THE JOURNAL OF PHYSICAL CHEMISTRY A LETTERS

2005, *109,* 2−5
Published on Web 12/13/2004

New Method for Direct Linear-Scaling Calculation of Electron Density of Proteins

Ye Mei,[†] Da W. Zhang,[‡] and J. Z. H. Zhang*[,†,‡]

Institute of Theoretical and Computational Chemistry, Nanjing University, Nanjing 210093, China, and Department of Chemistry, New York University, New York, New York 10003

Received: October 27, 2004; In Final Form: November 24, 2004

A new scheme for direct linear-scaling quantum mechanical calculation of electron density of protein systems is developed. The new scheme gives much improved accuracy of electron density for proteins than the original MFCC (molecular fractionation with conjugate caps) approach in efficient linear-scaling calculation for protein systems. In this new approach, the error associated with each cut in the MFCC approach is estimated by computing the two neighboring amino acids in both cut and uncut calculations and is corrected. Numerical tests are performed on six oligopeptide taken from PDB (protein data bank), and the results show that the new scheme is efficient and accurate.

Recently, a highly efficient approach for full quantum mechanical computation of electronic properties of polymers such as proteins has been developed.[1−3] In this molecular fractionation with conjugate caps (MFCC) approach, a protein, for example, is decomposed into amino acid-based fragments and pairs of conjugate caps (concap) are inserted at the cuts to properly cap the fragments. By employing the MFCC approach, one can compute electronic properties of protein systems such as protein−ligand interaction energy through an efficient linear scaling scheme using a variety of methods such as HF, DFT, MP2, etc.[4−7]

The MFCC approach has recently been further developed for efficient linear scaling computation of total electron density of proteins.[8] Accurate determination of electron density and electrostatic potential of molecular systems is of both fundamental importance and practical utility. For example, the spatial distribution of density or electrostatic potential is often used to understand chemical structure, reaction, binding, catalysis, and solvation.[9−16] The electrostatic potential is of particular interest in rational drug design for optimization of lead drug candidates and pharmacophore search.[17,18] It is also widely useful in areas such as force field parametrization[19−23] and quantitative structure−activity relationship (QSAR).[24,25] Furthermore, accurate determination of total electron density of proteins provides the basis for accurate quantum mechanical calculation of total energy of protein.

A number of methods that divide protein into smaller fragments for full quantum mechanical computation of electron density have been proposed before.[26−28] These methods have various degrees of success as well as problems. In comparison, the MFCC approach is numerically efficient and straightforward to apply to large protein systems. In this Letter, we develop a new scheme to provide improved accuracy of the computed electronic properties such as density, etc. from the MFCC calculation. The current correction scheme, which is named the MFCC II method, can be easily applied within the MFCC approach with modest amount of computational cost. In particular, the MFCC II correction method provides an easy means to estimate the error associated with each cut of the covalent bond along the protein backbone. Numerical calculations are carried out to test the current MFCC II scheme on several oligopeptide taken from the protein data bank (PDB).

To simplify discussion without loss of generality, we first take on a binary system composed of A and B components. By applying the MFCC approach,[1] this system is cut into A and B fragments that are capped with two conjugate caps (concap). Now the original system AB is replaced by three new subsystems or fragments, A−C, C*−B, and the concap C*−C. By employing a MFCC ansatz for total electron density, we can obtain, to a good approximation, the total electron density of the original AB system ρ by the following relation[8]

$$\rho = \rho_A + \rho_B - \rho^{cc} \tag{1}$$

where ρ_A and ρ_B are, respectively, electron densities of the capped A (A−C) and B (C*−B) fragments and ρ^{cc} is the electron density of the concap (C*−C) fragment. Equation 1 shows that by calculating electron density of individual protein fragment separately and independently, we can directly obtain *total* electron density of the whole system through eq 1. The above MFCC method is obviously linear scaling and is therefore applicable to calculations of large protein systems. In fact, the method can be easily paralleled in a highly efficient manner. It is important to note that eq 1 would give the exact electron density, if one of the conjugate caps C (or C*) includes the entire B (or A). In practical application, the trick is to employ smaller molecular caps that can faithfully simulate the original local environment of the said fragment.

* Corresponding author. E-mail john.zhang@nyu.edu.
[†] Nanjing University.
[‡] New York University.

10.1021/jp045109y CCC: $30.25

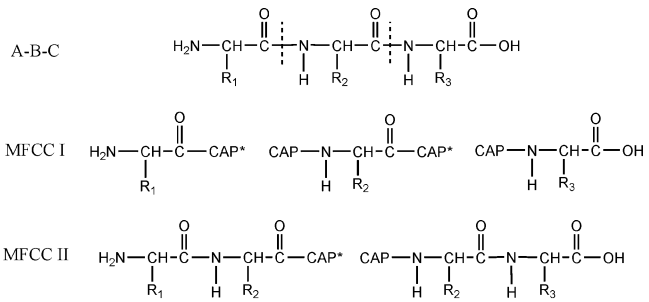

Figure 1. Illustration of the MFCC I and MFCC II schemes.

Equation 1 can be easily generalized to a polymer such as protein composed of many components. For a protein molecule with N fragments, eq 1 is easily generalized to

$$\rho = \sum_{k=1}^{N} \rho_k - \sum_{k=1}^{N-1} \rho_k^{cc} \tag{2}$$

where ρ_k is the density of the kth protein fragment, ρ_k^{cc} is the density of the kth concap, and ρ_k^{d} is the density of the kth disulfide concap.[3] Thus the total electron density of a protein can be obtained through simple combination of individual densities of amino acid fragments and concap species that can be calculated independently. It is useful to point out that the computational cost for concap species is minimal compared to that for larger capped protein fragments. Exactly the same relation holds for electrostatic potential

$$\phi = \sum_{k=1}^{N} \phi_k - \sum_{k=1}^{N-1} \phi_k^{cc} \tag{3}$$

and the total dipole moment

$$\mu = \sum_{k=1}^{N} \mu_k - \sum_{k=1}^{N-1} \mu_k^{cc} \tag{4}$$

where electrostatic potentials and dipole moments of the fragments are obtained from individual fragment calculations.

Because the MFCC approach is an approximate method, there are errors associated with the cutting of covalent bonds and the capping of fragments in the computed electronic properties such as electron density. It is therefore desirable to introduce a correction scheme to improve the accuracy of the numerically computed MFCC results including electron density, dipole moment, interaction energy, etc. The improved accuracy in electron density is of particular interest because it is directly related to the accuracy of the total protein energy obtained through DFT calculation. Here we present a correction scheme named MFCC II method for that purpose. To simplify discussion of the MFCC II method (correction scheme), we take as an example a tripeptide composed of A, B, and C amino acids as shown in Figure 1. Applying the MFCC approach, we can cut this peptide at peptide linkages between A−B and between B−C, and conjugate caps are inserted to cap the fragments as shown in Figure 1. Using the MFCC method (MFCC I in Figure 1), the electron density is calculated by the following equation:[8]

$$\rho_0 = \rho_A + \rho_B - \rho^{cc1} + \rho_C - \rho^{cc2} \tag{5}$$

where ρ^{cc1} and ρ^{cc2} are the densities of the concap inserted between A−B and B−C, as shown in Figure 1.

Here a correction scheme is introduced to estimate the error arising from the cut between A and B. By treating A−B as one group (without cut), we can apply the MFCC approach to A−B−C system by cutting the system only once (between B and C). By doing this, the electron density of A−B−C system can be evaluated as

$$\rho' = \rho_{AB} + \rho_C - \rho^{cc2} \tag{6}$$

where ρ_{AB} is the density of the AB fragment with capped ends. The difference between ρ' and ρ_0 is the approximate error from the cut between A and B and is given by

$$\Delta\rho_{AB} = \rho' - \rho_0 = \rho_{AB} - \rho_A - \rho_B + \rho^{cc1} \tag{7}$$

Similarly, the error associated with the cut between B and C can be evaluated by the difference

$$\Delta\rho_{BC} = \rho'' - \rho_0 = \rho_{BC} - \rho_B - \rho_C + \rho^{cc2} \tag{8}$$

If we add these error corrections to ρ_0, we then obtain an error-corrected electron density of the A−B−C system given by

$$\begin{aligned} \rho &= \rho_0 + \Delta\rho_{AB} + \Delta\rho_{BC} \\ &= \rho_{AB} + \rho_{BC} - \rho_B \end{aligned} \tag{9}$$

Equation 9 is the result of the MFCC II scheme for direct linear-scaling calculation of electron density for a three component system ABC.

For a protein or polypeptide containing N amino acids, it is not difficult to prove that the corrected electron density is given by

$$\rho = \sum_{i=1}^{N-1} \rho_{i,i+1} - \sum_{i=2}^{N-1} \rho_i \tag{10}$$

where $\rho_{i,i+1}$ is the density of the connected i, $i+1$ fragment and ρ_i is the density of the single ith fragment. We can also obtain the dipole moment in the same manner

$$\mu = \sum_{i=1}^{N-1} \mu_{i,i+1} - \sum_{i=2}^{N-1} \mu_i \tag{11}$$

and similar equation holds for the electrostatic potential

$$\phi = \sum_{i=1}^{N-1} \phi_{i,i+1} - \sum_{i=2}^{N-1} \phi_i \tag{12}$$

In this study, the concap is chosen to be $R_1CH_2CO-NHCH_2R_2$, where R_1 and R_2 are two side chains of the neighboring amino acids (cf. Figure 1). To balance the efficiency with accuracy, we perform calculations at the HF/6-31G* level using the Gaussian03 package. To measure the deviation of the computed electron density, we generate computed density within a cubic box with the center of the box located at the center of each peptide molecule, The box contains $50 \times 50 \times 50$ grids with the step-size of 0.2 Å in each dimension. The results from the corresponding full system HF/6-31G* calculations are taken as the standard for comparison.

We first compare dipole moments computed using both MFCC I and II methods with results from the corresponding full system calculation. Table 1 lists these calculated dipole moments for all six peptides. The result in Table 1 shows the following: (1) The MFCC I method already gives quite accurate

TABLE 1: Comparison of Dipole Moments from MFCC I, MFCC II, and Full System Calculations for Six Peptides at Crystal Structures from PDB

PDB ID	MFCC I	MFCC II	full system
1AB9	11.31	11.08	10.78
1QVO	9.26	9.45	9.56
1MHC	9.23	8.85	8.68
1R0T	10.44	10.78	10.80
1P7V	4.68	4.48	4.48
1BXX	4.87	4.64	4.55

TABLE 2: Comparison of rmsd of the Computed Densities Using MFCC I and MFCC II Method Relative to the Standard Full System Calculation

PDB ID	MFCC I	MFCC II
1AB9	0.111×10^{-3}	0.463×10^{-4}
1QVO	0.125×10^{-3}	0.260×10^{-4}
1MHC	0.252×10^{-3}	0.254×10^{-3}
1R0T	0.378×10^{-3}	0.361×10^{-3}
1P7V	0.687×10^{-4}	0.215×10^{-4}
1BXX	0.605×10^{-4}	0.184×10^{-5}

dipole moment as compared to the standard full system calculation. The errors are generally within just a few percent. (2) The MFCC II method clearly improves the accuracy of the computed dipole moment for all six peptides. Taking 1AB9 (a 10 amino acid peptide) as an example, the dipole from MFCC I calculation is larger than the standard result by 0.53 D (4.9% deviation). Using the MFCC II correction scheme, however, the error is reduced to 0.3 D (2.8% deviation). As seen from Table 1, the improvement in accuracy is uniform and across all six peptides, with even better accuracy for other peptides, especially 1P7V.

It is useful to point out that although the MFCC II method can correct errors associated with every cut along the protein backbone, it still does not give an exact result. This is because the current correction scheme can only correct errors of short range nature. However, the MFCC II scheme does provide a very useful measure of errors associated with individual cut and gives an overall improvement in numerical accuracy of the MFCC approach.

We next examine the electron density by comparing the rmsd (root-mean-square deviation) of the computed density relative to the standard full system result. The rmsd is defined by

$$\text{rmsd} = \sqrt{\frac{\sum_{k=1}^{N}(\Delta\rho_k)^2}{N}} \tag{13}$$

where $\Delta\rho_k$ is the deviation of the MFCC calculated density relative to the full system result and the summation is over all the grid points k within a box. We choose a $10 \times 10 \times 10$ (Å^3) cubic box with evenly spaced grids to calculate rmsd. Table 2 lists the rmsd of the computed electron density relative for all six peptides. We can see from Table 2 that rmsd from both MFCC results are quite small, indicating that the computed density from the MFCC method is quite accurate. Furthermore, the MFCC II method gives an overall improved density as measured by the rmsd shown in Table 2. Except for 1MHC and 1R0T, for which the MFCC II method does not show obvious improvement, there are significant improvements in density for the other four peptides.

To see more globally the effect of MFCC II correction on electron density, we show the absolute deviation between MFCC computed density and the density from the standard full system

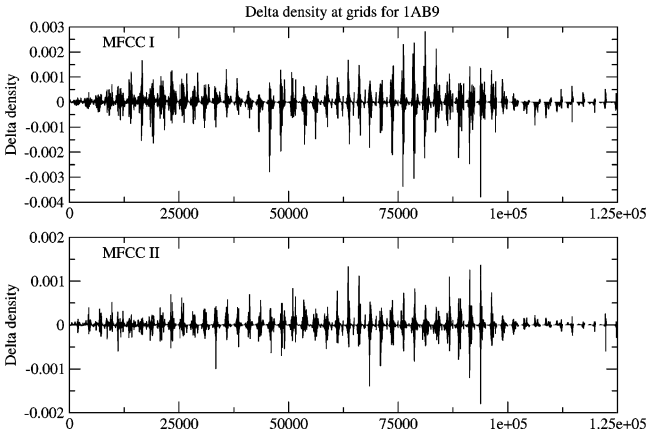

Figure 2. Absolute deviation of electron density of 1AB9 from MFCC I and MFCC II calculations relative to the standard full system result at numerical grids.

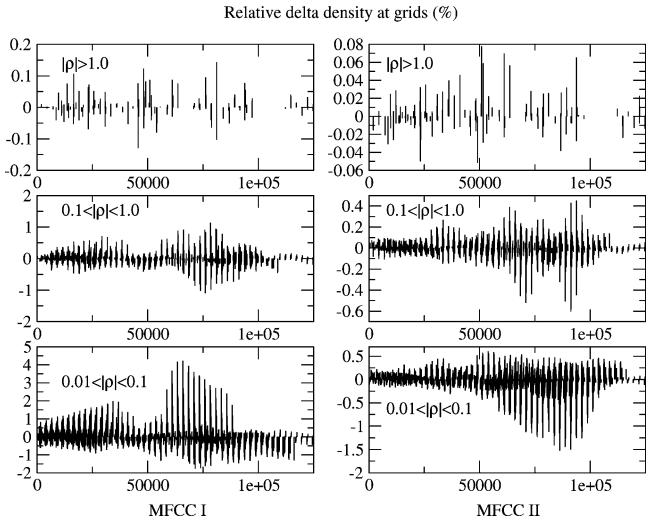

Figure 3. Relative (%) deviation of electron density of 1AB9 from MFCC I and MFCC II computations relative to the standard full system result at numerical grids.

calculation on $50 \times 50 \times 50$ individual grid points. Figure 2 shows the absolute deviation of electron from both MFCC I and MFCC II calculations for peptide 1AB9. We can see from Figure 2 that the MFCC II scheme shows clear improvement in the accuracy of the computed electron density. In the original MFCC I approach, the absolute deviation of density from the standard result is between -0.004 to $+0.003$. By employing MFCC II, the deviation is reduced to between -0.002 to $+0.0015$, which is just about half as much as shown in Figure 2.

Besides absolute deviation of density, we also show relative deviation of density defined as

$$re = \frac{\Delta\rho_k}{\rho_k} \tag{14}$$

It should be noted, however, that the relative deviation of density is generally larger if the value of density is small. Figure 3 compares the results from MFCC I and from MFCC II. In the MFCC I calculation, the relative errors are -0.2% to $+0.2\%$, -2% to $+2\%$, and -2% to $+5\%$, respectively, for $|\rho| < 1.0$, $0.1 < |\rho| < 1.0$, and $0.01 < |\rho| < 0.1$, which are about 3−10 times those from MFCC II approach. The MFCC II approach gives smaller relative density deviation especially for ρ in the range of 0.01 to 1.0.

In this Letter, we presented a new scheme called the MFCC II method that gives improved accuracy of electron density and related quantities for direct linear-scaling computation of protein systems in the MFCC approach. The MFCC II method provides a simple means to correct numerical errors in the standard MFCC approach with modest amount of computational cost. Numerical tests carried out on six peptides from PDB show systematic improvement of electron density and dipole moment over the results from the standard MFCC calculation. This should be very useful in a variety of applications, in particular, accurate computation of total protein energy using density functional theory (DFT) in future study.

For real protein systems, one needs to include more interactions in the MFCC approach. For example, in protein with helix or β structures, the effect of the hydrogen bonding network should be considered. It turns out that simple modification can be introduced to treat these hydrogen bondings in the MFCC approach, and work in this direction is already in progress with very encouraging results. The solvent effect could be handled by using implicit solvent models with the MFCC approach and work in this direction will commence shortly.

References and Notes

(1) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599.

(2) Zhang, D. W.; Chen, X. H.; Zhang, J. Z. H. *J. Comput. Chem.* **2003**, *24* 1846.

(3) Chen, X. H.; Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2004**, *120*, 839−844.

(4) Zhang, D. W.; Zhang, J. Z. H. *J. Theo. Comput. Chem.* **2004**, *3*, 43.

(5) Zhang, D. W.; Xiang, Y.; Zhang, J. Z. H. *J. Phys. Chem. B* **2003**, *107*, 12039.

(6) Zhang, D. W.; Xiang, Y.; Gao, A. M.; Zhang, J. Z. H. *J. Chem. Phys.* **2004**, *120*, 1145.

(7) Chen, X. H.; Zhang, J. Z. H. *J. