Chapter 2

A Modified Divide-and-Conquer Linear-Scaling Quantum Force Field with Multipolar Charge Densities

Timothy J. Giese and Darrin M. York

Center for Integrative Proteomics Research, BioMaPS Institute for Quantitative Biology and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854-8087, USA york@biomaps.rutgers.edu

2.1 Introduction

Recent advances in biomolecular modeling have emphasized the importance of inclusion of explicit electronic polarizabilty, and a description of electrostatic interactions that includes atomic multipoles; however, these additional levels of treatment necessarily increase a model's computational cost. Ultimately, the decision as to whether inclusion of these more rigorous levels are justified rests on the degree to which they impact the specific application areas of interest, balanced with the overhead of their computational cost. The purpose of this book is to stimulate the exchange of effective

Many-Body Effects and Electrostatics in Biomolecules

Edited by Qiang Cui, Markus Meuwly, and Pengyu Ren

Copyright © 2015 Pan Stanford Publishing Pte. Ltd.

ISBN 978-981-4613-92-7 (Hardcover), 978-981-4613-93-4 (eBook)

www.panstanford.com

strategies used to describe many-body effects and electrostatics across the quantum, classical, and coarse-grained modeling regimes.

In this chapter, we describe a linear-scaling quantum force field based on a modified divide-and-conquer (mDC) procedure and discuss the practical consequence of including (or excluding) multipolar electrostatic interactions with a few illustrative examples. These observations are then used to rationalize some of the hydrogen bond geometries produced by other models, including the standard DFTB3 semiempirical Hamiltonian, which include multipoles within its tight-binding matrix elements but limit the second-order electrostatic interactions to monopoles. Furthermore, we assess the ability of a recent mDC parametrization to reproduce nucleobase dimer binding energies relative to highlevel ab initio calculations and we compare nucleobase trimer formation enthalpies to experimental estimates.

The description of the mDC method in the present work is supplemented with mathematical details that we Have used to introduce multipolar densities efficiently into the model. In particular, we describe the mathematics needed to construct atomic multipole expansions from atomic orbitals (AOs) and interact the expansions with point-multipole and Gaussian-multipole functions. With that goal, we present the key elements required to use the spherical tensor gradient operator (STGO) and the real-valued solid harmonics; perform multipole translations for use in the Fast Multipole Method (FMM); electrostatically interact point-multipole expansions; interact Gaussian-multipoles in a manner suitable for real-space Particle Mesh Ewald (PME) corrections; and we list the relevant real-valued spherical harmonic Gaunt coefficients for the expansion of AO product densities into atom-centered multipoles.

Section 2.2 discusses the obstacles encountered in producing a linear-scaling quantum force field and the methods used to overcome them. The linear-scaling quantum force field energy is described in Section 2.3. Section 2.4 discusses the consequences of including higher-order multipoles into the model and assess the quality of the mDC method in reproducing nucleobase interactions. The mathematical details used in the mDC model are collected into a series of small appendices at the end of the chapter (Sections 2.6.1-2.6.5) to facilitate the narrative.

Linear-scaling quantum force fields 57

2.2 Linear-scaling quantum force fields

Conventional ab initio electronic structure methods have computational/memory requirements that scale non-linearly (typically N^3/N^2 or higher) with number of particles. This restricts the size of the systems that these methods can be applied. There is a rich literature associated with the development of electronic structure methods that scale "linearly" with system size, both at the ab initio and semiempirical levels, that allow them to be extended to very large systems [22]. These methods have traditionally involved introduction of carefully chosen approximations that allow reformulation of the equations so that computation can be achieved with computational cost and memory requirements that increase in linear proportion to the number of particles and size of the system. By adjustment of control parameters, these methods can be made to systematically converge to the full non-linear scaling result. The simplest and most widely applied linear-scaling electronic structure methods are based on single-determinant wave function methods such as Hartree-Fock Kohn-Sham density-functional theory or semiempirical/tight-binding models. With these classes of methods, the most critical challenge involves circumventing the need for a globally orthonormal set of molecular orbitals (MOs) or, equivalently, an exactly idempotent single-particle density-matrix.

A "linear-scaling quantum force field" is a model that abandons the goal of being able to recover the full nonlinear quantum result, but instead takes recourse into additional layers of empiricism to achieve much greater efficiency and even higher accuracy. Typically, these force fields invoke a construct whereby a large system is divided into predetermined localized fragments (or residues), and different models may be employed for intra- and inter-residue interactions. One strategy has been to develop electron density-based quantum force fields [8, 34] that do not require the construction or orthogonalization of molecular orbitals (MOs). While this class of force fields has demonstrated considerable promise for molecular simulations, it has limitations in its ability to model reactive chemical processes involving formation and cleavage of chemical bonds. A different strategy, which we have taken here, involves using

localized MOs to describe intra-residue interactions, and empirical density-based models to describe inter-residue interactions. This framework borrows ideas from "orbital-free" density-functional methods [48, 49, 56], but with the added element that non-bonded inter-residue interactions are much simpler and empirically fine-tuned to obtain accuracy required for biological applications. There are a number of recent models that have used this strategy. The X-Pol method replaces the explicit inter-residue orbital coupling with empirical Lennard–Jones or Buckingham potentials [9, 14, 26, 45, 50–52, 55] or through perturbative corrections [7, 27]. In Ref. [16], we used a charge-dependent density-overlap van der Waals model as means of combining the density- and MO-based quantum force field strategies.

All quantum force fields, regardless of the specific form of the intra- and inter-residue interactions, involve long-range electrostatic interactions that must be computed with linear-scaling methods. The two most common linear-scaling methods for electrostatic interactions are tree codes and FMMs [2, 20, 23, 46] for non-periodic systems, and linear-scaling Ewald methods such as PME [10, 12, 36] for periodic systems. In brief, FMM is founded upon the physical interpretation of the Laplace expansion of the Coulomb kernel [Eqs. (2.36)–(2.37)], i.e., the Taylor series expansion of 1/r: If each of two charge densities are circumscribed by non-intersecting spheres, then the Coulomb interaction between the two densities can be computed from a single point-multipole interaction between the sphere centers. Linear-scaling is achieved by introducing hierarchy, i.e., the system is recursively divided, the multipole moments of a region are computed from the moments of its children, and the electrostatic interaction is performed at the most "ancient level" possible. Linear-scaling Ewald methods, on the other hand, split the electrostatic interactions into a short-ranged "direct-space" [see e.g., Eq. (2.62) and surrounding discussion] and long-ranged "reciprocal-space" components, the former which can be computed using a distance cut-off, and the latter computed efficiently with $O[N \log(N)]$ computational scaling using Fast Fourier Transforms.

In the Methods section that follows, we describe an mDC method that is based on the DFTB3 Hamiltonian [15] and uses a simple Lennard–Jones model for the non-electrostatic non-bonded inter-

Methods 59

actions between residues. However, unlike the DFTB3 Hamiltonian, which uses a monopole approximation in its treatment of secondorder electrostatics, the mDC model employs atomic multipoles to compute the inter-region interactions. As demonstrated in the Results and Discussion section, the use of a multipolar representation for electrostatics is key for obtaining robust, hydrogen bond angles, and hydrogen bond and base stacking interactions for nucleobase dimers and trimers. The appendices contain further key technical details that are needed for implementation of the mDC method with linear-scaling electrostatic methods such as FMM and PME.

