Status of the Gaussian Electrostatic Model, a Density-Based Polarizable Force Field

Jean-Philip Piquemal, G. Andrés Cisneros

HAL Id: hal-01114075

https://hal.archives-ouvertes.fr/hal-01114075

Submitted on 3 Nov 2015

Contents

Chapter 1

Status of the Gaussian Electrostatic Model, a Density–Based Polarizable Force Field

Jean-Philip Piquemal[a] and G. Andrés Cisneros[b]

[a]UPMC Univ Paris 06, Sorbonne Universités, UMR 7616 Laboratoire de Chimie Théorique case courrier 137, 4 place Jussieu 75005, Paris (France); CNRS, UMR 7616 Laboratoire de Chimie Théorique case courrier 137, 4 place Jussieu 75005, Paris (France)
[b]Department of Chemistry, Wayne State University, Detroit, MI 48202
Email address: jpp@lct.jussieu.fr, andres@chem.wayne.edu

1.1 Abstract

We present the status and advances of the Gaussian Electrostatic Model (GEM) force field. GEM is a polarizable potential that relies on fitted molecular electronic densities to calculate intermolecular interactions. The philosophy of GEM is to separately account for each of the inter–molecular interaction contributions including Coulomb, exchange–repulsion, polarization, charge–transfer and dispersion. To evaluate each of these terms, GEM relies on fitted molecular electronic densities and molecular properties calculated with these densities. Recently, we have also developed hybrid force

fields combining GEM with polarizable force fields. GEM* combines the Coulomb and exchange–repulsion terms from GEM with the polarization, van der Waals (modified) and bonded terms from AMOEBA. We have also introduced a multi–scale implementation denoted S/G–1 that combines GEM and SIBFA demonstrating the advantage of the incorporation of the electron density force field when applied to the description of metals in biomolecular systems. Overall, the GEM framework and its hybrid approaches offer a new strategy for molecular dynamics with enhanced accuracy at a cost that will enable the possibility of adequate sampling.

1.2 Introduction

The use of classical potentials for simulations of chemical and biochemical systems with molecular dynamics has been a field of intense research. Currently, it is possible to simulate systems with millions of atoms and millisecond time scales [Schulten et al. (2008); Shaw et al. (2010)]. With exa–scale computing, i.e. 10^{18} floating point operations per second (FLOPs), on the horizon it is necessary to evaluate the performance of the current potentials. Indeed, long–time biomolecular simulations have revealed some issues already. For example Raval et al. carried out a study on 24 proteins (both homology models and experimental structures) used in recent CASP competitions involving at least $100\mu s$ MD simulations [Raval et al. (2012)]. For most systems, the structures drifted away from the native state, even when starting from the experimental structure. Although only two conventional force fields were employed, the authors concluded that this is most likely a limitation of the available point–charge force fields. As simulations on these and longer scales grow more widespread with improvements in computing power, node inter–connect, and graphical processing unit (GPU) hardware [Stone et al. (2007)], the accuracy of these classical potentials will be further tested.

In this context, there has been a recent impetus to develop more accurate force fields. One of the main thrusts has been to improve the description of the bonded interactions by including anharmonicity and of the non–bonded interactions by introducing ex-

plicit polarization and a better description of the charge anisotropy terms. Several force fields that employ distributed multipoles and use explicit polarization (or are QM based using simplified MO schemes) have been proposed including AMOEBA, SIBFA, EFP, X–Pol, mDC and NEMO among others [Hermida-Ramón et al. (2003); Gresh et al. (2007); Ponder et al. (2010); Day et al. (1996); Xie and Gao (2007); Xie et al. (2009); Mills and Popelier (2012); Popelier (2012); Giese et al. (2013); Babin et al. (2014); Giese et al. (2014)]. The use of distributed multipoles results in an improved description of the charge density anisotropy and provides more accurate electrostatic interactions [Stone (2000); Price (1999); Popelier (2000); Kosov and Popelier (2000); Popelier et al. (2001a); Popelier and Kosov (2001); McDaniel and Schmidt (2014)]. However, distributed multipoles suffer from one drawback since they cannot describe the overlap of charge density as two molecules get close to each other. This is known as the penetration effect [Stone (2000); Freitag et al. (2000)]. It is possible to reduce the penetration error by employing empirical damping functions [Kairys and Jensen (1999); Freitag et al. (2000); Piquemal et al. (2003); Cisneros et al. (2008); Wang and Truhlar (2010); Stone (2011)]. It is also possible to include this effect via the use of neural networks [Handley and Popelier (2010)].

Another possibility to avoid the charge penetration and anisotropy shortcomings is to use a continuous description of the molecular charge density. Several methods that describe the electronic distribution explicitly have been proposed [Wheatley (2011); Gavezzotti (2002); Eckhardt and Gavezzotti (2007); Volkov and Coppens (2004); Coppens and Volkov (2004); Paricaud et al. (2005)]. We have introduced the Gaussian Electrostatic Model (GEM) [Cisneros et al. (2005a); Piquemal et al. (2006a); Cisneros et al. (2006b)]. GEM uses density fitting (DF) techniques [Boys and Shavit (1959); Dunlap et al. (1979); Köster et al. (2002)] to reproduce the molecular electronic density using Hermite Gaussian auxiliary basis sets (ABSs). These fitted densities are employed to calculate each intermolecular component as obtained from energy–decomposition (EDA) prodecures. The reason for the use of EDA methods for the parametrization of GEM is that it enables the separation of each of the components of the intermolecular inter-

4 | Status of the GEM Force Field

actions. There are a variety of EDA approaches that can be employed including symmetry–adapted perturbation theory (SAPT), Kitaura–Morokuma (KM), restricted variational space (RVS), constrained space orbital variations (CSOV), to name a few [Eisenschitz and London (1930); Hirshfelder (1967a,b); Murrel and Shaw (1967); Kitaura and Morokuma (1976); Bagus et al. (1984); Stevens and Fink (1987); Jeziorski et al. (1994); Glendening and Streitwieser (1994); Glendening (1994); Mo et al. (2000); Heßelmann et al. (2005); Piquemal et al. (2005); Khaliullin et al. (2006); Wu et al. (2009); Lu et al. (2011)].

In this contribution we present the theory behind the GEM method and recent advances and results on the application of two hybrid GEM potentials. In section 1.3 we provide a brief review of the analytical and numerical density fitting methods and it's implementation including the methods employed to control numerical instabilities. This is followed by a review of the procedure to obtain distributed site multipoles from the fitted Hermite coefficients in section 1.4. Section 1.5 describes the extension of reciprocal space methods for continuous densities. Section 1.6 describes the complete form for GEM and a novel hybrid force field, GEM*, which combines term from GEM and AMOEBA for MD simulations. Finally, section 1.7 describes the implementation and initial applications of a multi–scale program that combines GEM and SIBFA.

1.3 Density fitting methods

The use of ABSs for density fitting is a field of intense study. This method relies on the use of auxiliary basis functions (ABS), generally Gaussians, to expand the molecular electron density

$$\tilde{\rho}(r) = \sum_k c_k \Lambda(r). \tag{1.1}$$

For GEM the ABSs consist of Hermite Gaussians, $\Lambda_{tuv}(r)$. The expansion coefficients c_k for the approximate density $\tilde{\rho}$ may be obtained by minimizing Eq. 1.2 using some metric $\hat{O}$ [Dunlap et al. (1979); Eichkorn et al. (1995); Köster (1996); Köster et al. (2002)].

$$E_{self} = <\rho(r) - \tilde{\rho}(r)|\hat{O}|\rho(r) - \tilde{\rho}(r)> \qquad (1.2)$$

Several operators $\hat{O}$ can be employed including the overlap operator $\hat{O} = 1$, the Coulomb operator $\hat{O} = 1/r$ or the damped Coulomb operator $\hat{O} = erfc(\beta r)/r$ [Jung et al. (2005)]. The minimization of Eq. 1.2 with respect to the expansion coefficients c_k leads to a linear system of equations:

$$\frac{\partial E_{self}}{\partial c_l} = -\sum_{\mu,\nu} P_{\mu\nu} <\mu\nu|\hat{O}|l> + \sum_k c_k <k|\hat{O}|l> \qquad (1.3)$$

The solution of Eq. 1.3 requires the inversion of a the ABS matrix $G = <k|\hat{O}|l>$. In principle this matrix should be positive definite and symmetric. In practice however, this matrix is almost singular and therefore the diagonalization to obtain its inverse must be done with care. To this end we have explored analytical and numerical procedures to obtain G and G^{-1}.

1.3.1 Analytical Fitting

The analytical fitting procedure involves the explicit evaluation of all the matrix elements of G and its subsequent inversion, which is achieved by diagonalization. We have implemented several methods for the diagonalization step. Initially we employed singular value decomposition (SVD) [Press et al. (1992)] by setting the inverse of the eigenvalue to zero if it is below a certain cutoff. However, this method produces undesirable numerical instabilities (noise) when the number of basis functions starts to grow as we and others have discussed previously [Cisneros et al. (2005b); Podeszwa et al. (2006)].

