PERSPECTIVE

NJC
www.rsc.org/njc

Continuing challenges in the parametrization of intermolecular force fields. Towards an accurate description of electrostatic and induction terms

Christophe Chipot*[a] and János G. Ángyán[b]

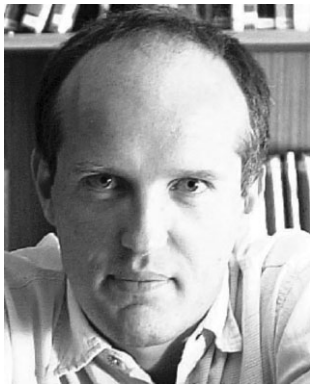
[a] *Equipe de Dynamique des Assemblages Membranaires (CNRS UMR 7565), Institut Nancéien de Chimie Moléculaire, Université Henri Poincaré, BP 239, 54506 Vandœuvre-lès-Nancy cedex, France. E-mail: Christophe.Chipot@edam.uhp-nancy.fr; Tel: +33 3 83 68 43 91*

[b] *Laboratoire de Cristallographie et de Modélisation des Matériaux Minéraux et Biologiques (CNRS UMR 7036), Université Henri Poincaré, BP 239, 54506 Vandœuvre-lès-Nancy cedex, France*

Received (in Montpellier, France) 11th September 2004, Accepted 23rd November 2004
First published as an Advance Article on the web 14th February 2005

The improvement in the description of molecular liquids by statistical mechanics simulations, going beyond the limitations of the pairwise additive approximation, has triggered the development of novel approaches for the design of tractable and reliable sets of electrostatic and induction parameters. Progress made in this direction over the past years is reviewed with an emphasis on the underlying theory of the formalism utilized. Generation of the models of distributed multipoles and polarizabilities relies upon the knowledge of the electrostatic potential and the induction energy mapped on a grid of points around the molecule. Exploitation of the symmetry and the transferability properties of the latter is achieved by means of local atomic frames of reference, which allow the distributed components to be expressed in a simple and compact form. Investigation of the prototypical cases of benzene and naphthalene illustrates the robustness of the methodology for the reproduction of electrostatic and induction interaction energies.

DOI: 10.1039/b414280m



Christophe Chipot received his BSc and MSc in physical chemistry from the University Henri Poincaré in Nancy, France. In the course of his PhD on the development of intermolecular potentials for the simulation of protein folding, with Bernard Maigret, Christophe Chipot collaborated fruitfully with the groups of Harold A. Scheraga at Cornell University, and Peter A. Kollman at the University of California in San Francisco, on the important problems of electrostatics parametrization and free energy calculations. After he obtained in 1994 his PhD with honors, cum laude, Christophe Chipot joined the group of Peter A. Kollman, where he pursued his investigations on molecular recognition via free energy calculations. He subsequently joined in 1995 the group of Andrew Pohorille at the NASA Ames Research Center, where he worked on the folding of short peptides at aqueous interfaces. In 1996, he joined the Centre National de la Recherche Scientifique and returned to Nancy, France, where he has been investigating ever since the interaction of proteins with membranes, using statistical mechanics. He develops new approaches for improving the description of intermolecular interactions in molecular simulations and is a contributor to the molecular dynamics program NAMD. In 1999, he received the Young Researcher Award of the Société Française de Chimie for his work on protein folding at aqueous interfaces. He is the recipient of the 2001 bronze medal of the Centre National de la Recherche Scientifique.

János Ángyán studied chemistry in Budapest, Hungary, and obtained his MSc and PhD in quantum chemistry from Eötvös Loránd University. In the early eighties, he joined the molecular modeling group of G. Náray-Szabó, at the Chinoin Research Centre, Budapest, and worked on protein electrostatics and quantum chemical solvent effects. With G. Náray-Szabó and P. R. Surján he co-authored the monograph 'Applied Quantum Chemistry' published by Kluwer in 1987. After postdoctoral stays in Toronto, Paris and a Humboldt Fellowship at the University of Bonn, in 1991 he joined the Centre National de Recherche Scientifique as a member of the Laboratoire de Chimie théorique at Nancy. During this period his main interest was the application of perturbation theory to solvent effect models, the use of distributed multipoles and polarizabilities to represent intermolecular forces and the development of bond order indices. In the late nineties, he started to work on solids, in particular zeolite catalysts, using ab initio DFT plane-wave methods. After a sabbatical year at the Computational Material Science College of the Institut für Materialphysik in Vienna, with J. Hafner, he joined the Crystallography Laboratory in Nancy, where he leads a group for the modeling of electron densities.

1. Introduction and general overview

One of the grand challenges tackled by modern theoretical chemistry is the accurate modelling of intermolecular interactions by means of appropriately designed sets of parameters that can be plugged directly into statistical mechanics simulation packages used to investigate a wide variety of chemical systems, ranging from simple molecular liquids to complex biological assemblies. Among the underlying assumptions that constitute the internal framework of the vast majority of intermolecular force fields, the concept of transferability is probably the most critical one. Whereas van der Waals parameters, determined from either accurate quantum chemical calculations or statistical simulations of neat liquids, have proven to transfer generally well between compounds of distinct chemical nature, the same cannot be said for both electrostatic and induction parameters. The stringent dependence of these terms on the chemical environment calls for a careful case-by-case treatment that reflects the specificity of any given molecule. Just like human fingerprints, electrostatic potential and induction energy maps of two chemical species are different and so should be the parameters that model these quantities.

From a practical perspective, it is tempting to parametrize electrostatic forces in terms of atomic quantities that can be used directly in molecular mechanics calculations. Conceptually, this approach relies upon some kind of partitioning of the total charge density into regions of Cartesian space that correspond to atoms, as well as a selection of leading terms in the classical expression of the forces exerted between atomic distributions. Historically, macromolecular force fields have restrained the description of the electrostatic potential to a monopole approximation, allowing electrostatic forces to be evaluated rapidly by means of simple Coulomb interactions, and thereby neglecting the possible influence of higher-order, local contributions, *viz*. dipole, quadrupole, *etc*.[1–4] The ambiguous nature of charge models, ascribable to the fact that a charge, *per se*, is not a quantum mechanical observable, has led to the development of numerous alternatives for their derivation.[5,6]

One of the most well-known methods for partitioning a molecular charge distribution into atomic pieces is based upon the idea of Mulliken,[7] which consists in dividing equally the overlap charge density, $\varrho_{\nu\mu}(\mathbf{r}) = \chi^a_\nu(\mathbf{r})\chi^b_\mu(\mathbf{r})$, between atoms a and b on which the orbitals $\chi^a_\nu(\mathbf{r})$ and $\chi^b_\mu(\mathbf{r})$ are centered. The sum of one-center overlap charge densities, together with half of the two-center overlap densities, constitutes an atomic density. Mulliken charges simply correspond to the truncation at the monopole level of the potential created by these densities. Sokalski and Poirier[8] generalized the multipole expansion of the Mulliken atomic densities and showed that the reproduction of electrostatic properties can be considerably improved by adding higher multipole moments. This model of cumulative atomic multipole moments (CAMM) has been successfully applied to calculate electrostatic interaction energies of intermolecular complexes, molecular crystals, *etc*.[9] It ought to be noted that even if halving the overlap densities may seem to be an arbitrary choice, it finds its justification in the mathematical simplicity of the partitioning in terms of density matrices $D_{\mu\nu}$:

$$\varrho_a(\mathbf{r}) = \sum_{\mu\in a}\sum_{\nu} D_{\mu\nu}\varrho_{\nu\mu}(\mathbf{r}) \qquad (1)$$

An electrostatically optimized generalization of the above scheme is due to Stone and Alderton,[10,11] and to Vigné-Maeder and Claverie,[12] who elaborated independently a scheme for partitioning densities expressed in Gaussian atomic orbital basis functions. Whereas the Mulliken, or the CAMM approach halves products of atomic orbitals, which are linear combinations of primitive Gaussian functions, the distributed multipole moments (DMM) are defined in terms of a weighted partition of the overlap density of the primitive functions themselves. The advantage of using primitive functions resides in the fact that a product of two Gaussian functions is again a Gaussian, located at the barycenter of the primitive functions. The multipole expansion of the potential of such a product is finite around the barycenter. The ensemble of multipole series forms the overlap multipole moment (OMM) model, which regenerates the exact multipolar part of the molecular electrostatic potential. The significantly more compact DMM model can be obtained without an appreciable loss of accuracy by translating each OMM onto the closest atomic center, or by distributing it among several closely lying centers. DMMs have encountered a considerable success: they are implemented in several standard quantum chemical packages and are widely used for estimating electrostatic interaction energies, for instance, in molecular crystals.[13]

Among the plethora of methods that have spread in the literature, the idea that simple, atom-centered point charges could be derived from the electrostatic potential has proven to be the most widely accepted approach for determining electrostatic interaction parameters. Over the past twenty years, the pioneering work of Cox and Williams[14] has fuelled the development of more sophisticated schemes capable of providing compact sets of point charges to reproduce the electrostatic potential as accurately and faithfully as possible. The target reference potential is usually evaluated on a finite grid of points from the solution of the Poisson equation with the charge density of the quantum mechanically determined wave function.[15,16] Alternatively, only its exact multipolar part, which satisfies the Laplace equation, can be considered.[17] The latter is generated in a straightforward fashion for charge densities expanded in Gaussian basis functions through an expansion of overlap multipole moments.