2.3 Methods

2.3.1 The Modified Divide-and-Conquer Method

The mDC total energy is a sum of fragment ab initio energies E_A which we compute using the DFTB3 Hamiltonian [15]; the inter-fragment multipolar electrostatics (second term); the inter-fragment Lennard–Jones (LJ) interactions (third term); and MM bond energies E_{bonded} (fourth term) for those bonds, angles, and torsions that cross fragment boundaries

$$E = \sum_{A} E_{A}(\mathbf{C}_{A}^{\sigma}; \mathbf{R}_{A}) + \frac{1}{2} \sum_{\substack{a \\ l_{\mu \in a}}} q_{l\mu} p_{l\mu} + \sum_{b>a}' E_{LJ}(R_{ab}) + E_{bonded}(\mathbf{R}).$$

$$(2.1)$$

 \mathbf{C}_{A}^{σ} are the σ -spin MO coefficients for the *A*'th fragment, and \mathbf{R}_{A} are the nuclear positions of the atoms in fragment *A*.

$$q_{l\mu\in a} = Z_a \delta_{l0} \delta_{\mu 0} - \int \rho_a(\mathbf{r}) \mathcal{C}_{l\mu} (\mathbf{r} - \mathbf{R}_a) d^3 r \qquad (2.2)$$

are atomic multipole moments on atom *a*, $C_{l\mu}(\mathbf{r})$ is a real regular solid harmonic [Eq. (2.43)], $\rho_a(\mathbf{r})$ is an atom-partitioned density, Z_a is a nuclear charge, and

$$p_{l\mu\in a} = \sum_{\substack{b\neq a\\j\kappa\in b}}' q_{j\kappa} \frac{C_{l\mu}(\nabla_a)}{(2l-1)!!} \frac{C_{j\kappa}(\nabla_b)}{(2j-1)!!} \frac{1}{R_{ab}}$$
(2.3)

is a "multipolar potential," i.e., the derivative of the interaction with respect to a multipole moment. The primed summations

indicate that intrafragment electrostatics are excluded because those Coulomb interactions are already considered in the ab initio calculation of E_A . $C_{l\mu}(\nabla_a)$ is a real-valued STGO acting on the coordinates of atom *a* (see Sec. 2.6.1). The expressions used to evaluate Eq. (2.3) are provided in Sec. 2.6.4. E_{bonded} includes corrections for those bonds, angles, and dihedrals that cross the boundary between two covalently bonded fragments; however, the present work will consider nonbonded interactions exclusively.

As discussed in the previous section, the relaxation of the orthonormality constraints allows one to solve for the MO coefficients through a series of small generalized eigenvalue problems (proportional to the size of a fragment)

$$\mathbf{F}_{A}^{\sigma} \cdot \mathbf{C}_{A}^{\sigma} = \mathbf{S}_{A} \cdot \mathbf{C}_{A}^{\sigma} \cdot \mathbf{E}_{A}^{\sigma}.$$
(2.4)

The inter-fragment coupling occurs through the interaction of their atomic multipoles which are determined from the fragment electron densities within the self-consistent-field (SCF) procedure. The σ -spin Fock matrix for region A with inclusion of this coupling is

$$F_{A,ij}^{\sigma} = \frac{\partial E_A}{\partial P_{A,ij}^{\sigma}} \bigg|_{\mathbf{q},\mathbf{p},\mathbf{R}} + \sum_{\substack{a \in A \\ l_{\mu \in a}}} p_{l\mu} \left| \frac{\partial q_{l\mu}}{\partial P_{A,ij}^{\sigma}} \right|_{\mathbf{p},\mathbf{R}},$$
(2.5)

where

$$P_{A,ij}^{\sigma} = \sum_{k} n_{A,k}^{\sigma} C_{A,ik}^{\sigma} C_{A,jk}^{\sigma}$$
(2.6)

is the spin-resolved AO density matrix of fragment *A*, and $n_{A,k}^{\sigma}$ is the occupation number of σ -spin orbital *k* in fragment *A*.

The atomic multipoles are computed from the DFTB3 density matrix

$$\rho_{a}(\mathbf{r}) = \sum_{ij \in a} P_{A,ij} \chi_{i}(\mathbf{r}) \chi_{j}(\mathbf{r}) + \sum_{b \neq a} f_{ab}(b_{ab}) \sum_{i \in a \atop i \in b} P_{A,ij} \chi_{i}(\mathbf{r}) \chi_{j}(\mathbf{r})$$
(2.7)

where $\chi_i(\mathbf{r}) = \chi_i(r)Y_{l_i\mu_i}(\Omega)$ is an AO basis function, $Y_{l\mu}(\Omega)$ is a real-valued spherical harmonic [Eq. (2.45)],

$$f_{ab}(b_{ab}) = f_{ab}^{s} + S_{on} \left(\frac{b_{ab} - b_{ab}^{s}}{b_{ab}^{d} - b_{ab}^{s}}\right) \left(f_{ab}^{d} - f_{ab}^{s}\right)$$
(2.8)

Methods 61

is a fraction between 0 and 1 and holding the property $f_{ab} = 1 - f_{ba}$, $b_{ab} = 2 \sum_{i \in a \atop j \in h} P_{A,ij} S_{A,ij}$ (2.9)

is a Mulliken bond-order, and

$$S_{\rm on}(x) = \begin{cases} 0 & \text{if } x < 0\\ 1 & \text{if } x > 1\\ 10x^3 - 15x^4 + 6x^5 & \text{otherwise} \end{cases}$$
(2.10)

is a smooth polynomial used to switch f_{ab} from f_{ab}^{s} to f_{ab}^{d} as the bond order increases.

The atomic multipoles are obtained by inserting Eq. (2.7) into Eq. (2.2). If we restrict the contributions of the two-center densities to charge only, then the charge on atom *a* is

$$q_{00} = Z_a - b_{aa}/2 - \sum_{\substack{b \in A \\ b \neq a}} f_{ab}(b_{ab})b_{ab}$$
(2.11)

and its higher-order multipole moments are

$$q_{lm} = \sum_{ij \in a} P_{ij}^{A} M_{ij}^{(l)} \sqrt{\frac{4\pi}{2l+1}} \int Y_{l\mu}(\Omega) Y_{l_{i}\mu_{i}}(\Omega) Y_{l_{j}\mu_{j}}(\Omega) d\Omega, \quad (2.12)$$

where the integral is a real-valued spherical harmonic Gaunt coefficient (Sec. 2.6.5) and the

$$M_{ij}^{(l)} = \int_0^\infty \chi_i(r) \chi_j(r) r^{2+l} dr$$
 (2.13)

are treated as a parameters. For an *sp*-basis, there are two parameters: $M_{sp}^{(1)}$ and $M_{pp}^{(2)}$, which control the magnitude of the dipole and quadrupole contributions, respectively. We restrict the $M_{sd}^{(2)}$, $M_{pd}^{(1)}$, and $M_{dd}^{(2)}$ parameters encountered in an *spd*-basis to the values of $M_{pp}^{(2)}$, $M_{sp}^{(1)}$, and $M_{pp}^{(2)}$, respectively.

2.3.2 Models

The different models compared and discussed in this paper include: **mDC**: The method described in the previous section and parametrized to the S22 [29], S66 [41], JSCH [28, 29, 43], SCAI [5] databases and to a database of sulfur containing molecules and water clusters.

Table 2.1Average moleculardipole and quadrupole moment errors of 10 nucleobasemonomers. All values in a.u..DipoleQuadrupolemDC0.071.22DFTB30.213.26GAFF0.252.29

The description of the parametrization procedure and a detailed and broad analysis of mDC performance are presented in a manuscript that is, at the time of this writing, in press.

mDC(q): The modified "charge-only" model described in Ref. [16]. This model does not expand the atomic densities to higher-order multipoles.

DFTB3: The 3ob parametrized version of DFTB3 [15], i.e., DFTB3-3ob.

GAFF/TIP3P: The general Amber force field [6, 44] and TIP3P water. **PM6**: The semiempirical method described in Ref. [38], as implemented in Gaussian 09 [13].

PM3_{BP}: The semiempirical method described in Ref. [17].

mPWPW91, **M062X**, and **B97D**: The *m*PWPW91/MIDI!, M062X/6-311++G**, and B97D/6-311++G** density functional methods implemented in Gaussian 09 [13]. The B97D model contains empirical long-range dispersion corrections [24].

2.3.3 Computational Details

Table 2.1 displays molecular dipole and quadrupole moment errors averaged over 10 nucleobases. The reference molecular moments where computed with B3LYP/6-311++ G^{**} . An error of these vector quantities is taken to be the magnitude of the difference vector between the model and reference moments. The average magnitude of the reference dipole and quadrupole moment vectors are 1.98 and 14.58 a.u., respectively.

The reference energies and geometries used in Table 2.2 were computed with counterpoise-corrected CCSD(T)/CBS//MP2/TZVPP or MP2/CBS//MP2/cc-pVTZ, which were taken from Refs. [28] and [43], whose naming convention we

Methods 63

austics, and a summary of geometrical errors.												
	Ref.	mDC	DFTB3	GAFF	PM6	M062X	B97D					
	H-bonded dimers											
AT wc	-16.86	-15.99	-8.90	-13.60	-8.90	-14.87	-15.56					
GC wc	-32.06	-32.50	-21.92	-26.15	-18.47	-28.22	-28.11					
GA 1	-19.40	-18.36	-11.32	-15.70	-10.13	-17.09	-18.14					
GA 2	-14.40	-14.11	-5.86	-12.40	-6.45	-11.23	-12.57					
GA 3	-18.80	-17.90	-9.60	-15.39	-9.18	-16.09	-16.83					
GA 4	-13.50	-14.43	-6.76	-12.85	-7.42	-12.07	-13.48					
GA 1 pl	-18.90	-17.98	-11.15	-11.20	-10.13	-17.09	-18.14					
GA 2 pl	-12.80	-13.28	-5.37	-8.20	-6.11	-11.22	-12.58					
A 17		0.72	0.22	2.00	0.74	2.26	1 4 1					
∆E mue		0.73	8.23	3.90	8.74	2.36	1.41					
∆E mse		0.27	8.23	3.90	8.74	2.36	1.41					
crms (A)		0.13	0.16	0.36	0.29	0.19	0.18					
∠ _{plane} (°)		4.91	13.50	17.48	12.00	10.62	9.66					
	Stacked dimers											
AT S1	-12.30	-13.34	-9.11	-13.20	-5.27	-13.86	-12.25					
mAmT S	-14.57	-15.29	-9.17	-14.66	-5.99	-16.46	-14.89					
GC S	-19.02	-18.62	-21.92	-26.15	-18.47	-28.25	-28.11					
mGmC S	-20.35	-20.02	-21.99	-22.01	-18.28	-27.94	-27.99					
Δ Γ		0.62	2.20	2.44	450	F 07	4.27					
∆E mue		0.62	3.28	2.44	4.56	5.07	4.27					
∆E mse		-0.26	1.01	-2.44	4.56	-5.07	-4.25					
crms (A)		0.28	2.70	0.80	1.56	1.41	1.42					
∠plane (°)		8.79	10.29	15.48	10.56	6.38	5.23					
	Combined statistics											
ΔE mue		0.70	6.58	3.42	7.35	3.26	2.37					
ΔE mse		0.10	5.83	1.79	7.35	-0.12	-0.47					
crms (Å)		0.18	1.00	0.51	0.71	0.60	0.59					
∠nlane (°)		6.20	12.43	16.81	11.52	9.21	8.18					

Table 2.2 Nucleobase dimer binding energies (kcal/mol), binding energystatistics, and a summary of geometrical errors.

adopt. "crms" is the average coordinate root mean square deviation of the optimized dimer geometry relative to the reference geometry. All atoms were included in the calculation of the crms. The row of angle mean signed errors was constructed by comparing the angles formed between the two planes of the nucleobases relative those in the reference geometry.

The geometrical errors shown in Table 2.3 include 6 N-H \cdots 0 bond lengths and angles and 11 N-H \cdots N bond lengths and angles.

Table 2.3Hydrogen bond length (Å) and angle (°) errors.

		mDC	DFTB3	GAFF	PM6	M062X	B97D
<i>R</i> _{N(-H)···0}	mse	-0.02	0.02	-0.01	0.07	0.04	0.01
R _{(N-)H0}	mse	-0.01	0.01	-0.00	0.07	0.04	-0.00
$\angle_{N\text{-}H\cdots0}$	mue	1.44	2.17	4.71	3.88	1.75	1.30
<i>R</i> _{N(-H)} N	mse	-0.10	0.08	-0.00	0.19	0.01	-0.03
$R_{(N-)H\cdots N}$	mse	-0.10	0.08	0.02	0.21	0.01	-0.04
$\angle_{N\text{-}H\cdots N}$	mue	2.93	1.97	6.69	9.40	1.59	2.20

Table 2.4Nucleobase trimer formation enthalpies(kcal/mol).Brackets represent a Boltzmann averaging of the conformations shown above it at 298K.

	Expt.	mDC	DFTB3	mPWPW91	PM3 _{BP}
UUA 1		28.3	15.3	21.0	25.2
UUA 2		28.0	15.1	21.4	25.1
UUA 3		24.7	16.1	17.0	20.5
UUA 4		26.7	17.5	17.4	20.6
$\langle UUA \rangle$	27-29	28.1	17.3	21.3	25.2
UUU 1		26.5	16.1	8.5	13.1
UUU 2		21.6	16.5	11.3	14.6
$\langle UUU \rangle$	20-22	26.5	16.4	11.3	14.5
UUT	23-25	21.6	15.6	7.1	12.7
CCC 4	33-38	34.0	20.0	22.0	28.9
(UUU 1 UUU 2 (UUU) UUT CCC 4	27-29 20-22 23-25 33-38	28.1 26.5 21.6 26.5 21.6 34.0	17.3 16.1 16.5 16.4 15.6 20.0	8.5 11.3 11.3 7.1 22.0	23.2 13.1 14.6 14.5 12.7 28.9

 $R_{X(-H)\cdots Y}$ and $R_{(X-)H\cdots Y}$ denote the bond length between X-Y and H-Y, respectively.

The experimental numbers appearing in Table 2.4 are taken from Ref. [53]. The *m*PWPW91/MIDI! and PM3_{BP} results are taken from Ref. [17], whose naming convention we adopt. The mDC, DFTB3, and *m*PWPW91 results include zero point and thermal corrections to the enthalpy at 298K using standard ideal-gas statistical mechanics and the rigid-rotor harmonic-oscillator approximation. The presence of two experimental numbers represents the two manners used to analyze the results in Ref. [53].



Figure 2.1 Optimized water dimer hydrogen bond angles.

2.4 Results and Discussion

In addition to providing the mathematical details of how atomic multipoles are incorporated into the mDC model, we wish to explain with some illustrative examples why we consider them necessary and then use these examples to interpret some observations made using the standard DFTB3 semiempirical Hamiltonian.

The DFTB3 Hamiltonian contains four components to the energy: (1) the MO-computed tight-binding interaction composed of the electron kinetic energy and the first-order interaction of the response density with the effective chemical potential caused by the neutral atom reference density, as modeled by a 1- or 2body approximation, (2) the second-order electrostatic interaction of the response density with itself, which has been parametrized to experimental hardness so as to effectively include nonclassical effects, (3) a short-range repulsive function to achieve good covalent bond lengths, and (4) a third-order response interaction which attempts to correct the second-order electrostatics to account for the fact that anionic electron densities should be more diffuse than the neutrals. The DFTB3 electrostatic interactions are computed from atomic charges only, even though the orbitals used to compute the first-order interactions contain higher-order multipoles. When we built the mDC method upon the DFTB3 Hamiltonian, we were thus faced with the choice of computing the inter-fragment interactions using the atomic charges that DFTB3 happens to use or construct our own representation of the charge density from



Figure 2.2 Hydrogen bond angle of water to asparagine at $N\cdots O$ separations of 3 and 4 Å.

the DFTB3 density matrix. Preliminary tests of a method using the DFTB3 charges, mDC(q), proved unsatisfactory upon examining the geometries of hydrogen bonded (H-bonded) clusters (see Fig. 2.1). One of our goals was to make sure that the mDC method was *at least* as good as DFTB3, but the H-bond angles produced by mDC(q) model were more similar to the TIP3P model than to either ab initio calculations or DFTB3. Considering that both DFTB3 and mDC(q) use atomic charges to compute the second-order electrostatics, we were left to hypothesize that it was the multipolar character in DFTB3's tight-binding matrix elements that caused it to achieve good H-bond angles. The inter-fragment tight-binding matrix elements are removed in the mDC model, so we chose to model the behavior by increasing the order of atomic multipoles used to compute the electrostatics [16]. The resulting method, mDC, yields water H-bond angles in good agreement with DFTB3.

This hypothesis is further supported upon considering the Hbond formed between water and the amine group of asparagine (see Fig. 2.2). Ab initio geometry optimizations produce a water that is angled relative to the plane of asparagine's amine group, whereas TIP3P water is consistently coplanar. We interpret this observation as resulting from TIP3P's lack of higher-order multipoles. DFTB3 produces an angled water near the energy minimum, but when the water is pulled away from the amine group, it reverts to a coplanar TIP3P-like structure. In other words, the DFTB3 geometries agree with ab initio when there is significant AO overlap between the July 2, 2015 12:10

Results and Discussion 67

molecules, but acts like a point-charge model when the overlap is small. This is consistent with the above hypothesis and suggests that one could improve DFTB3 by extending its second-order electrostatic interactions to include higher-order multipoles. The mDC model uses higher-order multipoles without making intrusive changes to the underlying DFTB3 Hamiltonian and produces H-bond angles in better agreement with ab initio for all separations.

