

CHAPTER 6

DESIGN OF NEXT GENERATION FORCE FIELDS FROM AB INITIO COMPUTATIONS: BEYOND POINT CHARGES ELECTROSTATICS

G.A. CISNEROS¹, T.A. DARDEN¹, N. GRESH², J. PILMÉ^{3,4,5},
P. REINHARDT^{4,5}, O. PARISEL^{4,5}, AND J.-P. PIQUEMAL^{4,5}

¹ *Laboratory of Structural Biology, National Institute of Environmental Health Sciences, MD F0-08, 111 TW. Alexander Dr., Research Triangle Park, NC 27709, USA*

² *Laboratoire de Pharmacochimie Moléculaire et Cellulaire, U648 INSERM, UFR Biomédicale, Université René-Descartes, 45, rue des Saints-Pères, 75006 Paris, France*

³ *Université de Lyon, Université Lyon 1, Faculté de pharmacie, F-69373 Lyon, Cedex 08, France*

⁴ *UPMC Univ Paris 06, UMR 7616, Laboratoire de Chimie Théorique, case courrier 137, 4 place Jussieu, F-75005, Paris, France, e-mail: jpp@lct.jussieu.fr*

⁵ *CNRS, UMR 7616, Laboratoire de Chimie Théorique, case courrier 137, 4 place Jussieu, F-75005 Paris, France*

Correspondance to J-P Piquemal, jpp@lct.jussieu.fr

Abstract: We present an overview of the energy functions used in two Anisotropic Polarizable Molecular Mechanics (APMM) procedures namely SIBFA (Sum of Interactions Between Fragments Ab initio computed) and GEM (Gaussian Electrostatic Model). As SIBFA is a second generation APMM scheme based on distributed multipoles, GEM is the first third generation APMM as it uses distributed hermite densities obtained from density fitting. The two approaches are formulated and calibrated on the basis of quantum chemistry. They embody nonclassical effects such as electrostatic penetration, exchange-polarization, and charge transfer. We address here the technical issues of anisotropy, nonadditivity, transferability and computational speedup of methods. In addition, we review the several ab initio intermolecular energy decomposition techniques that can be used to refine polarisable force fields. As we summarize their differences and similarities, we present our own scheme based on Fragment Localized Kohn-Sham orbitals through a Singles-Configuration Interaction (CI) procedure. We also present a chemically intuitive method based on the Electron Localization Function (ELF) which allows to unravel the local electrostatic properties beyond atomic centers: i.e., on bonds, lone pairs and π system, an useful asset to understand bonding in molecules in order to build models

Keywords: Polarizable force fields, Intermolecular interactions, Energy decomposition, Density fitting, Electron localization function, Topological analysis, Localized orbitals, Multipolar moments

6.1. INTRODUCTION

Nowadays, modern molecular modelling techniques propose numerous potential applications, from material sciences to protein structure prediction and drug design. Indeed, classical Molecular dynamics (MD) is now able to provide useful information to experimentalist as simulations are getting closer and closer to relevant biological timescales. Nevertheless, if MD is now able to produce microsecond trajectories, one should ask about the possible improvements of such simulations. At this point, two directions can be taken. The first consists in increasing the speed of MD softwares by coupling improved sampling methodology to massively parallel computers, having as goal to reach the second timescale. However, if this strategy will probably offer some interesting insights about biophysical process, there is no doubt that the question of the accuracy of the used empirical energy functions, the so-called force fields, should be raised. Indeed, current simulations are mainly aimed to compute free energies and despite success, actual data are already sufficient to demonstrate that current molecular mechanics (MM) potentials have serious shortcomings [1]. This can be easily understood when considering that free energy required an accurate evaluation of both enthalpic and entropic contributions. If entropy can be recovered through sampling efforts, enthalpy needs to be approximate from their quantum mechanical expression. In a way, classical MD can be simply seen as an approximate quantum Born-Oppenheimer MD approach treating the atomic nuclei as classical particles subject to interatomic forces. Presently, these latter remain obtained from empirical potentials far to reproduce first principles results. Therefore, MD should not be able to quantitatively describe vast numbers of systems dominated by difficult weak interactions such as H-bonds networks, metalloproteins and metal clusters, highly charges systems etc., where Chemistry and electron correlation/relativity dominate. For these systems, the right tools are required. In this context, Anisotropic Polarizable Molecular Mechanics (APMM) procedures have been developed (see Reference [2] and references therein). These approaches share the common characteristic of including a more evolved representation of the electrostatic contribution to the interaction energy compared to the usual point charge approximation, allowing a close reproduction of the anisotropic features of the *ab initio* Coulomb contribution. As we will see, some of them use distributed multipoles (sometime damped in order to include short-range penetration effects) or electronic Hermite densities for the latest generation. The philosophy of such approaches relies on an extensive use of quantum mechanics defining the so-called: “bottom-up strategy” [2]. First, the electrostatic moments or hermites densities are directly obtained from an *ab initio* calculation of the considered gas phase isolated molecule and stored in a library. Second, all intermolecular components of the force field should faithfully reproduce their *ab initio* counterpart as obtained from energy-decomposition procedures at the

Hartree-Fock, DFT or CCSD levels. Because they can reproduce such quantities, APMM procedures should account for an accurate description of the interactions including polarization cooperative effects and charge transfer. They should also enable the reproduction of local electrostatic properties such as dipole moments and also facilitate hybrid Quantum Mechanical/Molecular Mechanical (QM/MM) embeddings.

In this contribution, we will review some aspects of this strategy. First, we will explore some recent *ab initio* techniques that can be used for the refinement of APMM approaches. Among them, we will discuss the differences between energy decomposition approaches namely: Kitaura-Morokuma [3], Constrained Space Orbital Variations (CSOV) [4–6], Reduced Variational Space (RVS) [7], Ziegler-Baerends [8, 9] and Symmetry Adapted Perturbation Theory [10] procedures. Moreover, we will focus on a newly developed energy decomposition approach using Fragment-localized Kohn-Sham orbitals through a Singles-Configuration Interaction (CI) procedure [11]; and on a general approach to unravel local electrostatic properties, the so-called DEMEP [12] (Distributed Electrostatic Moments based on the Electron localization function Partition). In a second part, we will detail two APMM approaches in development in our labs. The first, called SIBFA (Sum of Interactions Between Fragments *Ab initio* computed) [2, 13] is a second APMM generation based on distributed multipoles. The second, named GEM (Gaussian Electrostatic Model) [2, 14–16] is the first APMM of the third generation based on electron density. Focusing on methodology, we will put in perspective the physical basis underlying the development of such MM energy functions and the possibility for a computation speedup, a key step to perform simulations.

6.2. AB INITIO TECHNIQUES: FROM INTERMOLECULAR INTERACTIONS TO LOCAL ELECTROSTATIC PROPERTIES

Intermolecular Energy decomposition analyses (EDA) are very useful approaches to calibrate force fields. Indeed, an evaluation of the different physical components of the interaction energy, especially of the many-body induction, is a key issue for the development of polarisable models.

6.2.1. Intermolecular Energy Decomposition Schemes: Equivalence Between Terms

However, due to the availability of numerous techniques, it is important to point out here the differences and equivalence between schemes. To summarize, two EDA families can be applied to force field parametrization. The first EDA type of approach is labelled SAPT (Symmetry Adapted Perturbation Theory). It uses non orthogonal orbitals and “recomputes” the total interaction upon perturbation theory. As computations can be performed up to the Coupled-Cluster Singles Doubles (CCSD) level, SAPT can be seen as a reference method. However, due to the cost of the use of non-orthogonal molecular orbitals, pure SAPT approaches remain limited

to small systems, even if Kohn-Sham orbitals based or local SAPT approaches tend to overcome such difficulties. The second family of methods is variational and based on the supermolecule approach ($\Delta E = E_{AB} - E_A - E_B$) following the early Kitaura-Morokuma (KM) and Ziegler schemes. It includes also the Constrained Space Orbital Variations (CSOV) (and the Reduced Variational Space (RVS), essentially similar to CSOV) approach. These methods are limited to the HF or DFT levels. Following a perturbation terminology, all EDA schemes can be partitioned between first, second and higher order terms:

$$\Delta E = E_1 - E_2 - \delta E_{higher-orders} + BSSE \quad (6-1)$$

All schemes furnish two first-order terms. The first is the electrostatic interaction of the frozen monomers denoted E_{es} in the variational approaches and which is strictly equivalent to the E_{pol}^{10} in SAPT. The second is an exchange-repulsion term $E_{exch-rep}$ (denoted E_{exch}^{10} in SAPT). The sum of them is sometime called frozen-core contribution (E_{FC}) like in the variational CSOV scheme of Bagus et al. This E_{FC} is also equivalent to the Heitler-London energy, employing the unperturbed monomers orbitals. At the HF level, despite a different use of operators ($\mathbf{V}$ for SAPT, vs. $\mathbf{H}$ for the variational methods), these terms should be equivalent for all approaches if a reasonable basis set is used. Second-Order terms are more problematic and can be divided into a so-called induction term and a dispersion component, each one of these terms being associated to a repulsive second-order exchange term. At the HF level, the SAPT induction term (E_{pol}^{20}) should be equivalent to the BSSE corrected Orbital relaxation term of the Ziegler scheme also called Orbital Interaction. This latter Orbital interaction term corresponds itself to the sum of polarization (E_{pol}), charge transfer (E_{CT}) and BSSE term in the CSOV or RVS approaches. The subtle question of the evaluation of the sole ab initio polarization energy (without charge transfer), so important for the evaluation of the accuracy of polarizable models, is important as its evaluation requires to conserve the antisymmetry of the wavefunction through relaxation of the monomers. Such computation remains limited to the CSOV and RVS scheme as the Morokuma scheme violates the antisymmetry leading to an overestimation of the polarization (and of the charge transfer term) (see References [14, 17–19] and reference therein). That way the CSOV and RVS E_{pol} (and E_{CT}) term embodies the $E_{pol-exch}$ term through conserved MOs orthogonality. It is also important to note that higher order coupling are not included in CSOV and RVS, that way, such polarization term can be seen as a lower bound for the evaluation of polarization. Strategies to use these schemes have been previously reported.

SAPT methods remain the only approaches allowing the evaluation of dispersion (Figure 6-1).

At this point, it is important to notice that in general, the sum of the contributions do not match exactly ΔE as higher order terms are present. The difference between the sum of contributions and ΔE is denoted δE . Concerning the variational schemes, δE is generally small in the CSOV (or RVS) approach thanks to the antisymmetry conservation and not present in the Ziegler scheme as the E_{OI} term is taking into account a fully relaxed wavefunction. It is not the case for the KM scheme which

$$\begin{aligned}
\Delta E_{\text{Ziegler-Baerends}}^{\text{HF,DFT}} &= E_{\text{es}} + E_{\text{exch-rep}} + E_{\text{OI}} = E_1 + E_2 + \text{BSSE} \\
\Delta E_{\text{CSOV}}^{\text{HF,DFT,MCSCF}} &= E_{\text{FC}} + E_{\text{pol/CSOV (A)}} + E_{\text{pol/CSOV (B)}} + E_{\text{CT/CSOV}_{A \rightarrow B}} + E_{\text{CT/CSOV}_{B \rightarrow A}} + \text{BSSE} + \delta E \\
\text{with : } E_{\text{FC}} &= E_{\text{es}} + E_{\text{exch-rep}} ; E_{\text{pol/CSOV}} = E_{\text{pol}} + E_{\text{Pol-exch}} \\
\Delta E_{\text{SAPT}}^{\text{HF}} &= E_{\text{es}}^{10} + E_{\text{exch-rep}}^{10} + E_{\text{ind,resp}}^{20} + E_{\text{exch-ind,resp}}^{20} + \delta E_{\text{resp}}^{\text{HF}} \\
\Delta E_{\text{SAPT}}^{\text{CCSD}} &= \Delta E_{\text{SAPT}}^{\text{HF}} + \text{correlation-corrections}
\end{aligned}$$

Figure 6-1. Notations for usual energy decomposition schemes

can embody very large δE (sometime denoted E_{mix}) in presence of charge species. For other reasons (especially due to some convergence difficulties of the perturbation series for induction, see Ref. [20–22] for details) the same problem can occur for SAPT. Table 6-1 summarizes these informations.

6.2.2. Beyond Two-Body Interaction: Fragment-Localized Kohn-Sham Orbitals via a Singles-CI Procedure

As discussed below, EDA schemes are generally limited to dimer interactions (up to small trimer for SAPT). If the RVS scheme allows an evaluation of contribution for more than two molecules at the HF level, EDA methods allowing the inclusion of electron correlation did not exist up to a very recent time (see Head-Gordon's Scheme [23]) for the computation of large assemblies of molecules. We present here the methodology at the basis of a new potentially linear scaling local approach based on Fragment-localized Kohn-Sham orbitals via a Singles-CI procedure [11].