On the road towards a fully quantum mechanical description of reasonably large ensembles of molecules, still out-of-reach for obvious computational cost-effectiveness reasons, statistical mechanics simulations have demonstrated their ability to capture the many-body essence of intermolecular interactions by means of appropriately designed polarizability parameters.[18–22] In sharp contrast with electrostatic forces, handling of induction forces, which reflect the mutual electric polarization of interacting molecules or fragments is, however, somewhat more intricate on account of technical difficulties intimately connected to the nature of these forces.[23,24] Representing the latter by utilizing atomic parameters inevitably leads to an ambiguous choice in the partitioning of the intrinsically nonlocal polarizability density into atomic contributions. Aside from the methodological aspects related to the determination of distributed polarizabilities, explicit inclusion of induction effects in statistical mechanics simulations results in a significant overhead, which has encouraged the modelling community to opt for more approximate, pairwise additive, effective potentials. For instance, taking advantage of the conspicuous tendency of the 6-31G* basis set to exaggerate molecular dipole moments, several authors have proposed to derive their sets of net atomic charges from electrostatic potentials evaluated at this level of approximation, thereby inflating artificially the polarity of the molecules and compensating, in an average sense, for missing induction effects. A more rational approximation advocates the use of wave functions determined from self-consistent reaction field (SCRF) calculations,[25] wherein mutual polarization of the solute and the solvent is accounted for by means of a dielectric continuum.[26–29]

Cognizant that effective potentials only provide a reasonable description of chemical systems in which polarizability phenomena can be neglected, a host of parametrization schemes have been devised to handle induction forces explicitly. The problem of a system of interacting polarizable particles has

been first discussed by Barker in the fifties.[30] Most of the approaches follow the pioneering efforts of Applequist to model many-body effects by means of interacting atomic dipole polarizabilities.[31] Similar in spirit to the method proposed by Cox and Williams to determine net atomic charges, Nakagawa suggested that distributed polarizabilities and hyperpolarizabilities could be derived from the perturbed potential of a molecule polarized by a point charge.[32–34] Closely related to this idea, the method of Alkorta *et al.* fits atomic polarizabilities to the reference induction energy mapped around the molecule of interest.[35] The computational bottleneck of this approach evidently lies in the evaluation of the induction energy over sufficiently large sets of points to sample the space around the molecule appropriately. Direct generation of induction energy maps using a finite perturbation (FP) scheme[36] undoubtedly represents the simplest, but also the most expensive, way to proceed. The induction energy, $\mathscr{U}_{\mathrm{ind},k}$, arises from the interaction of a nonpolarizable unit charge, q_k, located at a given point, k, of the grid, with the molecule of interest:

$$\mathscr{U}_{\mathrm{ind},k} = \mathscr{E}_{\mathrm{total},k} - \mathscr{E}^0 - q_k V_k \qquad (2)$$

where $\mathscr{E}_{\mathrm{total},k}$ denotes the energy of the molecule in the presence of the polarizing charge. $\mathscr{E}^0$ and V_k are the ground-state energy and the electrostatic potential at point k of the isolated molecule. As can be seen, mapping the induction energy over a grid of N_p points inevitably requires N_p distinct quantum mechanical calculations, generally using intramolecular electron correlation and double- to triple-ζ type basis sets, thereby limiting the applicability of the method to small enough chemical systems. To avoid undesirable hyperpolarizability effects and penetration of the charge densities, grids of points are constructed with particular care, usually increasing purposely the size of the van der Waals envelope of the molecule.[17,36–39]

As an alternative to the costly FP approach, approximate, infinitesimal second-order perturbation theory can be applied to obtain induction energies,[40,41] using restricted Hartree–Fock (RHF) wave functions:

$$\mathscr{U}_{\mathrm{ind}} \simeq \sum_{a\in\mathrm{occ}}\sum_{r\in\mathrm{vir}} \frac{1}{\varepsilon_a - \varepsilon_r}\left(\sum_\mu\sum_\nu c^*_{\mu a}c_{\nu r}\left\langle \varphi_\mu \left| \sum_k \frac{q_k}{|\mathbf{r}_k - \mathbf{r}|}\right| \varphi_\nu \right\rangle\right)^2 \qquad (3)$$

Here, ε_a and ε_r denote, respectively, the energy of occupied, a, and virtual, r, molecular orbitals of the isolated molecule. $|\varphi_\mu\rangle$ stands for a basis function and $c_{\mu a}$ is the coefficient of atomic orbital μ in the occupied molecular orbital a. The main advantage of this method resides in the fast mapping of the space around the molecule, based on a single quantum mechanical calculation. Its obvious drawback is rooted in the underlying approximations of the method, making use of an uncoupled form of the perturbed Hartree–Fock equations and the limitation to the RHF level, thus necessitating an *a posteriori* scaling of the computed induction energies to yield the expected answer.[39] A more rigorous approach has been devised recently by Williams and Stone for the direct evaluation of the induction energy on the basis of coupled-perturbed density functional theory (DFT) computations.[42]

One of the major dilemmas faced by the molecular modelling community is to increase the level of sophistication of current potential energy functions and thus limit the sampling in statistical simulations, or, alternatively, content itself with minimalist force fields to explore chemical and biophysical phenomena that occur over significant time scales. This choice will evidently be dictated by the problem investigated and the nature of the chemical systems. Whereas for large, complex assemblies of atoms and molecules, a rough representation of electrostatic forces by Coulomb interactions may prove to be satisfactory at a qualitative level, it may be desirable, in small enough systems, to improve the description of the constituent molecules by means of more sophisticated potential energy functions, if quantitative results are sought.[43] A major step forward on the road to more accurate intermolecular interaction models was made in the seventies by Claverie,[44] who showed the undeniable advantages of a piecewise parametrization of the physically distinct contributions of intermolecular energies and forces:

$$\mathscr{U}_{\mathrm{total}} = \mathscr{U}_{\mathrm{elec}} + \mathscr{U}_{\mathrm{ind}} + \mathscr{U}_{\mathrm{disp}} + \mathscr{U}_{\mathrm{rep}} \qquad (4)$$

where $\mathscr{U}_{\mathrm{elec}}$, $\mathscr{U}_{\mathrm{ind}}$, $\mathscr{U}_{\mathrm{disp}}$ and $\mathscr{U}_{\mathrm{rep}}$, respectively, stand for the electrostatic, the induction, the dispersion and the repulsion contributions to the total interaction energy, $\mathscr{U}_{\mathrm{total}}$.

The elegant and compact formalism developed by Stone for the determination of electrostatic and induction parameters is based on quantum mechanical monomer wave functions.[11,45] The key concept of distributed multipoles[10,11] and distributed polarizabilities[45] constitutes, at least in principle, the optimal route towards the calculation of extremely precise electrostatic and induction energies. Unfortunately, this DMM partitioning scheme is doomed to fail on account of the distributed response properties of the molecule, prone to be basis-set-dependent and leading to physically unacceptably high individual polarizability components.[46,47] Moreover, the constructed models soon become excessively complicated and intractable, as a consequence of the myriad of terms involved. Considering that a reasonably high level of theory is necessary to guarantee accurate distributed electric properties, the computational effort required to handle large molecular fragments is in many instances sufficiently overwhelming to force one to turn to the key concept of transferability, prevalent in most macromolecular force fields.