The use of atomic multipoles improves mDC's description of electrostatic potentials. Upon comparing the electrostatic potentials generated by mDC and DFTB3, we've found that the most significant improvements occur in molecules containing π -bonds, sp^3 oxygen and sulfur lone pair electrons, and sp^2 nitrogen lone pairs. In comparison to DFTB3 and GAFF, mDC also shows an overall statistical improvement in the molecular dipole and quadrupole moments (see e.g., Table 2.1).

The above assessment of mDC focused on examples that highlight the influence of including higher-order multipoles. We now assess the quality of mDC H-bonded and stacked nucleobase interactions and make comparison to other commonly used methods. There are many small variations and parametrizations of semiempirical models [1, 3, 17, 18, 30-32, 35, 37, 39, 40, 42, 54], but for brevity we limit our comparison to those which have seen widespread use and implementation into common software packages. Firstly, mDC produces the smallest energetic and geometrical errors of any method in Table 2.2. Generally speaking, the high-level reference binding energies are much stronger than those of the predicted by the standard semiempirical models. The GAFF force field energies are better than the other semiempirical methods and often prevents the stacked dimers from devolving into H-bonded complexes. The DFTB3 method reproduce H-bonded geometries more accurately than GAFF even though GAFF's H-bonded energetic errors are nearly twice as small. The ab initio H-bond interactions are superior to those of the semiempirical models, but do not show a significant improvement for stacked interactions. This is, in part, due to the "destacking" of some dimers upon geometry optimization.

Our primary measure of quantifying geometrical errors is through the coordinate root mean square overlays (crms). The \angle_{plane} errors measure the angle formed between the vectors normal to

Figure 2.3 mGmC S stacked dimer (left) and GA 4 hydrogen bonded dimer (right) coordinate root mean square overlay of the mDC structure (ball-and-stick, colored) on to the reference structures (lines, black). These two dimers are the *worst* mDC structures in the set of molecules listed in Table 2.2.

the plane of the two bases, which is computed from diagonalizing their moment of inertia tensors. We place greater meaning to the H-bonded \angle_{plane} errors than we do for the stacked dimers because the angle in a stacked dimer is approximately zero, but if the geometry optimization de-stacks the structure, then the angle within the resulting (incorrect) H-bonded structure is also approximately zero.

Table 2.3 compares the H-bond distance and angle errors. We note that the mDC N-N distances are 0.1Å too short, which may explain why mDC was capable of reproducing the high-level dimer interaction energies. The mDC errors listed in Table 2.3 and Fig. 2.3 are not disturbing considering that the overall errors in the mDC geometries are significantly better than the other methods.

Table 2.4 compares the experimental trimer enthalpies of formation to mDC, *m*PWPW91, and $PM3_{BP}$. mDC is in much better agreement with the experimental results than the other methods, which underpredict the strength of the H-bonds in the trimer. We suspect that the added strength afforded by mDC is largely a result of the slightly reduced N-N distances, as seen in Table 2.3; however, comparison between the *m*PWPW91 and mDC geometries show an overall agreement in geometries (see e.g., Fig. 2.4.)



Figure 2.4 UUA 1 trimer coordinate root mean square overlay of the mDC structure (ball-and-stick, colored) on to the *m*PWPW91 structure (lines, black). Hydrogen bond lengths are listed in Å.

2.5 Conclusion

This chapter has sought to stimulate the exchange of effective strategies used to describe many-body effects and electrostatics within the context of a linear-scaling quantum force field. In particular, we've provided the mathematical details required to implement the multipolar densities used in the mDC model and highlighted the importance of using multipoles in our method with some illustrative examples.

We observe that the water dimer H-bond angles are reproduced when higher-order multipoles are included, whereas a chargeonly model causes the dimer to revert into a TIP3P-like structure. Furthermore, we observe that standard DFTB3 H-bond angles are quite good when there is significant overlap between the AOs of two molecules, but it too can revert to a TIP3P like structure when the AO overlap is small. We attribute this phenomenon to DFTB3's use of multipoles in the AO tight-binding matrix and not the second-order electrostatics.

In addition, we provided a brief comparison between nucleobase dimer and trimer binding energies and geometries as computed with mDC, other semiempirical models, a molecular mechanical force field, and several ab initio methods. mDC was shown to

reproduce the high-level ab initio and experimental results with the greatest accuracy.

Further tests with the mDC model will be necessary to fully realize the benefits of a linear-scaling quantum force field. We are currently implementing a generalized PME method for condensed phase calculations using our treatment of atomic multipoles (further details of which are described in the appendices). Incorporation of mDC and the generalized PME method is ongoing and will be described in more detail in future work.

2.6 Appendices

2.6.1 Complex Harmonics and the Spherical Tensor Gradient Operator

The complex spherical harmonic $Y_{lm}(\Omega)$ is related to the associated Legendre polynomial $P_{lm}(x)$ by

$$Y_{lm}(\Omega) = (-1)^m \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_{lm}(\cos\theta) e^{im\phi}$$
(2.14)

$$P_{lm}(x) = (1 - x^2)^{m/2} \left(\frac{d}{dx}\right)^m P_l(x)$$
(2.15)

$$P_l(x) = \frac{1}{2^l l!} \left(\frac{d}{dx}\right)^l (x^2 - 1)^l,$$
(2.16)

where $P_l(x)$ is a Legendre polynomial. The complex-valued regular $C_{lm}(\mathbf{r})$ and irregular $Z_{lm}(\mathbf{r})$ solid harmonics and the complex-valued *scaled* regular $R_{lm}(\mathbf{r})$ and irregular $I_{lm}(\mathbf{r})$ solid harmonics are

$$C_{lm}(\mathbf{r}) = r^l \sqrt{\frac{4\pi}{2l+1}} Y_{lm}(\Omega)$$
(2.17)

$$Z_{lm}(\mathbf{r}) = r^{-l-1} \sqrt{\frac{4\pi}{2l+1}} Y_{lm}(\Omega)$$
 (2.18)

$$R_{lm}(\mathbf{r}) = C_{lm}(\mathbf{r})/a_{lm} \tag{2.19}$$

$$I_{lm}(\mathbf{r}) = a_{lm} Z_{lm}(\mathbf{r}) \tag{2.20}$$

$$a_{lm} = \sqrt{(l+m)!(l-m)!}$$
 (2.21)

The spherical tensor gradient operator (STGO) is a solid harmonic whose Cartesian coordinate arguments have been replaced by PSP Book - 9in x 6in

Appendices **71**

Cartesian derivatives. Hobson's theorem [25] is the result of acting a STGO upon any spherical function $f(r^2)$

$$C_{lm}(\nabla)f(r^2) = 2^l C_{lm}(\mathbf{r}) \left(\frac{d}{dr^2}\right)^l f(r^2)$$
(2.22)

$$R_{lm}(\nabla)f(r^2) = 2^l R_{lm}(\mathbf{r}) \left(\frac{d}{dr^2}\right)^l f(r^2).$$
 (2.23)

The STGO obeys the product rule [11]

$$C_{lm}(\nabla) [f(\mathbf{r})g(\mathbf{r})] = \sum_{jk} \frac{a_{lm}}{a_{jk}a_{l-j,m-k}}$$

$$\times [C_{l-j,m-k}(\nabla)f(\mathbf{r})] [C_{jk}(\nabla)g(\mathbf{r})]$$
(2.24)

$$R_{lm}(\nabla) \left[f(\mathbf{r})g(\mathbf{r}) \right] = \sum_{jk} \left[R_{l-j,m-k}(\nabla) f(\mathbf{r}) \right] \left[R_{jk}(\nabla)g(\mathbf{r}) \right] \quad (2.25)$$

And when acted upon another solid harmonic, one obtains the following STGO differentiation rules [4, 47]

$$C_{jk}(\nabla)C_{lm}(\mathbf{r}) = \frac{(2j-1)!!a_{lm}}{a_{jk}a_{l-j,m+k}}(-1)^k C_{l-j,m+k}(\mathbf{r})$$
(2.26)

$$C_{jk}^{*}(\nabla)C_{lm}(\mathbf{r}) = \frac{(2j-1)!!a_{lm}}{a_{jk}a_{l-j,m-k}}C_{l-j,m-k}(\mathbf{r})$$
(2.27)

$$C_{jk}(\nabla)Z_{lm}(\mathbf{r}) = \frac{(2j-1)!!a_{l+j,m+k}}{a_{lm}a_{jk}}(-1)^{j}Z_{l+j,m+k}(\mathbf{r}) \qquad (2.28)$$

$$C_{jk}^{*}(\nabla)Z_{lm}(\mathbf{r}) = \frac{(2j-1)!!a_{l+j,m-k}}{a_{lm}a_{jk}}(-1)^{j+k}Z_{l+j,m-k}(\mathbf{r}) \quad (2.29)$$

$$\frac{a_{jk}^2}{(2j-1)!!}R_{jk}(\nabla)R_{lm}(\mathbf{r}) = (-1)^k R_{l-j,m+k}(\mathbf{r})$$
(2.30)

$$\frac{a_{jk}^2}{(2j-1)!!}R_{jk}^*(\nabla)R_{lm}(\mathbf{r}) = R_{l-j,m-k}(\mathbf{r})$$
(2.31)

$$\frac{a_{jk}^2}{(2j-1)!!}R_{jk}(\nabla)I_{lm}(\mathbf{r}) = (-1)^j I_{l+j,m+k}(\mathbf{r})$$
(2.32)

$$\frac{a_{jk}^2}{(2j-1)!!} R_{jk}^*(\nabla) I_{lm}(\mathbf{r}) = (-1)^{j+k} I_{l+j,m-k}(\mathbf{r})$$

= $(-1)^{j+m} I_{l+j,k-m}^*(\mathbf{r}).$ (2.33)

The utility of the above rules are numerous; however, the reader may gain a better appreciation upon considering two brief examples. We can express the translation of a regular or irregular solid harmonic with a Taylor series expansion.

$$C_{lm}(\mathbf{r} + \mathbf{a}) = e^{\mathbf{a}^{\top} \cdot \nabla} C_{lm}(\mathbf{r})$$

= $\sum_{jk} \frac{C_{jk}(\mathbf{a})}{(2j-1)!!} C_{jk}^*(\nabla) C_{lm}(\mathbf{r})$
= $\sum_{jk} \frac{a_{lm}}{a_{jk}a_{l-j,m-k}} C_{l-j,m-k}(\mathbf{r}) C_{jk}(\mathbf{a})$ (2.34)

The second line made use of the fact that $\nabla^2 C_{lm}(\mathbf{r}) = 0$, and the third line used Eq. (2.27). This result is known as the addition theorem of solid harmonics. Applying this same procedure to the other harmonics produces

$$R_{lm}(\mathbf{r} + \mathbf{a}) = \sum_{jk} R_{l-j,m-k}(\mathbf{r}) R_{jk}(\mathbf{a})$$
(2.35)

$$Z_{lm}(\mathbf{r} + \mathbf{a}) = \sum_{jk} \frac{a_{l-j,k-m}}{a_{lm}a_{jk}} (-1)^{j+m} C_{jk}(\mathbf{a}) Z_{l+j,k-m}^*(\mathbf{r})$$
(2.36)

$$I_{lm}(\mathbf{r} + \mathbf{a}) = \sum_{jk} (-1)^{j+m} R_{jk}(\mathbf{a}) I_{l+j,k-m}^*(\mathbf{r}).$$
(2.37)

For the special case l = m = 0 and $\mathbf{a} = -\mathbf{r}'$, Eqs. (2.36)–(2.37) are known as the Laplace expansion.

2.6.2 Real-Valued Harmonics

The scaled solid harmonics are decomposed into their real (c) and imaginary (s) components $R_{lm}(\mathbf{r}) = R_{lm}^c(\mathbf{r}) + i R_{lm}^s(\mathbf{r})$ from which one defines the real-valued scaled solid harmonics

$$R_{l\mu}(\mathbf{r}) = \begin{cases} R_{l|m|}^{c}(\mathbf{r}), & \mu \ge 0\\ R_{l|m|}^{s}(\mathbf{r}), & \mu < 0, \end{cases}$$
(2.38)

where a negative μ represents the sine-component of positive *m* and is used only to simplify notation where appropriate. When -m's

PSP Book - 9in x 6in

July 2, 2015 12:10

Appendices 73

are encountered in formula, the reader is implicitly instructed to apply the symmetry property $R_{l,-m}^{c/s}(\mathbf{r}) = \pm (-1)^m R_{lm}^{c/s}(\mathbf{r})$, which follows directly from $R_{lm}^*(\mathbf{r}) = (-1)^m R_{l,-m}(\mathbf{r})$, where the sign \pm corresponds to the cosine/sine designation. The solid harmonics are efficiently computed from recursion [46]

$$R_{mm}^{c/s}(\mathbf{r}) = -\frac{x R_{m-1,m-1}^{c/s}(\mathbf{r}) \mp y R_{m-1,m-1}^{s/c}(\mathbf{r})}{2m}$$
(2.39)

$$R_{lm}^{c/s}(\mathbf{r}) = \frac{(2l-1)zR_{l-1,m}^{c/s}(\mathbf{r}) - r^2R_{l-2,m}^{c/s}(\mathbf{r})}{(l+m)(l-m)}$$
(2.40)

$$I_{mm}^{c/s}(\mathbf{r}) = -\frac{(2m-1)}{r^2} \left(x I_{m-1,m-1}^{c/s}(\mathbf{r}) \mp y I_{m-1,m-1}^{s/c}(\mathbf{r}) \right)$$
(2.41)

$$I_{lm}^{c/s}(\mathbf{r}) = \frac{(2l-1)}{r^2} z I_{l-1,m}^{c/s}(\mathbf{r}) - \frac{(l-1)^2 - m^2}{r^2} I_{l-2,m}^{c/s}(\mathbf{r})$$
(2.42)

which are initiated from $R_{00}^{c}(\mathbf{r}) = 1$, $R_{00}^{s}(\mathbf{r}) = 0$, $I_{00}^{c}(\mathbf{r}) = 1/r$, and $I_{00}^{s}(\mathbf{r}) = 0$. The real-valued regular and irregular solid harmonics and real-valued spherical harmonics are then

$$C_{l\mu}(\mathbf{r}) = A_{l\mu} R_{l\mu}(\mathbf{r}) \tag{2.43}$$

$$Z_{l\mu}(\mathbf{r}) = I_{l\mu}(\mathbf{r}) / A_{l\mu} \tag{2.44}$$

$$Y_{l\mu}(\Omega) = \sqrt{\frac{2l+1}{4\pi}} C_{l\mu}(\hat{\mathbf{r}})$$
(2.45)

where

$$A_{l\mu} = (-1)^{\mu} \sqrt{(2 - \delta_{\mu,0})(l + \mu)!(l - \mu)!}.$$
 (2.46)

One can construct a real-valued STGO by replacing the Cartesian coordinate arguments of $C_{l\mu}(\mathbf{r})$ with their Cartesian gradients. By using the complex-valued STGO differentiation rules and the relationship between the complex- and real-valued harmonics, one obtains the gradients [33]

$$\frac{dR_{lm}^{c/s}(\mathbf{r})}{dx} = \frac{1}{2} \left(R_{l-1,m+1}^{c/s}(\mathbf{r}) - R_{l-1,m-1}^{c/s}(\mathbf{r}) \right)$$
(2.47)

$$\frac{dR_{lm}^{c/s}(\mathbf{r})}{dy} = \pm \frac{1}{2} \left(R_{l-1,m+1}^{s/c}(\mathbf{r}) + R_{l-1,m-1}^{s/c}(\mathbf{r}) \right)$$
(2.48)

$$\frac{dR_{lm}^{c/s}(\mathbf{r})}{dz} = R_{l-1,m}^{c/s}(\mathbf{r})$$
(2.49)

and

$$\frac{dI_{lm}^{c/s}(\mathbf{r})}{dx} = \frac{1}{2} \left(I_{l+1,m+1}^{c/s}(\mathbf{r}) - I_{l+1,m-1}^{c/s}(\mathbf{r}) \right)$$
(2.50)

$$\frac{dI_{lm}^{l/s}(\mathbf{r})}{dy} = \pm \frac{1}{2} \left(I_{l+1,m+1}^{s/c}(\mathbf{r}) + I_{l+1,m-1}^{s/c}(\mathbf{r}) \right)$$
(2.51)

$$\frac{dI_{lm}^{c/s}(\mathbf{r})}{dz} = -I_{l+1,m}^{c/s}(\mathbf{r}).$$
(2.52)

The real solid harmonics obey the translation theorems [19, 46]

$$R_{l\mu}(\mathbf{r} - \mathbf{R}_b) = \sum_{j\kappa} W_{l\mu,j\kappa}(\mathbf{R}_{ab}) R_{j\kappa}(\mathbf{r} - \mathbf{R}_a)$$
(2.53)

$$C_{l\mu}(\mathbf{r} - \mathbf{R}_b) = \sum_{j\kappa} \bar{W}_{l\mu, j\kappa}(\mathbf{R}_{ab}) C_{j\kappa}(\mathbf{r} - \mathbf{R}_a)$$
(2.54)

where

$$W_{lm,jk}^{c/s,c}(\mathbf{r}) = \left[R_{l-j,m-k}^{c/s}(\mathbf{r}) + (-1)^k R_{l-j,m+k}^{c/s}(\mathbf{r}) \right] / 2^{\delta_{k,0}}$$
(2.55)

$$W_{lm,jk}^{c/s,s}(\mathbf{r}) = \mp R_{l-j,m-k}^{s/c}(\mathbf{r}) \pm (-1)^k R_{l-j,m+k}^{s/c}(\mathbf{r})$$
(2.56)

and

$$\overline{W}_{l\mu,j\kappa}(\mathbf{r}) = (A_{l\mu}/A_{j\kappa})W_{l\mu,j\kappa}(\mathbf{r}).$$
(2.57)

Consider a system composed of atomic multipoles, i.e., $\rho(\mathbf{r}) = \sum_{a,l\mu\in a} q_{l\mu}\chi_{l\mu}(\mathbf{r} - \mathbf{R}_a)$, where $\chi_{l\mu}(\mathbf{r} - \mathbf{R}_a)$ is any function satisfying

$$\int \chi_{l\mu} (\mathbf{r} - \mathbf{R}_a) C_{j\kappa} (\mathbf{r} - \mathbf{R}_a) d^3 r = \delta_{lj} \delta_{\mu\kappa}; \qquad (2.58)$$

then the multipole moments of $\rho(\mathbf{r})$ evaluated about the origin \mathbf{R}_o is

$$q_{l\mu} = \int \rho(\mathbf{r}) C_{l\mu} (\mathbf{r} - \mathbf{R}_o) d^3 r$$

= $\sum_{a, j\kappa \in a} \bar{W}_{l\mu, j\kappa} (\mathbf{R}_{ao}) q_{j\kappa}.$ (2.59)

The translation of multipoles, as in Eq. (2.59), is a key component to the Fast Multipole Method [20, 46].

The gradients of $\mathbf{W}(\mathbf{r})$ can be expressed in terms of the matrix elements themselves in a manner analogous to Eq. (2.47)–(2.49), e.g. $d/dz W_{lm,jk}^{c/s,c/s}(\mathbf{r}) = W_{(l-1,m),jk}^{c/s,c/s}(\mathbf{r})$. The translation matrix is efficiently computed using the identity $W_{(l+1,m),(j+1,k)}^{c/s,c/s}(\mathbf{r}) = W_{lm,jk}^{c/s,c/s}(\mathbf{r})$.

Appendices 75

2.6.3 Gaussian Multipole Expansions

One deduces the form of a Gaussian multipole upon considering Eq. (2.58) and the orthogonality of the spherical harmonics [19]

$$\chi_{l\mu}(\mathbf{r} - \mathbf{R}_{a}; \zeta) = \left(\frac{\zeta}{\pi}\right)^{\frac{3}{2}} \frac{(2\zeta)^{l}}{(2l-1)!!} e^{-\zeta|\mathbf{r}-\mathbf{R}|^{2}} C_{l\mu}(\mathbf{r} - \mathbf{R}_{a})$$

$$= \frac{C_{l\mu}(\nabla_{a})}{(2l-1)!!} \chi_{00}(\mathbf{r} - \mathbf{R}_{a}; \zeta)$$
(2.60)

The interaction of two Gaussian multipoles via operator $\hat{O}(\mathbf{r} - \mathbf{r}')$ is [19]

$$E = \sum_{l \downarrow \in a \atop j_{\kappa \in b}} q_{l \downarrow \mu} q_{j \kappa} O_{l \downarrow \mu, j \kappa} (\mathbf{R}_{ab}), \qquad (2.61)$$

where

$$O_{l\mu,j\kappa}(\mathbf{R}_{ab}) = \frac{C_{l\mu}(\nabla_{a})}{(2l-1)!!} \frac{C_{j\kappa}(\nabla_{b})}{(2j-1)!!} \int \int \chi_{00}(\mathbf{r} - \mathbf{R}_{a}; \zeta_{a}) \\ \times \hat{O}(\mathbf{r} - \mathbf{r}')\chi_{00}(\mathbf{r}' - \mathbf{R}_{b}; \zeta_{b})d^{3}rd^{3}r' \\ = (-1)^{j} \sum_{u=0}^{\min(l,j)} O_{l+j-u} \frac{2^{l}(2u-1)!!2^{j}}{(2l-1)!!2^{u}(2j-1)!!} \qquad (2.62) \\ \times \sum_{\nu=-u}^{u} \bar{W}_{l\mu,u\nu}(\mathbf{R}_{ab})\bar{W}_{j\kappa,u\nu}(\mathbf{R}_{ab})$$

and

$$O_n = \left(\frac{d}{dR_{ab}^2}\right)^n O_{00,00}(R_{ab}^2)$$
(2.63)

is an "auxiliary vector." For example, when $O_0 = \text{erfc}(\zeta R_{ab})/R_{ab}$, Eq. (2.62) is the real-space Ewald correction for point-multipole interactions. The gradients of Eq. (2.62) can be expressed as a linear combination of auxiliary matrix elements [19], where the auxiliary matrix is Eq. (2.62) evaluated with one extra derivative applied to Eq. (2.63). The beauty of Eq. (2.62) is exhibited when contracted Gaussian functions are used; in which case, only Eq. (2.63) depends on the contraction coefficients and primitive exponents. This property was exploited in Ref. [21] which demonstrated how Eq. (2.62) is used to efficiently rotate the pretabulated overlap and tight-binding matrix elements encountered in the DFTB2 and DFTB3 semiempirical Hamiltonians.

2.6.4 Point Multipole Expansions

A point multipole $\delta_{l\mu}(\mathbf{r} - \mathbf{R}_a)$ is best described as a Gaussian multipole [Eq. (2.60)] in the limit of infinite exponent [19, 47]

$$\delta_{l\mu}(\mathbf{r} - \mathbf{R}_{a}) = \lim_{\zeta \to \infty} \chi_{l\mu}(\mathbf{r} - \mathbf{R}_{a}; \zeta)$$

$$= \frac{C_{l\mu}(\nabla_{a})}{(2l-1)!!} \lim_{\zeta \to \infty} \left(\frac{\zeta}{\pi}\right)^{\frac{3}{2}} e^{-\zeta |\mathbf{r} - \mathbf{R}_{a}|^{2}} \qquad (2.64)$$

$$= \frac{C_{l\mu}(\nabla_{a})}{(2l-1)!!} \delta(\mathbf{r} - \mathbf{R}_{a}).$$

By writing the real-valued STGO as a linear combination of the complex-valued STGO and applying the product and differentiation rules, one derives the Coulomb interaction energy between two point multipole expansions [46]

$$E = \sum_{\substack{l\mu \in a \\ j\kappa \in b}} q_{l\mu} q_{j\kappa} \frac{C_{l\mu}(\nabla_a)}{(2l-1)!!} \frac{C_{j\kappa}(\nabla_b)}{(2j-1)!!} \frac{1}{R_{ab}}$$

= $\mathbf{q}_a^{\top} \cdot \bar{\mathbf{T}}(\mathbf{R}_{ab}) \cdot \mathbf{q}_b$ (2.65)

where

$$\bar{T}_{l\mu,j\kappa}(\mathbf{R}_{ab}) = T_{l\mu,j\kappa}(\mathbf{R}_{ab})/(A_{l\mu}A_{j\kappa}), \qquad (2.66)$$

$$T_{lm,jk}^{c,c/s}(\mathbf{r}) = \left[I_{l+j,m+k}^{c/s}(\mathbf{r}) \pm (-1)^k I_{l+j,m-k}^{c/s}(\mathbf{r}) \right] \frac{2(-1)^l}{2^{\delta_{m,0}+\delta_{k,0}}}$$
(2.67)

and

$$T_{lm,jk}^{s,c/s}(\mathbf{r}) = \left[(-1)^k I_{l+j,m-k}^{s/c}(\mathbf{r}) \pm I_{l+j,m+k}^{s/c}(\mathbf{r}) \right] \frac{2(-1)^l}{2^{\delta_{m,0}+\delta_{k,0}}}.$$
 (2.68)

Eqs. (2.66)–(2.68) are a special case of the more general Eq. (2.62). The gradients of **T**(**r**) can be expressed in terms of the matrix elements themselves in a manner analogous to Eq. (2.50)–(2.52), e.g. $d/dzT_{lm,jk}^{c/s,c/s}(\mathbf{r}) = -T_{(l+1,m),jk}^{c/s,c/s}(\mathbf{r})$. The interaction matrix is efficiently computed using the identities $T_{j\kappa,l\mu}(\mathbf{r}) = (-1)^{l+j}T_{l\mu,j\kappa}(\mathbf{r})$ and $T_{(l-1,m),(j+1,k)}^{c/s,c/s}(\mathbf{r}) = -T_{lm,jk}^{c/s,c/s}(\mathbf{r})$. By having written a point-multipole as the spherical tensor gra-

By having written a point-multipole as the spherical tensor gradients passing through a point, one easily derives the particle mesh Ewald method for point multipoles. The main differences occur in the calculation of the structure factor, which requires spherical tensor gradients of the Cardinal B-spline weight, and the calculation of the short-range real-space correction (see Section 2.6.3).

Appendices 77

Table2.5 Theuniquenonzero real-valued spherical harmonic Gauntcoefficients for expanding atomic orbital prod-ucts to quadrupole. $G^{l\mu}_{(l_a\mu_a),(l_b\mu_b)} = s |G^{l\mu}_{(l_a\mu_a),(l_b\mu_b)}|$,where the sign *s* and the magnitude are listed inthe table.

$ G^{l\mu}_{(l_a\mu_a),(l_b\mu_b)} $	S	1	μ	la	μ_a	l_b	μ_b
/ 1	+	1	0	0	0	1	0
$\sqrt{4\pi}$	+	1	1	0	0	1	1
	+	1	-1	0	0	1	-1
	+	2	0	0	0	2	0
	+	2	1	0	0	2	1
	+	2	-1	0	0	2	-1
	+	2	2	0	0	2	2
	+	2	-2	0	0	2	-2
$\sqrt{\frac{1}{4\pi}\frac{4}{5}}$	+	2	0	1	0	1	0
/ 1 1	_	2	0	1	1	1	1
$\sqrt{\frac{1}{4\pi}\frac{1}{5}}$	_	2	0	1	-1	1	-1
/ 1.2	+	2	1	1	0	1	1
$\sqrt{\frac{1}{4\pi}\frac{5}{5}}$	+	2	$^{-1}$	1	0	1	-1
	+	2	2	1	1	1	1
	_	2	2	1	$^{-1}$	1	$^{-1}$
	+	2	-2	1	1	1	-1
F / 1 4	+	2	0	2	0	2	0
$\frac{5}{7}\sqrt{\frac{1}{4\pi}\frac{4}{5}}$	_	2	0	2	2	2	2
	-	2	0	2	-2	2	-2
- /	+	2	0	2	1	2	1
$\frac{5}{7}\sqrt{\frac{1}{4\pi}\frac{1}{5}}$	+	2	0	2	-1	2	-1
, <u> </u>	т.	2	2	2	1	2	1
$\frac{5}{7}\sqrt{\frac{1}{4\pi}\frac{3}{5}}$	т _	2	2	2	_1	2	_1
·	+	2	_2 _2	2	-1	2	-1
		-	-	-	-	-	-

2.6.5 Real-Valued Spherical Harmonic Gaunt Coefficients

A real-valued spherical harmonic Gaunt coefficient corresponds to the integral

$$G_{(l_{a}\mu_{a}),(l_{b}\mu_{b})}^{l\mu} = \int Y_{l\mu}(\Omega)Y_{l_{a}\mu_{a}}(\Omega)Y_{l_{b}\mu_{b}}(\Omega)d\Omega$$

= $G_{(l_{b}\mu_{b}),(l_{a}\mu_{a})}^{l\mu}$
= $G_{(l\mu),(l_{b}\mu_{b})}^{l_{a}\mu_{a}} = G_{(l_{b}\mu_{b}),(l_{\mu})}^{l_{a}\mu_{a}}$
= $G_{(l\mu),(l_{a}\mu_{a})}^{l_{b}\mu_{b}} = G_{(l_{a}\mu_{a}),(l_{\mu})}^{l_{b}\mu_{b}},(l_{\mu}),(l$

which has a six-fold degeneracy. The values of the these coefficients are different than those encountered in text books, which tend to list those for complex-valued harmonics. Most combinations of indices produce a zero result. The unique nonzero values used to perform the auxiliary expansion of the DFTB3 density are listed in Table 2.5.

Acknowledgments

The authors are grateful for financial support provided by the National Institutes of Health (GM62248). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575.

References

- 1. Acevedo, O., and Jorgensen, W. L. (2010). Advances in Quantum and Molecular mechanical (QM/MM) simulations for organic and enzymatic reactions, *Acc. Chem. Res.* **43**, 142–151.
- 2. Barnes, J., and Hut, P. (1986). A hierarchical *O*(*Nlog N*) force-calculation algorithm, *Nature* **324**, 446–449.
- Barnett, C. B., and Naidoo, K. J. (2010). Ring puckering: A metric for evaluating the accuracy of AM1, PM3, PM3CARB-1, and SCC-DFTB carbohydrate QM/MM Simulations, *J. Phys. Chem. B* **114**, 17142–17154.
- 4. Bayman, B. F. (1978). A generalization of the spherical harmonic gradient formula, *J. Math. Phys.* **19**(12), 2558–2562.
- Berka, K., Laskowski, R., Riley, K. E., Hobza, P., and Vondrášek, J. (2009). Representative amino acid side chain interactions in proteins. A comparison of highly accurate correlated ab initio quantum chemical and empirical potential procedures, *J. Chem. Theory Comput.* 5, 982–992.
- 6. Case, D. A., Darden, T. A., Cheatham III, T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Götz, A. W., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wolf, R. M., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.-J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Salomon-Ferrer, C., R. Sagui,

July 2, 2015 12:10

Appendices 79

Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2012). *AMBER 12*, University of California, San Francisco, San Francisco, CA.

- Cembran, A., Bao, P., Wang, Y., Song, L., Truhlar, D. G., and Gao, J. (2010). On the interfragment exchange in the X-Pol method, *J. Chem. Theory Comput.* 6(8), 2469–2476.
- Cisneros, G. A., Piquemal, J., and Darden, T. A. (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.
- 9. Dahlke, E. E., and Truhlar, D. G. (2007). Electrostatically embedded many-body correlation energy, with applications to the calculation of accurate second-order Møller–Plesset perturbation theory energies for large water clusters, *J. Chem. Theory Comput.* **3**(4), 1342–1348.
- Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N log(N) method for Ewald sums in large systems, *J. Chem. Phys.* 98, 10089–10092.
- 11. Dunlap, B. I. (2001). Direct quantum chemical integral evaluation, *Int. J. Quantum Chem.* **81**, 373–383.
- Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Hsing, L., and Pedersen, L. G. (1995). A smooth particle mesh Ewald method, *J. Chem. Phys.* **103**(19), 8577–8593.
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, M., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J. A., J., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Straverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, O., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2009). *Gaussian 09, Revision A.02*, Gaussian, Inc., Wallingford, CT.
- 14. Gao, J., and Wang, Y. (2012). Communication: Variational many-body expansion: Accounting for exchange repulsion, charge delocalization, and dispersion in the fragment-based explicit polarization method, *J.*

Chem. Phys. **136**, 071101.