In the current implementation we employ the Tikhonov regularization method [Press et al. (1992)]. This approach is similar to the constrained density fitting algorithm of Misquitta and Stone [Misquitta and Stone (2006)]. Here, the redundant basis set contributions are penalized by minimizing $E_{self} + \lambda \sum_k x_k^2$, resulting in a more stable diagonalization procedure [Cisneros et al. (2006a)].

For problematic systems we also implemented the damped Coulomb operator $\hat{O} = erfc(\beta r)/r$ proposed by Jung et al. [Jung

et al. (2005)] to attenuate the near singular behavior due to long range interactions present in G. In our studies we observed that noise is still a problem for some systems. The noise in the fit is known to arise due to the attempt of the ABSs to fit the density at the nuclear cores [Cisneros et al. (2005b, 2006a)].

The use of Hermite Gaussians with angular moment greater than 0 requires the rotation of the fitting coefficients. In GEM this is addressed in a similar manner to multipolar force fields by defining a global molecular frame and a reference (local) site frame [Toukmaji et al. (2000); Sagui et al. (2004); Cisneros et al. (2006a)]. The use of Hermite Gaussians provides a straightforward solution to the rotation since they are defined by partial derivatives of a spherical Gaussian which can be taken either with respect to the local (reference) frame or with respect to the global coordinates [Cisneros et al. (2006a)]. Moreover, the rotation frames are the same for the distributed multipoles.

1.3.2 Numerical fitting

As mentioned above, the numerical instabilities in the fit arise from the attempt to fit the nuclear cores. Thus, if the density at the cores is discarded then the fit should become more stable. This can be achieved by using numerical grids to evaluate a given molecular property and discarding points at and near the core. This can be achieved by minimizing the following fitting function:

$$\chi^2 = \sum_i W(r_i) \left(y(\mathrm{r}_i) - \tilde{y}(\mathrm{r}_i, c_k) \right)^2, \qquad (1.4)$$

where $y(\mathrm{r}_i)$ denotes the ab initio molecular property of interest at point i and $\tilde{y}(\mathrm{r}_i, c_k)$ is the same property evaluated with the *kth* ABS element at the same point on the grid. Finally, $W(r)$ is the weighting function for the point on the grid, which can be defined in several ways [Bayly et al. (1993); Hu et al. (2007)]. Hu et al have proposed a weighting function that provides a smooth cutoff near the cores and at long distance to avoid any discontinuities [Hu et al. (2007)]:

$$W(r_i) = exp[-\sigma(log\ \rho_{promol}(\mathrm{r}) - log\ \rho_{ref})^2], \qquad (1.5)$$

where ρ_{promol} is a reference promolecular atomic density, and σ and ρ_{ref} are adjustable parameters. It has been shown that the surface for σ and ρ_{ref} is relatively flat [Hu et al. (2007); Elking et al. (2010)]. We have implemented a modified version of this weighting function previously [Elking et al. (2010)]. The main differences between the original Hu et al. weight and our implementation are the re–optimization of the promolecular atomic electron densities at the MP2/aug–cc–pVQZ level and the values for σ and ρ_{ref} which correspond to 0.42 and -7.0 respectively.

The minimization of Equation 1.5 leads to a linear system of equations that can be expressed as: $c - c^0 = -H^{0,-1}g^0$. As was the case for the analytic DF, we employ Tikhonov regularization for the inversion of the Hessian that arises for the linear–least–squares procedure. In our initial implementation of numerical fitting we explored different molecular properties including electronic density, molecular electrostatic potential (mESP), and the three components of the electric field [Cisneros et al. (2007)]. All the properties were gridded on rectangular grids. Subsequently we showed that the use of spherical molecular grids based on the scheme proposed by Becke [Becke (1988)] significantly reduce the number of fitting points [Elking et al. (2010); Cisneros (2012)].

1.4 Distributed multipoles

In this subsection we present the methodology to obtain Cartesian point multipoles from the Hermite coefficients obtained in the fitting procedure. In all our work we have purposefully employed ABSs with a maximum angular momentum of 2, which results in distributed multipoles only up to quadrupoles. This ensures that the distributed multipoles obtained can be directly employed in the AMOEBA force field. However, higher order multipoles can be obtained if an ABS with higher angular momentum is used.

Briefly, we have expanded on the work by Challacombe et al., who have shown that Hermite Gaussians have a simple relation to elements of the Cartesian multipole tensor [Challacombe et al. (1996)]. Once the Hermite coefficients have been determined, they may be employed to calculate point multipoles centered at the ex-

8   Status of the GEM Force Field

pansion sites. Thus, if hc_{tuv} represents the coefficient of a Hermite Gaussian of order Λ_{tuv}, then if this Hermite is normalized we have:

$$hc_{000} \int \Lambda_0 dr = hc_{000} \qquad (1.6)$$

This guarantees that higher order multipole integrals will integrate to integer numbers, for example, for the dipole integral in the z direction, d_z:

$$
\begin{aligned}
hc_{001} \int \int \int z\Lambda_{001} dxdydz &= hc_{001} \int z\frac{\partial}{\partial S_z}\Lambda_0 dz \\
&= -hc_{001} \int z\frac{\partial}{\partial z}\Lambda_0 dz = hc_{001} \quad (1.7)
\end{aligned}
$$

For quadrupole and higher order integrals the same relationships hold, although different cases need to be considered (see ref. [Cisneros et al. (2006a)]). In practice, following Stone's definition [Stone (2000)], we have used traceless quadrupoles. Furthermore, the use of GEM distributed multipoles (GEM–DM) for multipolar force fields provides a straightforward way to determine the penetration error in the site–site Coulomb interaction energy due to the connection with the GEM Hermites. Thus, this connection provides a natural way to generate damping functions to lessen the penetration error [Piquemal et al. (2003); Cisneros et al. (2008)].

A further advantage of this approach to distributed multipoles is that, unlike some conventional multipole expansions [Stone (2005); Popelier et al. (2001b)], the (spherical) multipole expansion obtained from Hermite Gaussians in this way is intrinsically finite of order $t + u + v$ (i.e. the highest angular momentum in the ABS) as shown in [Cisneros et al. (2006a)], similar to the multipoles obtained by Volkov and Coppens [Volkov and Coppens (2004)].

1.5 Reciprocal space methods for integral evaluation

The use of molecular densities results in the need to compute a large number of two center integrals for the intermolecular interaction. A significant computational speedup can be achieved by using reciprocal space methods based on Ewald sums. In this way,

the integrals are calculated in direct or reciprocal space depending on the exponent of the Gaussian Hermites.

Here we describe how the Ewald formalism can be extended to take into account the Gaussian distribution. Let U denote a unit cell whose edges are given by the vectors a_1, a_2 and a_3. An idealized infinite crystal can be generated by all periodic translations $n = n_1 a_1 + n_2 a_2 + n_3 a_3$ for all integer triples (n_1, n_2, n_3), with n_1, n_2, n_3 not all zero. Now, consider a collection of N normalized spherical Gaussian charge densities $\rho_1 \ldots \rho_N$ (e.g. GEM–0) centered at $\{R_1 \ldots R_N\} \in U$ with exponents α_i, i.e. $\rho_i(r) = q_i(\alpha_i/\pi)^{3/2} exp(-\alpha_i(r - R_i)^2)$, and let $q_1 + \ldots + q_N = 0$. Note that N need not be limited only to atomic positions, e.g., GEM–0 includes sites on the oxygen lone pairs and the bisector line between the two hydrogens [Piquemal et al. (2006a)]. The Coulomb energy of the central unit cell within a large spherical crystal, due to the interactions of the Gaussian charge distributions with each other and all periodic images within the crystal can be calculated using Ewald methods.

In particular, to determine the reciprocal part in the Ewald sum it is necessary to grid the Gaussian densities. However, this can become intractable for Gaussian functions with large exponents (compact Gaussians). In the initial implementation the charge densities were classified into compact or diffuse Hermite Gaussians based on a given Ewald exponent β. Therefore, if the exponent of a given Hermite was above the cutoff it was considered compact, and diffuse ($\alpha_i < \beta$) otherwise. With this, the contributions involving diffuse Hermites can be calculated in reciprocal space exclusively [Cisneros et al. (2006b)].