Only a handful of schemes targeted at partitioning the molecular volume into atomic or possibly larger regions of space possess acceptable transferability properties. For instance, simple net atomic charges derived from quantum mechanically determined electrostatic potentials are only partially transferable.[38,48,49] Whereas polar moieties, which are well-described by atom-centered point charges, transfer generally well between chemically homologous compounds, the same cannot be said for nonpolar fragments, which require higher-order local multipole moments for their accurate representation.[50,51] In sharp contrast with the previously discussed partitioning techniques that are based on the assignment of atomic orbital basis functions onto atoms or functional groups, partitioning the 3D physical space into mutually exclusive domains can also be contemplated. The formalism of atoms-in-molecules proposed by Bader,[52] which relies upon the topological partitioning of molecular charge densities, constitutes one possible appealing solution to circumvent the shortcomings of more rudimentary approaches like electrostatic potential derived charges and its numerous variants. The atomic domains, separated by the zero-flux surfaces, behave like quantum mechanical open systems and obey a number of fundamental quantum mechanical principles—for example, the virial theorem holds within each atomic domain. From a chemical standpoint, the most important feature is that these domains have very good transferability properties. Atomic charges and multipole moments are reasonably constant for atoms in different molecules found in chemically analogous environments. These multipole moments should be evaluated by numerical integrations within the atomic domains of often quite complicated form. The multicentered multipole expansion of the electrostatic potential based on the atoms-in-molecule scheme is slightly less effective than the DMM models. As has been demonstrated by Popelier and Kosov, the convergence properties of the multipole expansion for electrostatic potentials and interaction energies based on

topological atoms are, nevertheless, more favorable than expected before.[53–56]

Combining the partitioning of the physical space into atomic regions with the powerful concepts of distributed multipoles[10,11] and distributed polarizabilities[45–47] is at the origin of a novel approach called topologically partitioned electric properties (TPEP), which, at the price of numerical integrations over atomic basins, yields accurate and transferable electrostatic and induction models.[57–60] In this framework, the induction energy is obtained from the double space integration of the charge density of atom B, $\rho^{\mathrm{B}}(\mathbf{s})$, and the susceptibility density of atom A, $\alpha^{\mathrm{A}}_{\ell m,\ell' m'}(\mathbf{r},\mathbf{r}')$, written as:

$$\mathscr{U}^{(2)}_{\mathrm{ind}}(\mathrm{A}\leftarrow\mathrm{B}) = \frac{1}{2}\int \mathrm{d}\mathbf{s}\int \mathrm{d}\mathbf{s}'\int \mathrm{d}\mathbf{r}\int \mathrm{d}\mathbf{r}'\rho^{\mathrm{B}}(\mathbf{s})T_{\ell m}(\mathbf{s},\mathbf{r})\alpha^{\mathrm{A}}_{\ell m,\ell' m'} \times (\mathbf{r},\mathbf{r}')T_{\ell' m'}(\mathbf{r}',\mathbf{s}')\rho^{\mathrm{B}}(\mathbf{s}') \tag{5}$$

where $T_{\ell m}(\mathbf{s},\mathbf{r})$ is the ℓm element of the charge-multipole interaction tensor. $\alpha^{\mathrm{A}}_{\ell m,\ell' m'}(\mathbf{r},\mathbf{r}')$ is related to the multipole polarizabilities, defined by:

$$\alpha^{\mathrm{A}}_{\ell m,\ell' m'} = \int \mathrm{d}\mathbf{r}\int \mathrm{d}\mathbf{r}' R^{\mathrm{A}}_{\ell m}(\mathbf{r})\alpha^{\mathrm{A}}_{\ell m,\ell' m'}(\mathbf{r},\mathbf{r}')R^{\mathrm{A}}_{\ell' m'}(\mathbf{r}') \tag{6}$$

and $R_{\ell m}{}^{\mathrm{A}}(\mathbf{r})$ stands for a regular spherical harmonic.

Just like standard models of distributed multipoles or distributed polarizabilities, TPEP models[57–60] are plagued by the same tendency to involve a cumbersome plethora of independent parameters that might be relevant for benchmark applications, but are evidently too complicated to be plugged directly into molecular statistical simulations. The complexity of these models, which are difficult to handle in a practical fashion, can be ascribed to countless small, nonlocal components, corresponding to tedious, lengthy multipole expansions. It is, therefore, legitimate to wonder whether compact, tractable models can be designed without jeopardizing accuracy. One possible route towards this goal consists in increasing moderately the complexity of the model and subsequently map around the molecule the induction energy, to which a reduced set of polarizability components are fitted numerically. In spirit, this solution is very similar to the method devised by Ferenczy and colleagues[50,61] to construct compact models of atomic multipoles based on elaborate distributed multipole analyses (DMA).[10,11]

Transforming intricate, large ensembles of components into finite sets of physically meaningful, optimally partitioned electric properties (OPEP) constitutes a promising alternative to conventional distributed multipole and polarizability analyses.[62] In essence, OPEP represents a new route towards the derivation of reliable interaction parameters from virtually any electrostatic and induction data set, evaluated on a grid of points that samples the space around the molecule. These interaction parameters, ranging from simple net atomic charges and charge flows to more sophisticated distributed multipoles and polarizabilities, can be readily plugged into standard statistical simulation packages. Starting from a knowledge of the molecular geometry and running the relevant quantum chemical calculations, compact sets of parameters can be determined seamlessly, using various fitting procedures that take advantage of the molecular symmetry through local, atomic frames of reference and, whenever needed, properly selected constraints.

In the present contribution, progress made in recent years towards the design of OPEP are reviewed. The following section describes the framework, in which electrostatic and induction parameters are derived, introducing the working equations and providing the necessary computational details for the construction of those grids of points over which the relevant electric properties are mapped, the definition of local frames of reference to account for the molecular symmetry, and the fitting of the distributed properties. The methodology is then illustrated and discussed critically through the example of two prototypical aromatic compounds, benzene and naphthalene.

2. Theoretical background and computational details

2.1. Construction of the grid

A common route for the determination of distributed multipoles and polarizabilities consists in evaluating quantum mechanically the relevant electric properties—either the electrostatic potential or the induction energy, using Cartesian grids of regularly spaced points in the x, y and z directions around the molecule. The total number of points, N_{p}, of the grid is dictated by two parameters—*viz.* a multiplicative factor, ξ, for the van der Waals radius of the participating atoms and the separation, Δr, between consecutive points. All points lying further than ξ times the van der Waals radius[63] from any nucleus are discarded. In addition, points located within the so-called van der Waals envelope are also rejected. Remembering that the electrostatic potential is the sum of a long-range, multipolar contribution and a short-range, exponentially decreasing contribution due to the penetration of electron clouds, it should be ascertained that the former is not contaminated by the latter.[17,37,38] By doubling the van der Waals radius of those atoms forming the molecular envelope, the "exact" multipolar part of the quantum mechanical electrostatic potential can be safely described by distributed multipoles. In the context of distributed polarizabilities, it is recommended to adopt a similar strategy to avoid hyperpolarizability effects, which are likely to modulate the fitted induction interaction parameters.[36,39,64]

Definition of the shape and the size of the grid is dictated by the electric property to which interaction parameters will be fitted. On the basis of a thorough analysis, it has been concluded that by setting $\xi \leq 4$, the space around the molecule is appropriately sampled when distributed multipoles are derived from the electrostatic potential.[37,38] Increasing ξ leads to an exaggerated weight of those points lying far from the nucleus, reinforcing artificially the contribution of long-range multipoles—*viz.* the charge and the dipole, at the expense of the more subtle, short-range components. The same study revealed that beyond a certain threshold of Δr, *ca.* 0.4 Å, increasing further the density of points by diminishing the step of the grid has virtually no effect on the fitted atomic multipoles. In the light of a similar investigation, it has been shown that regions of space appreciably far from the nuclei ought to be sampled adequately to warrant a faithful reproduction of molecular dipole polarizabilities. This stringent condition is usually fulfilled by choosing $\xi = 5$ when building the grid over which the induction energy is mapped.[39] Compared to the molecular electrostatic potential, the density of the grid, however, appears to be of lesser importance when mapping the induction energy, thereby limiting the number of points to a reasonable value—a crucial requirement in the context of costly FP calculations. Put together, electrostatic potential grids are characterized by a high density of points lying close enough to the nuclei, whereas induction energy grids exhibit a lesser compactness, but a greater spatial extension from the molecule.

