401
$\alpha_{11s,11s}$	89.154	89.279	89.934	87.576	81.401

	Cartesian grid				
r_{cut}	$3.0 \times R_{vdW}$	$5.0 \times R_{vdW}$	$5.0 \times R_{vdW}$	$5.0 \times R_{vdW}$	
Δr	0.50 Å	0.50 Å	0.55 Å	0.65 Å	MP2/Sadlej
$\alpha^{CC}_{00,00}$	−1.722	−1.614	−1.614	−1.626	
$\alpha^{CH}_{00,00}$	−0.602	−0.524	−0.523	−0.504	
$\alpha^{CC}_{1m,1m}$	7.982	7.666	7.666	7.658	
RMSD (10^{-3} a.u.)	0.242	0.184	0.189	0.188	
$\Delta\epsilon$ (%)	2.66	2.14	2.14	2.12	
$\alpha_{10,10}$	47.894	45.996	45.995	45.945	45.171
$\alpha_{11c,11c}$	91.373	86.246	86.253	86.211	81.401
$\alpha_{11s,11s}$	91.373	86.246	86.253	86.211	81.401

	Cartesian grid				
r_{cut}	$5.0 \times R_{vdW}$	$5.0 \times R_{vdW}$	$5.0 \times R_{vdW}$	$5.0 \times R_{vdW}$	
Δr	0.75 Å	1.00 Å	1.50 Å	2.00 Å	MP2/Sadlej
$\alpha^{CC}_{00,00}$	−1.677	−1.582	−1.583	−1.527	
$\alpha^{CH}_{00,00}$	−0.514	−0.549	−0.558	−0.636	
$\alpha^{CC}_{1m,1m}$	7.664	7.685	7.659	7.658	
RMSD (10^{-3} a.u.)	0.187	0.185	0.190	0.192	
$\Delta\epsilon$ (%)	2.12	2.12	2.11	2.09	
$\alpha_{10,10}$	45.986	46.111	45.955	45.949	45.171
$\alpha_{11c,11c}$	86.191	86.021	86.003	85.790	81.401
$\alpha_{11s,11s}$	86.191	86.021	86.003	85.790	81.401

[a] Models of distributed polarizabilities include charge flow plus isotropic dipole polarizabilities. [b] All molecular dipole polarizabilities in atomic units (a.u.).

be inferred that grids of induction energies mapped by means of TPEPs probably contain fewer points corresponding to large errors that could deteriorate globally the determination of distributed polarizabilities. It should be remembered, indeed, that for all neutral molecules, the *r*-dependent scaling factor defined in eq 9 provides only an average correction of the PT induction energies, neglecting possible angular dependencies. The most striking difference between the two sets of models lies in the systematic exaggeration of the atomic dipole polarizabilities by the PT approach, while charge flow polarizabilities are conspicuously underestimated; the almost vanishing $\alpha^{CH}_{00,00}$ component in C_6H_6 is a glaring example of the witnessed tendency. This conspicuous behavior is at the origin of overestimated molecular dipole polarizabilities, suggesting that a topological partitioning of the response charge density supplies somewhat more reliable induction energies than does the PT scheme.

The deficiencies characteristic of the latter approach are magnified in the special case of anion $HCOO^-$. In particular, whereas component $\alpha_{11c,11c}$ of the molecular dipole polarizability is reproduced accurately, $\alpha_{11s,11s}$ and $\alpha_{10,10}$ are exaggerated by ca. 15 and 6 a.u., respectively. Although the introduction of a θ,φ-angular dependence in the scaling of PT induction energies improved the model significantly, it should be acknowledged that, at this stage, only the use of TPEPs can guarantee a faithful description of polarization phenomena involving anions.

4. Conclusion

Aside from the computational cost overrun implied in statistical simulations, one of the limitations to the generalized use of polarizable potential energy functions lies in the difficulty to design compact, nonheuristic models of distributed polarizabilities based on sophisticated QM calculations and capable of accounting for through-space induction effects appropriately. In essence, the main obstacle to the derivation of distributed polarizabilities is the fast yet accurate computation of induction energies to which the latter will be fitted. Preliminary attempts to determine models of atomic polarizabilities based on FP calculations emphasized that the applicability of this approach was restrained to sufficiently small chemical species,[19,20] on account of the tremendous computational effort it involves, i.e., mapping the induction energy over a grid of N_p points requires N_p separate calculations. At this point, earlier studies illuminated the preponderant role of the basis set in the faithful reproduction of experimental molecular dipole polarizabilities,[27] suggesting that only the use of large and flexible basis sets can guarantee reliable results. Careful examination of the behavior of the ELP,[24] the Sadlej,[25] and the Spackman[26] basis sets led to the conclusion that the second offers the best possible compromise in terms of accuracy and computational investment. Whereas the geometries optimized at the MP2/6-311++G(2d,2p) level of approximations yield molecular dipole polarizabilities in

TABLE 4: Models of Charge Flow Plus Isotropic Dipole Polarizabilities[a] and the Regenerated Molecular Dipole Polarizabilities[a] for a Series of Small Prototypical Molecules at the MP2/Sadlej//MP2/6-311++G(2d,2p) Level of Approximation

		distributed polarizabilities			molecular polarizabilities		
		PT	TPEPs		PT	TPEPs	MP2/Sadlej
C_2H_6	$\alpha_{00,00}^{C_1C_2}$	−0.267	−1.714	$\alpha_{10,10}$	28.740	33.612	31.111
	$\alpha_{00,00}^{C_1H_1}$	−0.493	−1.279	$\alpha_{11c,11c}$	30.359	29.294	27.252
	$\alpha_{1m,1m}^{C_iC_i}$	12.435	7.520	$\alpha_{11s,11s}$	30.359	29.294	27.252
	RMSD (10^{-3} a.u.)	0.319	0.119				
	$\Delta\epsilon$ (%)	5.33	3.03				
C_6H_6	$\alpha_{00,00}^{C_1C_2}$	−1.252	−1.614	$\alpha_{10,10}$	52.638	45.996	45.171
	$\alpha_{00,00}^{C_iH_i}$	−0.269	−1.614	$\alpha_{11c,11c}$	82.146	86.246	81.401
	$\alpha_{1m,1m}^{C_iC_i}$	8.773	7.666	$\alpha_{11s,11s}$	82.146	86.246	81.401
	RMSD (10^{-3} a.u.)	0.299	0.184				
	$\Delta\epsilon$ (%)	4.10	2.14				
$HCONH_2$	$\alpha_{00,00}^{CH}$	−0.516	−0.760	$\alpha_{10,10}$	25.051	21.235	20.686
	$\alpha_{00,00}^{CO}$	−1.491	−1.915	$\alpha_{11c,11c}$	35.887	37.965	36.683
	$\alpha_{00,00}^{CN}$	−0.711	−1.566	$\alpha_{11s,11s}$	31.884	30.112	27.696
	$\alpha_{00,00}^{NH}$	−0.408	−0.262	$\alpha_{10,11c}$	−0.072	0.099	0.178
	$\alpha_{1m,1m}^{CC}$	7.243	6.013	$\alpha_{11s,11c}$	−2.548	−1.167	−0.911
	$\alpha_{1m,1m}^{NN}$	6.582	4.806				
	$\alpha_{1m,1m}^{OO}$	11.197	10.395				
	RMSD (10^{-3} a.u.)	0.214	0.139				
	$\Delta\epsilon$ (%)	5.15	2.48				
CH_3OH	$\alpha_{00,00}^{CH_1}$	−0.791	−0.960	$\alpha_{10,10}$	23.142	21.076	19.775
	$\alpha_{00,00}^{CH_{2,3}}$	−0.939	−1.213	$\alpha_{11c,11c}$	25.103	22.421	20.258
	$\alpha_{00,00}^{CO}$	−0.697	−1.158	$\alpha_{11s,11s}$	24.579	24.684	23.095
	$\alpha_{00,00}^{OH}$	−0.890	−0.828	$\alpha_{11s,11c}$	0.449	0.198	0.530
	$\alpha_{1m,1m}^{CC}$	11.632	8.944				
	$\alpha_{1m,1m}^{OO}$	6.229	5.262				
	RMSD (10^{-3} a.u.)	0.201	0.173				
	$\Delta\epsilon$ (%)	5.76	4.06				
HCOOH	$\alpha_{00,00}^{CH}$	−0.557	−0.880	$\alpha_{10,10}$	19.879	16.922	16.388
	$\alpha_{00,00}^{CO_1}$	−1.242	−1.294	$\alpha_{11c,11c}$	28.719	28.575	28.004
	$\alpha_{00,00}^{CO_2}$	−0.575	−1.212	$\alpha_{11s,11s}$	25.824	25.568	23.725
	$\alpha_{00,00}^{O_2H}$	−0.669	−0.588	$\alpha_{11s,11c}$	−1.343	0.281	0.753
	$\alpha_{1m,1m}^{CC}$	6.309	5.114				
	$\alpha_{1m,1m}^{O_1O_1}$	6.744	5.704				
	$\alpha_{1m,1m}^{O_2O_2}$	6.826	6.105				
	RMSD (10^{-3} a.u.)	0.159	0.087				
	$\Delta\epsilon$ (%)	4.89	2.22				
$HCOO^-$	$\alpha_{00,00}^{CH}$	−2.477	−2.201	$\alpha_{10,10}$	42.598	38.900	36.551
	$\alpha_{00,00}^{CO}$	−3.581	−2.104	$\alpha_{11c,11c}$	24.208	24.732	24.684
	$\alpha_{1m,1m}^{CC}$	4.936	10.459	$\alpha_{11s,11s}$	57.564	44.321	42.080
	$\alpha_{1m,1m}^{OO}$	8.551	7.134				
	RMSD (10^{-3} a.u.)	0.502	0.265				
	$\Delta\epsilon$ (%)	19.33	2.70				