This was later improved by the realization that the Ewald exponent, β may be different for each pair ij [Darden (2007)]. Thus, β is chosen to be infinite for ij pairs where at least one of the Gaussians is diffuse. In this way, all pairs that involve diffuse Hermites are evaluated in reciprocal space. That is, given $\theta > 0$ a Gaussian distribution $q_i\rho_i$ is classified as compact ($i \in c$) if $\alpha_i \geq 2\theta$ and diffuse ($i \in d$) otherwise. Subsequently, for $i, j \in \{c\}$, select β so that $1/\theta = 1/\alpha_i + 1/\alpha_j + 1/\beta$, otherwise $\beta = \infty$. In the case of GEM, the fitted densities are expanded in a linear combination of Hermite Gaussians $\Lambda_{tuv}(r, \alpha, R)$. Thus, the charge distribution is given by

10 | Status of the GEM Force Field

$\rho_i(\mathrm{r}, \alpha, \mathrm{R}) = \sum_{l=1}^{L} \sum_{tuv} \mathrm{c}_{i,l,tuv} \Lambda_{tuv}(\mathrm{r}, \alpha_l, \mathrm{R_i})$, where $\mathrm{c}_{i,l,tuv}$ are the Hermite coefficients and L denotes the different ABS exponents on center i. With this, the Ewald expression becomes

$$
\begin{aligned}
E(\rho^{\{N\}}) = {} & \frac{1}{2} \sum_{\mathrm{n}}{}' \sum_{i=1}^{N} \sum_{l_i \in c} \sum_{t_i u_i v_i} \mathrm{c}_{i,l_i,t_i u_i v_i} \\
& \sum_{j=1}^{N} \sum_{l_j \in c} \sum_{t_j u_j v_j} (-1)^{(t_j+u_j+v_j)} \mathrm{c}_{j,l_j,t_j u_j v_j} \\
& \times \left(\frac{\partial}{\partial \mathrm{R}_{ijx}} \right)^{t_i+t_j} \left(\frac{\partial}{\partial \mathrm{R}_{ijy}} \right)^{u_i+u_j} \left(\frac{\partial}{\partial \mathrm{R}_{ijz}} \right)^{v_i+v_j} \\
& \times \left\{ \frac{erfc(\theta^{1/2}|\mathrm{R}_{ij}-\mathrm{n}|) - erfc(\mu_{l_i l_j}^{1/2}|\mathrm{R}_{ij}-\mathrm{n}|)}{|\mathrm{R}_{ij}-\mathrm{n}|} \right\} \\
& + \frac{1}{2\pi V} \sum_{\mathrm{m} \neq 0} \frac{1}{\mathrm{m}^2} \exp(-\pi^2 \mathrm{m}^2/2\theta) \sum_{l_1 \in c} S_{l_1}(\mathrm{m}) \\
& \times \exp(-\pi^2 \mathrm{m}^2/2\theta) \sum_{l_2 \in c} S_{l_2}(-\mathrm{m}) \hspace{2cm} (1.8) \\
& + \frac{1}{2\pi V} \sum_{\mathrm{m} \neq 0} \frac{1}{\mathrm{m}^2} \sum_{(l_1,l_2) \notin c \times c}^{N} \exp(-\pi^2 \mathrm{m}^2/\alpha_{l_1}) \\
& \times \exp(-\pi^2 \mathrm{m}^2/\alpha_{l_2}) S_{l_1}(\mathrm{m}) S_{l_2}(-\mathrm{m}) \\
& - \frac{\pi}{2V} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l_1 \in c} \sum_{l_2 \in c} \mathrm{c}_{i,l_1,000} \mathrm{c}_{j,l_2,000} \left(\frac{1}{\theta} - \frac{1}{\alpha_{l_1}} - \frac{1}{\alpha_{l_2}} \right) \\
& - \sum_{i=1}^{N} E_{self}(\rho_i) + \frac{2\pi \mathrm{D}^2}{3V} + \varepsilon(K),
\end{aligned}
$$

where the first term corresponds to the direct part of the Ewald sum, the second and third terms to the reciprocal part, $\mathrm{R}_{ij} = \mathrm{R}_i - \mathrm{R}_j$, the structure factors $S_l(\mathrm{m})$ involve derivatives of the Fourier exponential with respect to the Hermite centers, $E_{self}(\rho_i)$ is the correction due to the self energy of each Hermite interacting with its replicate, the term involving the unit cell dipole $\mathrm{D} = q_1 \mathrm{R}_1 + \ldots + q_N \mathrm{R}_N$ is the surface term, $\varepsilon(K)$ denotes a quantity that converges to 0 as $K \longrightarrow \infty$, m denotes the reciprocal lattice vectors, and $1/\mu_{l_i l_j} = 1/\alpha_{l_i} + 1/\alpha_{l_j}$ [Cisneros et al. (2006b); Darden (2007)]..

Since the ABSs include Hermites with $l > 0$, the direct space contributions can be efficiently evaluated by using the McMurchie-Davidson (MD) recursion [McMurchie and Davidson (1978)]. This recursion has been used to calculate the required erfc and higher derivatives for multipole interactions [Sagui et al. (2004)]. This approach was also employed for the Hermite Gaussians [Cisneros et al. (2006b)], where it was shown that the MD recursion is applicable to other types of operators besides $1/r$. For the reciprocal sums three methods were implemented: full Ewald [Ewald (1921)], sPME [Essmann et al. (1995)] and FFP [York and Yang (1994)]. The latter two methods rely on the use of fast fourier transforms to approximate the structure factors that arise in the reciprocal term, which results in the efficient evaluation of this term and has been shown to scale as $O(NlogN)$ for sPME [Essmann et al. (1995)].

1.6 The GEM and GEM* force fields

The initial implementation of the full GEM potential involved the use of spherical type Hermites only, resulting in what was termed GEM–0 [Piquemal et al. (2006b)]. This initial parametrization included the terms described below.

1.6.1 The GEM functional form

The idea for GEM is to employ the fitted Hermite Gaussians to evaluate each term in

$$E_{Total} = E_{Coulomb}^{GEM} + E_{exch-rep}^{GEM} \\ + E_{polarization}^{GEM} + E_{charge-transfer}^{GEM}, \qquad (1.9)$$

where the Coulomb term is given by

12   Status of the GEM Force Field

$$E_{Coulomb}^{GEM} = \sum_{A>B} \frac{Z_A Z_B}{r_{AB}} + \sum_{A>B} \int \frac{Z_A \tilde{\rho}^B(r_B)}{r_{AB}} dr$$
$$+ \sum_{A>B} \int \frac{Z_B \tilde{\rho}^A(r_A)}{r_{AB}} dr$$
$$+ \sum_{A>B} \int \frac{\tilde{\rho}^A(r_A)\tilde{\rho}^B(r_B)}{r_{AB}} dr. \qquad (1.10)$$

The exchange–repulsion term is calculated by means of the charge density overlap following the Wheatley–Price overlap model [Wheatley and Price (1990); Domene et al. (2001)]:

$$E_{exch-rep}^{GEM} = K \sum_{A>B} \int \tilde{\rho}^A(r_A)\tilde{\rho}^B(r_B). \qquad (1.11)$$

The polarization term is approximated by the use of dipole polarizabilities, which yield a very good results for the polarization energies (if the electric fields are not large) [Böttcher (1993)]. To this end, the electric fields are calculated with the fitted densities and interacted with distributed dipolar polarizabilidies with the Garmer and Steven's approach [Day et al. (1996)]

$$E_{polarization}^{GEM} = \frac{1}{2} \sum_{j}^{xyz} \Delta\mu(i)(\gamma E_0(j)), \qquad (1.12)$$

where $\Delta\mu(i) = \alpha(i)\sum_{j}^{xyz} E(\Delta\mu(i)) + (\gamma E_0(j))$, and γ is a scaling factor for the permanent electric fields [Piquemal et al. (2006b)].

Finally, the charge transfer term is evaluated using the semiempirical formalism implemented in the SIBFA force field [Gresh et al. (1979); Piquemal et al. (2007)]:

$$E_{charge-transfer}^{GEM} = 2C \sum_{L_\alpha} \frac{(I_{\alpha\beta}^*)^2}{\Delta E_{\alpha\beta}^*}, \qquad (1.13)$$

where $C = 3.5$ is a constant parametrized to reproduce the value of the charge–transfer energy (obtained with CSOV) for the canonical water dimer at equilibrium distance. $I_{\alpha\beta}^*$ is a function

of the overlap between the localized molecular orbital (LMO) for
the donor lone pair and antibonding LMO of the acceptor, as well
as the electrostatic potential on site A arising from all other sites
obtained with the GEM densities. $\Delta E^*_{\alpha\beta}$ is a difference between the
ionization potential of A and the electron affinity of the acceptor
site.

In our initial implementations of GEM–0 and GEM we have not
introduced an explicit term for the dispersion interactions. This is
because these force fields have been originally parametrized using
the CSOV method at the DFT level, which, by definition, does
not include a dispersion contribution. However, it is possible to
include this term in a similar way to the SIBFA potential [Gresh
et al. (1979); Piquemal et al. (2007)].

1.6.2 GEM*: molecular dynamics with fitted densities

After our implementation of GEM–0, we extended the Coulomb
and exchange–repulsion terms to enable the use or arbitrary angu-
lar momentum Hermites [Cisneros et al. (2006b)]. However, both
implementations only enabled energy calculations. In order to carry
out MD simulations it is necessary to evaluate the associated forces
efficiently. Until recently, it was impractical to do this since the
analitical form of the force for the charge-transfer term was un-
available.