2.2. Distributed multipoles

The electrostatic potential created by an isolated molecule is evaluated quantum mechanically on a grid of N_{p} points. Using a multipole expansion, this reference quantum mechanical (QM) potential, V_k, can be approximated by:

$$\tilde{V}_k = \sum_a \sum_{\ell=0}^{N_\ell(a)} \sum_\kappa T^{ka}_{\ell\kappa} Q^a_{\ell\kappa}, \quad \forall k = 1,\ldots,N_{\mathrm{p}} \tag{7}$$

where $Q^a_{\ell\kappa}$ is the $\ell\kappa$ real spherical harmonic component of a multipole located at point $\mathbf{r}_a$. Here, $T^{ka}_{\ell\kappa} \equiv T_{\ell\kappa}(\mathbf{r}_k - \mathbf{r}_a)$. The optimal series of distributed multipoles is obtained by minimizing the functional:

$$f(\{Q^a_{\ell\kappa}\}) = \sum_{k=1}^{N_p} \left(V_k - \tilde{V}_k\right)^2 \qquad (8)$$

The number of distributed multipole components to be determined is equal to $N_c = \sum_a [2N_\ell(a) + 1]$. The solution, $\mathbf{Q} \equiv (Q^a_1, \cdots, Q^a_{N_c})$, which minimizes f($Q^a_{\ell\kappa}$), satisfies the so-called normal equations of the least-squares problem:

$$\mathbf{T}^T\mathbf{T}\mathbf{Q} = \mathbf{T}^T\mathbf{V} \qquad (9)$$

where $\mathbf{T}$ is an $N_p \times N_c$ matrix, the elements of which are $T^{ka}_{\ell\kappa}$. $\mathbf{V}$ is an N_p vector containing the different values of the electrostatic potential, V_k. Provided that the rank of $\mathbf{T}$ is equal to N_c, the matrix product $\mathbf{T}^T\mathbf{T}$ can, in principle, be inverted,[65] thus leading to the optimal, least-squares solution, $\mathbf{Q}$. Translation of the distributed quantities, $Q_{\ell_a\kappa_a}{}^a$, using the translation function defined by Stone,[66] yields the molecular moments at the origin.

In a number of cases, it may be desirable to constrain the total sum of partial charges to a predefined value. Although the use of reasonably dense grids of points guarantees a satisfactory reproduction of the total charge, small systematic errors can be advantageously avoided by imposing charge conservation. Furthermore, following the philosophy adopted by several statistical mechanics codes, in which a group-based truncation of electrostatic forces is implemented, the charge of selected fragments can be constrained.[62] From a more general perspective, this approach could be employed for the piecewise construction of macromolecular charge density models.

Aside from the molecular multipole moments, which allow the modeller to appraise the quality of the derived distributed moments by direct comparison to the corresponding reference QM expectation values, a global assessment of the fitting procedure is possible through an evaluation of the root-mean-square deviation (RMSD) and the mean relative deviation, $\Delta\varepsilon$, between the quantum mechanically determined and the regenerated electrostatic potentials:

$$\begin{cases} \mathrm{RMSD} = \left[\frac{1}{N_p}\sum_k \left(V_k - \tilde{V}_k\right)^2\right]^{1/2} \\ \Delta\varepsilon = \frac{1}{N_p}\sum_k \left|\frac{V_k - \tilde{V}_k}{V_k}\right| \end{cases} \qquad (10)$$

2.3. Distributed polarizabilities

As has been shown previously, the induction energy resulting from the polarization of the molecule by a nonpolarizable charge can be expressed by eqn. (2). Usually, a default value of the polarizing charge equal to +1.0 electron charge unit (e.c.u.) is chosen to map the induction energy over the grid of points. This choice, however, is not expected to affect significantly the derived models of distributed polarizabilities.[32–34] As was demonstrated in a recent study, provided that the points of the grid lie sufficiently far from the nuclei to avoid possible hyperpolarizability effects, variation of the charge from −0.7 to +1.2 e.c.u. results in the expected change of the induction energy.[64] In principle, the latter is anticipated to be invariant with respect to the sign of the polarizing charge, q_k. In reality, however, the electron cloud is polarized differently with $+q_k$ and $-q_k$, which is an artifact of the FP approach.[36] Polarizing charges of identical magnitude, but of opposite sign are, therefore, expected to yield slightly different results, albeit of comparable quality and accuracy.

2.3.1. Implicitly and explicitly interacting distributed polarizability models. The response of a many-body system to an external polarizing field is essentially nonlocal. Two possibilities may be considered, which differ in the way they take into account the coupling of the distributed subunits. The explicitly nonlocal polarizability parameters, $\alpha^{ab}_{\ell\kappa,\ell'\kappa'}$ describe the response of a given subunit—or more specifically, a pair of subunits—to the *external* field, that is the field created by the permanent charge distribution of the whole system. In this case, the interaction of the subunits, leading to the coupling of the polarizability response of different subunits of the molecular system, is included implicitly, at the quantum mechanical level. This model will be referred to as the implicitly interacting polarizability model.

The reference induction energy, $\mathscr{U}_{\mathrm{ind},k}$, can then be approximated at any given order of a distributed polarizability expansion:

$$\tilde{\mathscr{U}}_{\mathrm{ind},k} = -\frac{1}{2}\sum_{a,b}\sum_{\ell=0}^{N_\ell(a)}\sum_{\kappa}\sum_{\ell'=0}^{N_{\ell'}(b)}\sum_{\kappa'} q_k T^{ka}_{\ell\kappa}\alpha^{ab}_{\ell\kappa,\ell'\kappa'}T^{bk}_{\ell'\kappa'}q_k, \quad \forall k = 1,\ldots,N_p \qquad (11)$$

where the many-body polarizability components, $\alpha^{ab}_{\ell\kappa,\ell'\kappa'}$, can be parametrized directly. $N_\ell(a)$ and $N_{\ell'}(b)$ denote the orders of the multipole expansion at centers a and b, respectively.

Most of the polarizable force fields adopt a different viewpoint, by reflecting the many-body nature of the polarizability response through an explicit but approximate interaction model. Such schemes with various degrees of sophistication have been proposed, ranging from the simple dipole–dipole interaction scheme of Applequist[31] to various screened multipole-multipole models, like those of Birge[67] or Thole.[68] The general form of such an explicitly interacting many-body polarizability model is:

$$\alpha = \mathbf{S}\mathbf{a} = (\mathbf{I} + \mathbf{a}\tilde{\mathbf{T}})^{-1}\mathbf{a} \qquad (12)$$

where $\mathbf{S}$, the screening matrix, depends on the "explicitly interacting" multipole polarizability parameters, $\mathbf{a}$, of the individual subunits as well as on their geometry-dependent interaction function $\tilde{\mathbf{T}}$. The advantage of this latter formulation is that the number of possible components to fit increases linearly, rather than quadratically, with the number of atoms. Furthermore, these polarizability parameters are supposed to be conformationally independent, since it is the explicit interaction that is supposed to absorb most of the conformation-dependence of the polarizability.[48] The price to pay, however, is the approximate and rudimentary character of the coupling interaction, compared to the full quantum mechanical description of the coupling in the implicitly interacting model.

2.3.2. Calculation of implicitly interacting polarizability models. Here, $\alpha^{ab}_{\ell\kappa,\ell'\kappa'}$ stand for the components of the many-body polarizability tensor α^{ab} that corresponds to sites a and b. The total number of distinct components of a, which contains all possible one- and two-center distributed polarizabilities, is $N_\zeta = \sum_{a\le b}[2N_\ell(a) + 1] \times [2N_\ell(b) + 1]$. Interestingly enough, eqn. (11) yields the total induction energy, wherein higher-order terms are strictly zero. This is a consequence of using a nonpolarizable perturbing charge, q_k. In practice, the number of non-zero components, $N_c \ll N_\zeta$, is kept as low as possible without compromising the accuracy of the fit. As has been commented on by Nakagawa,[32–34] sampling of the space around the molecule might be improved by considering in eqn. (11) all pairs of grid points, $\{k,k'\}$, at which the polarizing charge is located, which increases the data base of induction energy values to $N_p \times (N_p - 1)/2$.

