

## Distributed Multipole Analysis: Stability for Large Basis Sets

Anthony J. Stone\*

*University Chemical Laboratory, Lensfield Road,  
Cambridge CB2 1EW, United Kingdom*

Received August 2, 2005

**Abstract:** The distributed multipole analysis procedure, for describing a molecular charge distribution in terms of multipole moments on the individual atoms (or other sites) of the molecule, is not stable with respect to a change of basis set, and indeed, the calculated moments change substantially and unpredictably when the basis set is improved, even though the resulting electrostatic potential changes very little. A revised procedure is proposed, which uses grid-based quadrature for partitioning the contributions to the charge density from diffuse basis functions. The resulting procedure is very stable, and the calculated multipole moments converge rapidly to stable values as the size of the basis is increased.

### I. Introduction

In many calculations involving the interactions between molecules, it is necessary to calculate the electrostatic interaction between molecular charge distributions. The exact expression, however, involves an integral over both charge distributions and is much too time consuming for most applications. Instead, it is usual to use a multipole approximation, and it is well-understood that, for all but the very smallest molecules, it is necessary to use some form of distributed multipole model, in which multipole moments are attached to several sites in each molecule, usually to each atom but sometimes to groups of atoms and sometimes to additional sites as well as the atoms. Often, only charges are used, but it is recognized that, for accurate work, atomic dipoles and perhaps higher moments are needed.

Distributed multipole analysis (DMA)<sup>1–3</sup> is a simple, fast procedure that assigns multipole moments to each atom or other specified site in a molecule. It is exact, in the sense that the overall multipole moments constructed from the distributed moments exactly reproduce the overall multipole moments of the charge distribution. However, it has become clear, in the 25 years since the method was first introduced, that it is not stable with respect to changes of the basis set. Very large basis sets with diffuse functions are commonly

used in accurate work nowadays, and these are particularly troublesome, yielding distributed multipoles that may not correspond to physical expectations. Moreover, there is no convergence as the basis set is increased; on the contrary, the distributed multipoles vary more and more wildly.

It should be emphasized that the overall multipole moments remain exact, but this is not helpful for short-range interactions, where the expansion in terms of overall multipoles diverges, and the interaction is determined by the multipole moments for a few sites on each molecule.

The DMA procedure works as follows. A Gaussian basis function centered at  $A$  is the product of a Gaussian function  $\exp[-\zeta_A(r - A)^2]$  with a low-degree polynomial (degree 0, i.e., constant, for  $s$  functions, 1 for  $p$  functions, and so on). Boys showed<sup>4</sup> that the product of two such functions, one at  $A$  and the other at  $B$ , is another Gaussian of the same form. The degree of its polynomial is the sum of the degrees of the original Gaussians, and the Gaussian part is  $\exp[-(\zeta_A + \zeta_B)(r - P)^2]$ , where the “overlap center”  $P = (\zeta_A A + \zeta_B B)/(\zeta_A + \zeta_B)$ . The electron density is a sum of such products, with coefficients determined from the density matrix. Any individual product can be described exactly in terms of a sum of multipole moments of ranks up to the degree of its polynomial; so, the product of two  $s$  functions can be expressed as a pure charge, the product of an  $s$  with a  $p$  as a charge plus dipole, and so on. Distributed multipole analysis evaluates these exact representations and ap-

\* Author to whom correspondence should be addressed. Phone: +44 1223 336375. Fax: +44 1223 336362. E-mail: [ajs1@cam.ac.uk](mailto:ajs1@cam.ac.uk).

proximates each of them by a multipole expansion (nonterminating, in general) about the nearest of a set of expansion sites, which comprises some or all of the atomic nuclei, and perhaps some additional sites.

The nature of the problem with this method can be easily understood if we contemplate a simple molecule such as carbon dioxide. In a large and diffuse basis, we may have a diffuse  $p_z$  primitive on the C atom (we take  $z$  along the molecular axis), which is very similar to an out-of-phase combination of diffuse  $s$  primitives on the O atoms. The densities associated with these two functions are  $(p_z^C)^2$  and  $1/2(s^A - s^B)^2 = 1/2(s^A)^2 - s^A s^B + 1/2(s^B)^2$ , respectively, where A and B denote the two oxygen atoms and C the carbon. In the distributed-multipole procedure, each product of primitives is represented by multipoles on the site nearest to its overlap center. If we assume that both functions are normalized, the density  $(p_z^C)^2$  becomes a charge  $-1$  plus a pure quadrupole  $\Theta$  on the C atom. The density  $1/2(s^A)^2 - s^A s^B + 1/2(s^B)^2$  becomes a set of charges,  $-1/2q$  on each O atom and  $+q - 1$  on the C.

Now if we change the basis, the exponents of the diffuse functions will be different. In one basis, the C  $p_z$  function may be variationally preferred as a description of this particular aspect of the charge distribution, while in another, the combination of  $s$  functions may be preferred. The actual charge distribution may change very little, especially if both basis sets are large, but the distributed-multipole description may be very different. As the size of the basis set is increased, the opportunities for this kind of ambiguity increase, so the problem gets worse. Far from approaching a converged description, the distributed multipoles vary ever more wildly as the basis size is increased.

At short distances (for example, in a close end-on contact with one of the oxygen atoms), it may be supposed, and often is assumed, that the electrostatic interaction obtained from the distributed multipoles could also vary significantly. In fact, we shall see that this is not the case, but the lack of convergence in the description is a very unsatisfactory feature.

## II. A New Version of Distributed Multipole Analysis

The difficulty with standard DMA arises because the partitioning of the density between atoms is carried out in basis-function space. It has been recognized for some time that a partition in real space is more appropriate. Probably the best method of this type that is currently available is the atoms-in-molecules procedure, originally due to Bader<sup>5</sup> (see also references therein) and more recently developed as a method for distributed multipoles by Popelier.<sup>6</sup> This method works well, but it involves a very time-consuming procedure for determining the surfaces that separate one atom basin from another. Another recent method<sup>7</sup> uses the Hirshfeld partitioning method,<sup>8,9</sup> apportioning electron density between atoms in proportion to the free atom densities at the same distances from the nuclei.

The purpose of this paper is to propose a version of distributed multipole analysis that overcomes the strong basis-set dependence of the original version but is simpler

to implement than other real-space partitioning schemes. It uses numerical quadrature for the diffuse functions, while retaining the original method for the more compact ones. This avoids any difficulties associated with the numerical quadrature of strongly peaked functions. The method used for the numerical quadrature is that of Becke,<sup>10,11</sup> with some minor differences. A grid of integration points is constructed around each atom, and the molecular grid is the union of these atom grids. Becke associates a set of atom weights with any point in space; the sum of the weights is unity, the nearest atom having a weight close to 1 and the others having weights close to zero. At the boundary between two atoms, the weight of one atom falls smoothly to zero as the weight of the other increases to 1. This is achieved by a "smoothing function", and the steepness of this function is a parameter of the quadrature algorithm. The density is partitioned in this way into overlapping regions, each assigned to one atom. In density-functional calculations, the parameter is usually chosen so that the smoothing function is quite steep; this means that many of the more distant points on each atom grid have weights close to zero and can be dropped, making the integration more efficient. For the distributed multipole analysis, however, it is advantageous to make it less steep, so that the regions associated with each atom interpenetrate and are more nearly spherical. The density assigned to each atom is integrated over the grid for that atom to yield atomic multipole moments referred to the atom nucleus as the origin.

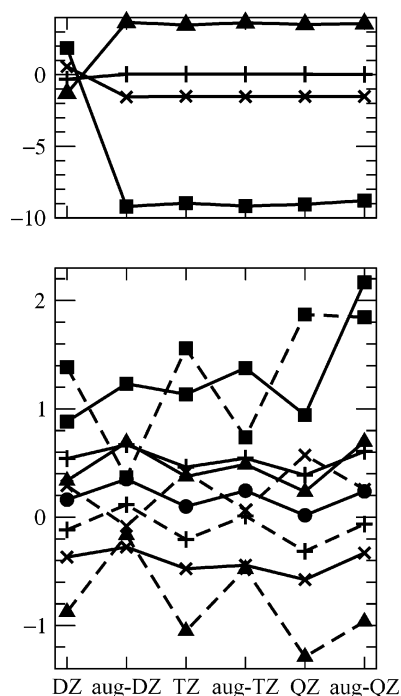
For calculations of multipole moments, the quadrature grid needs to extend to large distances, because the multipole moment functions (the regular spherical harmonics) include radial factors of  $r^k$  for rank  $k$ , and the products of basis functions that contribute to moments of rank  $k$  themselves include radial factors  $r^n$  with  $n \geq k$ . Moreover, the regular spherical harmonics have a strong angular dependence at high rank, as do the high-angular-momentum basis functions, so the angular integral also needs to be more accurate than is usually necessary for Kohn–Sham calculations, where the integral involves the total density.

These considerations have prompted the development of a new program for distributed multipole analysis. It is called GDMA2, where the G stands for Gaussian and reflects the fact that the program is intended for use with the Gaussian suite of quantum chemistry programs, and it supersedes the older GDMA program, which uses the original distributed multipole analysis.

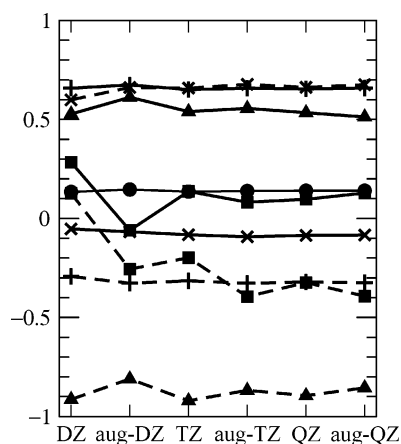
## III. Examples and Discussion

**III.1. Carbon Monoxide.** The features that led to dissatisfaction with the original distributed multipole analysis are illustrated in Figure 1, where the multipole moments are plotted against the basis sets. It is clear that the distributed multipoles vary substantially with basis set and, in particular, that there is a systematic difference between the augmented and nonaugmented sets. Moreover, the values seem to be diverging rather than converging as the basis set is improved.

In contrast, the behavior of the total moments is just what one would expect. The results are poor for the unaugmented double- $\zeta$  basis, but the remaining values rapidly settle down to consistent values. The hexadecapole values are not quite



**Figure 1.** Distributed multipoles (bottom) and total multipole moments (top) for CO, in atomic units, using the original DMA algorithm. ● denotes charge, + denotes dipoles, × denotes quadrupoles, ▲ denotes octopoles, and ■ denotes hexadecapoles. In the lower diagram, solid lines refer to C and dashed lines to O. “nZ”,  $n = D, T, Q$ , are abbreviations for “cc-pVnZ”.



**Figure 2.** Distributed multipoles for CO, in atomic units, using the new DMA algorithm. ● denotes charge, + denotes dipoles, × denotes quadrupoles, ▲ denotes octopoles, and ■ denotes hexadecapoles. Solid lines refer to C and dashed lines to O. “nZ”,  $n = D, T, Q$ , are abbreviations for “cc-pVnZ”.

converged with the QZ basis, but they are affected by the high angular-momentum ( $g$ ) basis functions that first appear in the QZ basis.

Figure 2 shows the distributed multipoles for the revised method. It is evident that the calculated values are converging steadily; the behavior as the basis set is improved is much more satisfactory.

At the same time, it is important to realize that the difference is a matter more of perception than reality. Although the distributed multipole moments may look very

**Table 1.** Multipole Moments on the C and O Atoms of Carbon Monoxide Using the Original Calculation Method (Switch = 0) and the New Method, with Grid-Based Quadrature for Products of Primitives with a Total Exponent Less than 4<sup>a</sup>

multipole	switch = 0	switch = 4	difference
C $Q_0$	0.244 75	−0.108 77	−0.353 52
C $Q_1$	0.544 70	0.367 55	−0.177 15
C $Q_2$	−0.443 94	−0.203 48	0.240 46
C $Q_3$	0.487 55	0.870 63	0.383 08
C $Q_4$	1.376 77	0.800 72	−0.576 05
O $Q_0$	−0.244 75	0.108 77	0.353 52
O $Q_1$	0.011 59	−0.564 95	−0.576 54
O $Q_2$	0.064 39	0.688 67	0.624 28
O $Q_3$	−0.478 34	−0.432 66	0.045 68
O $Q_4$	0.738 07	−0.857 87	−1.595 94
$V_{\max}/V$	0.073 34	0.074 23	0.001 23
$V_{\min}/V$	−0.146 99	−0.148 23	−0.001 06