6.2.2.1. Method: Fragment-Localized Kohn-Sham Orbitals

In the literature we may find the procedure for creating localized Hartree-Fock orbitals via an energy minimization based on a CI procedure employing mono-excitations (see for instance Reference [24]). The scheme starts from a set of given (guess) orbitals and solves iteratively the Hartree-Fock equations via the steps:

1. Symmetric (Löwdin) orthogonalisation of the orbitals via $\mathbf{S}^{-1/2}$
2. Construction of the Fock matrix
3. Calculation of the total energy
4. Construction and diagonalisation of an approximate Singles-CI matrix
5. Use in first order of the CI coefficients to correct the occupied and virtual molecular orbitals
6. Return to step 1

In step 3, a criterion of convergence may be introduced to terminate the iterations. Two other points should be mentioned: instead of taking the correct Singles-CI matrix, we may resort to a simpler one, omitting single bi-electronic integrals and using only Fock-matrix elements as:

$$\langle \Phi_i^a | \mathbf{H} | \Phi_i^b \rangle \approx F_{ab} \delta_{ij} - F_{ij} \delta_{ab} \quad (6-2)$$

Table 6-1. Contribution to the total interaction energy from different energy decomposition schemes

ΔE	First order (or Frozen Core)			Second Order		Higher orders (δE)		
Contributions	Electrostatics	Exch.-rep.	Correlation corrections	Induction	Exchange-induction	Ind. and exch-Ind. Correlation corrections	Dispersion	
Methods								
SAPT	Yes	Yes	Yes (up to CCSD)	Yes	Yes	Yes Yes (up to CCSD)	Yes (+exchange dispersion; up to CCSD)	Yes (up to third order)
Kitaura-Morokuma	Yes	Yes	No	Yes ($E_{ind}=E_{pol}+E_{ct}$)	No	No	No	Yes (by difference from ΔE)
CSOV	Yes	Yes	Through DFT or MCSCF	Yes ($E_{ind}=E_{pol}+E_{ct}$)	Yes (included in E_{pol} and E_{ct})	Through DFT or MCSCF	No	Yes (by difference from ΔE)
Ziegler/Baerends	Yes	Yes	Through DFT	Yes ($E_{ind}=E_{OI}$)	Yes (included in E_{OI})	Through DFT	No	No (included in E_{OI})

From the obtained wavefunction:

$$\Psi = \Phi_0 + \sum_i^a c_i^a \Phi_i^a \quad (6-3)$$

we use the coefficients for correcting the orbitals as:

$$\varphi_i' = \varphi_i + \sum_a c_i^a \varphi_a (\text{occupied orbitals}) \quad (6-4)$$

$$\varphi_a' = \varphi_a - \sum_i c_i^a \varphi_i (\text{virtual orbitals}) \quad (6-5)$$

Including the correction for the virtual orbitals ensures the orthogonality between occupied and virtual orbitals. Nevertheless, within the two separate orbital spaces, the orbitals must be re-orthogonalized in each iteration.

The advantage of the scheme lies in possibility to cut indices with a distant dependent selection criterion, rendering the method potentially linear scaling. As a consequence, orbitals for periodic structures may be created in this way (see References [25, 26]).

We may ask now, whether the same procedure may be applied to density-functional theory, just by replacing the Fock operator by the corresponding Kohn-Sham operator. To this end we have to look at the minimization of the total energy with respect to the density of a multi-determinantal wavefunction Ψ . We write the density as:

$$\begin{aligned} \Psi &= c_0 \Phi_0 + \sum_i c_i \Phi_i \\ \rho(\vec{r}) &= N \int \dots \int d^3 r_1 \dots d^3 r_{N-1} \left| \Psi^2(\vec{r}_1, \dots, \vec{r}_{N-1}, \vec{r}) \right| \\ &= c_0^2 \rho_{\Phi_0}(\vec{r}) + \sum_i c_i^2 \rho_{\Phi_i}(\vec{r}) + 2 \sum_{I < J} c_I c_J \varphi_k^I(\vec{r}) \varphi_l^J(\vec{r}) \end{aligned} \quad (6-6)$$

Following Reference [27], we may write the variation of the exchange-correlation energy as:

$$\begin{aligned} \int v^{XC}(\vec{r}) \rho_I(\vec{r}) d^3 r &= \langle \Phi_I | V^{XC} | \Phi_I \rangle \\ \int v^{XC}(\vec{r}) \varphi_k^I(\vec{r}) \varphi_j^I(\vec{r}) d^3 r &= \langle \Phi_I | V^{XC} | \Phi_J \rangle \end{aligned} \quad (6-7)$$

$$\frac{\partial \rho(\vec{r})}{\partial c_I} = 2 c_I \rho_{\Phi_I}(\vec{r}) + 2 \sum_{J \neq I} c_J \varphi_k^I(\vec{r}) \varphi_l^J(\vec{r}) \quad (6-8)$$

$$\frac{\delta E^{XC}[\rho]}{\delta c_I} = 2c_I \langle \Phi_I | V^{XC} | \Phi_I \rangle + 2 \sum_{I \neq J} c_J \langle \Phi_I | V^{XC} | \Phi_J \rangle \quad (6-9)$$

As the same construction holds for the Coulomb energy and the mono-electronic part, we obtain equations completely analogous to the system of linear equations for the Singles-CI:

$$\begin{aligned} E &= \langle \Phi_0 | K | \Phi_0 \rangle + \sum_I c_I \langle \Phi_0 | K | \Phi_I \rangle \\ c_I E &= \langle \Phi_0 | K | \Phi_I \rangle + c_I \langle \Phi_I | K | \Phi_I \rangle + \sum_{J \neq 0, I} c_J \langle \Phi_I | K | \Phi_J \rangle \end{aligned} \quad (6-10)$$

which we have to solve for the coefficients c_I at each SCF iteration. Indeed, the only difference to Hartree-Fock theory lies in the use of the Kohn-Sham operator $K = T + Z + J + V^{XC}$ instead of the usual Hamiltonian $H = T + Z + 1/r_{12}$, reduced in the CI matrix to Fock-matrix elements.

6.2.2.2. Usefulness: From Energy Decomposition to Local Properties

Apart from the question of linear scaling methods, we may employ the so-constructed orbitals for studying weakly interacting complexes. Of course, usual functionals do not include the important dispersion terms, but such an approach remains effective to study induction in large assemblies of molecules and, as we will see, for extracting monomer properties and interaction-induced changes of these.

(a) Application to energy decomposition: We first tested the accuracy our approach by implementing a Ziegler-Baerends type scheme by separately computing the electrostatic, exchange-repulsion and Orbital Interaction components of the interaction energy.

As expected we observed the invariance to orbitals localization of our decompositions scheme on a previously investigated linear water dimer configuration: localized and canonical orbitals lead to rigorously the same energy contributions, which is not the case for all decomposition schemes (due to projections or approximative orbital rotations). Concerning force field parametrization, it is interesting to observe the influence of the addition of the second monomer basis functions. It clearly affects the energy components by diminishing the value of electrostatic and increasing the value of exchange-repulsion energy values. This “BSSE-like” (BSSE stands for Basis Set Superposition Error) effect is then clearly pronounced for Frozen Core (or first order in the SAPT terminology) as BSSE clearly acts on the other components. Table 6-2 displays such effect on the canonical water dimer. The decompositions have been easily extended beyond dimer systems [11], allowing the calculation of many-body contributions in contrast to SAPT or CSOV calculations, often restricted to the implementation of 2-body terms [14]. The scheme is also interesting to compute local fragment properties such as dipoles moments. As an example we may look at the dipole moment of two interacting NH_3 molecules. For each molecule we may

Table 6-2. Effect of dimer basis set on the components of the interaction energy

Method	E_s	$E_{\text{exch-rep}}$	E_{FC}	$E_{\text{OI}} + \text{BSSE}$	ΔE
Monomers in the respective monomer basis					
HF(6d)	-8.27	6.91	-1.36	-2.18	-3.55
B3LYP(6d)	-8.03	7.48	-0.55	-3.30	-3.86
Monomers in the dimer basis					
HF(6d)	-8.30	7.02	-1.28	-2.27	-3.55
B3LYP(6d)	-8.18	6.87	-1.31	-3.04	-4.35

Table 6-3. Comparison of the dipoles of the isolated individual monomers (dipole M) compare to the dipole moments of molecules within the dimer (dipole D) via the interaction, calculated with different functionals. Units are atomic units, and we give as well the difference in length and orientation

NH3-NH3	Dipole(M)	Dipole(D)	Difference	angle
HF	0.613	0.664	0.051	8.8
	0.613	0.690	0.077	0.1
BLYP	0.552	0.635	0.0803	15.5
	0.535	0.677	0.142	1.2

calculate a dipole moment separately, and look for the deformation of the monomer orbitals (M) when constructing the dimer orbitals (D) via the described Singles-CI procedure. We have a good trace for the deformation, as the iterations only deform in a minimal sense the starting guess orbitals (Table 6-3).

6.2.3. Distributed Electrostatic Moments Based on the Electron Localization Function Partition

(a) **Theory:** In addition, as a fine understanding of cooperative effects is required in order to test the validity and the transferability of force fields parameters, some of us have been developing methodologies enabling the evaluation of local chemically intuitive distributed electrostatic moments using the topological analysis of the Electron Localization Function (ELF) [12].

For over a decade, the topological analysis of the ELF has been extensively used for the analysis of chemical bonding and chemical reactivity. Indeed, the Lewis' pair concept can be interpreted using the Pauli Exclusion Principle which introduces an effective repulsion between same spin electrons in the wavefunction. Consequently, bonds and lone pairs correspond to area of space where the electron density generated by valence electrons is associated to a weak Pauli repulsion. Such a property was noticed by Becke and Edgecombe [28] who proposed an expression of ELF based on the laplacian of conditional probability of finding one electron of spin σ at r_2 , knowing that another reference same spin electron is present at r_1 . Such a function

was later linked by Savin [29] to a local excess of kinetic energy due to the Pauli repulsion and reformulated by taking the homogenous electron gas as reference. That way, the ELF function (denoted η) can be interpreted as a measure of the Pauli repulsion in the atomic or molecular space and allows an access to the probability of finding two same spin electrons:

$$\eta(\mathbf{r}) = \frac{1}{1 + (\frac{D}{D_0})^2} \quad (6-11)$$

where D is a measure of kinetic energy excess and D_0 is the kinetic energy of a same density homogenous electron gas. ELF is defined to have values restricted between 0 and 1 in order to tend to 1 where parallel spins are highly improbable (there is therefore a high probability of opposite-spin pairs), and to zero in regions where there is a high probability of same-spin pairs. The ELF function can be interpreted as a signature of the electronic-pair distribution but, in contrast to pair functions, it can be more easily calculated and interpreted.

Once computed on a 3D grid from a given ab initio wave function, the ELF function can be partitioned into an intuitive chemical scheme [30]. Indeed, core regions, denoted $C(X)$, can be determined for any atom, as well as valence regions associated to lone pairs, denoted $V(X)$, and to chemical bonds ($V(X,Y)$). These ELF regions, the so-called basins (denoted Ω), match closely the domains of Gillespie's VSEPR (Valence Shell Electron Pair Repulsion) model. Details about the ELF function and its applications can be found in a recent review paper [31].

It has been recently shown [12] that the ELF topological analysis can also be used in the framework of a distributed moments analysis as was done for Atoms in Molecules (AIM) by Popelier and Bader [32, 33]. That way, the $M_0(\Omega)$ monopole term corresponds to the opposite of the population (denoted N):

$$M_0(\Omega) = - \int_{\Omega} \rho(\mathbf{r}) d\tau = -N(\Omega) \quad (6-12)$$

The first moments or dipolar polarization components of the charge distribution are defined by three-dimensional integrals for a given basin Ω according to:

$$\begin{aligned} M_{1,x}(\Omega) &= - \int_{\Omega} (x - X_c) \rho(\mathbf{r}) d\tau \\ M_{1,y}(\Omega) &= - \int_{\Omega} (y - Y_c) \rho(\mathbf{r}) d\tau \\ M_{1,z}(\Omega) &= - \int_{\Omega} (z - Z_c) \rho(\mathbf{r}) d\tau \end{aligned} \quad (6-13)$$

where X_c , Y_c , and Z_c are the Cartesian coordinates of the basin centres.