Since the number of components of α to be determined is substantially smaller than the number of points of the grid, the

over-determined system of linear equations expressed in eqn. (11) can be conveniently rewritten in a matrix form similar to that proposed for distributed multipoles:

$$\mathbf{B}\alpha = \mathbf{U}_{\mathrm{ind}} \tag{13}$$

where $\mathbf{U}_{\mathrm{ind}}$ is the vector that contains the induction energies computed at various locations of the polarizing charge. $\mathbf{B}$ is an $N_{\mathrm{p}} \times N_{\mathrm{c}}$ matrix constructed with the $N_{\mathrm{p}} \times N_{\zeta}$ electrostatic tensor, $\mathbf{T}$. The elements of matrix $\mathbf{B}$ are defined by:

$$B_{k,s} = -\frac{1}{2} q_k^2 T_{\ell\kappa}^{ka} T_{\ell'\kappa'}^{bk}, \quad \forall k, s = 1, \ldots, N_{\mathrm{c}} \tag{14}$$

where $s \equiv (ab; \ell\kappa; \ell'\kappa')$. The computational effort involved in the fitting procedure can be reduced appreciably by considering the symmetry relationship $\alpha_{\ell\kappa,\ell'\kappa'}^{ab} = \alpha_{\ell'\kappa',\ell\kappa}^{ba}$, so that only the upper triangle of matrix $\mathbf{B}$ is evaluated. The dimensionality of the least-squares problem is lowered accordingly to, at the most, $N_{\mathrm{p}} \times N_{\zeta}(N_{\zeta} + 1)/2$.

The matrix block that corresponds to charge flow polarizabilities requires particular care. The solution provided by the fitting procedure may turn out to be mathematically correct, but not necessarily physically so. A crucial property that the algorithm ought to enforce is the absence of an induced net charge in the molecule of interest.

The over-determined system of linear equations, summarized in eqn. (11), is solved by minimizing the functional:

$$g(\{\alpha_{\ell\kappa,\ell'\kappa'}^{ab}\}) = \sum_k \left(\mathscr{U}_{\mathrm{ind},k} - \tilde{\mathscr{U}}_{\mathrm{ind},k}\right)^2 \tag{15}$$

Its optimal solution, in a least-squares sense, satisfies the normal equations:

$$\mathbf{B}^{\mathrm{T}}\mathbf{B}\alpha = \mathbf{B}^{\mathrm{T}}\mathbf{U}_{\mathrm{ind}} \tag{16}$$

Provided that the rank of matrix $\mathbf{B}$ is equal to N_{c}, the matrix product $\mathbf{B}^{\mathrm{T}}\mathbf{B}$ can, in principle, be inverted, hence, yielding the set of N_{c} polarizability components.

In the spirit of the translation of atomic multipole moments, molecular polarizabilities may be regenerated from the distributed components, employing the formulae devised by Stone.[66]

The accuracy of the fit may further be probed by means of the RMSD and the corresponding mean relative deviation, $\Delta\varepsilon$, introduced in eqn. (10), replacing the target electrostatic potential by the induction energy mapped on the grid of N_{p} points.

2.4. Molecular symmetry and local frames of reference

In order to be useful, electrostatic and induction parameters should reflect the symmetry of the molecule and the transferability properties that are expected for a given family of chemically similar compounds.[37] These requirements are of paramount importance when designing multipurpose, all-atom, macromolecular force fields, in which the number of different types of atoms is necessarily limited for obvious practical reasons. Discretization of electric properties on a grid of points unavoidably results in slight variations in the fitted parameters of symmetry-related atoms. It is, therefore, recommended to preserve molecular symmetry by imposing *exact* transformations between tensor properties of these symmetry-related atoms. In contrast, transferability exploits the *approximate* equivalence of the tensor properties common to those atoms that share similar chemical environments. Whereas the transformation of zeroth-order properties—atom-centered point charges and charge-flow polarizabilities—is trivial, equivalence of higher-order properties cannot be captured without appropriate transformation into a common frame of reference, which requires a preliminary analysis of the local environments in the molecule.

Once the local frames of reference are determined, distributed multipoles or polarizabilities can be derived from the relevant electric property over families rather than individual atoms. To achieve this goal, use is made of the formalism of charge-multipole interaction tensors, $T_{00,\ell\kappa}^{ka}(\Omega_a)$, that intrinsically depend upon the orientation, Ω_a, of the local frames of reference, in which the multipole moments, $Q_{\ell\kappa}^a$, are calculated.[66] Adopting this philosophy, the approximation of the QM electrostatic potential defined in eqn. (7) may be readily restated into:

$$\tilde{V}_k = \sum_{A=1}^{N_{\mathrm{f}}} \sum_{\ell}^{N_\ell(A)} \sum_{\kappa} Q_{\ell\kappa}^A \sum_{a\in A}^{N_a(A)} T_{\ell\kappa}^{ka}(\Omega_a), \quad \forall k = 1, \ldots, N_{\mathrm{p}} \tag{17}$$

Here, N_{f} denotes the number of families of symmetry-related atoms, and $N_a(A)$ is the number of atoms belonging to family A. The approximate expression of the induction energy, *viz.* eqn. (11), can be rearranged in a similar spirit, allowing one to work with a non-redundant set of polarizability parameters.

2.5. Computational details

The prototypical cases of benzene and naphthalene were chosen to analyze the relative performance of simple models of distributed multipoles and polarizabilities. Molecular geometries were optimized at the second-order Møller–Plesset level of theory (MP2), using the Gaussian suite of programs with the triple-ζ-like 6-311++G** basis set. The open-source code OPEP was employed to construct the grids of points (see Fig. 1), over which the molecular electrostatic potential and the induction energy were subsequently mapped. In the case of the former, sufficiently dense Cartesian grids were obtained with $\zeta = 3$ and a grid step of 0.42 Å. This corresponds to 8008 and 9851 points for benzene and naphthalene, respectively. As has been underlined previously, mapping the induction energy requires somewhat more diffuse Cartesian grids, with a greater spatial extent of the points. This is achieved by setting ξ to 5, and Δr to *ca.* 0.75 Å. Taking advantage of the molecular

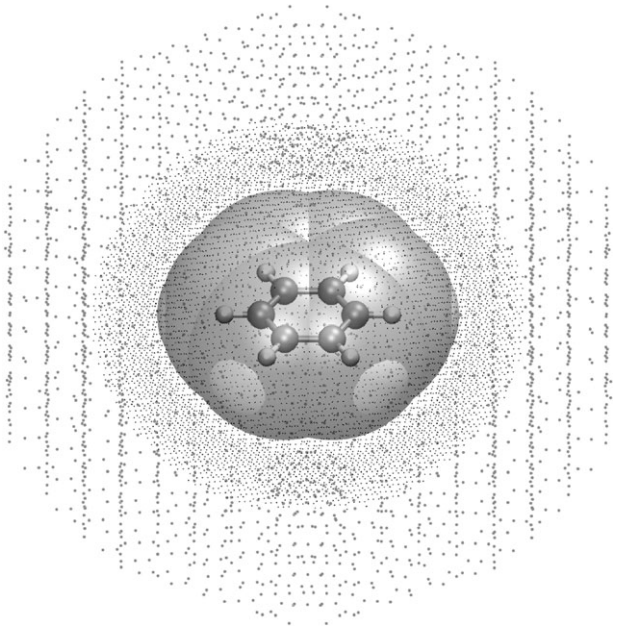


Fig. 1 Construction of a grid of points around benzene. In sharp contrast with quantum mechanical electrostatic potentials, induction energies are mapped over points with a sufficient spatial extent from any nucleus—*viz.* typically up to five times the participating van der Waals radii. To avoid spurious penetration and nonlinear hyperpolarizability effects, the van der Waals radius of those atoms forming the molecular envelope has been doubled.

symmetry, only 115 points were actually considered for benzene, and 214 for naphthalene. The full grids of points sampling the space around the molecules can then be recovered by applying the relevant symmetry operations. This step, however, may not be necessary, should the electric properties be fitted for families of symmetry-related atoms instead of the complete set of atoms forming the molecule. In the first instance, the non-redundant information contained in a reduced portion of the molecular space is sufficient to account for induction phenomena over the entire molecule. The electrostatic potential was computed using the Gaussian suite of programs[69] at the MP2 level of approximation with the basis set of Sadlej,[70], which was shown to represent a middle ground between the number of basis functions and the accuracy to reproduce electric properties. Induction energies were derived from an FP scheme, at the same level of theory.