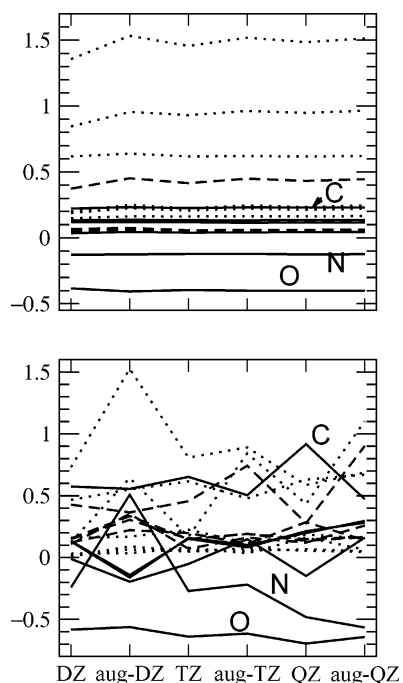
<sup>a</sup> The last two lines of the table show the maximum and minimum electrostatic potential on the  $\text{vdW} \times 2$  surface and the maximum and minimum difference.

different for the two methods, the resulting electrostatic potentials are almost identical. In Table 1, the aug-cc-pVQZ multipole moments are shown for the cases illustrated in Figures 1 and 2, together with the differences. The atom charges differ by  $0.35 e$  and are different in sign between old and new, and the higher multipoles also differ substantially. However, the electrostatic potential on the  $\text{vdW} \times 2$  surface (i.e., the surface of the solid that is the union of spheres of twice the van der Waals radius around each atom) differs by less than  $5 \times 10^{-5}$  au, or about 1.2 mV, everywhere. Note particularly that the values shown at the bottom of the last column are maximum and minimum potentials, on the  $\text{vdW} \times 2$  surface, arising from the set of multipole differences shown in the last column.

This corresponds to energy differences, for a unit charge at the  $\text{vdW} \times 2$  surface, of about 50 microHartrees. Although the *multipole moments* differ considerably, the *electrostatic potentials* hardly differ at all. The wide variation in multipole descriptions provided by different basis sets is not as troublesome as it appears at first sight—any set of multipoles will serve, provided that they are calculated with a reasonably good basis set.

Nevertheless, the more stable description provided by the new method has the advantage that it converges more satisfactorily as the basis set is improved, so that it becomes possible to assess when the basis set is of adequate quality, and multipole moments of related systems are more comparable and transferable.

**III.2. Formamide.** As a second example, we consider formamide. Here, there are many more multipole moments, too many to tabulate or even to display diagrammatically in detail. Figure 3 shows the behavior of the atom moments as the basis set is improved. The solid lines denote the atom charges; the dashed lines the magnitudes of each atom dipole,  $(Q_{10}^2 + Q_{11c}^2 + Q_{11s}^2)^{1/2}$ ; and the dotted lines the magnitudes of the atom quadrupoles,  $(\sum_m Q_{2m}^2)^{1/2}$ . The C, N, and O charges are identified, but the other moments have not been identified individually, as the object is just to contrast the



**Figure 3.** Distributed multipoles for formamide, in atomic units, using the original DMA algorithm (bottom) and the new (top). Solid lines denote atom charges, dashed lines denote atom dipoles, and dotted lines denote quadrupoles. “ $nZ$ ”,  $n = D, T, Q$ , are abbreviations for “cc-pV $nZ$ ”.