- Gaus, M., Goez, A., and Elstner, M. (2013). Parametrization and benchmark of DFTB3 for organic molecules, *J. Chem. Theory Comput.* 9, 338–354.
- 16. Giese, T. J., Chen, H., Dissanayake, T., Giambaşu, 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. Theory Comput.* 9, 1417–1427.
- 17. Giese, T. J., Sherer, E. C., Cramer, C. J., and York, D. M. (2005). A semiempirical quantum model for hydrogen-bonded nucleic acid base pairs, *J. Chem. Theory Comput.* **1**(6), 1275–1285.
- Giese, T. J., and York, D. M. (2005). Improvement of semiempirical response properties with charge-dependent response density, *J. Chem. Phys.* **123**(16), 164108.
- 19. Giese, T. J., and York, D. M. (2008). Contracted auxiliary Gaussian basis integral and derivative evaluation, *J. Chem. Phys.* **128**(6), 064104.
- 20. Giese, T. J., and York, D. M. (2008). Extension of adaptive tree code and fast multipole methods to high angular momentum particle charge densities, *J. Comput. Chem.* **29**(12), 1895–1904.
- 21. Giese, T. J., and York, D. M. (2008). Spherical tensor gradient operator method for integral rotation: A simple, efficient, and extendable alternative to Slater–Koster tables, *J. Chem. Phys.* **129**(1), 016102.
- 22. Goedecker, S., and Scuseria, G. E. (2003). Linear scaling electronic structure methods in chemistry and physics, *IEEE Comput. Sci. Eng.* **5**, 14–21.
- 23. Greengard, L., and Rokhlin, V. (1987). A fast algorithm for particle simulations, *J. Comput. Phys.* **73**, 325–348.
- 24. Grimme, S., Ehrlich, S., and Goerigk, L. (2011). Effect of the damping function in dispersion corrected density functional theory, *J. Comput. Chem.* **32**(7), 1456–1465.
- 25. Hobson, E. W. (1892). On a theorem in differentiation, and its application to spherical harmonics, *Proc. London Math. Soc.* **24**(1), 55–67.
- 26. Isegawa, M., Gao, J., and Truhlar, D. G. (2011). Incorporation of charge transfer into the explicit polarization fragment method by grand canonical density functional theory, *J. Chem. Phys.* **135**, 084107.
- 27. Jacobson, L. D., and Herbert, J. M. (2011). An efficient, fragment-based electronic structure method for molecular systems: Self-consistent

July 2, 2015 12:10

polarization with perturbative two-body exchange and dispersion, *J. Chem. Phys.* **134**, 094118.

- 28. Jurečka, P., and Hobza, P. (2003). True stabilization energies for the optimal planar hydrogen-bonded and stacked structures of guanine...cytosine, adenine...thymine, and their 9- and 1-methyl derivatives: Complete basis set calculations at the MP2 and CCSD(T) levels and comparison with experiment, *J. Am. Chem. Soc.* **125**, 15608–15613.
- 29. Jurečka, P., Šponer, J., Černý, J., and Hobza, P. (2006). Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs, *Phys. Chem. Chem. Phys.* **8**, 1985–1993.
- Korth, M. (2010). Third-generation hydrogen-bonding corrections for semiempirical QM methods and force fields, *J. Chem. Theory Comput.* 6, 3808–3816.
- 31. Martin, B., and Clark, T. (2006). Dispersion treatment for NDDO-based semiempirical MO techniques, *Int. J. Quantum Chem.* **106**, 1208–1216.
- 32. McNamara, J. P., and Hillier, I. H. (2007). Semi-empirical molecular orbital methods including dispersion corrections for the accurate prediction of the full range of intermolecular interactions in biomolecules, *Phys. Chem. Chem. Phys.* **9**, 2362–2370.
- Pérez-Jordá, J., and Yang, W. (1996). A concise redefinition of the solid spherical harmonics and its use in fast multipole methods, *J. Chem. Phys.* 104, 8003–8006.
- 34. Piquemal, J., Cisneros, G., Reinhardt, P., Gresh, N., and Darden, T. A. (2006). Towards a force field based on density fitting, *J. Chem. Phys.* 124(10), 104101.
- 35. Rocha, G. B., Freire, R. O., Simas, A. M., and P. Stewart, J. J. (2006). RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I, *J. Comput. Chem.* **27**(10), 1101–1111.
- 36. 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 simulations, *J. Chem. Phys.* **120**(1), 73–87.
- Sattelmeyer, K. W., Tubert-Brohman, I., and Jorgensen, W. L. (2006). NO-MNDO: Reintroduction of the overlap matrix into MNDO, *J. Chem. Theory Comput.* 2, 413–419.
- 38. Stewart, J. J. P. (2007). Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements, *J. Mol. Model.* **13**, 1173–1213.

- 39. Tuttle, T., and Thiel, W. (2008). OMx-D: semiempirical methods with orthogonalization and dispersion corrections. Implementation and biochemical application, *Phys. Chem. Chem. Phys.* **10**, 2125–2272.
- Řezáč, J., and Hobza, P. (2012). Advanced corrections of hydrogen bonding and dispersion for semiempirical quantum mechanical methods, *J. Chem. Theory Comput.* 8, 141–151.
- Řeżač, J., Riley, K. E., and Hobza, P. (2011). S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures, *J. Chem. Theory Comput.* 7, 2427–2438.
- 42. Řezáč, J., Fanfrlík, J., Salahub, D., and Hobza, P. (2009). Semiempirical quantum chemical PM6 method augmented by dispersion and Hbonding correction terms reliably describes various types of noncovalent complexes, *J. Chem. Theory Comput.* **5**, 1749–1760.
- 43. Šponer, J., Jurečka, P., and Hobza, P. (2004). Accurate interaction energies of hydrogen-bonded nucleic acid base pairs, *J. Am. Chem. Soc.* **126**, 10142–10151.
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field, *J. Comput. Chem.* 25, 1157–1174.
- 45. Wang, Y., Sosa, C. P., Cembran, A., Truhlar, D. G., and Gao, J. (2012). Multilevel X-Pol: A fragment-based method with mixed quantum mechanical representations of different fragments, *J. Phys. Chem. B* 116(23), 6781–6788.
- Watson, M. A., Sałek, P., Macak, P., and Helgaker, T. (2004). Linear-scaling formation of Kohn–Sham Hamiltonian: Application to the calculation of excitation energies and polarizabilities of large molecular systems, *J. Chem. Phys.* **121**(7), 2915–2931.
- Weniger, E. J., and Steinborn, E. O. (1983). New representations for the spherical tensor gradient and the spherical delta function, *J. Math. Phys.* 24(11), 2553–2563.
- 48. Wesolowki, T. A. (2008). Embedding a multideterminantal wave function in an orbital-free environment, *Phys. Rev. A.* 77, 012504–012513.
- Wesolowski, T. A., and Warshel, A. (1993). Frozen density functional approach for ab Initio calculations of solvated molecules, *J. Phys. Chem.* 97, 8050–8053.
- 50. Xie, W., and Gao, J. (2007). Design of a next generation force field: The X-Pol potential, *J. Chem. Theory. Comput.* **3**(6), 1890–1900.

July 2, 2015 12:10

Appendices 83

- 51. 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. Theory Comput.* 5, 459–467.
- 52. Xie, W., Song, L., Truhlar, D. G., and Gao, J. (2008). The variational explicit polarization potential and analytical first derivative of energy: Towards a next generation force field, *J. Chem. Phys.* **128**, 234108.
- 53. Yanson, I. K., Teplitsky, A. B., and Sukhodub, L. F. (1979). Experimental studies of molecular interactions between nitrogen bases of nucleic acids, *Biopolymers* **18**, 1149–1170.
- 54. Zhang, P., Fiedler, L., Leverentz, H. R., Truhlar, D. G., and Gao, J. (2011). Polarized molecular orbital model chemistry. 2. The PMO method, *J. Chem. Theory Comput.* **7**, 857–867.
- 55. Zhang, P., Truhlar, D. G., and Gao, J. (2012). Fragment-based quantum mechanical methods for periodic systems with Ewald summation and mean image charge convention for long-range electrostatic interactions, *Phys. Chem. Chem. Phys.* **14**(21), 7821–7829.
- Zhou, B., Ligneres, V. L., and Carter, E. A. (2005). Improving the orbitalfree density functional theory description of covalent materials, *J. Chem. Phys.* **122**(4), 044103.

July 2, 2015 12:10

02-Qiang-Cui-c02