3. Results and discussion

The models of distributed multipoles for benzene and naphthalene are presented in Table 1. From the onset, it can be observed that, whereas models limited to a monopole approximation (model 1b) provide a reasonable reproduction of the target MP2/Sadlej electrostatic potential, inclusion of point dipoles onto the participating carbon atoms (model 1a) improves the agreement markedly. It may be safely inferred that, for these aromatic compounds, models consisting of atomic charges and dipoles probably constitute the best compromise between accuracy and tractability.[51] From a transferability point-of-view, it should be noted that only the C_β carbon atoms of naphthalene bear a point charge comparable to that found in benzene, which reflects the similar chemical environments of these atoms. Interestingly enough, usage of local frames of reference allows distributed moments to be written

Table 1 Models of distributed multipoles and regenerated molecular multipole moments of benzene and naphthalene, at the MP2/Sadlej//MP2/6-311++G(2d,2p) level of approximation

	Distributed multipoles		Molecular multipoles		MP2/Sadlej
Benzene					
Model 1a	Q_{00}^{C}	0.2195	Q_{20}	−6.5508	−6.5657
	Q_{00}^{H}	−0.2195	Q_{40}	178.7703	182.2422
	Q_{10}^{C}	0.2080			
	RMSD/10^{-3} a.u.	−0.061			
	% $\Delta\varepsilon$	5.586			
Model 1b	Q_{00}^{C}	−0.1437	Q_{20}	−6.4436	−6.5657
	Q_{00}^{H}	0.1437	Q_{40}	139.5256	182.2422
	RMSD/10^{-3} a.u.	0.218			
	% $\Delta\varepsilon$	28.849			
Naphthalene					
Model 1a	$Q_{00}^{C_\alpha}$	−0.9079	Q_{20}	−10.2908	−10.3035
	$Q_{00}^{C_\beta}$	−0.1969	Q_{22c}	0.0554	0.0684
	$Q_{00}^{C_\gamma}$	1.2704	Q_{40}	564.6107	567.1866
	$Q_{00}^{H_\alpha}$	0.2106	Q_{42c}	−260.9683	−261.8359
	$Q_{00}^{H_\beta}$	0.2590	Q_{44c}	97.9001	96.5461
	$Q_{10}^{C}{}_{\alpha}$	0.2209			
	$Q_{11c}^{C_\alpha}$	−0.7924			
	$Q_{10}^{C_\beta}$	0.4282			
	$Q_{11c}^{C_\beta}$	0.3969			
	$Q_{10}^{C_\gamma}$	−1.4968			
	RMSD/10^{-3} a.u.	0.046			
	% $\Delta\varepsilon$	2.093			
Model 1b	$Q_{00}^{C_\alpha}$	−0.3826	Q_{20}	−10.2466	−10.3035
	$Q_{00}^{C_\beta}$	−0.1426	Q_{22c}	0.0953	0.0684
	$Q_{00}^{C_\gamma}$	0.3356	Q_{40}	539.3711	567.1866
	$Q_{00}^{H_\alpha}$	0.1941	Q_{42c}	−264.8325	−261.8359
	$Q_{00}^{H_\beta}$	0.1633	Q_{44c}	119.2445	96.5461
	RMSD/10^{-3} a.u.	0.116			
	% $\Delta\varepsilon$	8.382			

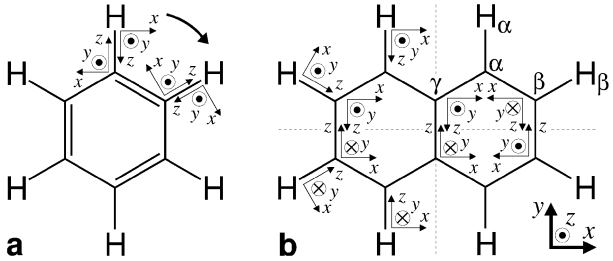


Fig. 2 Definition of the local frames of (a) benzene and (b) naphthalene utilized to derive the models of distributed multipoles and polarizabilities. Passage between local frames is achieved by applying the symmetry operations appropriate to the molecule.

in a convenient and compact form. Expression of these moments for the different atoms of the molecule can be readily recovered by applying the appropriate symmetry operations, as suggested in Fig. 2. A closer examination of the regenerated molecular moments in Table 1 reveals that models limited to net atomic charges offer a surprisingly remarkable description of molecular quadrupoles.[71] This result rationalizes the success of molecular mechanical force fields for investigating π–π association, wherein quadrupole interactions can play a significant role.[72]

Explicit inclusion of induction effects may turn out to be a critical issue in the case of aromatic compounds interacting with highly polarizable species like ions. Such is the case, for instance, of cation–π interactions, which are commonly found in a host of molecular assemblies, in particular those of biological interest.[73] Modelling of cation–π complexes by means of molecular mechanical force fields illustrates the shortcomings of the pairwise additive approximation.[74,75] In Table 2, four distinct models of distributed polarizabilities are proposed for benzene and naphthalene, at diverse levels of sophistication. The simplest description corresponds to charge flow polarizabilities only (model 2a) and yields a rather poor reproduction of the target MP2/Sadlej induction energies. Discrepancies originate mostly from the absence of out-of-plane components, which tends to exaggerate the in-plane, $\alpha_{11c,11c}$ and $\alpha_{11s,11s}$, contributions.[36] Representation of out-of-plane molecular polarizability components can, nonetheless, be enforced by utilizing atomic dipolar models. If the latter are

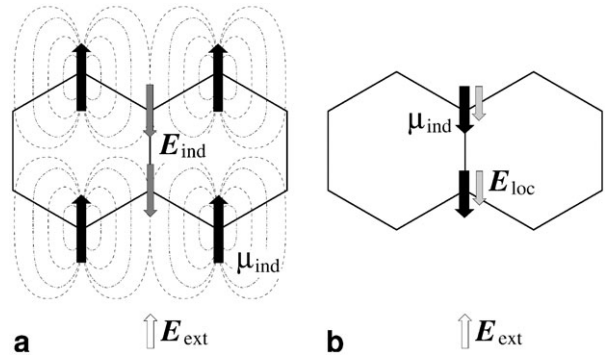


Fig. 3 Illustration of how negative local components occur in a model of implicitly interacting distributed polarizabilities. In the prototypical case of naphthalene, the external polarizing field creates induced dipole moments on the outermost carbon atoms (black arrows). The local field felt by the central, C_γ, carbon atoms corresponds to the sum of the external field and that generated by the induced moments. This effect is dominated by the superposition of the induced fields, the direction of which is reversed with respect to that of the external field (a). The induced dipole moments are parallel to the local field, but anti-parallel to the external field. Since in the model of implicitly interacting distributed polarizabilities, polarizability components describe the proportionality between the induced moments and the external polarizing field, the central C_γ carbon atoms necessarily bear negative atomic dipole polarizabilities (b).

Table 2 Models of distributed polarizabilities and regenerated molecular polarizabilities of benzene and naphthalene, at the MP2/Sadlej//MP2/6-311++G(2d,2p) level of approximation