**Table 2.** Maximum and Minimum Electrostatic Potentials (V) on the vdW  $\times 2$  Surface of Formamide, for Original and New Methods of Calculation and for Several Basis Sets, and the Maximum Differences between Old and New Potentials for Each Basis

basis	switch = 0		switch = 4		difference max
	min	max	min	max	
cc-pVDZ	-0.744	0.714	-0.745	0.712	0.0034
aug-cc-pVDZ	-0.800	0.748	-0.800	0.749	0.0073
cc-pVTZ	-0.782	0.744	-0.782	0.743	0.0041
aug-cc-pVTZ	-0.796	0.745	-0.782	0.743	0.0072
cc-pVQZ	-0.794	0.749	-0.794	0.748	0.0047
aug-cc-pVQZ	-0.797	0.750	-0.797	0.754	0.0069

behaviors of the two methods. It is clear that the original DMA shows no sign of settling down to a converged value as the basis set is improved; on the contrary, some values, especially the C and N charges, appear to be diverging. The behavior of the new method is considerably better.

As in the case of CO, the electrostatic potential is well-described by most of the basis sets, whichever method is used. In this case, the original DMA leads to very large charges on the atoms when the cc-pVQZ basis is used. Nevertheless, the differences between the old and new methods in the electrostatic potential on the vdW  $\times 2$  surface are still less than 1% of the potential itself, as Table 2 shows. The corresponding energy differences, for a unit charge at the vdW  $\times 2$  surface, are a few meV, or less than 0.5 kJ mol<sup>-1</sup>, and are significantly smaller than the differences between the better basis sets when the same method of calculation is used for both.

Nevertheless, the much greater stability of the multipole description with respect to changes of basis set, as illustrated in Figure 3, is sufficient reason for preferring the new method.

#### IV. The Program

The GDMA program is written in Fortran90 and is self-contained. It requires as input a Gaussian03<sup>12</sup> (or earlier) formatted check-point file, together with a small controlling data file. It can currently handle basis functions up to  $g$ ; an extension to handle  $h$  basis sets is planned. The program can, in principle, determine distributed multipole moments up to rank 10, but the higher moments are likely to be inaccurate unless very large angular integration grids are used. (The original method gives results to the full accuracy of the input density matrix and may be preferred if very high ranks are needed.) Various parameters can be controlled in the data file, though the default values will usually be adequate. The default number of radial grid points is 80; more may be needed if moments higher than hexadecapole are required. The angular integration is usually carried out on a 590-point Lebedev grid;<sup>13</sup> as noted above, a large angular grid is needed if high-rank moments are required, because of the strong angle-dependence of the multipole operators. The number of grid points may be specified; any number may be requested, and the program will use the next-larger size of grid. Inadequacies in the angular grid are usually evident in small spurious nonzero values for multipole moments that should be zero by symmetry. The program can also use Gauss–Legendre integration over the angular coordinates if preferred.

The radial integration is carried out using Euler–Maclaurin quadrature, as formulated by Murray et al.<sup>11</sup> Following normal practice, the radial grid is scaled by reference to Bragg–Slater covalent radii,<sup>14</sup> but with the hydrogen radius taken to be 0.5 Å, twice the Bragg–Slater value. For the cutoff between atoms, we use the scheme proposed by Becke.<sup>10</sup> The softness or sharpness of the boundary between atoms is controlled by Becke’s cutoff parameter  $k$ . The effect of various values is illustrated in Figure 1 of Becke’s paper. Becke recommends a value of 3, which gives a fairly soft cutoff; sharper cutoff functions have been used,<sup>11</sup> but the value of 3 seems to be satisfactory for the present purpose and is the default for the GDMA program. Other values can be specified if required. The position of the cutoff is controlled, again as proposed by Becke,<sup>10</sup> by assigning radii to each site and transforming the cutoff function by reference to the ratio of radii. It is customary to use the Bragg–Slater radii in the Becke partitioning, but this leads to rather implausible multipole moments—rather large and not in keeping with chemical intuition—though, like the multipoles from the original method, they lead to accurate electrostatic potentials. The default procedure in the GDMA program is to set all the radii equal for this purpose, so that the boundary between neighboring atoms comes halfway between them. Any additional multipole sites are also assigned the same radius. However, any of the radii may be changed if required, in the data file controlling the program, and a somewhat smaller radius is probably appropriate for hydrogen atoms.



The results given above for formamide were obtained with the hydrogen radius set to 0.35 Å and the rest to 0.65 Å.

The other main controlling parameter determines the switch between old and new integration methods. If the sum  $\zeta_A + \zeta_B$  of exponents in a product of primitives is larger than the specified switch value, the product function is regarded as compact and can be efficiently and accurately integrated using Gauss–Hermite quadrature. This is the method used in the original DMA and GDMA programs. If the sum of exponents is smaller than the switch value, the product function is relatively diffuse, and it is evaluated on the quadrature grid to accumulate the diffuse contribution to the total density. Finally, the multipole moments are evaluated from the diffuse density, the integration being carried out for each atom grid separately and the moments assigned to the corresponding atom. A value of 4 has been found satisfactory for the switch between methods. A value of zero causes the old method to be used throughout.

Integration over a grid is a slow process, and treating every product of primitives individually makes it even slower, so the new method is considerably slower than the old—by a factor of 2 orders of magnitude or more—but the computation times are still short enough not to be a concern. For formamide, with a cc-pVQZ basis set (255 basis functions) and a large integration grid, the time needed is a few minutes on a Pentium 4 workstation. The CPU time scales roughly as the square of the number of diffuse functions in the basis.

The program can be downloaded from my web site, [www.stone.ch.cam.ac.uk](http://www.stone.ch.cam.ac.uk), and full documentation is provided.

## V. Conclusions

Although the distributed multipole analysis procedure, as originally defined, was entirely satisfactory 25 years ago, when small basis sets were the norm and few calculations included diffuse functions, it becomes less satisfactory for modern large basis sets with many diffuse functions. As the basis set size is increased, the distributed multipole moments do not approach converged values but oscillate in a divergent fashion. The overall molecular moments, and the electrostatic potential derived from the distributed multipoles, do converge satisfactorily with increasing basis set size, but the lack of converged distributed multipoles is, nevertheless, an unsatisfactory feature.

The revised method described in this paper deals with this problem by integrating the diffuse contributions to the electron density by grid-based quadrature, while the more compact functions are handled by exact Gauss–Hermite quadrature as in the original procedure. The resulting method

is very stable with respect to increasing basis set size and, indeed, reaches convergence for relatively modest basis sets.

## References

- (1) Stone, A. J. *Chem. Phys. Lett.* **1981**, 83, 233–239.
- (2) Stone, A. J.; Alderton, M. *Mol. Phys.* **1985**, 56, 1047–1064.
- (3) Stone, A. J. *The Theory of Intermolecular Forces*, International Series of Monographs in Chemistry; Clarendon Press: Oxford, U. K., 1996.
- (4) Boys, S. F. *Proc. R. Soc. London, Ser. A* **1950**, 200, 542–554.
- (5) Bader, R. *Atoms in Molecules*; Clarendon Press: Oxford, U. K., 1990.
- (6) Popelier, P. L. A. *Atoms in Molecules: An Introduction*; Prentice-Hall: Harlow, U. K., 2000.
- (7) Harrison, J. F. *J. Chem. Phys.* **2003**, 119, 8763–8764.
- (8) Hirshfeld, F. L. *J. Mol. Struct.* **1985**, 130, 125–141.
- (9) Hirshfeld, F. L. *Theor. Chim. Acta* **1977**, 44, 129–138.
- (10) Becke, A. D. *J. Chem. Phys.* **1988**, 88, 2547–2553.
- (11) Murray, C. W.; Handy, N. C.; Laming, G. J. *Mol. Phys.* **1993**, 78, 997–1014.
- (12) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Pittsburgh, PA, 2003. For a catalogue of basis sets, see [http://www.gaussian.com/g\\_ur/m\\_basis\\_sets.htm](http://www.gaussian.com/g_ur/m_basis_sets.htm).
- (13) Lebedev, V. I.; Laikov, D. N. *Dokl. Math.* **1999**, 59, 477–481.
- (14) Slater, J. C. *J. Chem. Phys.* **1964**, 41, 3199–3204.

CT050190+