The five second-moment spherical tensor components can also be calculated and are defined as the quadrupolar polarization terms. They can be seen as the ELF basin equivalents to the atomic quadrupole moments introduced by Popelier [32] in the case of an AIM analysis:

$$\begin{aligned}
 M_{2,zz}(\Omega) &= -\frac{1}{2} \int_{\Omega} (3(z - Z_c)^2 - \mathbf{r}^2) \rho(\mathbf{r}) d\tau \\
 M_{2,x^2-y^2}(\Omega) &= -\frac{\sqrt{3}}{2} \int_{\Omega} [(x - X_c)^2 - (y - Y_c)^2] \rho(\mathbf{r}) d\tau \\
 M_{2,xy}(\Omega) &= -\sqrt{3} \int_{\Omega} (x - X_c)(y - Y_c) \rho(\mathbf{r}) d\tau \\
 M_{2,xz}(\Omega) &= -\sqrt{3} \int_{\Omega} (x - X_c)(z - Z_c) \rho(\mathbf{r}) d\tau \\
 M_{2,yz}(\Omega) &= -\sqrt{3} \int_{\Omega} (y - Y_c)(z - Z_c) \rho(\mathbf{r}) d\tau
 \end{aligned} \tag{6-14}$$

The first- or second-moment basin magnitude is then defined as the square root of the sum of squared components:

$$|\mathbf{M}(\Omega)| = \sqrt{\sum_i M_i(\Omega)^2} \tag{6-15}$$

Thanks to the invariance of the magnitude of any multipole rank ($|\mathbf{M}1|$ or $|\mathbf{M}2|$) with respect to the axis for a given bond or lone pair, the approach allows us to compare the dipolar or quadrupole polarization of a given basin in different chemical environments.

That way, the Distributed Electrostatic Moments based on the ELF Partition (DEMEP) allows computing of local moments located at non-atomic centres such as lone pairs, σ bonds and π systems. Local dipole contributions have been shown to be useful to rationalize inductive polarization effects and typical hydrogen bond interactions. Moreover, bond quadrupole polarization moments being related to a π character enable to discuss bond multiplicities, and to sort families of molecules according to their bond order.

(b) Applications: It is then possible to compute a chemically intuitive distributed analysis of electrostatic moments based on ELF basins. As this partition of the total charge density provides an accurate representation of the molecular moments (dipole, quadrupole etc...), the distributed ELF electrostatic moments allows the computation of local moments located at non-atomic centers such as lone pairs, bonds and π systems. It has been recently shown [12] that local dipole contributions

Figure 6-2. ELF moments for the canonical water dimer (dipole moments in Debye, M_1 in au.)

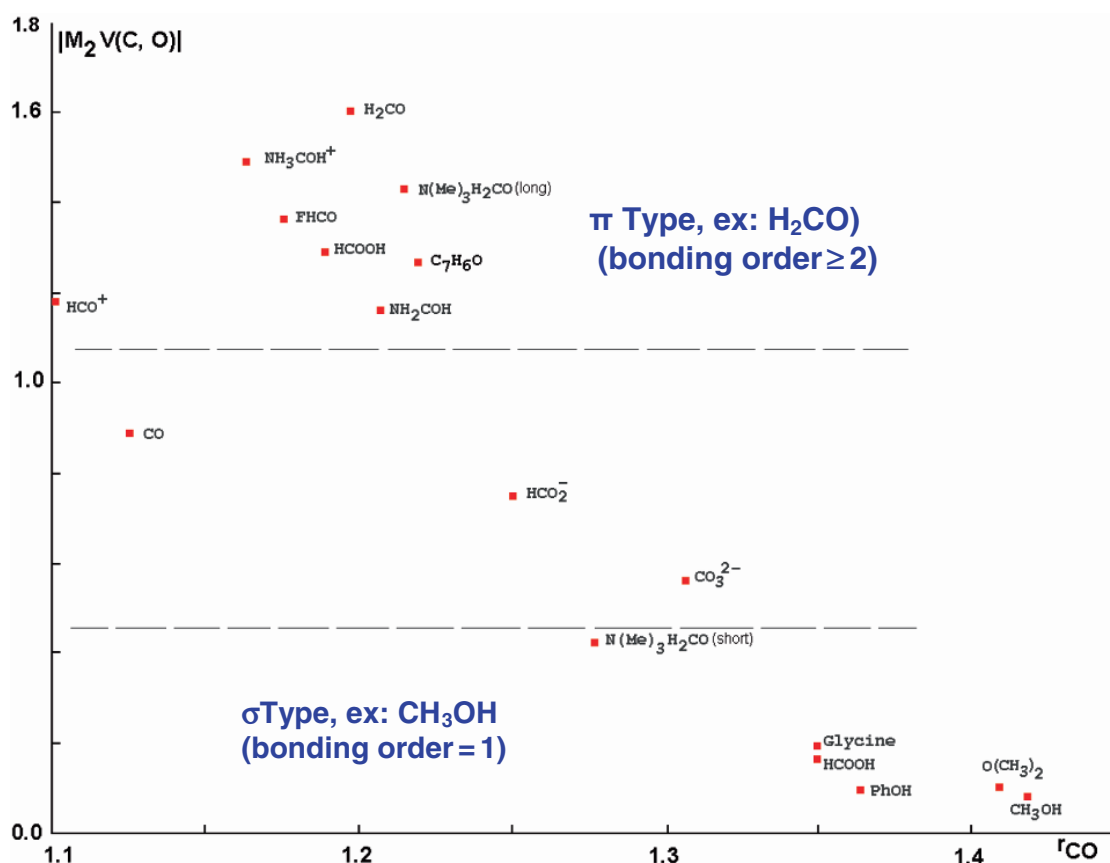
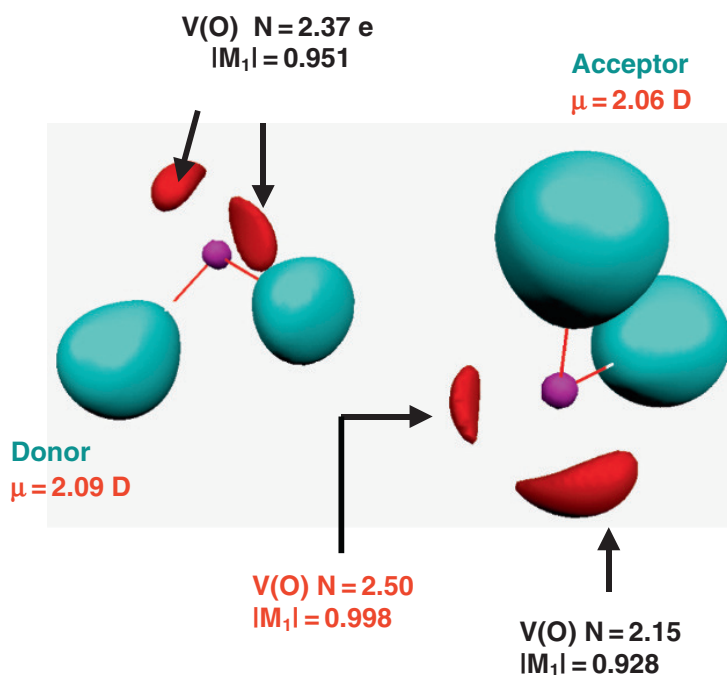


Figure 6-3. Correlation between the quadrupolar polarization ($|M_2|$) of CO bonds in selected molecule and of the bond multiplicity

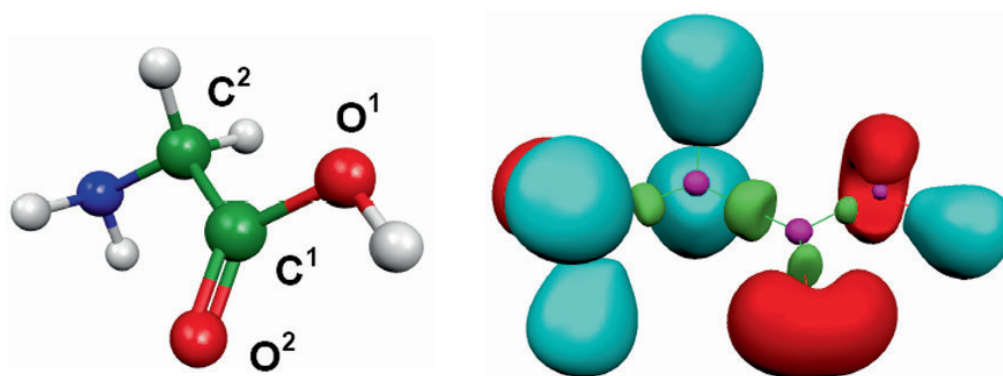


Figure 6-4. Electron localization function domains (concentration of electrons) in glycine. Lone pair domains are displayed in red

can be used to rationalize inductive polarization effects and so should be able to give some new insight towards a better understanding of local density modifications due to the cooperative effects. Figure 6-2 shows the difference of local dipole moment within the canonical water dimer. We can clearly distinguish the acceptor molecule from the donor one. Moreover, the local M1 value of the lone pair involved in the hydrogen bond (i.e. Acceptor molecule) is clearly higher than for the lone pairs of the donor molecule.

Following the same idea, it has been shown that the quadrupole moment of a bond was related to its π character, allowing the discussion on its multiplicity. Then it becomes possible to discuss the influence of the intra- or inter-molecular environment on a given constituent of a molecule. Figure 6-3 displays such influence on a $C=O$ bond through a large set of molecules.

In a recent study of the transferability of moments, it has shown that stable trends are actually observed for the chemical bond features along investigated test peptide chains (Figure 6-4 and Table 6-4).

Such results are interesting for force field development as they clearly establish the existence of conserved “electrostatic blocks” within amino acids, an encouraging step for transferability of force field parameters.

6.3. DEVELOPMENT OF NEXT GENERATION POLARIZABLE FORCE FIELDS: FROM SIBFA TO GEM

6.3.1. Sum of Interaction Between Fragments Ab Initio (SIBFA)

SIBFA [2, 13] is a polarizable molecular mechanics procedure, formulated as a sum of five energy contributions, each of which is destined to reproduce its counterpart from reference EDA ab initio computations. The intermolecular interaction energy is formulated as:

$$\Delta E_{\text{int}} = E_{\text{MTP}^*} + E_{\text{rep}^*} + E_{\text{pol}} + E_{\text{ct}} + E_{\text{disp}}(+E_{\text{LF}}) \quad (6-16)$$

Table 6-4. Local dipole contributions and magnitude of the first and second moments of some basins of typical basins involved in the main chain $C^2H(NH_2)C^1O^1O^2H$ of amino acid

Amino acid	$V(C^1, O^1)$		$V(C^1, O^2)$		$V(C^2, N)$		$V(N)$		$V(O^2)$		Total	
	$ M_1 $	$ M_2 $	$ M_1 $	$ M_2 $	$ M_1 $	$ M_2 $	$ M_1 $	$ M_2 $	$ M_1 $	$ M_2 $	$ \mu $	$ \mu ^{a,b}$
Glycine ^{c,d}	0.255 (0.247)	1.249 (1.221)	0.052 (0.050)	0.257 (0.270)	0.183 (0.207)	0.166 (0.177)	0.952 (1.055)	0.125 (0.100)	11.6 (12.0)	3.308 (2.955)	11.6 (11.5)	0.50 (0.48)
Valine ^c	0.281	1.241	0.059	0.256	0.177	0.173	0.931	0.167	11.5	3.296	2.417	0.46 (0.46)
Tyrosine ^c	0.270	1.210	0.059	0.268	0.177	0.172	0.929	0.161	11,5	3.286	2.454	0.55 (0.55)
											12,0	0.95 (0.92)

^a Total molecular dipole of the amino acid in a.u.

^b The values given in parentheses are obtained from the SCF calculation provided by the GAUSSIAN03 software.

^c Optimized at B3LYP/6-31+G(d,p) level of computation.

^d The values given in parentheses correspond to a single point calculation at the B3LYP/Aug-cc-pVTZ level of computation.

Which denotes respectively the short-range penetration corrected electrostatic multipolar (E_{MTP^*}) energy, short-range repulsion (E_{rep^*}), polarization (E_{pol}), charge-transfer (E_{ct}), and dispersion (E_{disp}) contributions. In presence of an open-shell cation, a ligand field correction is introduced (E_{LF}).

The connectedness of SIBFA to quantum chemistry stems from the use of distributed multipoles and polarizabilities. They are derived from the molecular orbitals of any given molecular fragment using procedures due to Claverie and coworkers [34] concerning the multipoles and by Garmer and Stevens concerning the polarizabilities [35, 36]. They are then stored in the SIBFA library of fragments along with the fragment internal geometry and types of successive atoms and used in subsequent inter- or intramolecular interactions that involve that fragment. SIBFA can be seen as a set of parametric equations aiming to reproduce the required integrals produced by Localized Molecular Orbitals Theory.

We have previously [2] emphasized the features that an MM methodology should have in view of a meaningful reproduction of QC, namely separability, anisotropy, non-additivity and transferability.