	Distributed polarizabilities		Molecular polarizabilities		Experimental[76]
Benzene					
Model 2a	α_{00}^{CC}	−2.281	$\alpha_{10,10}$	0.000	47.7
	α_{00}^{CH}	−4.017	$\alpha_{11c,11c}$	97.91	79.2
			$\alpha_{11s,11s}$	97.91	79.2
	RMSD/10^{-3} a.u.	0.383			
	% $\Delta\varepsilon$	28.537			
Model 2b	$\alpha_{1m,1m}^{C}$	11.467	$\alpha_{10,10}$	68.80	47.7
			$\alpha_{11c,11c}$	68.80	79.2
			$\alpha_{11s,11s}$	68.80	79.2
	RMSD/10^{-3} a.u.	0.214			
	% $\Delta\varepsilon$	16.346			
Model 2c	α_{00}^{CC}	−1.822	$\alpha_{10,10}$	47.54	47.7
	α_{00}^{CH}	−0.280	$\alpha_{11c,11c}$	89.09	79.2
	$\alpha 00_{1m,1m}^{C}$	7.923	$\alpha_{11s,11s}$	89.09	79.2
	RMSD/10^{-3} a.u.	0.025			
	% $\Delta\varepsilon$	2.737			
Model 2d	$\alpha_{10,10}^{C}$	10.163	$\alpha_{10,10}$	47.48	47.7
	$\alpha_{11c,11c}^{C}$	19.347	$\alpha_{11c,11c}$	88.53	79.2
	$\alpha_{11s,11s}^{C}$	7.913	$\alpha_{11s,11s}$	88.53	79.2
	RMSD/10^{-3} a.u.	0.027			
	% $\Delta\varepsilon$	2.674			
Naphthalene					
Model 2a	$\alpha_{00}^{C\alpha C\beta}$	22.374	$\alpha_{10,10}$	0.00	70.9
	$\alpha_{00}^{C\alpha C\gamma}$	−32.955	$\alpha_{11c,11c}$	427.51	162.0
	$\alpha_{00}^{C\alpha C\gamma}$	54.428	$\alpha_{11s,11s}$	88.51	119.5
	$\alpha_{00}^{C\alpha C\gamma}$	−104.671			
	$\alpha_{00}^{C\alpha H\alpha}$	1.187			
	$\alpha_{00}^{C\alpha H\alpha}$	−12.516			
	RMSD/10^{-3} a.u.	0.217			
	% $\Delta\varepsilon$	16.442			
Model 2b	$\alpha_{1m,1m}^{C\alpha}$	49.563	$\alpha_{10,10}$	106.04	70.9
	$\alpha_{1m,1m}^{C\beta}$	20.248	$\alpha_{11c,11c}$	106.04	162.0
	$\alpha_{1m,1m}^{C\gamma}$	−86.602	$\alpha_{11s,11s}$	106.04	119.5
	RMSD/10^{-3} a.u.	0.100			
	% $\Delta\varepsilon$	7.797			
Model 2c	$\alpha_{00}^{C\alpha C\beta}$	−2.895	$\alpha_{10,10}$	75.38	70.9
	$\alpha_{00}^{C\gamma}$	−2.031	$\alpha_{11c,11c}$	174.18	162.0
	$\alpha_{00}^{C\beta C\beta}$	−2.249	$\alpha_{11s,11s}$	121.59	119.5
	$\alpha_{00}^{C\gamma C\gamma}$	3.156			
	$\alpha_{00}^{C\alpha H\alpha}$	−0.296			
	$\alpha_{00}^{C\beta H\beta}$	0.351			
	$\alpha_{1m,1m}^{C\alpha}$	21.063			
	$\alpha_{1m,1m}^{C\beta}$	6.708			
	$\alpha_{1m,1m}^{C\beta}$	−17.852			
	RMSD/10^{-3} a.u.	0.026			
	% $\Delta\varepsilon$	1.993			
Model 2d	$\alpha_{10,10}^{C\alpha}$	24.098	$\alpha_{10,10}$	74.66	70.9
	$\alpha_{11c,11c}^{C\alpha}$	38.469	$\alpha_{11c,11c}$	175.65	162.0
	$\alpha_{11s,11s}^{C\alpha}$	18.599	$\alpha_{11s,11s}$	121.34	119.5
	$\alpha_{10,11c}^{C\alpha}$	3.169			
	$\alpha_{10,10}^{C\beta}$	12.398			
	$\alpha_{11c,11c}^{C\beta}$	10.722			
	$\alpha_{11s,11s}^{C\beta}$	8.309			
	$\alpha_{10,11c}^{C\beta}$	−1.960			
	$\alpha_{10,10}^{C\gamma}$	−12.316			
	$\alpha_{11c,11c}^{C\gamma}$	−10.563			
	$\alpha_{11s,11s}^{C\gamma}$	−16.488			
	RMSD/10^{-3} a.u.	0.025			
	% $\Delta\varepsilon$	1.948			

limited to an isotropic atomic description (model 2b), a somewhat better agreement with the MP2/Sadlej induction energies is obtained. Yet, such models remain inappropriate because they impose the equivalence of in-plane and out-of-plane polarizability components. This effect is particularly glaring in the case of naphthalene, which exhibits a strong anisotropy, manifested in the components of its molecular dipole polarizability. Anisotropy effects may be mimicked by either considering anisotropic polarizabilities explicitly (model 2d) or by incorporating charge flow polarizabilities in the isotropic model (model 2c). For both benzene and naphthalene, these two representations behave similarly, leading to equivalent reproductions of the target induction energies. Compared to a fully isotropic description, introduction of anisotropy reduces by

two orders of magnitude the RMSD between the reproduced and the quantum mechanically calculated induction energy. Furthermore, the regenerated molecular dipole polarizabilities for these two models agree quantitatively, and approach the corresponding experimental values.[76] It is worth noting that, contrary to the C_α and C_β carbon atoms, the atomic polarizability borne by C_λ is negative, as predicted by Fig. 3. This phenomenon is rooted in the implicit interaction model adopted in the present scheme to account for induction effects. In the classical view, the polarizability parameter describes the proportionality between the induced dipole moment and the local electric field—that is, the superposition of the external field and that created by the induced moments—rather than the external one, as assumed in the implicitly interacting polarizability models. The local field felt by the central carbon atom, C_γ, is inverted with respect to the external field, owing to the cumulative contributions of the participating induced moments borne by the surrounding atoms.

4. Concluding remarks

Improving the accuracy and the reliability of the electrostatic and the induction contributions of multipurpose potential energy functions undeniably constitutes one of the major challenges faced today by the modelling community, as statistical simulations of increasingly complex molecular assemblies are becoming more precise and realistic. Design of well-behaved models of distributed multipoles and polarizabilities, for which accuracy and compactness are optimally balanced, is now within reach using the set of tools described herein. This tool kit relies upon robust numerical approaches that take advantage of both the symmetry and the transferability properties of the functional groups by means of local atomic frames of reference.[62] It epitomizes the progress made in recent years in the development of adapted methods for the derivation of distributed polarizabilities,[45] following a non-heuristic scheme and based on an accurate determination of the quantum mechanical induction energy, which has been discretized on a grid of points.[36] Whereas the resulting models, in principle, may be plugged directly into a simulation program capable of handling polarizable systems, great care should be taken when considering how induction phenomena are treated. In particular, the full quantum mechanical approach proposed here, which has been referred to as the implicitly interacting polarizable model, is not suited for the statistical simulation of flexible molecules. The model is parametrized for a given geometry and, therefore, cannot account for the variation in the mutual interaction of the individual subunits when the conformation of the molecule changes. On the other hand, an explicitly interacting polarizable model, relying on a heuristic description of the intramolecular interactions, is able to follow variations in the molecular response caused by conformational modifications. Careful selection of non-redundant components also represents a critical step in the derivation of physically meaningful polarizability parameters able to mimic faithfully the anisotropy of the molecule.[36,64] It should be recalled that this route towards the parametrization of electrostatic and induction energy terms is bereft of underlying empirical assumptions and is solely based on quantum mechanical charge and response densities. This paves the road to the study of chemical systems of any sort, ranging from organic and inorganic compounds to more complex, biologically relevant molecular assemblies, amenable to quantum chemical calculations.

Acknowledgements

The authors would like to thank Drs François Dehez, Claude Millot (Vandœuvre-lès-Nancy, France), Christof Hättig (Karlsruhe, Germany), Georg Jansen (Düsseldorf, Germany) and F. Javier Luque (Barcelona, Spain) for many stimulating and fruitful discussions. The Centre Charles Hermite (CRVHP; Vandœuvre-lès-Nancy, France) and the CINES (Montpellier, France) are gratefully acknowledged for provision of generous amounts of CPU time on their SGI Origin 3000 and IBM Power 4 array.