To enable the performance of MD simulations, a hybrid force
field called GEM* was developed. GEM* combines the Coulomb
and exchange–repulsion terms from GEM with the polarization,
van der Waals (modified) and bonded terms from AMOEBA. The
functional form for GEM* is thus:

$$E_{Total} = E^{GEM}_{Coulomb} + E^{GEM}_{exch-rep} + E^{AMOEBA}_{polarization}$$
$$+E^{AMOEBA}_{VdW} + E^{AMOEBA}_{bonded}. \tag{1.14}$$

The Coulomb and exchange–repulsion terms for GEM* are eval-
uated with the same expressions as for GEM (Eqns. 1.10 and 1.11).
Since GEM* includes an explicit term for exchange, it was neces-
sary to modify the original van der Waals function implemented

in AMOEBA. In this case, we have modified the buffered Halgren function (modHalgren) by removing the repulsive term as follows:

$$E_{modHalgren} = -\epsilon_{ij} \left[\frac{1.07 R_{ij}^*}{(R_{ij} + 0.07 R_{ij}^*)} \right]^7 . \qquad (1.15)$$

The polarization and bonded terms are the same as those in the original AMOEBA implementation [Ren and Ponder (2003)]. The only difference in the polarization is that the permanent electric fields for the calculation of the induced dipoles are calcualted with the distributed multipoles obtained from the fitted Hermites for consistency between the Coulomb and polarization terms [Cisneros (2012)].

The initial implementation of GEM* was tested by fitting parameters for a water potential [Duke et al. (2014)]. These parameters were compared to reference ab initio values for total intermolecular interactions corrected for basis set superposition error via the counterpoise correction . The reference data was calculated at the MP2(full)/aug–cc–pVTZ level to match the original AMOEBA parametrization [Ren and Ponder (2003, 2002); Ponder et al. (2010); Ren et al. (2011)]. The molecular density used to obtain the fitting coefficients for GEM* was calculated at the same level of theory as above for a water molecule at the AMOEBA equilibrium geometry.

Three parametrizations were investigated, termed models 1–3 in the discussion below. The difference among the three models involves the use of different ABSs, A1 or A2 [Andzelm and Wimmer (1992); Godbout and Andzelm (1999)], and/or the dataset of water oligomers used for the parametrization. Model 1 was fitted using the A2 ABS to reproduce intermolecular interaction energies for the canonical water dimer (see Figure 1.1.), several random dimers, and selected water clusters from [Temelso et al. (2011)]. Models 2 and 3 were parametrized to reproduce intermolecular energies for the canonical water dimer only using the A2 (model 2) and A1 (model 3) ABSs. All calculations for GEM* were performed with a modified pmemd version in the AMOEBA suite of programs [Case et al. (2005)].

The comparison of the QM reference for energies and forces calculated with GEM* models 2 and 3 showed that both models reproduce the total intermolecular interactions well. However, model 3 deviated in the forces due to the limited accuracy because of the small number of Gaussians in the ABS employed for model 3. In addition, both models were observed to produce significant errors in the intermolecular energies for random dimers and binding energies for different oligomers. These errors are due to the improper description of these two models to describe interactions between H atoms since both models were fitted only to the canonical water dimer.

Conversely, all results for model 1 showed good agreement with the QM references for dimers as well as larger clusters. The results show that a better parametrization can be obtained once a slightly larger data set that included different dimer orientations was considered. Recently Babin et al. have developed a novel water model parametrized only from QM data [Babin et al. (2012, 2013)] using results from 40000 dimers calculated at the CCSD/CBS level.

The performance of GEM* was tested by performing 100 MD steps in the NVE ensemble with a series of water boxes of increasing size (216, 512, 1024, 2048 and 4096 molecules). All MD calculations were done on a single Xeon X5550 CPU with 12 GB of memory at 300 K with an 8 Å cutoff for Van der Waals interactions, using the Beeman integrator, a 1 fs time step and a dipole tolerance for the SCF convergence of 10^{-6}. The calculation of the polarization with the induced dipoles was performed using the PME method with a B–spline order of 5 and a grid–size of 24. During the parametrization and testing, it was realized that the overlap integrals for both sets of ABSs employed (A1 and A2) tend to 0 at distances greater than 5 Å. Therefore, although it is also possible to perform the Overlap integrals in reciprocal space as described in [Cisneros et al. (2006b); Darden (2007)], the exchange integrals were evaluated only in direct space with the same cutoff as the van der Waals interactions (8 Å), or with a reduced cutoff of 6 Å.

Timings for all the tested systems are shown in Figure 1.2.. For comparison we performed the calculations for all cases with the Coulomb integrals evaluated completely in direct space (all Gaus-

sians set as compact), or by placing two Gaussians in the diffuse set and employing sPME or FFP with two different exchange cutoffs. As discussed previously, the evaluation of the Hermite Gaussians in reciprocal space requires significantly larger grids and B–spline orders [Cisneros et al. (2006b); Darden (2007)]. For the three smallest boxes the calculations are faster when all the integrals are evaluated in direct space. This is due to the large overhead for the FFTs due to the fine grids required for accurate evaluation of the energies and forces in reciprocal space. As the system grows in size, the calculation becomes faster by using the compact/diffuse density split method with PME. The smallest water box comprising 216 molecules takes 100 s for the evaluation of the 100 energy/force calculations including the evaluation of Coulomb and overlap integrals for 15120 basis functions. That is, our code is able to evaluate all Coulomb and overlap integrals for 15120 basis functions for a single step in 1 s. For the 4096 box, comprising 286720 basis functions, when the exchange cutoff is reduced to 6 Åthe time is 2363 s. Moreover, after an initial optimization of the code, this time is reduced to 789 s. This is only 4 times slower than the same 4096 system calculated with AMOEBA on the same CPU. We expect more gains in performance as the code is further optimized.

1.7 Combining SIBFA and GEM: S/G–1

In this initial implementation the direct coupling between GEM and SIBFA has been only performed at the induction level (polarization and charge transfer energies) in the spirit of QM/MM techniques [Chaudret et al. (2014)]. Indeed, the GEM equations for exchange–repulsion involve overlap integrals between densities of both interacting fragments. Thus a mixed S/G–1 scheme is not possible for this term since the overlap between a GEM density and SIBFA's multipoles would be zero. Therefore, in the present S/G–1 implementation, the electrostatic, exchange–repulsion and dispersion energies are computed at the sole SIBFA level and include SIBFA's short range corrections. In a forthcoming work the full multiscale implementation including full Gaussian electrostatics first order energy will be reported.

Therefore in the first implementation, GEM is only used to compute the second–order polarization and charge–transfer contributions between the cation and its bound ligands. Finally, the dispersion equations are the same for both methods, as they do not depend on electric fields or potentials. As a proof of concept, we limited ourselves to a Hartree–Fock level parametrization (no dispersion) of the method.

Briefly, within S/G–1, the evaluation of the polarization and charge–transfer energies use the same formalism. The differences between the two levels of computations are linked to the level of calculation of electric fields and potentials (i.e. using electronic densities for GEM or using distributed multipoles for SIBFA). Indeed, differences between GEM and SIBFA electric fields and electrostatic potentials can arise at short distances since GEM is identical to an ab initio field. Both procedures converge to the same solutions upon increasing the distances when the multipolar approximation starts to be valid as GEM densities act as a continuous electrostatic model. Therefore, in a similar spirit as in QM/MM approaches, specific fragments can be defined so as to be handled with either the GEM density or the SIBFA multipoles. For the first implementation, the use of GEM densities were restricted to the metal cations whereas the rest of molecules were described using SIBFA.

In order to try to include the different previously discussed physical effects within a MM scheme, we show here some results focusing on the polarization contribution in the case of the Ca(II)-H2O complex.

Figure 1.3. displays four curves, namely the reference ab initio CSOV polarization contribution, the undamped full GEM polarization energy, the full GEM + damping approach, and results obtained upon computing the polarization energy obtained with the exact ab initio undamped field values extracted from a quantum mechanical Gaussian 09 computation. The damping procedure is identical to the one used by SIBFA and is detailed in the technical appendix in ref. [Chaudret et al. (2014)].

It is important to point out that the GEM fields alone, in spite of their quasi-perfect match with their ab initio counterparts, do not provide a good reproduction of ab initio results at short-range:

a damping of the fields is required to gain accuracy at very short distances. Such conclusions are confirmed as our results obtained from undamped exact ab initio fields (i.e. computed using the original molecular orbitals) are basically identical to the undamped GEM results (see Figure I). This clearly shows that inclusion of short-range quantum effects, inherently present within QM and GEM fields, is not sufficient to reproduce the true polarization energy. This is because the final ab initio CSOV polarization energy embodies both penetration and exchange-polarization effects. The first quantity is present in GEM (as in QM), however, the exchange-polarization arises from the required orthogonalization of molecular orbitals of both Ca(II) and H2O fragments within the constrained self-consistent field procedure. Therefore, since GEM does not include this repulsive effect, the computed polarization energy is overestimated. A straightforward solution to the problem is to apply the exact same field damping procedure that is used for the SIBFA polarization contribution. As can be sene from Figure I, the GEM+damping approach accurately reproduces the CSOV reference by selectively including the different effects.