Non-additivity and anisotropy of the interaction potential are critical features in molecular recognition and docking. Non-additivity in SIBFA stems from both second-order contributions, E_{pol} and E_{ct} . That of E_{pol} stems from the vector addition of the polarizing fields on a given centre and the use of the square of its norm. Iteratively accounting for the effects of the induced dipoles further enhances non-additivity. That of E_{ct} is conferred by the modulation of the ionization potential of the electron donor on the one hand, and of the electronic affinity of the electron acceptor on the other hand, by the electrostatic potential that each undergoes in a multimolecular complex. Moreover, such potentials embody components due to the induced dipoles, whose amplitudes themselves depend non-additively upon the fields. An additional coupling to nonadditive polarization effects stems from the increase of the effective radius of the electron donor, intervening in the exponential of E_{ct} , by a term proportional to the magnitude of the field undergone by the electron donor.

The anisotropy of E_{MTP^*} stems for the use of distributed multipoles on atoms and on the barycentres of the chemical bonds, thus advancing beyond the assumption of spherical symmetry incurred by the use of atom-centred point-charges. That of E_{pol} stems from: (i) the multipolar nature of the polarizing field; (ii) the use of lone-pair polarizabilities that are off-centred, being located on the barycentres of the Boys localized lone-pair orbitals; (iii) and the use of polarizability tensors instead of scalars. The anisotropies of both short-range contributions, E_{rep} and E_{ct} , which are overlap-dependent terms, is conferred by the use of localized lone-pairs accounting for hybridization. The anisotropy of E_{disp} is, similarly, conferred by the introduction of fictitious atoms on the localized lone pairs. We detail here the methodology used for each one of the components of the SIBFA intermolecular interaction energy. Such equations have been shown to be transferable for intermolecular interactions, see Reference [2] and references therein.

• **Multipolar Electrostatic contribution: penetration corrected EMTP***

In SIBFA, electrostatics is computed upon using distributed multipoles (monopoles, dipoles, quadrupole) located on atoms and bond midpoints as:

$$E_{\text{MTP}} = E_{\text{mono-mono}} + E_{\text{mono-dip}} + E_{\text{mono-quad}} + E_{\text{dip-dip}} + E_{\text{dip-quad}} + E_{\text{quad-quad}} \quad (6-17)$$

If we review the separated components of EMTP, the electrostatic energy appears mainly dominated by the terms involving the charge (Table 6-5).

However, if we analyse the functional form used to compute the charge–charge interaction,

$$E_{\text{mono-mono}} = q_i \cdot q_j / r \quad (6-18)$$

where r is the distance between q_i and q_j , we can easily see that it remains very different from the quantum chemistry formulation [6]:

$$\begin{aligned} E_c = & -2 \sum_i \sum_v Z_v \int (|\varphi_i(1)|^2) / (r_{1v}) d\mathbf{r}_1 - 2 \sum_j \sum_\mu Z_\mu \int (|\varphi_j(2)|^2) / (r_{2\mu}) d\mathbf{r}_2 \\ & + 4 \sum_i \sum_j \int (|\varphi_i(1)|^2 |\varphi_j(2)|^2) / (r_{12}) d\mathbf{r}_1 d\mathbf{r}_2 + \sum_\mu \sum_v Z_\mu \cdot Z_v / r_{\mu v} \end{aligned} \quad (6-19)$$

where μ and φ_i are respectively the nucleus and the unperturbed MOs of monomer A ; and v and φ_j , those of monomer B.

Indeed, the ab initio integrals exhibit an exponential decay at short-range which is not present in any of the EMTP energy terms. This comportment of integrals is at the origin of the so-called penetration energy, an overlap dependant term which is, by definition, absent of the long-range multipolar approximation.

In our approach [18, 37], we have modified the formulation of the terms involving the charges (mono-mono, mono-dip and mono quad term) to screen the electrostatic interaction.

Table 6-5. Contributions to the multipolar electrostatic energy (kcal/mol) for various complexes at their equilibrium geometry

Complexes	mono-mono.	mono-dip.	mono-quad.	dip-dip.	dip-quad.	Quad-quad.
(H ₂ O) ₂ linear	−3.3	−2.6	−1.1	−0.5	−0.1	−0.3
(HCONH ₂) ₂ linear	−7.3	−2.8	1.6	0.1	−0.7	0.6
Cu ²⁺ – H ₂ O	−46.2	−27.9	0.9	0.0	0.0	0.0
HCOO [−] – H ₂ O	−11.3	−4.9	−0.3	−0.1	−0.7	0.6
monodentate						
H ₃ CNH ₃ ⁺ – H ₂ O	−12.5	−7.6	0.1	−0.2	−0.1	0.0

First, we have modified the mono-mono term (now denoted $E_{\text{mono-mono}^*}$) to propose a functional form mimicking the three terms present in ab initio, namely the nucleus–nucleus repulsion, the electron–nucleus attraction and the electron–electron repulsion. For two interacting centers i and j , the modified mono-mono term is:

$$E_{\text{mono-mono}^*} = [Z_i Z_j - \{Z_i (Z_j - q_j)(1 - \exp(-\alpha_i r)) + Z_j (Z_i - q_i)(1 - \exp(-\alpha_j r))\} + (Z_i - q_i)(Z_j - q_j)(1 - \exp(-\beta_i r))(1 - \exp(-\beta_j r))]^* (1/r) \quad (6-20)$$

Where Z_i and Z_j are the valence electrons for the i and j atoms. This number is set to 0 for sites located on bonds. α_i and β_i are parameters depending on effective van der Waals radii (denoted r_{vdw}) and given by:

$$\alpha_i = \gamma / r_{\text{vdw } i} \text{ and } \beta_i = \delta / r_{\text{vdw } i}$$

γ and δ are fixed parameters depending on the reference ab initio level (methodology and chosen basis set). They are transferable to any atom and are evaluated once and for all upon fitting on a set of H_2 or H_2O dimers geometries. For bonds monipoles, the r_{vdw} values are given by the arithmetic mean of the radii forming the bond.

From a physical point of view, this new formulation includes exponential terms that are in agreement with the observed ab initio and experimental results. Moreover, it is easy to verify that the new expression converges to the classical one when r increases. That way, at long range, where the multipolar approximation is valid, the exponential part dies whereas, at short distances, the monopole–monopole interaction embodies a part of the penetration energy. Consequently, $E_{\text{mono-mono}^*}$ has the correct dependence at any range.

The second modification acts on the monopole-dipole term.

$$E_{\text{mono-dip}} = -\mu_j \cdot \xi \quad (6-21)$$

Where ξ , the electric field due to a charge q_i located at a point j is equal to:

$$\xi = q_i \mathbf{r}_{ij} / r_{ij}^3 \quad (6-22)$$

where $\mathbf{r}_{ij}$ is the vector oriented along r from i towards j .

We chose here to only modify ξ to obtain the $E_{\text{mono-dip}^*}$ component:

$$E_{\text{mono-dip}^*} = -\mu_j \cdot \xi^* \quad (6-23)$$

where

$$\xi^* = \{Z_i - (Z_i - q_i)(1 - \exp(-\eta r))\} \cdot \mathbf{r}_{ij} / r_{ij}^3 \quad (6-24)$$

η is given by:

$$\eta = \chi / ((r_{vdw\ i} + r_{vdw\ j})) / 2$$

As for the charge-charge term, χ is a constant depending on the chosen level of reference ab initio computation and converges to the classical form as r increases.

At this level, our formulation includes a penetration correction for terms varying like R^{-1} (monopole-monopole), like R^{-2} (monopole-dipole) but does not include any correction for terms varying like R^{-3} (dipole-dipole and monopole-quadupole). We then added a correction for the monopole-quadupole interaction.

For the evaluation of this term, we used a non usual formalism (see Ref. [20] and references therein), the so-called axial quadupole. Indeed, it is possible to define any quadupole as the sum of 3 axial quadupoles oriented towards the main axis (e_1, e_2, e_3) of a local frame. We then obtain:

$$Q = \sum_{i=1}^3 Q_{ij}(e_i \otimes e_j) \quad (6-25)$$

Thanks to tensors mathematic properties, it is possible to add the same constant to each one of the diagonal terms, which allows the elimination of one of the axial quadupoles.

That way, the mono-quad interaction energy is given by:

$$E_{\text{mono-quad}} = E_{mq1} + E_{mq2} \quad (6-26)$$

where E_{mq1} and E_{mq2} are respectively the monopole-axial quadupole interaction (E_{mq1} and E_{mq2} are different and represent the true quadupole) given by:

$$E_{mq1} = q^*(Q_1/2r^3)[3(\mathbf{a} \cdot \mathbf{r}/r)^2 - 1] \quad (6-27)$$

Where $\mathbf{a}$ is the unit vector defined by the local frame defining the axial quadupole, $\mathbf{r}$ is the vector oriented towards r from the monopole to the axial quadupole. Q_a is the corresponding matrix element. Following the modification of the charge-dipole interaction, we introduced a modified mono-quad interaction, namely $E_{\text{mono-quad}^*}$:

$$E_{\text{mono-quad}^*} = E_{mq1^*} + E_{mq2^*} \quad (6-28)$$

with:

$$E_{mq1^*} = \{Z_i - (Z_i - q_i)(1 - \exp(-\varphi r))\}^* \varphi (Q_a/2r^3)[3(\mathbf{a} \cdot \mathbf{u})^2 - 1] \quad (6-29)$$

φ is given by:

$$\varphi = \Omega / (r_{vdw\ i} + r_{vdw\ j}) / 2$$

Ω is a constant dependant on the chosen level of reference ab initio computation.

To conclude, the penetration corrected E_{MTP^*} interaction energy is then computed as:

$$E_{\text{MTP}^*} = E_{\text{mono-mono}^*} + E_{\text{mono-dip}^*} + E_{\text{mono-quad}^*} + E_{\text{dip-dip}} + E_{\text{dip-quad}} + E_{\text{quad-quad}}$$

• **Short Range exchange-repulsion: E_{rep}^***

To determine an expression of E_{rep}^* , we chose to follow the theoretical results of Murrell [38, 39] who proposed a simplified ab initio perturbation scheme to represent the exchange-repulsion energy based on an overlap expansion of localized Molecular Orbitals (MOs). Studying the interaction between hydrogen atoms, they shown that a KS^2/R relation, S being the overlap between MOs, was not able to accurately reproduce the exchange-repulsion energy. They then proposed an extended $S^2(AR^{-1}+BR^{-2})$ expression that we have used to formulate E_{rep}^* . Following early expression (see References [2, 13] and references therein) based on LMOs approximation derived by Claverie, we have expressed E_{rep}^* [18, 40, 41] as a sum of bond-bond, bond-lone pair and lone pair-lone pair repulsion:

$$\begin{aligned} E_{\text{rep}}^* = C_1 & \left(\sum_{AB} \sum_{CD} \text{rep}^*(AB, CD) + \sum_{AB} \sum_{L\gamma} \text{rep}^*(AB, L\gamma) + \right. \\ & \left. \sum_{L\alpha} \sum_{CD} \text{rep}^*(L\alpha, CD) + \sum_{L\alpha} \sum_{L\gamma} \text{rep}^*(L\alpha, L\gamma) \right) \end{aligned} \quad (6-30)$$

Where each one of the repulsion term includes two components: one varying like $1/R$, the other like $1/R^2$:

$$\begin{aligned} \text{Rep}^*(AB, CD) = N_{\text{occ}}(AB)N_{\text{occ}}(CD)S^2(AB, CD)/R_{AB,CD} + N_{\text{occ}}(AB)N_{\text{occ}} \\ (CD)S^2(AB, CD)/(R_{AB,CD})^2 \end{aligned} \quad (6-31)$$

AB and CD denoted the center positions of the bonds formed by atoms A and B ; and C and D respectively. L_α and L_γ represent the lone pair positions. As we will see, this formulation takes into account bonds and lone pairs hybridation, each one of the term depending of an overlap functional. $N_{\text{occ}}(AB)$ and $N_{\text{occ}}(CD)$ are the electron occupation numbers of the AB and BC bonds. Therefore, N_{occ} is equal to 2 for usual bonds and lone pairs. $R_{AB,CD}$ is the distance between the barycenters of the AB and CD bonds.

The S overlap expression relies on the general situation where 4 atoms form 2 bonds having their valence electrons involved in sp^n hybrid Mos. In the context of Slater orbitals, c_s et c_p are the hybridation coefficients and, for example, S is then formulated for the bond-bond term as:

$$\begin{aligned} S(AB, CD) = (c_s I c_{sk} \langle 2s_l 2s_k \rangle + c_p I c_{sk} \langle 2p_{\sigma l} 2s_k \rangle \cos(\mathbf{IJ}, \mathbf{IK}) \\ + c_p K c_{sl} \langle 2p_{\sigma K} 2s_l \rangle \cos(\mathbf{KL}, \mathbf{KI}) + c_p I c_{pK} \langle 2p_{\sigma l} 2p_{\sigma K} \rangle \cos(\mathbf{IJ}, \mathbf{IK}) \cos(\mathbf{KL}, \mathbf{KI})) \end{aligned} \quad (6-32)$$

To simplify the problem, we introduce the following approximation:

$$\langle 2p_{\sigma A} 2s_C \rangle = m_{AC} \langle 2s_A 2s_C \rangle \quad (6-33)$$

m_{AC} is a parameter obtained from computations of overlap integrals between atoms A and B obtained from Mulliken [42] and Roothaan [43] approximations using Slater orbitals. Their values are tabulated and depend on a given atoms couple.

$\langle 2s_A 2s_C \rangle$ can be approximate by an exponential depending on the distance separating atoms A and C and modulated on effective van der Waals radii:

$$\langle 2s_A 2s_C \rangle = M_{AC} \exp(-\alpha \rho_{AC}) \quad (6-34)$$

with

$$\rho_{AC} = R_{AC}/4\sqrt{W_A W_C}$$

and

$$M_{AC} = \sqrt{K_{AC}(1 - Q_A/N_{VAL}^A)(1 - Q_C/N_{VAL}^C)}$$

Q_A and Q_C are the charges obtained from the multipolar expansion of the interacting A and C molecular charge distributions, N_{VAL}^A and N_{VAL}^C being their respective number of valence electrons. W_A and W_C are the A and C atoms effective van der Waals radii. K_{AC} is a proportionality factor tabulated upon the atomic numbers of the A and C atoms. α is a constant fixed to 12.35. The same treatment is applied to the others terms of the repulsion energy.

It is important to point out that recent results on density based overlap integrals [16] confirm the interest of the formulation of E_{rep}^* as a sum of bond-bond, bond-lone pair and lone pair-lone pair repulsion: indeed, core electrons do not contribute to the value of the overlap integrals.

• Polarization contribution

E_{pol} also relies on a local picture as it uses polarizabilities distributed at the Boys LMOs centroids [44] on bonds and lone pairs using a method due to Garmer et al. [35]. In this framework, polarizabilities are distributed within a molecular fragment and therefore, the induced dipoles do not need to interact together (like in the Applequist model) within a molecule as their value is only influenced by the electric fields from the others interacting molecules.

The general expression of the polarization energy at center I located at the centroid of an LMO of a A molecule is:

$$\begin{aligned}
E_{pol}(i) &= -0.5 \sum_j \Delta\mu(i) E_0(j) \\
\Delta\mu(i) &= \alpha(i) \sum_j^{xyz} E(\Delta\mu(i)) + E_0(j)
\end{aligned}
\tag{6-35}$$

E_0 and $E(\Delta\mu(i))$ are respectively the electric fields generated by the permanent and induced multipoles moments. $\alpha(i)$ represents the polarisability tensor and $\Delta\mu(i)$ is the induced dipole at a center i . This computation is performed iteratively, as E_{pol} generally converges in 5–6 iterations. It is important to note that in order to avoid problems at the short-range, the so-called polarization catastrophe, it is necessary to reduce the polarization energy when two centers are at close contact distance. In SIBFA, the electric fields equations are “dressed” by a Gaussian function reducing their value to avoid such problems.

The initial electric field generated by a centre i of molecule A on a center j of molecule B is denoted $E_{i \rightarrow j}^{init}$ is modified by a Gaussian function denoted S [45] as:

$$\begin{aligned}
E_{i \rightarrow j}^{final} &= (1 - S(i, j)) E_{i \rightarrow j}^{init} \\
S(i, j) &= \Theta_i \exp(-\Gamma(R_{ij}^2)/(r_{vdw}(i) + r_{vdw}(j)))
\end{aligned}
\tag{6-36}$$

R_{ij} is the distance between centers i and j . Θ_i is the monopole associated to center i . E and Γ are empirical parameters associated to each atom types as r_{vdw} are the atom effective radii.

It is important to point that parametrization procedure of the short-range damping is really important. In SIBFA, in order to embody short-range penetration and exchange-polarization effects, the fit is performed upon CSOV (or RVS) polarization energy which embodies exchange effects (see Section 6.1). To do so, the SIBFA polarization energy “prior iterating” is adjusted to the CSOV value which corresponds to the relaxation of a molecule A in the field of a frozen B molecule (before the computation of higher-orders induction terms δE). Details can be found in Refs. [18, 19].

• Charge transfer contribution

As for E_{rep}^* , E_{ct} is derived from an early simplified perturbation theory due to Murrell [46]. Its formulation [47, 48] also takes into account the L_α lone pairs of the electron donor molecule (denoted molecule A). Indeed, they are the most exposed in this case of interaction (see Section 6.2.3) and have, with the π orbital, the lowest ionization potentials. The acceptor molecule is represented by bond involving an hydrogen (denoted BH) mimicking the set, denoted ϕ_{*BH} , of virtual bond orbitals involved in the interaction.

E_{ct} is expressed as:

$$E_{tc} = -2C \sum_{L\alpha} N_{occ}(\alpha) (T_{\alpha\beta^*})^2 / \Delta E_{\alpha\beta^*}
\tag{6-37}$$

C is a constant which has been calibrated in order to reproduce E_{ct} at the equilibrium geometry of the water dimer. This value is transferable for all non metallic acceptors. $N_{occ}(\alpha)$ is the occupation number L_α lone pair.

$T_{\alpha\beta^*}$ is a function of:

- i) The transition density overlap of the L_α donor lone pairs and the bond BH centroid, expressed with the same approximations as E_{rep^*} .
- ii) The electrostatic potential applied on A by all the other molecules ($\sum_C V_{C \rightarrow A}$). It

is important to point out that the fields due to the polarization converged induced dipoles are taken into account in order to introduce an explicit link between polarisation and charge transfer.

$\Delta E_{\alpha\beta^*}$ is the energy is the energy required to allow an electron transfer from an orbital α of molecule A towards a virtual β^* orbital on molecule B. It can be expressed as:

$$\Delta E_{\alpha\beta^*} = (I_{L_\alpha} + \sum_C V_{C \rightarrow A}) - (A_{\beta^*} + \sum_C V_{C \rightarrow B}) \quad (6-38)$$

I_{L_α} is the ionization potential of the L_α lone pair as A_{β^*} is the electronic affinity of the electron acceptor. Here also, the introduction of the final iterated field of induce dipole allow to take into account the many-body properties of E_{ct} .

Here also, E_{ct} has its ab initio counterpart within the CSOV framework. The sum of E_{pol} and E_{ct} matches the E_{OI} contribution at the HF and DFT level and the SAPT induction when δE remains small (see Section 6.1).

• Dispersion contribution

E_{disp} [49, 50] is coupled to an exchange-dispersion term and is computed as an expansion of C_n terms: C_6/Z^6 , C_8/Z^8 et C_{10}/Z^{10} , Z being expressed as:

$$Z = r_{ij} / \sqrt{vdw_A \cdot vdw_B}$$

r_{ij} is the distance between atoms i and j; vdw_A and vdw_B are the effective radii of the involved atoms. The C_6 , C_8 and C_{10} coefficients are empirical parameters adjusted on H_2 dimers SAPT computations. Each one of the C_n terms are damped at short-range by the following function:

$$E_{damp}(n) = (1/R^n) L_{ij} \exp(-a_{damp}(n) D(i, j)) \quad (6-39)$$

where:

$$D(i, j) = ((vdw_A + vdw_B) b_{damp} / R_{ij}) - 1$$

L_{ij} , a_{damp} and b_{damp} are parameters; n can be 6, 8 ou 10.

This dispersion energy is coupled to an exchange-dispersion component:

$$E_{exch-disp} = L_{ij} (1 - Q_i / N_{val}(i)) (1 - Q_j / N_{val}(j)) C_{exch} \exp(-\beta_{ecxc} Z) \quad (6-40)$$

Q_i and Q_j are the net charges of atoms i and j ; $N_{\text{val}}(i)$ and $N_{\text{val}}(j)$ their number of valence electrons. C_{exch} and β_{exch} are empirical parameters. Some additional refinements exist within SIBFA as explicit addition of lone pairs for the exchange term [50].

• Ligand field contribution

To correct the energy function from ligand field effects (presently in the case of open-shell cations), SIBFA uses the formalism of the Angular Overlap Model (AOM) [51–53]. The AOM [52] is based on the fact that the relative changes of **d** orbitals energies caused by ligand field effects can be associated to the overlap of these orbitals with the ligands orbitals. As intermolecular overlap integral can be factorized into radial and angular parts, it is then possible to consider the radial part as a constant for a given intermolecular distance. That way, it can be introduced as a parameter, specific of a given metal/ligand couple, whereas the angular part can be exactly computed as it depends only on the relative orientation of the metal **d** orbitals and of those of the ligands. More precisely, the AOM treatment can be seen as an effective Hamiltonian built on the basis of **d** orbitals. Its evaluation uses spherical coordinates (θ , ϕ) and requires the diagonalization of an energy matrix. For each computation, each ligand is considered separately as the total matrix reflects the sum of the local perturbation of the **d** orbitals due to ligands as each matrix element is the sum of the contribution of each ligands. The construction of the energy matrix uses angular coefficient denoted D_i (see Table 6-6) which give the values of overlap of a ligand involved in a σ interaction with the metal.

In order to be able to evaluate the radial part in all point of space and to adapt the AOM to the SIBFA intermolecular potential, we have introduced an exponential dependence of the radial overlap following a procedure introduced by Woodley et al. [54]:

$$e_{\lambda} = a + b \cdot \exp(-\alpha \cdot r) \quad (6-41)$$

r is the metal-ligand distance; a , b and α are parameters, specific of a given metal-ligand couple.

It is then possible to construct the energy matrix [51, 53]:

Table 6-6. angular coefficients for the different **d** orbitals involved **d** in a σ interaction with ligands

i	$D_i(\theta_i, \phi_i)$
z^2	$1/2 (3 \cos^2 \theta - 1)$
yz	$1/2 (\sqrt{3} \sin(2\theta) \sin \phi)$
xz	$1/2 (\sqrt{3} \sin(2\theta) \sin \phi)$
xy	$1/4 (\sqrt{3} (1 - \cos 2\theta) \sin 2\phi)$
$x^2 - y^2$	$1/4 (\sqrt{3} (1 - \cos 2\theta) \cos 2\phi)$

$$H_{dd'} = \sum_l e_{\lambda}^l \langle d|l \rangle \langle l|d' \rangle \quad (6-42)$$

Its eigenvalues (ϵ_i) correspond to the relative energies of the **d** orbitals of the metal. These energies are used to compute the ligand field energy contribution denoted E_{LF} .

For a d^n system:

$$E_{LF} = -2 \sum_{i=1}^5 \epsilon_i + \sum_{i=1}^n \rho_i \epsilon_i \quad (6-43)$$

Where ρ_i are the d orbital occupation numbers (0, 1 or 2).

6.3.2. The Gaussian Electrostatic Model (GEM)

As the SIBFA approach relies on the use of distributed multipoles and on approximation derived from localized MOs, it is possible to generalize the philosophy to a direct use of electron density. That way, the Gaussian electrostatic model (GEM) [2, 14–16] relies on ab initio-derived fragment electron densities to compute the components of the total interaction energy. It offers the possibility of a continuous electrostatic model going from distributed multipoles to densities and allows a direct inclusion of short-range quantum effects such as overlap and penetration effects in the molecular mechanics energies.