References

1 G. Nemethy, K. D. Gibson, K. A. Palmer, C. N. Yoon, G. Paterlini, A. Zagari, S. Rumsey and H. A. Scheraga, *J. Phys. Chem.*, 1992, **96**, 6472–6484.
2 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. C. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.
3 A. D. MacKerell, Jr, D. Bashford, M. Bellott, R. L. Dunbrack, Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B.*, 1998, **102**, 3586–3616.
4 GROMOS force field: W. F. van Gunsteren, X. Daura and A. E. Mark, in *Encyclopedia of Computational Chemistry*, eds. P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III and P. R. Schreiner, Wiley and Sons, Chichester, 1998, **vol. 2**, pp. 1211–1216.
5 S. M. Bachrach, in *Reviews in Computational Chemistry*, eds. K. Lipkowitz and D. B. Boyd, VCH, New York, 1994, **vol. 5**, pp. 171–228..
6 W. D. Cornell and C. Chipot, in *Encyclopedia of Computational Hhemistry*, eds. P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III and P. R. Schreiner, Wiley and Sons, Chichester, 1998, **vol. 1**. pp. 258–263.
7 R. S. Mulliken, *J. Chem. Phys.*, 1955, **23**, 1833–1840.
8 W. A. Sokalski and R. A. Poirier, *Chem. Phys. Lett.*, 1983, **98**, 86–92.
9 W. A. Sokalski and A. Sawaryn, *J. Mol. Struct. (Theochem)*, 1992, **256**, 91–112.
10 A. J. Stone, *Chem. Phys. Lett.*, 1981, **83**, 233–239.
11 A. J. Stone and M. Alderton, *Mol. Phys.*, 1985, **56**, 1047–1064.
12 F. Vigné-Maeder and P. Claverie, *J. Chem. Phys.*, 1988, **88**, 4934–4948.
13 D. S. Coombes, S. L. Price, D. J. Willock and M. Leslie, *J. Phys. Chem.*, 1996, **100**, 7352–7360.
14 S. R. Cox and D. E. Williams, *J. Comput. Chem.*, 1981, **2**, 304–323.
15 L. E. Chirlian and M. M. Francl, *J. Comput. Chem.*, 1987, **8**, 894–905.
16 M. M. Francl and L. E. Chirlian, *Rev. Comput. Chem.*, 2000, **14**, 1–31.
17 F. Colonna, E. Evleth and J. G. Ángyán, *J. Comput. Chem.*, 1992, **13**, 1234–1245.
18 P. Ahlström, A. Wallquist, S. Engströms and B. Jönsson, *Mol. Phys.*, 1989, **68**, 563–581.
19 L. X. Dang, J. E. Rice, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1991, **113**, 2481–2486.
20 D. N. Bernardo, Y. Ding, K. Krogh-Jesperesen and R. M. Levy, *J. Phys. Chem.*, 1994, **98**, 4180–4187.
21 M. H. New and B. J. Berne, *J. Am. Chem. Soc.*, 1995, **117**, 7172–7179.
22 J. C. Soetens, C. Millot, C. Chipot, G. Jansen, J. G. Ángyán and B. Maigret, *J. Phys. Chem. B*, 1997, **101**, 10910–10917.
23 A. J. Stone, *Chem. Phys. Lett.*, 1989, **155**, 102–110.
24 S. W. Rick and S. J. Stuart, in *Reviews in Computational Chemistry*, eds. K. Lipkowitz and D. B. Boyd, VCH, New York, 2002, **vol. 18**, pp. 89–146.
25 C. J. Cramer and D. G. Truhlar, *Chem. Rev.*, 1999, **99**, 2161–2200.
26 F. J. Luque, J. M. Bofill and M. Orozco, *J. Chem. Phys.*, 1995, **103**, 10183–10191.
27 (*a*) J. G. Ángyán, *J. Chem. Phys.*, 1995, **103**, 10183–10191; (*b*) J. G. Ángyán, *J. Chem. Phys.*, 1997, **107**, 1–3.
28 (*a*) F. J. Luque, J. M. Bofill and M. Orozco, *J. Chem. Phys.*, 1997, **107**, 1–3; (*b*) F. J. Luque, J. M. Bofill and M. Orozco, *J. Chem. Phys.*, 1997, **107**, 5573–5582.
29 C. Chipot, *J. Comput. Chem.*, 2003, **24**, 409–415.
30 J. A. Barker, *Proc. Roy. Soc. A*, 1953, **219**, 367–372.
31 J. Applequist, *Acc. Chem. Res.*, 1977, **10**, 79–85.
32 S. Nakagawa and N. Kosugi, *Chem. Phys. Lett.*, 1993, **210**, 180–186.
33 S. Nakagawa, *Chem. Phys. Lett.*, 1997, **278**, 272–277.

34 S. Nakagawa, *Chem. Phys. Lett.*, 1995, **246**, 256–262.
35 I. Alkorta, M. Bachs and J. J. Perez, *Chem. Phys. Lett.*, 1994, **224**, 160–165.
36 N. Celebi, J. G. Ángyán, F. Dehez, C. Millot and C. Chipot, *J. Chem. Phys.*, 2000, **112**, 2709–2717.
37 C. Chipot, B. Maigret, J. L. Rivail and H. A. Scheraga, *J. Phys. Chem.*, 1992, **96**, 10276–10284.
38 J. G. Ángyán and C. Chipot, *Int. J. Quantum Chem.*, 1994, **52**, 17–37.
39 C. Chipot, F. Dehez, J. G. Ángyán, C. Millot, M. Orozco and F. J. Luque, *J. Phys. Chem. A.*, 2001, **105**, 11505–11514.
40 F. J. Luque and M. Orozco, *J. Comput. Chem.*, 1998, **19**, 866–881.
41 C. Chipot and F. J. Luque, *Chem. Phys. Lett.*, 2000, **19**, 190–198.
42 G. J. Williams and A. J. Stone, *J. Chem. Phys.*, 2003, **119**, 4620–4628.
43 P. Cieplak, J. W. Caldwell and P. A. Kollman, *J. Comp. Chem.*, 2001, **22**, 1048–1057.
44 P. Claverie, in *Intermolecular Interactions: From Diatomics to Biopolymers*, ed. B. Pullman, Wiley–Interscience, New York, 1978, pp. 69–306..
45 A. J. Stone, *Mol. Phys.*, 1985, **56**, 1065–1082.
46 C. R. Le Sueur and A. J. Stone, *Mol. Phys.*, 1973, **78**, 1267–1291.
47 C. R. Le Sueur and A. J. Stone, *Mol. Phys.*, 1994, **83**, 293–308.
48 F. Colonna and E. Evleth, *Chem. Phys. Lett.*, 1993, **212**, 665–670.
49 C. Chipot, J. Ángyán, B. Maigret and H. A. Scheraga, *J. Phys. Chem.*, 1993, **97**, 9797–9807.
50 C. Chipot, J. G. Ángyán, G. G. Ferenczy and H. A. Scheraga, *J. Phys. Chem.*, 1993, **97**, 6628–6636.
51 C. Chipot, J. G. Ángyán and C. Millot, *Mol. Phys.*, 1998, **97**, 881–895.
52 R. F. W. Bader, *Atoms in Molecules–A Quantum Theory*, Oxford University Press, London, 1990.
53 D. Kosov and P. L. A. Popelier, *J. Phys. Chem. A.*, 2000, **104**, 7339–7345.
54 D. Kosov and P. L. A. Popelier, *J. Chem. Phys.*, 2000, **113**, 3969–3974.
55 P. L. A. Popelier and D. Kosov, *J. Chem. Phys.*, 2001, **114**, 6539–6547.
56 P. L. A. Popelier and D. Kosov, *J. Phys. Chem. A.*, 2001, **105**, 8254–8261.
57 J. G. Ángyán, G. Jansen, M. Loos, C. Hättig and B. A. Heß, *Chem. Phys. Lett.*, 1994, **219**, 267–273.
58 C. Hättig, G. Jansen, B. A. Heß and J. G. Ángyán, *Can. J. Chem.*, 1996, **74**, 976–987.
59 G. Jansen, C. Hättig, B. A. Heß and J. G. Ángyán, *Mol. Phys.*, 1996, **88**, 69–92.
60 C. Hättig, G. Jansen, B. A. Heß and J. G. Ángyán, *Mol. Phys.*, 1997, **91**, 145–160.
61 G. G. Ferenczy, *J. Comput. Chem.*, 1991, **12**, 913–917.
62 J. G. Ángyán, C. Chipot, F. Dehez, C. Hättig, G. Jansen and C. Millot, *J. Comput. Chem.*, 2003, **24**, 997–1008.
63 A. Bondi, *J. Phys. Chem.*, 1964, **68**, 441–451.
64 F. Dehez, J. C. Soetens, C. Chipot, J. G. Ángyán and C. Millot, *J. Phys. Chem. A*, 2000, **104**, 1293–1303.
65 M. M. Francl, C. Carey, L. E. Chirlian and D. Gange, *J. Comput. Chem.*, 1996, **17**, 367–383.
66 A. J. Stone, in *Theoretical Models of Chemical Bonding*, ed. H. Maksić, Springer–Verlag, Berlin, 1991, **vol. 4**, pp. 103–131.
67 R. R. Birge, *Chem. Phys.*, 1980, **72**, 5312–5319.
68 B. T. Thole, *Chem. Phys.*, 1981, **59**, 341–350.
69 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *GAUSSIAN 03 (Revision A.1)*, Gaussian Inc., Pittsburgh, PA, 2003.
70 A. J. Sadlej, *Collec. Czech. Chem. Commun.*, 1988, **53**, 1995–2016.
71 C. A. Hunter, *Chem. Soc. Rev.*, 1994, **23**, 101–109.
72 C. Chipot, R. Jaffe, B. Maigret, D. A. Pearlman and P. A. Kollman, *J. Am. Chem. Soc.*, 1996, **118**, 11217–11224.
73 J. C. Ma and D. A. Dougherty, *Chem. Rev.*, 1997, **97**, 1303–1324.
74 C. Chipot, B. Maigret, D. A. Pearlman and P. A. Kollman, *J. Am. Chem. Soc.*, 1996, **118**, 2998–3005.
75 H. Minoux and C. Chipot, *J. Am. Chem. Soc.*, 1999, **121**, 10366–10372.
76 M. Okruss, R. Müller and A. Hese, *J. Chem. Phys.*, 1999, **110**, 10393–10402.