[a] All molecular dipole polarizabilities in atomic units (a.u.).

excellent agreement with experiment, it has been observed that those obtained with the smaller 6-31G(d) basis set could lead to equally remarkable results.

Even at the MP2/Sadlej level of theory, FP calculations are still too expensive, thus making it necessary to search for swift, inexpensive methods to compute induction energies and to derive distributed polarizabilities. With this objective in mind, two alternatives were examined, one relying upon second-order perturbation theory and an uncoupled form of the perturbed HF equations,[21,22] the other upon a topological analysis of the response charge density,[12,23] in the spirit of Bader's theory of "atoms in molecules".[13] The presdent contribution analyzes critically these alternatives and proposes a series of guidelines and recommendations for the design of models of distributed polarizabilities.

Based on a single QM calculation, both the PT approach and that making use of the TPEPs supply induction energies of comparable quality in terms of error with respect to the reference FP values for all neutral molecules. Because the PT scheme described here makes use of an uncoupled form of the HF equations,[43] scaling of the induction energies by means of an appropriately parameterized, *r*-dependent function is necessary to compare with the MP2/Sadlej quantities. The main advantage of this method resides in its simplicity of implementation and the very low computational effort it implies, the single QM computation being performed at the HF level of approximation. This also constitutes the weakness of the approach, considering that an appropriate scaling is necessary to yield values of MP2/Sadlej quality, a key requirement that appears to be impossible to fulfill completely in the special case of anions, even by increasing the complexity of the scaling function through an angular dependence. In contrast, the method relying on a relocalization of the TPEPs, although more complex in its implementation and more computationally demanding on ac-

count of the integration of the electron density over atomic domains, proves to be more rigorous as it does not call for any a posteriori adjustment of the determined quantities.

Among the recommendations for designing models of distributed polarizabilities, an important issue concerns the definition of the grid over which the induction energy is mapped. In light of a thorough analysis, using distinct spatial extent, r_{cut}, and separation of first-neighbor points, Δr, it was found that increasing the density is not as pivotal as sampling the space surrounding the molecule sufficiently far from the nuclei. The choice of r_{cut} corresponding to approximately five times the van der Waals radius of the participating atoms proved to yield a reliable reproduction of the molecular dipole polarizabilities. Finally, a critical test for probing the induction energies evaluated using the PT approach and an approach that relies upon a partitioning of the response charge density consists of computing the distributed polarizabilities from which molecular quantities can be regenerated and confronted to the corresponding experimental estimates. For all neutral molecules, the derived models of distributed polarizabilities unambiguously reflect the higher quality of the induction energies calculated from the TPEPs. The PT scheme, nevertheless, proved to provide reasonable results and could thus constitute, on account of both its ease of implementation and limited computational cost, a very appealing alternative. Yet, the description of induction phenomena involving anions illustrates the limitations of this approach. Enhancements in the scaling of the induction energies by means of an angular dependence partially alleviate the deficiencies of the method, while not correcting them fully. In sharp contrast, the approach based on a relocalization of the TPEPs is robust and rigorous enough and should therefore be preferred for handling negatively charged species.

Acknowledgment. We are indebted to the Institut du Développement et des Ressources en Informatique Scientifique (IDRIS), Orsay, France, the Centre Charles Hermite (CCH), Vandœuvre-lès-Nancy, France, the Centre de Supercomputació de Catalunya (CESCA), Barcelona, Spain, for generous provision of computer time. We also acknowledge financial support from the Ministerio de Educacion y Cultura (grants PB98-1222 and PM99-0046) and the Direccio General de Recerca (grant ACI2000-2).

References and Notes

(1) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.

(2) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab initio molecular orbital theory*; Wiley Interscience: New York, 1986.

(3) Carlson, H. A.; Nguyen, T. B.; Orozco, M.; Jorgensen, W. L. *J. Comput. Chem.* **1993**, *14*, 1240.

(4) Colonna, F.; Evleth, E. *Chem. Phys. Lett.* **1993**, *212*, 665.

(5) Chipot, C.; Ángyán, J. G.; Maigret, B.; Scheraga, H. A. *J. Phys. Chem.* **1993**, *97*, 9797.

(6) Applequist, J. *Acc. Chem. Res.* **1977**, *10*, 79.

(7) Thole, B. T. *J. Chem. Phys.* **1981**, *59*, 341.

(8) Stone, A. J. *Mol. Phys.* **1985**, *56*, 1065.

(9) Le Sueur, C. R.; Stone, A. J. *Mol. Phys.* **1994**, *83*, 293.

(10) Maaskant, W. J. A.; Oosterhof, L. J. *Mol. Phys.* **1964**, *8*, 319.

(11) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233.

(12) Ángyán, J. G.; Jansen, G.; Loos, M.; Hättig, C.; Hess, B. A. *Chem. Phys. Lett.* **1994**, *219*, 267.

(13) Bader, R. F. W. *Atoms in Molecules − A Quantum Theory*; Oxford University Press: London, 1990.

(14) Bonaccorsi, R.; Petrongolo, C.; Scrocco, J. *Theor. Chim. Acta* **1971**, *20*, 331.

(15) Momany, F. A. *J. Phys. Chem.* **1978**, *82*, 592.

(16) Cox, S. R.; Williams, D. E. *J. Comput. Chem.* **1981**, *2*, 304.

(17) Nakagawa, S.; Kosugi, N. *Chem. Phys. Lett.* **1993**, *210*, 180.

(18) Alkorta, I.; Bachs, M.; Perez, J. J. *Chem. Phys. Lett.* **1994**, *224*, 160.

(19) Celebi, N.; Ángyán, J. G.; Dehez, F.; Millot, C.; Chipot, C. *J. Chem. Phys.* **2000**, *112*, 2709.

(20) Dehez, F.; Soetens, J. C.; Chipot, C.; Ángyán, J. G.; Millot, C. *J. Phys. Chem. A* **2000**, *104*, 1293.

(21) Luque, F. J.; Orozco, M. *J. Comput. Chem.* **1998**, *19*, 866.

(22) Chipot, C.; Luque, F. J. *Chem. Phys. Lett.* **2000**, *332*, 190.

(23) Dehez, F.; Chipot, C.; Millot, C.; Ángyán, J. G. *Chem. Phys. Lett.* **2001**, *338*, 180.

(24) Liu, S. Y.; Dykstra, C. E. *J. Phys. Chem.* **1987**, *91*, 1749.

(25) Sadlej, A. J. *Collect. Czech. Chem. Commun.* **1988**, *53*, 1995.

(26) Spackman, M. A. *J. Chem. Phys.* **1986**, *85*, 6587.

(27) Voisin, C.; Cartier, C.; Rivail, J. L. *J. Phys. Chem.* **1992**, *96*, 7966.

(28) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *GAUSSIAN 98*, revision A.7; Gaussian Inc.: Pittsburgh, PA, 1999.

(29) Francl, M. M. *J. Phys. Chem.* **1985**, *89*, 428.

(30) Stone, A. J. Classical electrostatics in molecular interactions. In *Theoretical models of chemical bonding*; Maksić, H., Ed.; Springer-Verlag: Berlin, 1991; vol. 4, pp. 103−131.

(31) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes*; Cambridge University: Cambridge, 1986.

(32) Chipot, C.; Maigret, B.; Rivail, J. L.; Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 10276.

(33) Chipot, C.; Ángyán, J. G.; Millot, C. *Mol. Phys.* **1998**, *94*, 881.

(34) Stone, A. J.; Tong, C. S. *Chem. Phys.* **1989**, *137*, 121.

(35) Chipot, C.; Ángyán, J. G. GRID Version 3.0: Point multipoles derived from molecular electrostatic properties. QCPE No. 655, 1994.

(36) Luque, F. J.; Alhambra, C.; Orozco, M. Unpublished version of the MOPETE computer programs. University of Barcelona, Barcelona, Spain, 1997.

(37) Anderson, K.; Blomberg, M. R. A.; Fülscher, M. P.; Lindh, R.; Malmquist, P.-Å.; Olsen, J.; Roos, B. O.; Sadlej, A. J.; Widmark, P. O. MOLCAS version 2; University of Lund, Sweden, 1992.

(38) Hättig, C. FOURIER. Institut für Physikalische und Theoretische Chemie, Universität Bonn, Wegelerstrasse 12, D-53115 Bonn, Germany, 1995.

(39) Chipot, C.; Ángyán, J. G.; Maigret, B.; Scheraga, H. A. *J. Phys. Chem.* **1993**, *97*, 9788.

(40) Colonna, F.; Evleth, E.; Ángyán, J. G. *J. Comput. Chem.* **1992**, *13*, 1234.

(41) Ángyán, J. G.; Chipot, C. *Int. J. Quantum Chem.* **1994**, *52*, 17.

(42) Chandrasekhar, J.; Andrade, J. G.; Schleyer, P. v. R. *J. Am. Chem. Soc.* **1981**, *103*, 5609.

(43) It should be underlined that the PT scheme can be extended in a straightforward fashion to the CPHF, the time-dependent HF (TDHF), or the time-dependent MP2[44] (TDMP2) levels of theory. Work in this direction is currently in progress.

(44) Hättig, C.; Hess, B. A. *Chem. Phys. Lett.* **1995**, *233*, 359.

(45) *CRC Handbook of Chemistry and Physics*, 80th ed.; Lide, D. R., Ed.; Chapman and Hall, CRC Press: Boca Raton, 1999.