6.3.2.1. From Density Matrices to GEM

This method relies on the use of an auxiliary gaussian basis set to fit the molecular electron density obtained from an ab initio one-electron density matrix:

$$\tilde{\rho} = \sum_{k=1}^N x_k k(r) \approx \rho = \sum_{\mu\nu} P_{\mu\nu} \phi_{\mu}(r) \phi_{\nu}^*(r) \quad (6-44)$$

Do so, we use the formalism of the variational density fitting method [55, 56] where the Coulomb self-interaction energy of the error is minimized:

$$E_2 = \frac{1}{2} \iint \frac{[\rho(r_1) - \tilde{\rho}(r_1)][\rho(r_2) - \tilde{\rho}(r_2)]}{|r_1 - r_2|} dr_1 dr_2 = \langle \rho - \tilde{\rho} | \rho - \tilde{\rho} \rangle \quad (6-45)$$

inserting the right hand of Eq. (6-44) into Eq. (6-45), we obtained:

$$E_2 = \frac{1}{2} \sum_{\mu,\nu} \sum_{\sigma,\tau} P_{\mu\nu} P_{\sigma\tau} \langle \mu\nu | \sigma\tau \rangle - \sum_l x_l \sum_{\mu,\nu} P_{\mu\nu} \langle \mu\nu | l \rangle + \frac{1}{2} \sum_k \sum_l x_k x_l \langle k | l \rangle \quad (6-46)$$

E_2 from Eq. (6-3) can be minimized with respect to the expansion coefficients x_l and a linear system of equations can be obtained:

$$\frac{\partial E_2}{\partial x_l} = - \sum_{\mu\nu} P_{\mu\nu} \langle \mu\nu \| l \rangle + \sum_k x_k \langle k \| l \rangle \quad (6-47)$$

Equation (6-5) is used to determine the coefficients:

$$\mathbf{x} = \mathbf{A}^{-1} \mathbf{b} \quad (6-48)$$

Where

$$b_l = \sum_{\mu\nu} P_{\mu\nu} \langle \mu\nu \| l \rangle \text{ and } A_{kl} = \langle k \| l \rangle$$

In a standard density fitting, the determination of the coefficients requires the use of a modified singular value decomposition (SVD) procedure in which the inverse of an eigenvalue is set to zero if it is below a certain cutoff. A cutoff value of 10^{-8} has been previously determined [14] to be acceptable for the molecules which will be under study. In addition to the SVD approach, we have also implemented noise reduction techniques for the fitting procedure as this method can produce numerical instabilities (noise) when the number of basis functions starts to grow and when higher angular momentum are used (these instabilities are also present when using only *s*-type spherical [14] functions albeit to a lower extent). Several strategies have been implemented [15]. Among them, we used the Tikhonov regularization formalism and a damped Coulomb operator $\hat{O} = \text{erfc}(\beta r)/r$ procedure in order to localize the integrals to increase the calculation speed. Alternatively to the DF procedure, it is possible to perform such a fit using density and electrostatic grids [16]. That way, the ab initio calculated properties (density, electrostatic potential, and/or electric field) are fitted via a linear or nonlinear-least-squares procedure to the auxiliary basis sets (ABS). Neglecting the core contributions allows to perform more robust fit of the coefficients compared to the numerical grids and allows to reduce the number of functions and so the computational cost.

Using fitted densities expressed in a linear combination of Gaussian functions has the advantage that the choice of Gaussian functions auxiliary basis set is not restricted to Cartesian Gaussians. To use higher order angular momenta, normalized *Hermite Gaussian functions* can be preferred [2, 15, 16]. Indeed, the use of Hermite Gaussians in integral evaluation improves efficiency by the use of the McMurchie-Davidson (McD) recursion [57] since the expensive Cartesian-Hermite transformation is avoided. Obtaining the Hermite expansion coefficients from the fitted Cartesian coefficients is straightforward since Hermite polynomials form a basis for the linear space of polynomials. Moreover, Hermite Gaussians have a simple relation to elements of the Cartesian multipole tensor and can be used to multipoles distributed at the expansion sites. This smooth connection leads to a continuous

electrostatic model that can be used directly into second generation APMM such as SIBFA. It is important to note that unlike conventional multipole expansions, the spherical multipole expansion obtained from Hermite Gaussians has an intrinsic finite order, namely, the highest angular momentum in the ABS. This connection between multipoles and Hermite densities has its importance as, unlike s-type functions ($l = 0$), fitting coefficients with $l > 0$ (sp, spd . . .) are not invariant by rotation. These coefficients must be transformed for each molecular fragment orientation in order to compute interaction energies. Such a transformation can be achieved [15] by defining both a *global* orthogonal coordinate system frame and a *local* orthogonal coordinate frame for each fragment fitting site.

6.3.2.2. Computing Integrals for Molecular Mechanics

The GEM force field follows exactly the SIBFA energy scheme. However, once computed, the auxiliary coefficients can be directly used to compute integrals. That way, the evaluation of the electrostatic interaction can virtually be exact for an perfect fit of the density as the three terms of the coulomb energy, namely the nucleus–nucleus repulsion, electron–nucleus attraction and electron–electron repulsion, through the use of $\tilde{\rho}$ [2, 14–16, 58].

$$E_{\text{coulomb}} = \frac{Z_A Z_B}{r_{AB}} - \int \frac{Z_A \tilde{\rho}^B(r_B)}{r_{AB}} dr - \int \frac{Z_B \tilde{\rho}^A(r_A)}{r_{AB}} dr + \int \frac{\tilde{\rho}^A(r_A) \tilde{\rho}^B(r_B)}{r_{AB}} dr \quad (6-49)$$

To complete the first order terms, the exchange–repulsion energy can be evaluated through an overlap model [14, 59] as:

$$E_{\text{exch/rep}} \approx K S_{\rho} \quad (6-50)$$

Where:

$$S_{\rho} = \int \rho a(r) \rho b(r) dr \approx \int \tilde{\rho} a(r) \tilde{\rho} b(r) dr$$

As electric fields and potential of molecules can be generated upon distributed $\tilde{\rho}$, the second order energies schemes of the SIBFA approach can be directly fueled by the density fitted coefficients. To conclude, an important asset of the GEM approach is the possibility of generating a general framework to perform Periodic Boundary Conditions (PBC) simulations. Indeed, such process can be used for second generation APMM such as SIBFA since PBC methodology has been shown to be a key issue in polarizable molecular dynamics with the efficient PBC implementation [60] of the multipole based AMOEBA force field [61].

6.3.2.3. Using Periodic Boundary Conditions to Increase Computational Efficiency

In this section we describe the methods to extend Ewald sum methodologies to accelerate the calculation of the intermolecular interactions using PBC. For simplicity, we begin with a generalization of Ewald sums to interacting spherical Hermite Gaussians (e.g. GEM-0 [14]). This is followed by the extension to arbitrary angular momentum. Finally, we describe the implementation of methods to speed up both the direct and reciprocal terms in the Ewald sum [62].

N.1 Spherical charge densities

To begin let U denote a unit cell such that it contains the set of points $\mathbf{r}$ with associated fractional coordinates s_1, s_2, s_3 satisfying $-1/2 \leq s_i \leq +1/2, i = 1, 2, 3$. Then the idealized infinite crystal C is generated by the union of all periodic translations $\mathbf{U}_\mathbf{n}$ of U , using the set of general lattice vectors $\mathbf{n}$. For the Ewald sum to be convergent, extra conditions need to be imposed. To that end, consider a large but finite crystal, i.e., let P denote a closed, bounded region in space, centered at the origin (e.g., sphere, cube, etc.). For a positive integer K , let $\Omega(P, K)$ denote the set of lattice vectors such that $|\mathbf{n}|/K$ is in P . The Coulomb interaction of a spherical Gaussian charge distribution ρ_1 in a unit cell $\mathbf{U}_0$ centered at point $\mathbf{R}_1 \in \mathbf{U}_0$ with an exponent α_1 , i.e. $\rho_1 = (\alpha_1/\pi)^{3/2} \exp(-\alpha_1(\mathbf{r} - \mathbf{R}_1))$, interacting with a second normalized Gaussian charge distribution ρ_2 centered at point $\mathbf{R}_2 \in \mathbf{U}_0$ with exponent α_2 together with all images in $\mathbf{U}_\mathbf{n}$ for $\mathbf{n} \in \Omega(P, K)$ centered at $\mathbf{R} + \mathbf{n}$ can be shown to be:

$$\begin{aligned}
 E_{12} = & \sum_{\mathbf{n} \in \Omega(P, K)} \frac{\text{erfc}(\mu_{12\alpha_0} |\mathbf{R}_{12} - \mathbf{n}|) - \text{erfc}(\mu_{12} |\mathbf{R}_{12} - \mathbf{n}|)}{|\mathbf{R}_{12} - \mathbf{n}|} \\
 & + \frac{1}{\pi V} \sum_{\mathbf{m} \neq 0} \frac{\exp(-\pi^2 \mathbf{m}^2 / \mu_{12\alpha_0})}{\mathbf{m}^2} \exp(-2\pi i \cdot (\mathbf{R}_{12})) \\
 & - \frac{\pi}{\mu_{12\alpha_0}} + \frac{1}{\pi} H_{P, K}(\mathbf{R}_{12}) + \varepsilon(K),
 \end{aligned} \tag{6-51}$$

where the first term corresponds to the direct part of the Ewald sum, the second to the reciprocal part, $H_{P, K}(\mathbf{R}_{12})$ is the surface term which depends on the dipole of the unit cell ($\mathbf{D}$), $\varepsilon(K)$ denotes a quantity that converges to 0 as $K \rightarrow \infty$, $\mathbf{m}$ denotes the reciprocal lattice vectors, $1/\mu_{12\theta} = 1/\alpha_1 + 1/\alpha_2 + 1/\alpha_0$, $1/\mu_{12} = 1/\alpha_1 + 1/\alpha_2$ and α_0 is the Ewald exponent [62].

Equation (6-51) can be generalized to calculate the energy $E_{P, K}$ of U interacting with the entire crystal P . Let $\rho_1 \dots \rho_N$ be normalized spherical Gaussian charge distributions (e.g. GEM-0) centered at $\{\mathbf{R}_1 \dots \mathbf{R}_N\} = \mathbf{R}^{\{N\}} \in U$, and let $q_1 + \dots + q_N = 0$ (neutral unit cell). Then the energy of the central unit cell $\mathbf{U}_0$ within a large spherical crystal, due to the interactions of the Gaussian charge distributions $q_i \rho_i$ with each other and all periodic images within the crystal is given by

$$\begin{aligned}
E_{S,K}(\mathbf{R}^{\{N\}}) = & \frac{1}{2} \sum_{\mathbf{n}}' \sum_{i,j=1}^N q_i q_j \left\{ \frac{\text{erfc}(\mu_{ij\alpha_0}^{1/2} |\mathbf{R}_{ij} - \mathbf{n}|) - \text{erfc}(\mu_{ij}^{1/2} |\mathbf{R}_{ij} - \mathbf{n}|)}{|\mathbf{R}_{ij} - \mathbf{n}|} \right\} \\
& + \frac{1}{2\pi V} \sum_{\mathbf{m} \neq 0} \sum_{i,j=1}^N q_i q_j \frac{\exp(-\pi^2 \mathbf{m}^2 / \mu_{ij\alpha_0})}{\mathbf{m}^2} \exp(-2\pi i \mathbf{m} \cdot (\mathbf{R}_{ij})) \\
& - \frac{\pi}{2V} \sum_{i,j=1}^N \frac{q_i q_j}{\mu_{ij\alpha_0}} - \sum_{i=1}^N q_i^2 \left(\frac{\mu_{ii\alpha_0}}{\pi} \right)^{1/2} + \frac{2\pi \mathbf{D}^2}{3V} + \varepsilon(K),
\end{aligned} \tag{6-52}$$

where $\mathbf{R}_{ij} = \mathbf{R}_i - \mathbf{R}_j$ and $\mathbf{D} = q_1 \mathbf{R}_1 + \dots + q_N \mathbf{R}_N$ is the unit cell dipole.

In order to be able to calculate the reciprocal contribution in Eq. (6-52) it is necessary to grid the Gaussian densities. However, for large exponents (compact Gaussians) this can become intractable. To overcome this problem, the first GEM implementation relied on a method inspired by Fusti-Molnar and Pulay [63]. In this method, the individual Gaussian charge densities are classified into compact and diffuse Hermite Gaussian functions for a given α_0 . Thus, all Hermites with an exponent $\alpha_i \geq \alpha_0$ are considered compact, and the rest are considered diffuse. In this way, the interaction energy expressions may be re-expressed in order to calculate the contributions involving diffuse Hermites completely in reciprocal space [64].