The initial implementation of the S/G–1 method has been developed to describe metal cations in a ligand environment. To this end, S/G–1 has been parametrized to model Zn(II) and Hg(II) cations. The Zn(II) parametrization was performed on a series of representative mono–ligand complexes and subsequently employed to calculate the polarization and charge–transfer energies for a series of multi–valent complexes as shown in Figure 1.4.. S/G–1 matches the anti–cooperative behavior of the polarization and charge–transfer (at the RVS level) energies. For example, for $[Zn(H_2O)_{4/2}]^{2+}$ the charge–transfer energy from S/G–1(RVS) corresponds to -26.8(-28.7) kcal/mol. Correspondingly, the polarization energy for S/G–1 is -103.8 kcal/mol compared to -101.7 kcal/mol for its RVS counterpart.

As a test of the applicability of S/G–1 to large systems, we applied this method to a Zn(II) dependent system recently studied using the SIBFA procedure. To this end, the Zn(II)–alcohol dehydrogenase (ADH) active site [de Courcy et al. (2008)] was calculated, GEM was used to model the Zn(II) cation and the remaining system was treated with SIBFA. As can be seen in Figure 1.4.,

S/G–1 successfully reproduced the RVS values for a complicated hetero–polyligated complex [Chaudret et al. (2014)].

To further test the performance of S/G–1 to model heavy metals where relativistic and correlation effects are important , a calibration for Hg(II) was performed. The cation polarization energy requires two components. The first arises from the cation dipolar polarizability and depends on the electrostatic field that the cation is subjected to. The second component arises from the quadrupolar polarizability and depends upon the field gradient [Buckingham (1975)]. The magnitude of the second component was found to be important in the case of some metal cations, such as Cu(I) and Hg(II), and this component had to be explicitly formulated in SIBFA [Gresh et al. (2002)]. However, although the reference quantum dipolar polarizability for Hg(II) can be easily obtained, its quadrupolar polarizability could not be derived by QM calculations using a small core pseudopotential. Therefore we resorted to theo available Cu(I) [Buckingham (1975)] value as a starting point. The values of the polarization energy for Hg(II) were obtained from RVS calculations on the $[Hg(H_2O)_2]^{2+}$ complex, with Hg(II) equidistant from the two water molecules. In this complex, the field undergone by Hg(II) is zero, but the gradient is non–zero. Subsequently, the Cu(I) quadrupolar polarizability was employed as a starting point for the Hg(II) and scaled to match the polarization energy of the complex.

Figure 1.5. shows the second–order polarization and charge–transfer energies and the total intermolecular interaction energies for a di–aquo mercury complex at different distances. Except at very short distances, the error between SIBFA/GEM with respect to the ab initio calculations are very small. The S/G–1 charge–transfer values are very close to one another. The main differences arise from the polarization contribution. Moreover, the use of the GEM densities for the calculation of the second–order components results in better agreement at short range with respect to RVS than the original SIBFA method. The close agreement found for the monoligated Hg(II) complex used for the parametrization is conserved in the polyligated complexes. This shows that the non-additivity of both the polarization and charge–transfer components

can be reproduced with S/G–1 for Hg(II) as it has been demonstrated for Zn(II).

1.8 Conclusion and Perspective

The present contribution reports the development of the ab initio GEM force field that uses Hermite Gaussian electrostatics to include short–range quantum mechanical effects into molecular mechanics. It also details the GEM* and S/G-1 approaches that blend together GEM with the AMOEBA and SIBFA polarizable potentials, both of which rely on distributed multipoles. As the capability to achieve high accuracy on the interaction energies is demonstrated, first application to molecular dynamics have been detailed as the potentialities of the GEM densities to accurately treat difficult systems such as metalloproteins have been exposed. In term of perspectives, all models should be able to be used directly in MD simulations as all the required gradients were recently coded. It should open the possibility of large–scale molecular dynamics using Gaussian Hermite functions as the models will benefit from recent advances in algorithmic and in hybrid MPI/OPEN–MP parallelism that use new scalable strategies with gains going from 2 to 3 orders in magnitude in time within the present framework [Lipparini et al. (2014)]. Overall, efforts will be devoted to propose a scalable integrate methodology incorporating both distributed Multipoles and Hermite Gaussian densities in popular packages such as Amber and Tinker.

1.9 Acknowledgments

Figure 1.1.   Total intermolecular interaction energy for the canonical water dimer calculated with the three GEM* models compared to ab initio and AMOEBA
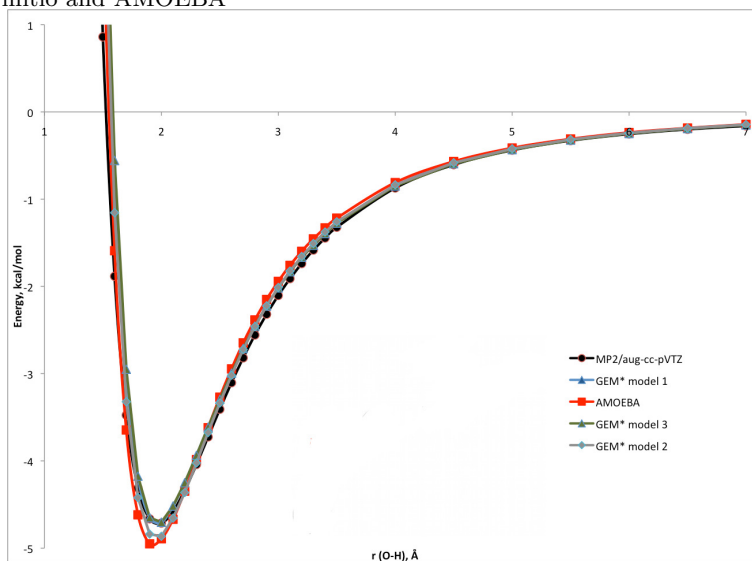

Figure 1.2.   Timings for different water boxes for GEM* using the model 1 parametrization. All times in seconds.
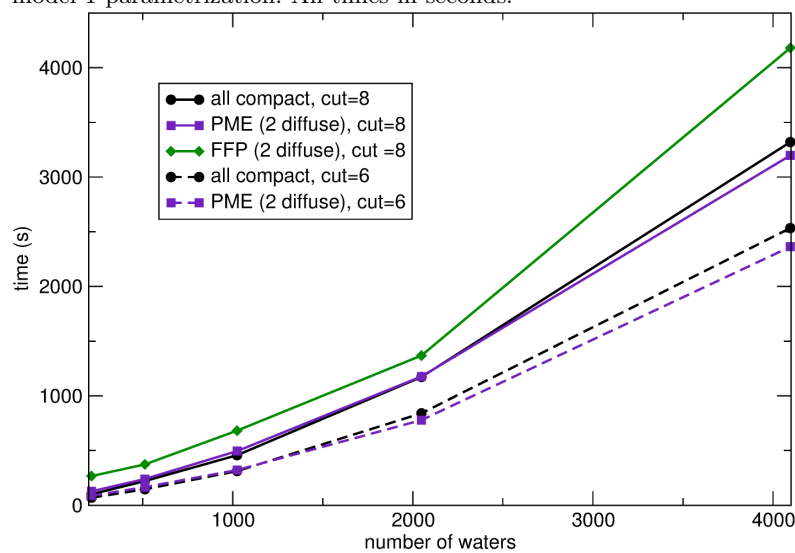
22 | Status of the GEM Force Field

Figure 1.3.   Ab initio polarization energies (kcal/mol) for the Ca(II)-H2O complex, computed using the CSOV procedure (blue), polarization energy computed using: i) distributed polarizabilities+ab-initio fields (grey),ii) distributed polarizabilities +GEM with (green) and iii) without fields damping (red).
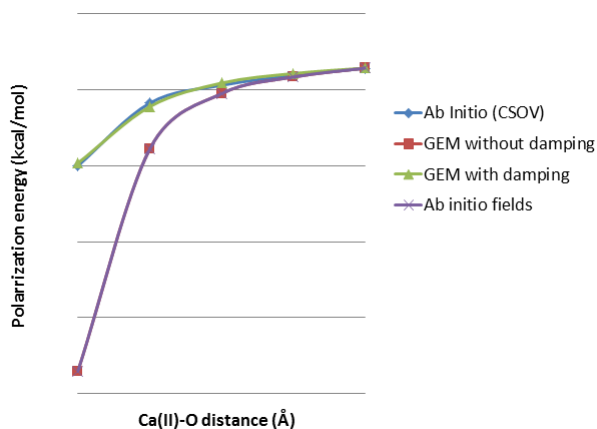
Figure 1.4. Polarization (A), charge–transfer (B) and second or-
der induction (C) (pol+c–t) energies in poly–ligated Zn(II) com-
plexes calculated with RVS and S/G–1. The RVS charge–transfer
and induction energies are BSSE corrected. The complexes are as
follows: 1: $[Zn(CH_3S)_3]^-$, 2: $[Zn(CH_3S)_4]^{2-}$, 3: $[Zn(imidazole)_3]^{2+}$,
4: $[Zn(imidazole)_4]^{2+}$, 5: $[Zn(H_2O)_6]^{2+}$, 6: $[Zn(H_2O)_{5/1}]^{2+}$, 7:
$[Zn(H_2O)_{4/2}]^{2+}$, 8: cluster model for alcohol dehydrogenase active site

24   Status of the GEM Force Field
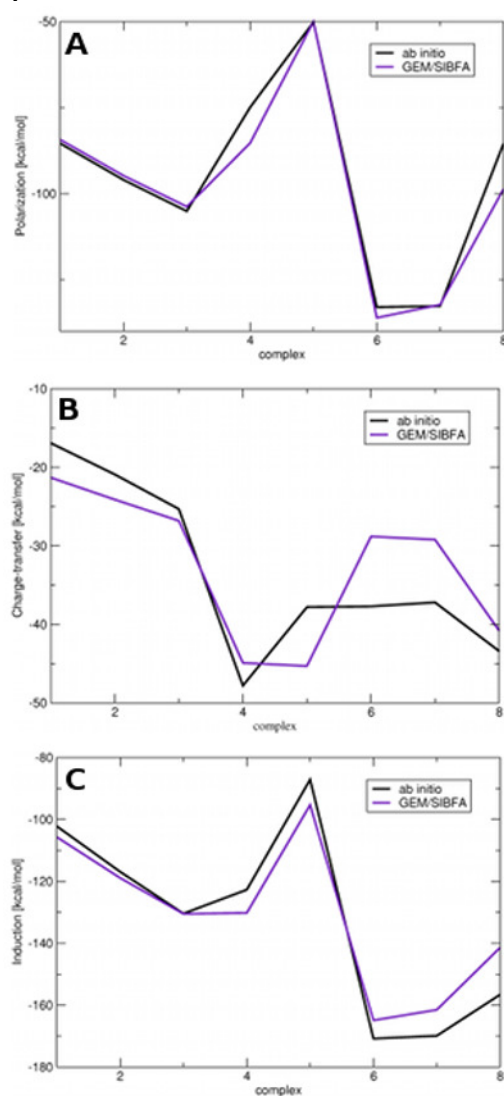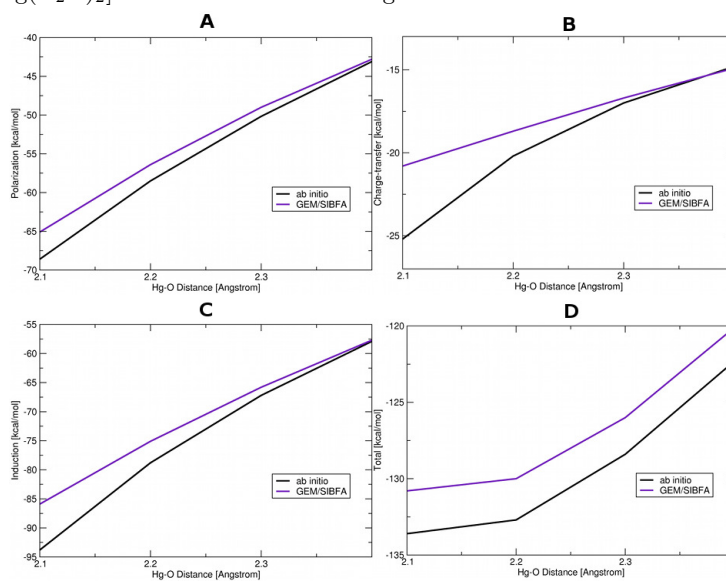
Figure 1.5.   Polarization (A), charge–transfer (B), second order induction (C), and total intermolecular interaction energy (D) for $[Hg(H_2O)_2]^{2+}$ as a function of the Hg–O distances.

Bibliography

Andzelm, J. and Wimmer, E. (1992). Density functional gaussian–type–orbital approach to molecular geometries, vibrations and reaction energies, J. Chem. Phys. 96, pp. 1280–1303.

Babin, V., Leforestier, C. and Paesani, F. (2013). Development of a first principles water potential with flexible monomers: Dimer potential energy surface, vrt spectrum, and second virial coefficient, J. Chem. Theo. Comp. 9, 12, pp. 5395–5403, doi:10.1021/ct400863t, URL `http://pubs.acs.org/doi/abs/10.1021/ct400863t`.

Babin, V., Medders, G. R. and Paesani, F. (2012). Toward a universal water model: First principles simulations from the dimer to the liquid phase, J. Phys. Chem. Lett. 3, 24, pp. 3765–3769, doi:10.1021/jz3017733, URL `http://pubs.acs.org/doi/abs/10.1021/jz3017733`.

Babin, V., Medders, G. R. and Paesani, F. (2014). Development of a first principles water potential with flexible monomers. ii: Trimer potential energy surface, third virial coefficient, and small clusters, J. Chem. Theo. Comp. 10, 4, pp. 1599–1607.

Bagus, P. S., Hermann, K. and C. W. Bauschlicher Jr. (1984). A new analysis of charge transfer and polarization for ligand–metal bonding: model studies for al$_4$co and al$_4$nh$_3$, J. Chem. Phys. 80, pp. 4378–4386.

Bayly, C. I., Cieplak, P., Cornell, W. D. and Kollman, P. A. (1993). A well–behaved electrostatic potential base method using charge restraints for deriving atomic charges: the resp method, J. Phys. Chem. 97, pp. 10269–10280.

Becke, A. D. (1988). A multicenter numerical integration scheme for polyatomic molecules, J. Chem. Phys. 88, 4, pp. 2547–2553.

Böttcher, C. (1993). Theory of Electric Polarization (Elsevier, Amsterdam).

Boys, S. F. and Shavit, I. (1959). A Fundamental Calculation of the Energy Surface for the System of Three Hydrogen Atoms (AD212985, NTIS, Springfield, VA).

Buckingham, A. (1975). multipolar expansion, Phil. Trahs. Roy. Soc. (London) B 272, p. 5.

26 | Bibliography

Case, D. A., T.E. Cheatham III, Darden, T. A., Gohlke, H., Luo,
    R., K.M. Merz Jr., Onufirev, A., Simmerling, C., Wang, B. and
    Woods, R. J. (2005). The amber biomolecular simulation pro-
    grams, J. Comput. Chem. 26, pp. 1668–1688.
Challacombe, M., Schwgler, E. and Almlöf, J. (1996). "Modern devel-
    opments in Hartree–Fock theory: Fast methods for computing the
    Coulomb matrix" in Computational Chemistry: Review of Current
    Trends (World Scientific Inc., Singapore).
Chaudret, R., Nohad Gresh, O. P., Darden, T. A., Cisneros, G. A.
    and Piquemal, J.-P. (2014). Towards improved treatment of metal
    cations in polarizable molecular mechanics using the hybrid gaus-
    sian electrostatics / distributed multipoles gem/sibfa approach, J.
    Chem. Theo. Comp. Submitted.
Cisneros, G. A. (2012). Application of gaussian electrostatic model
    (gem) distributed multipoles in the amoeba force field, J. Chem.
    Theo. Comp. 12, pp. 5072–5080, URL http://pubs.acs.org/doi/
    abs/10.1021/ct300630u.
Cisneros, G. A., Elking, D. M., Piquemal, J.-P. and Darden, T. A.
    (2007). Numerical fitting of molecular properties to hermite gaus-
    sians, J. Phys. Chem. A 111, pp. 12049–12056.
Cisneros, G. A., Piquemal, J. P. and Darden, T. A. (2005a). Intermolec-
    ular electrostatic energies using density fitting, J. Chem. Phys.
    123, p. 044109.
Cisneros, G. A., Piquemal, J.-P. and Darden, T. A. (2005b). Intermolec-
    ular electrostatic energies using density fitting, J. Chem. Phys.
    123, p. 044109.
Cisneros, G. A., Piquemal, J.-P. and Darden, T. A. (2006a). Gener-
    alization of the gaussian electrostatic model: extension to arbi-
    trary angular momentum, distributed multipoles and computa-
    tional speedup with reciprocal space methods, J. Chem. Phys.
    125, p. 184101.
Cisneros, G. A., Piquemal, J. P. and Darden, T. A. (2006b). Generali-
    ization of the gaussian electrostatic model: Extension to arbitrary
    angular momentum, distributed multipoles, and speedup with re-
    ciprocal space methods, J. Chem. Phys. 125, p. 184101.
Cisneros, G. A., Tholander, S. N.-I., Parisel, O., Darden, T. A., Elking,
    D., Perera, L. and Piquemal, J.-P. (2008). Simple formulas for im-
    proved point–charge electrostatics in classical force fields and hy-
    brid quantum mechanical/molecular mechanical embedding, Int.
    J. Quantum Chem. 108, pp. 1905–1912.
Coppens, P. and Volkov, A. (2004). The interplay between experiment
    and theory in charge–density analysis, Acta Cryst. A 60, 5, pp.
    357–364.