This method was subsequently improved by noting that the α_0 in $\mu_{ij\alpha_0}$ can be different for each pair ij [15]. In this way, the Hermite charge distributions $q_i \rho_i$ are separated into compact and diffuse sets based on their exponents α_i . Subsequently, α_0 is chosen to be infinite for ij pairs where at least one of the two Gaussians is diffuse. This ensures that all pairs involving diffuse Hermites are evaluated in reciprocal space. For all compact ij pairs, α_0 is chosen so that $\mu_{ij\alpha_0}$ is constant, that is, given $\beta > 0$, a Gaussian distribution $q_i \rho_i$ is classified as compact if $\alpha_i \geq 2\beta$ (C set) and diffuse otherwise (D set). Then, for $i, j \in C$, choose α_0 so that $1/\mu_{ij\alpha_0} = 1/\alpha_i + 1/\alpha_j + 1/\theta = 1/\beta$. Otherwise, α_0 is set to infinity. From this, the Coulomb energy of the spherical unit cell can be re-expressed as:

$$\begin{aligned}
E_{S,K}(\mathbf{R}^N) = & \frac{1}{2} \sum_{\mathbf{n}}' \sum_{(i,j) \in C \times C} q_i q_j \left\{ \frac{\text{erfc}(\beta^{1/2} |\mathbf{R}_{ij} - \mathbf{n}|) - \text{erfc}(\mu_{ij}^{1/2} |\mathbf{R}_{ij} - \mathbf{n}|)}{|\mathbf{R}_{ij} - \mathbf{n}|} \right\} \\
& + \frac{1}{2\pi V} \sum_{\mathbf{m} \neq 0} \sum_{(i,j) \in C \times C} q_i q_j \frac{\exp(-\pi^2 \mathbf{m}^2 / \beta)}{\mathbf{m}^2} \exp(-2\pi i \mathbf{m} \cdot (\mathbf{R}_{ij})) \\
& + \frac{1}{2\pi V} \sum_{\mathbf{m} \neq 0} \sum_{(i,j) \notin C \times C} q_i q_j \frac{\exp(-\pi^2 \mathbf{m}^2 / \mu_{ij})}{\mathbf{m}^2} \exp(-2\pi i \mathbf{m} \cdot (\mathbf{R}_{ij})) \\
& - \frac{\pi}{2V} \sum_{(i,j) \notin C \times C} q_i q_j \left(\frac{1}{\beta} + \frac{1}{\alpha_i} + \frac{1}{\alpha_j} \right)
\end{aligned}$$

$$\begin{aligned}
& - \sum_{i \in C}^N q_i^2 \left(\frac{\beta}{\pi} \right)^{1/2} - \sum_{i \notin C}^N q_i^2 \left(\frac{\alpha_i}{\pi} \right)^{1/2} \\
& + \frac{2\pi \mathbf{D}^2}{3V} + \varepsilon(K),
\end{aligned} \tag{6-53}$$

N.2 Higher angular momentum charge densities

In the case of GEM, the auxiliary bases employed for the fitting of the molecular fragment include higher angular momentum Gaussians. In this case, Eq. (6-53) can be extended to account for the higher order Gaussians. As explained above, the fitted densities are expanded in a linear combination of Hermite Gaussians $\Lambda_{tuv}(\mathbf{r}, \alpha, \mathbf{R})$. Here, the Gaussian charge distribution is given by $\rho_i(\mathbf{r}, \mathbf{R}_i, \alpha) = \sum_{l=1}^L \sum_{tuv} \mathbf{c}_{i,l,tuv} \Lambda_{tuv}(\mathbf{r}, \alpha_l, \mathbf{R}_i)$, where $\mathbf{c}_{i,l,tuv}$ are the Hermite coefficients, and L denotes the different exponents in the ABS on center i . Based on this, the Coulomb energy of the total density within the spherical crystal is given by

$$\begin{aligned}
E_{S,K}(\rho^{\{N\}}) &= \frac{1}{2} \sum_{\mathbf{n}}' \sum_{i=1}^N \sum_{l_i \in C} \sum_{t_i u_i v_i} \mathbf{c}_{i,l_i,t_i u_i v_i} \sum_{j=1}^N \sum_{l_j \in C} \sum_{t_j u_j v_j} (-1)^{(t_j+u_j+v_j)} \mathbf{c}_{j,l_j,t_j u_j v_j} \\
&\times \left(\frac{\partial}{\partial \mathbf{R}_{ijx}} \right)^{t_i+t_j} \left(\frac{\partial}{\partial \mathbf{R}_{ijy}} \right)^{u_i+u_j} \left(\frac{\partial}{\partial \mathbf{R}_{ijz}} \right)^{v_i+v_j} \\
&\times \left\{ \frac{\text{erfc}(\beta^{1/2} |\mathbf{R}_{ij} - \mathbf{n}|) - \text{erfc}(\mu_{l_i l_j}^{1/2} |\mathbf{R}_{ij} - \mathbf{n}|)}{|\mathbf{R}_{ij} - \mathbf{n}|} \right\} \\
&+ \frac{1}{2\pi V} \sum_{\mathbf{m} \neq 0} \frac{1}{\mathbf{m}^2} \exp(-\pi^2 \mathbf{m}^2 / 2\beta) \sum_{l_1 \in C}^N S_{l_1}(\mathbf{m}) \\
&\times \exp(-\pi^2 \mathbf{m}^2 / 2\beta) \sum_{l_2 \in C}^N S_{l_2}(-\mathbf{m}) \\
&+ \frac{1}{2\pi V} \sum_{\mathbf{m} \neq 0} \frac{1}{\mathbf{m}^2} \sum_{(l_1, l_2) \notin C \times C}^N \exp(-\pi^2 \mathbf{m}^2 / \alpha_{l_1}) \\
&\times \exp(-\pi^2 \mathbf{m}^2 / \alpha_{l_2}) S_{l_1}(\mathbf{m}) S_{l_2}(-\mathbf{m}) \\
&- \frac{\pi}{2V} \sum_{l_1 \in C}^N \sum_{l_2 \in C}^N \sum_{i=1}^N \sum_{j=1}^N \mathbf{c}_{i,l_1,000} \mathbf{c}_{j,l_2,000} \left(\frac{1}{\beta} - \frac{1}{\alpha_{l_1}} - \frac{1}{\alpha_{l_2}} \right) \\
&- \sum_{i=1}^N E_{\text{self}}(\rho_i) + \frac{2\pi \mathbf{D}^2}{3V} + \varepsilon(K),
\end{aligned} \tag{6-54}$$

where the structure factors $S_l(\mathbf{m})$ are given by

$$S_l(\mathbf{m}) = \sum_{i=1}^N \sum_{tuv} \mathbf{c}_{i,l,tuv} \left(\frac{\partial}{\partial \mathbf{R}_{ijx}} \right)^{t_i+t_j} \left(\frac{\partial}{\partial \mathbf{R}_{ijy}} \right)^{u_i+u_j} \left(\frac{\partial}{\partial \mathbf{R}_{ijz}} \right)^{v_i+v_j} \exp(-2\pi i \mathbf{m} \cdot (\mathbf{R}_i)) \quad (6-55)$$

$E_{self}(\rho_i)$ is given by

$$\begin{aligned} & - \lim_{\mathbf{R} \rightarrow 0} \sum_{t_1 u_1 v_1} \sum_{t_2 u_2 v_2} (-1)^{(t_j+u_j+v_j)} \mathbf{c}_{j,l_j,t_j u_j v_j} \\ & \times \left(\frac{\partial}{\partial \mathbf{R}_x} \right)^{t_i+t_j} \left(\frac{\partial}{\partial \mathbf{R}_y} \right)^{u_i+u_j} \left(\frac{\partial}{\partial \mathbf{R}_z} \right)^{v_i+v_j} \\ & \times \left\{ \sum_{(l_1, l_2) \in C \times C} \mathbf{c}_{i,l_1,t_1 u_1 v_1} \mathbf{c}_{i,l_2,t_2 u_2 v_2} \frac{\text{erfc}(\beta^{1/2} |\mathbf{R}|)}{|\mathbf{R}|} \right. \\ & \left. - \sum_{(l_1, l_2) \notin C \times C} \mathbf{c}_{i,l_1,t_1 u_1 v_1} \mathbf{c}_{i,l_2,t_2 u_2 v_2} \frac{\text{erfc}(\mu_{12}^{1/2} |\mathbf{R}|)}{|\mathbf{R}|} \right\} \end{aligned} \quad (6-56)$$

where, similar to the previous section above, $1/\mu_{12} = 1/\alpha_{l_1} + 1/\alpha_{l_2}$ and $\mathbf{D}$ is the unit cell dipole [62].

N.2 Computational speedup for the direct and reciprocal sums

Computational speedups can be obtained for both the direct and reciprocal contributions. In the direct space sum, the issue is the efficient evaluation of the erfc function. One method proposed by Sagui et al. [64] relies on the McMurchie-Davidson [57] recursion to calculate the required erfc and higher derivatives for the multipoles. This same approach has been used by the authors for GEM [15]. This approach has been shown to be applicable not only for the Coulomb operator but to other types of operators such as overlap [15, 62].

In the case of the reciprocal sum, two methods have been implemented, smooth particle mesh Ewald (SPME) [65] and fast Fourier Poisson (FFP) [66]. SPME is based on the realization that the complex exponential in the structure factors can be approximated by a well behaved function with continuous derivatives. For example, in the case of Hermite charge distributions, the structure factor can be approximated by

$$\begin{aligned}
S_l(\mathbf{m}) \approx & \lambda(z_1)\lambda(z_2)\lambda(z_3) \sum_{k_1=-\frac{M_1}{2}+1}^{\frac{M_1}{2}} \sum_{k_2=-\frac{M_2}{2}+1}^{\frac{M_2}{2}} \sum_{k_3=-\frac{M_3}{2}+1}^{\frac{M_3}{2}} \sum_{i=1}^N \sum_{tuv} \mathbf{d}_{i,l,tuv} \\
& \times \sum_{l_1=-\infty}^{\infty} \left(\frac{\partial}{\partial w_{1j}} \right)^t B(w_{1i} - k_1 - l_1 M_1) \\
& \times \sum_{l_2=-\infty}^{\infty} \left(\frac{\partial}{\partial w_{2i}} \right)^u B(w_{2i} - k_2 - l_2 M_2) \\
& \times \sum_{l_3=-\infty}^{\infty} \left(\frac{\partial}{\partial w_{3i}} \right)^v B(w_{3i} - k_3 - l_3 M_3) \\
& \times \exp \left(2\pi i \left(\frac{m_1 k_1}{M_1} + \frac{m_2 k_2}{M_2} + \frac{m_3 k_3}{M_3} \right) \right)
\end{aligned} \tag{6-57}$$

where $\mathbf{d}_{i,l,tuv}$ are the transformed Hermite coefficients obtained from $\mathbf{c}_{i,l,tuv}$ by change of variable and $B(w)$ are B-splines [15]. Equation (6-57) is the approximation to $S_l(\mathbf{m})$ as a three dimensional discrete Fourier transform (3DFFT) times $\lambda(z_1)\lambda(z_2)\lambda(z_3)$. If the function $B(w)$ has finite support, then the structure factors can be calculated in $O(N(\log(N)))$ time.

The FFP method relies on the fact that the structure factors can be approximated by rewriting the reciprocal sum such that the structure factor is re-expressed as the 3DFFT evaluated at $\mathbf{m}$ of a Gaussian density $\rho(\mathbf{r}) = \sum_{i=1}^N q_i \rho_{2\alpha_0}(\mathbf{r} - \mathbf{r}_i)$ where $\rho_{2\alpha_0}$ is a normalized Gaussian with exponent $2\alpha_0$. This is very similar to the FFT methods used to accelerate structure factor and density map calculations in macromolecular structure determinations.

The efficiency of the methods outlined above has been tested by calculating the intermolecular Coulomb energies and forces for a series of water boxes (64, 128, 256, 512 and 1024) under periodic boundary conditions [15, 62]. The electron density of each monomer is expanded on five sites (atomic positions and bond mid-points) using two standard ABSs, A2 and P1. These sets were used to fit QM density of a single water molecule obtained at the B3LYP/6-31G* level. We have previously shown that the A1 fitted density has an 8% RMS force error with respect to the corresponding ab initio results. In the case of P1, this error is reduced to around 2% [15, 16]. Table 6-1 shows the results for the 5 water boxes using both ABSs (Table 6-7).

6.4. CONCLUSION

As we have seen, Anisotropic Polarizable Molecular Mechanics (APMM) procedures such as SIBFA or GEM are more complex than usual classical approaches.

Table 6-7. Timing (in seconds) of the calculated water boxes using full Ewald, PME and FFP for two different accuracies. All calculations were performed on a single 3.6 GHz Xeon processor^a

	RMS force = 10^{-3}					
	A1			P1		
	Ewald	PME	FFP	Ewald	PME	FFP
64	0.365	0.106	0.142	2.321	0.310	0.399
128	1.336	0.218	0.271	7.706	0.591	0.832
256	5.239	0.387	0.528	35.178	1.186	1.920
512	17.881	0.837	1.100	119.863	2.549	3.825
1024	71.513	1.701	2.236	486.384	4.953	6.890
	RMS force = 10^{-4}					
	A1			P1		
	Ewald	PME	FFP	Ewald	PME	FFP
64	0.520	0.144	0.274	3.858	0.478	0.688
128	1.869	0.287	0.380	11.056	0.923	1.576
256	7.256	0.517	0.846	49.107	1.805	2.736
512	25.511	1.104	1.534	183.487	3.794	5.684
1024	108.158	2.249	3.152	714.644	6.947	11.307

^a Additional speedups can be gained by reducing the size of the mesh for the sampling of the Gaussians by using the Gaussian split Ewald approach [67].

However, their parametrization is performed upon first principle energy decomposition schemes therefore lies on solid ground as any physical ingredients of the interaction can be added. We have seen that despite their variety, EDA schemes present significant convergence and can be easily used to calibrate MM approaches, especially as linear scaling techniques will allow performing large reference computations. In addition, methods able to unravel local electrostatic properties such as the ELF based DEMEP approach should help force field developers to build more realistic models. To conclude, we have seen that despite their different formulation, SIBFA and GEM share a common philosophy. That way, the GEM continuous electrostatic model will be used to replace SIBFA's distributed multipoles to produce a multiscale SIBFA-GEM approach. It will use the newly developed density based Periodic boundary conditions techniques giving access to an $N \cdot \log(N)$ evaluation of integrals, a key issue to perform fast and accurate polarizable molecular dynamics simulations. As perspectives, it is important to point out that such APMM approaches should clearly be an asset for hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) computations [2, 68, 69] as they embody short-range electrostatics and full induction. They will also help improving classical force fields by performing higher level reference calculations on relevant system [69].

ACKNOWLEDGMENTS

We thank the computer centers of CINES (Montpellier, France), CRIHAN (Rouen, France). This research was supported in part by the Agence National de la Recherche (ANR, France) on projects Lasihmodo and Wademecom; and by the Intramural Research Program of the NIH and NIEHS (U.S.A.).

REFERENCES

1. Yoda T, Sugita Y, Okamoto Y (2004) Secondary-structure preferences of force fields for proteins evaluated by generalized-ensemble simulations. *Chem Phys* 307:269
2. Gresh N, Cisneros GA, Darden TA, Piquemal J-P (2007) Anisotropic, polarizable molecular mechanics studies of inter-, intra-molecular interactions, and ligand-macromolecule complexes. A bottom-up strategy. *J Chem Theory Comput* 3:1960
3. Kitaura K, Morokuma K (1976) A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation. *Int J Quantum Chem* 10:325
4. Bagus PS, Hermann K, Bauschlicher CW, Jr (1984) A new analysis of charge transfer and polarization for ligand-metal bonding: Model studies of Al_4CO and Al_4NH_3 . *J Chem Phys* 80:4378
5. Bagus PS, Illas F (1992) Decomposition of the chemisorption bond by constrained variations: order of the variations and construction of the variational spaces. *J Chem Phys* 96:8962
6. Piquemal J-P, Marquez A, Parisel O, Giessner-Prettre C (2005) A CSOV Study of the difference between HF and DFT Intermolecular Interaction Energy Values: the importance of the charge transfer contribution. *J Comput Chem* 26:1052
7. Stevens WJ, Fink WH (1987) Frozen fragment reduced variational space analysis of hydrogen bonding interactions. Application to water dimer. *Chem Phys Lett* 139:15
8. Ziegler T, Rauk A (1979) CO, CS, N₂, PF₃ and CNCH₃ as donors and acceptors. A theoretical study by the Hartree-Fock-Slater transition state method. *Inorg Chem* 18:1755;
9. Bickelhaupt FM, Baerends EJ (2000) In: Lipkowitz KB, Boyd DB (ed) *Rev Comp Chem* 15(1), Wiley-VCH, New York
10. Jeziorski B, Moszynski R, Szalewicz K (1994) Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes. *Chem Rev* 94:1887
11. Reinhardt P, Piquemal J-P, Savin A (2008) Fragment-localized Kohn-Sham orbitals via a Singles-CI procedure and application to local properties and intermolecular energy decomposition analysis. *J Chem Theory Comput* 4:2020
12. Pilmé J, Piquemal J-P (2008) Advancing beyond charge analysis using the electronic localization function: Chemically intuitive distribution of electrostatic moments. *J Comput Chem* 29:1440
13. Gresh N, Claverie P, Pullman A (1984) Theoretical studies of molecular conformation. Derivation of an additive procedure for the computation of intramolecular interaction energies. Comparison with ab initio SCF computations. *Theor Chim* 66(1):1-20
14. Piquemal J-P, Cisneros GA, Reinhardt P, Gresh N, Darden TA (2006) Towards a force field based on density fitting. *J Chem Phys* 124:104101
15. Cisneros GA, Piquemal J-P, Darden TA (2006) Generalization of the Gaussian electrostatic model: extension to arbitrary angular momentum, distributed multipoles and speedup with reciprocal space methods. *J Chem Phys* 125:184101
16. Cisneros GA, Elking D, Piquemal J-P, Darden TA (2007) Numerical fitting of molecular properties to Hermite Gaussians. *J Phys Chem A* 111:12049

17. Gordon MS, Jensen JH (1998) Wavefunctions and chemical bonding. In: Schleyer PvR, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer III, HF, Schreiner PR (eds) *The encyclopedia of computational chemistry*, 5:3198 John Wiley and Sons, Chichester
18. Piquemal J-P, Chevreau H, Gresh N (2007) Towards a separate reproduction of the contributions to the Hartree-Fock and DFT intermolecular interaction energies by polarizable molecular mechanics with the SIBFA potential. *J Chem Theory Comput* 3:824
19. Piquemal J-P, Chelli R, Procacci P, Gresh N (2007) Key role of the polarization anisotropy of water in modeling classical polarizable force fields. *J Phys Chem A* 111:8170
20. Claverie P, PhD Thesis, CNRS library number A.O. 8214, Paris (1973)
21. Claverie P (1976) Localization and delocalization in quantum chemistry. In: Chalvet O, Daudel R, Diner S, Malrieu JP (eds) 21:127, Reidel, Dordrecht
22. Hess O, Caffarel M, Huiszoon C, Claverie P (1990) Second-order exchange effects in intermolecular interactions. The water dimer. *J Chem Phys* 92:6049
23. Khaliullin RZ, Cobar EA, Lochan RC, Bell AT, Head-Gordon M (2007) Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals. *J Phys Chem A* 111:8753
24. Daudey J-P (1974) Direct determination of localized SCF orbitals. *Chem Phys Lett* 24:574
25. Ochsenfeld C, Kussmann J, Lambrecht DS (2007) Linear-scaling methods in quantum chemistry. In: Lipkowitz KB, Cundari TR (ed) *Rev Comp Chem* 23(1), VCH, New York
26. Rubio J, Povill A, Malrieu J-P, Reinhardt PJ (1997) Direct determination of localized Hartree-Fock orbitals as a step toward Nscaling procedures. *J Chem Phys* 107:10044
27. Pople JA, Gill PMW, Johnson BG (1992) Kohn-Sham density-functional theory within a finite basis set. *Chem Phys Lett* 199:557
28. Becke AD, Edgecombe KE (1990) A simple measure of electron localization in atomic and molecular systems. *J Chem Phys* 92:5397
29. Savin A, Flad OJ, Andersen J, Preuss H, Von Schering HG, Angew (1992) Electron localization in solid-state. Structures for the elements: the diamond. *Chem Int* 31:187
30. Silvi B, Savin A (1994) Classification of chemical bonds based on topological analysis of electron localization functions. *Nature* 371:683
31. Piquemal J-P, Pilmé J, Parisel O, Gérard H, Fourré I, Bergès J, Gourlaouen C, de la Lande A, van Severen MC, Silvi B (2008) What can be learnt on biological or biomimetic systems with the topological analysis of the electron localization function? *Int J Quant Chem* 108:1951
32. Popelier PLA (2000) *Atoms in molecules: An introduction*, Prentice-Hall, Harlow, UK
33. Bader RFW (1990) *Atoms in molecules: A quantum theory*, Oxford University Press, Oxford
34. Vigné-Maeder F, Claverie P (1988) The exact multicenter multipolar part of a molecular charge distribution and its simplified representations. *J Chem Phys* 88:4934
35. Garmer DR, Stevens WJ (1989) Transferability of molecular distributed polarizabilities from a simple localized orbital based method. *J Phys Chem* 93:8263
36. Piquemal J-P *int. J. Quant Chem.* (2010) submitted (Theobio 09, special issue)
37. Piquemal J-P, Gresh N, Giessner-Prettre C (2003) Improved formulas for the calculation of the electrostatic contribution to intermolecular interaction energy from multipolar expansion of the electronic distribution. *J Phys Chem A* 107:10353
38. Murrell JN, Teixeira D (1970) *J Mol Phys* 19:521
39. Williams DR, Schaad LJ, Murrell JN (1967) *J Chem Phys* 47:4916
40. Piquemal J-P (2004) Evaluation of molecular interactions in bioinorganic systems: from ab initio computations to polarizable force fields. PhD Thesis, UPMC number 2004PA066267, Université Pierre et Marie Curie, Paris, France

41. Gresh N, Piquemal J-P, Krauss M (2005) Representation of Zn(II) complexes in polarizable molecular mechanics. Further refinements of the electrostatic and short-range contribution of the intermolecular interaction energy. Comparisons with parallel ab initio computations. *J Comput Chem* 26:1113
42. Mulliken R, Rieke C, Orloff D, Orloff H (1948) Formulas and numerical tables for overlap integrals. *J Chem Phys* 17:1248
43. Roothan CJ (1951) A study of two-center integrals useful in calculations on molecular structure. *J Chem Phys* 19:1445
44. Foster JM, Boys SF (1960) Canonical configurational interaction procedure. *Rev Mod Phys* 32:300
45. Gresh N (1995) Energetics of Zn²⁺ binding to a series of biologically relevant ligands: A molecular mechanics investigation grounded on ab initio self-consistent field supermolecular computations. *J Comput Chem* 16:856
46. Murrell JN, Randic M, Williams DR (1966) *Proc Roy Soc A* 284:566 (London)
47. Gresh N, Claverie P, Pullman A (1982) Computations of intermolecular interactions: Expansion of a charge-transfer energy contribution in the framework of an additive procedure. Applications to hydrogen-bonded systems. *Int J Quant Chem* 22:199
48. Gresh N, Claverie P, Pullman A (1986) Intermolecular interactions: Elaboration on an additive procedure including an explicit charge-transfer contribution. *Int J Quant Chem* 29:101
49. Creuzet S, Langlet J, Gresh N (1991) *J Chim Phys* 88:2399
50. Piquemal J-P, Gresh N (to be published)
51. Piquemal J-P, Williams-Hubbard B, Fey N, Deeth RJ, Gresh N, Giessner-Prettre C (2003) Inclusion of the ligand field contribution in a polarizable molecular mechanics: SIBFA LF. *J Comput Chem* 24:1963
52. Schäffer CE, Jørgensen CK (1965) The angular overlap model, an attempt to revive the ligand field approaches. *Mol Phys* 9:401
53. Deeth RJ (2001) The ligand field molecular mechanics model and the stereoelectronic effects of d and s electrons. *Coord Chem Rev* 212:11
54. Woodley M, Battle PD, Catlow CRA, Gale JD (2001) Development of a new interatomic potential for the modeling of ligand field effects. *J Phys Chem B* 105:6824
55. Boys SF, Shavitt I (1959) A fundamental calculation of the energy surface for the system of three hydrogen atoms, NTIS, Springfield, VA, AD212985
56. Dunlap BI, Connolly JWD, Sabin JR (1979) On first row diatomic molecules and local density models. *J Chem Phys* 71:4993
57. McMurchie LE, Davidson ER (1978) One- and two-electron integrals over cartesian gaussian functions. *J Comp Phys* 26:218
58. Cisneros GA, Piquemal J-P, Darden TA (2005) Intermolecular electrostatic energies using density fitting. *J Chem Phys* 123:044109
59. Wheatley RJ, Price S (1990) An overlap model for estimating the anisotropy of repulsion. *Mol Phys* 69:507
60. Piquemal J-P, Perera L, Cisneros GA, Ren P, Pedersen LG, Darden TA (2006) Towards accurate solvation dynamics of divalent cations in water using the polarizable Amoeba force field: from energetics to structure. *J Chem Phys* 125:054511
61. Ren P, Ponder JW (2003) Temperature and pressure dependence of AMOEBA water model. *J Phys Chem B* 107:5933
62. Darden TA (2008) Dual bases in crystallographic computing. In *International Tables of Crystallography*, vol. B, 3rd edn. Kluwer Academic Publishers, Dordrecht, The Netherlands, (in press)
63. Fusti-Molnar L, Pulay P (2002) The Fourier transform Coulomb method: Efficient and accurate calculation of the Coulomb operator in a Gaussian basis. *J Chem Phys* 117:7827

64. Sagui C, Pedersen LG, Darden TA (2004) Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. *J Chem Phys* 120:73–87
65. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577
66. York D, Yang W (1994) The fast Fourier Poisson method for calculating Ewald sums. *J Chem Phys* 101:3298
67. Shan YB, Klepeis JL, Eastwood MP, Dror RO, Shaw DE (2005) Gaussian split Ewald: A fast Ewald mesh method for molecular simulation. *J Chem Phys* 122:054101
68. Cisneros GA, Piquemal J-P, Darden TA (2006) QM/MM electrostatic embedding with continuous and discrete functions. *J Phys Chem B* 110:13682
69. Cisneros GA, Na-Im Tholander S, Elking D, Darden TA, Parisel O, Piquemal J-P (2008) Simple formulas for improved point-charge electrostatics in classical force fields and hybrid Quantum Mechanical/Molecular Mechanical embedding. *Int J Quant Chem* 108:1905