Darden, T. A. (2007). "Dual bases in crystallographic computing" in International Tables of Chrystallography, Vol. B (Kluwer Academic Publishers, Dordrecht, The Netherlands).

Day, P. N., Jensen, J. H., Gordon, M. S., Webb, S. P., Stevens, W. J., Krauss, M., Garmer, D., Basch, H. and Cohen, D. (1996). An effective fragment method for modeling solvent effects in quantum mechanical calculations, J. Chem. Phys. 105, pp. 1968–1986.

de Courcy, B., Piquemal, J.-P. and Gresh, N. (2008). Energy analysis of zn polycoordination in a metalloprotein environment and of the role of a neighboring aromatic residue. what is the impact of polarization? J. Chem. Theo. Comp. 4, 10, pp. 1659–1668.

Domene, C., Fowler, P. W., Wilson, M., Madden, P. and Wheatley, R. J. (2001). Overlap–model and ab initio cluster calculations of ion properties in distorted environments, Chem. Phys. Lett. 333, pp. 403–412.

Duke, R. E., Starovoytov, O. N., Piquemal, J.-P. and Cisneros, G. A. (2014). Gem*: A molecular electronic density–based force field for classical molecular dynamics simulations, J. Chem. Theo. Comp. 10, pp. 1361–1365.

Dunlap, B. I., Connolly, J. W. D. and Sabin, J. R. (1979). On first–row diatomic molecules and local density models, J. Chem. Phys. 71, pp. 4993–4999.

Eckhardt, C. J. and Gavezzotti, A. (2007). Computer simulations and analysis of structural and energetic features of some crystalline energetic materials, J. Phys. Chem. B 111, 13, pp. 3430–3437.

Eichkorn, K., Treutler, O., Öhm, H., Häser, M. and Ahlrichs, R. (1995). Auxiliary basis sets to approximate coulomb potentials, Chem. Phys. Lett. 240, pp. 283–290.

Eisenschitz, R. and London, F. (1930). Perturbation theory, Z. Phys. 60, pp. 491–527.

Elking, D. M., Cisneros, G. A., Piquemal, J.-P., Darden, T. A. and Pedersen, L. G. (2010). Gaussian multipole model (gmm), J. Chem. Theo. Comp. 6, pp. 190–202.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. and Pedersen, L. G. (1995). A smooth particle mesh Ewald method, J. Chem. Phys. 103, pp. 8577 – 8593.

Ewald, P. (1921). Die Berechnung optischer und elektrostatischer Gitterpotentiale, Ann. Phys. 64, pp. 253 – 287.

Freitag, M. A., Gordon, M. S., Jensen, J. H. and Stevens, W. J. (2000). Evaluation of charge penetration between distributed multipolar expansions, J. Chem. Phys. 112, pp. 7300–7306.

Gavezzotti, A. (2002). Calculation of intermolecular interaction energies by direct numerical integration over electron densities i. electrostatic and polarization energies in molecular crystals, J. Phys.

28   Bibliography

Chem. B 106, pp. 4145–4154.

Giese, T. J., Chen, H., Dissanayake, T., Giambau, G. M., Heldenbrand, H., Huang, M., Kuechler, E. R., Lee, T.-S., Panteva, M. T., Radak, B. K. and York, D. M. (2013). A variational linear-scaling framework to build practical, efficient next-generation orbital-based quantum force fields, J. Chem. Theo. Comp. 9, 3, pp. 1417–1427.

Giese, T. J., Chen, H., Huang, M. and York, D. M. (2014). Parametrization of an orbital-based linear-scaling quantum force field for noncovalent interactions, J. Chem. Theo. Comp. 10, 3, pp. 1086–1098.

Glendening, E. D. (1994). Natural energy decomposition analysis: explicit evaluation of electrostatic and polarization effects with application to aqueous clusters of alkali metal cations and neutrals, J. Am. Chem. Soc. 118, pp. 2473–2482.

Glendening, E. D. and Streitwieser, A. (1994). Natural energy decomposition analysis: An energy partitioning procedure for molecular interactions with application to weak hydrogen bonding strong ionic, and moderate donor–acceptor interactions, J. Chem. Phys. 100, pp. 2900–2909.

Godbout, N. and Andzelm, J. (1999). DGauss Version 2.0, 2.1, 2.3, 4.0: the file that contains the A1, A2 and P1 auxiliary basis sets can be obtained from the CCL WWW site at http://www.ccl.net/cca/data/basis-sets/DGauss/basis.v3.html (Computational Chemistry List, Ltd., Ohio).

Gresh, N., Cisneros, G. A., Darden, T. A. and Piquemal, J.-P. (2007). Anisotropic, polarizable molecular mechanics studies of inter–, intra–molecular interactios, and ligand–macromolecule complexes. a bottom–up strategy, J. Chem. Theo. Comp. 3, pp. 1960–1986.

Gresh, N., Claverie, P. and Pullman, A. (1979). Sibfa, Int. J. Quantum Chem. , p. 253Symp 11.

Gresh, N., Policar, C. and Giessner-Prettre, C. (2002). Modeling copper(i) complexes: Sibfa molecular mechanics versus ab initio energetics and geometrical arrangements, J. Phys. Chem. A 106, 23, pp. 5660–5670.

Handley, C. M. and Popelier, P. L. A. (2010). Potential energy surfaces fitted by artificial neural networks, J. Phys. Chem. A 114, 10, pp. 3371–3383, doi:10.1021/jp9105585.

Hermida-Ramón, J. M., Brdarski, S., Karlström, G. and Berg, U. (2003). Inter- and intramolecular potential for the n-formylglycinamide-water system. a comparison between theoretical modeling and empirical force fields, J. Comput. Chem. 24, 2, pp. 161–176.

Heßelmann, A., Jansen, G. and Schütz, M. (2005). Dft–sapt with density fitting: a new efficient method to study intermolecular interaction energies, J. Chem. Phys. 122, pp. 14103–14120.

Hirshfelder, J. O. (1967a). Perturbation theory for exchange forces, i, Chem. Phys. Lett. 1, pp. 325–329.

Hirshfelder, J. O. (1967b). Perturbation theory for exchange forces, ii, Chem. Phys. Lett. 1, pp. 363–368.

Hu, H., Lu, Z. and Yang, W. (2007). Fitting molecular electrostatic potentials from quantum mechanical calculations, J. Chem. Theo. Comp. 3, pp. 1004–1013.

Jeziorski, B., Moszynski, R. and Szalewicz, K. (1994). Perturbation theory approach to intermolecular potential energy surfaces of van der waals complexes, Chem. Rev. 94, pp. 1887–1930.

Jung, Y., Sodt, A., P. M. W. Gill and Head-Gordon, M. (2005). Auxiliary basis expansions for large scale electronic structure calculations, Proc. Natl. Acad. Sci. 102, pp. 6692–6697.

Kairys, V. and Jensen, J. H. (1999). Evaluation of the charge penetration energy between non-orthogonal molecular orbitals using the spherical gaussian overlap approximation, Chem. Phys. Lett. 315, 1-2, pp. 140–144.

Khaliullin, R. Z., Head-Gordon, M. and Bell, A. T. (2006). An efficient self-consistent field method for large systems of weakly interacting components, J. Chem. Phys. 124, 20, 204105.

Kitaura, K. and Morokuma, K. (1976). A new energy decomposition scheme for molecular interactions within the hartree-fock approximation, Int. J. Quantum Chem. 10, pp. 325–340.

Kosov, D. S. and Popelier, P. L. A. (2000). Atomic partitioning of molecular electrostatic potentials, J. Phys. Chem. A 104, pp. 7339 – 7345.

Köster, A. M. (1996). Efficient recursive computation of molecular integrals for density functional methods, J. Chem. Phys. 104, pp. 4114–4124.

Köster, A. M., Calaminici, P., Gómez, Z. and Reveles, U. (2002). "Density functional theory calculation of transition metal clusters", in Reviews of Modern Quantum Chemistry, A Celebration of the Contribution of Robert G. Parr (World Scientific, Singapore).

Lipparini, F., Lagardère, L., Stamm, B., Cancès, E., Schnieders, M., Ren, P., Maday, Y. and Piquemal, J.-P. (2014). Scalable evaluation of polarization energy and associated forces in polarizable molecular dynamics: I. toward massively parallel direct space computations, J. Chem. Theo. Comp. 10, 4, pp. 1638–1651.

Lu, Z., Zhou, N., Wu, Q. and Zhang, Y. (2011). Directional dependence of hydrogen bonds: A density-based energy decomposition analysis and its implications on force field development, J. Chem. Theo. Comp. 7, 12, pp. 4038–4049.

McDaniel, J. G. and Schmidt, J. R. (2014). First-principles many-body force fields from the gas phase to liquid: A universal approach, J.

Phys. Chem. B 0, 0, p. null, doi:10.1021/jp501128w.

McMurchie, L. and Davidson, E. (1978). One- and two-electron integrals over cartesian gaussian functions, J. Comput. Phys 26, pp. 218 – 231.

Mills, M. and Popelier, P. (2012). Polarisable multipolar electrostatics from the machine learning method kriging: an application to alanine, Theo. Chem. Acc. 131, 3, pp. 1–16, doi:10.1007/s00214-012-1137-7.

Misquitta, A. J. and Stone, A. J. (2006). Distributed polarizabilities obtained using a constrained density–fitting algorithm, J. Chem. Phys. 124, p. 024111.

Mo, Y., Gao, J. and Peyerimhoff, S. D. (2000). Energy decomposition analysis of intermolecular interactions using a block-localized wave function approach, J. Chem. Phys. 112, 13, pp. 5530–5538.

Murrel, J. N. and Shaw, G. (1967). Intermolecular forces in the region of small orbital overlap, J. Chem. Phys. 46, pp. 1768–1772.

Paricaud, P., Predota, M., Chialvo, A. A. and Cummings, P. T. (2005). From dimer to condensed phases at extreme conditions: Accurate predictions of the properties of water by a gaussian charge polarizable model, J. Chem. Phys. 122, 24, p. 244511.

Piquemal, J.-P., Chevreau, H. and Gresh, N. (2007). Toward a separate reproduction of the contributions to the hartree–fock and dft intermolecular energies by polarizable molecular mechanics with the sibfa potential, J. Chem. Theo. Comp. 3, pp. 824–837.

Piquemal, J. P., Cisneros, G. A., Reinhardt, P., Gresh, N. and Darden, T. A. (2006a). Towards a force field based on density fitting, J. Chem. Phys. 124, p. 104101.

Piquemal, J.-P., Cisneros, G. A., Reinhardt, P., Gresh, N. and Darden, T. A. (2006b). Towards a force field based on density fitting, J. Chem. Phys. 124, p. 104101.

Piquemal, J.-P., Gresh, N. and Giessner-Prettre, C. (2003). Improved formulas for the calculation of the electrostatic contribution to the intermolecular interaction energy from multipolar expansion of the electronic distribution, J. Phys. Chem. A 107, pp. 10353–10359.

Piquemal, J.-P., Marquez, A., Parisel, O. and 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, pp. 1052–1062.

Podeszwa, R., Bukowski, R. and Szalewicz, K. (2006). Density-fitting method in symmetry–adapted perturbation theory based on kohn–sham description of monomers, J. Chem. Theo. Comp. 2, pp. 400–412.

Ponder, J. W., Wu, C., Ren, P., Pande, V. S., Chodera, J. D., Schnieders, M. J., Haque, I., Mobley, D. L., Lambrecht, D. S., Robert A. DiStasio Jr., Head-Gordon, M., Clark, G. N. I., Johnson, M. E. and Head-Gordon, T. (2010). Current status of the amoeba polarizable force field, J. Phys. Chem. B 114, pp. 2549–2564.

Popelier, P. (2000). Atoms in Molecules: An Introduction (Prentice Hall, Harlow, England).

Popelier, P. (2012). A generic force field based on quantum chemical topology, in C. Gatti and P. Macchi (eds.), Modern Charge-Density Analysis (Springer Netherlands), ISBN 978-90-481-3835-7, pp. 505–526, doi:10.1007/978-90-481-3836-4˙14.

Popelier, P. L. A., Joubert, L. and Kosov, D. S. (2001a). Convergence of the electrostatic interaction based on topological atoms, J. Phys. Chem. A 105, pp. 8254 – 8261.

Popelier, P. L. A., Joubert, L. and Kosov, D. S. (2001b). The convergence of the electrostatic interaction based on topological atoms, J. Phys. Chem. A 105, pp. 8524–8261.

Popelier, P. L. A. and Kosov, D. S. (2001). Atomatom partitioning of intramolecular and intermolecular coulomb energy, J. Chem. Phys. 114, pp. 6539 – 6547.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). Numerical recipes in fortran77; the art of scientific computing, 2nd Ed. (Cambridge University Press, New York, NY).

Price, S. (1999). Toward more accurate model intermolecular potentials for organic molecules, in K. Lipkowitz and D. B. Boyd (eds.), Reviews in Computational Chemistry, Vol. 14 (VCH Publishers, New York, NY), pp. 225–289.

Raval, A., Piana, S., Eastwood, M. P., Dror, R. O. and Shaw, D. E. (2012). Refinement of protein structure homology models via long, all-atom molecular dynamics simulations, Prot. Struct. Func. Bioinf. 80, 8, pp. 2071–2079, doi:10.1002/prot.24098, URL http://dx.doi.org/10.1002/prot.24098.

Ren, P. and Ponder, J. W. (2002). A consistent treatment of inter– and intramolecular polarization in molecular mechanics calculations, J. Comput. Chem. 23, pp. 1497–1506.

Ren, P. and Ponder, J. W. (2003). Polarizable atomic multipole water model for molecular mechanics simulation, J. Phys. Chem. B 107, pp. 5933–5947.

Ren, P., Wu, C. and Ponder, J. W. (2011). Polarizable atomic multipole-based molecular mechanics for organic molecules, J. Chem. Theo. Comp. 7, 10, pp. 3143–3161, doi:10.1021/ct200304d.

Sagui, C., Pedersen, L. G. and Darden, T. A. (2004). Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular

32 | Bibliography

   simulations, J. Chem. Phys. 120, pp. 73 – 87.

Schulten, K., Phillips, J. C., Kale, L. V. and Bhatele, A. (2008). Biomolecular modeling in the era of petascale computing, in D. Bader (ed.), Petascale Computing: Algorithms and Applications (Chapman & Hall / CRC Press), pp. 165–181.

Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., Bank, J. A., Jumper, J. M., Salmon, J. K., Shan, Y. and Wriggers, W. (2010). Atomic-Level Characterization of the Structural Dynamics of Proteins, Science 330, 6002, pp. 341–346.

Stevens, W. J. and Fink, W. H. (1987). Frozen fragment reduced variational space analysis of hydrogen bonding interactions. applications to the water dimer, Chem. Phys. Lett. 139, pp. 15–22.

Stone, A. J. (2000). The theory of intermolecular forces (Oxford University Press, Oxford, UK).

Stone, A. J. (2005). Distributed multipole analysis: Stability for large basis sets, J. Chem. Theo. Comp. 1, pp. 1128–1132.

Stone, A. J. (2011). Electrostatic damping functions and the penetration energy, J. Phys. Chem. A 115, 25, pp. 7017–7027.

Stone, J. E., Phillips, J. C., Freddolino, P. L., Hardy, D. J., Trabuco, L. G. and Schulten, K. (2007). Accelerating molecular modeling applications with graphics processors, J. Comput. Chem. 28, 16, pp. 2618–2640, doi:{10.1002/jcc.20829}.

Temelso, B., Archer, K. A. and Shields, G. C. (2011). Benchmark structures and binding energies of small water clusters with anharmonicity corrections, J. Phys. Chem. A 115, 43, pp. 12034–12046.

Toukmaji, A., Sagui, C., Board, J. A. and Darden, T. (2000). Efficient PME-based approach to fixed and induced dipolar interactions, J. Chem. Phys. 113, pp. 10913 – 10927.

Volkov, A. and Coppens, P. (2004). Calculation of electrostatic interaction energies in molecular dimers from atomic multipole moments obtained by different methods of electron density partitioning, J. Comput. Chem. 25, pp. 921–934.

Wang, B. and Truhlar, D. G. (2010). Including charge penetration effects in molecular modeling, J. Chem. Theo. Comp. 6, 11, pp. 3330–3342.

Wheatley, R. (2011). Gaussian multipole functions for describing molecular charge distributions, Mol. Phys. 7, 3, pp. 761–777, doi: 10.1021/ct100530r.

Wheatley, R. J. and Price, S. L. (1990). An overlap model for estimating the anisotropy of repulsion, Mol. Phys. 69, pp. 507–533.

Wu, Q., Ayers, P. W. and Zhang, Y. (2009). Density-based energy decomposition analysis for intermolecular interactions with variationally determined intermediate state energies, J. Chem. Phys. 131, 16, 164112.

Xie, W. and Gao, J. (2007). The design of a next generation force field:
        The x-pol potential, J. Chem. Theo. Comp. 3, 6, pp. 1890–1900.
Xie, W., Orozco, M., Truhlar, D. G. and Gao, J. (2009). X-pol potential:
        An electronic structure-based force field for molecular dynamics
        simulation of a solvated protein in water, J. Chem. Theo. Comp.
        5, 3, pp. 459–467.
York, D. and Yang, W. (1994). The Fast Fourier Poisson (FFP) method
        for calculationg Ewald sums, J. Chem. Phys. 101, pp. 3298 – 3300.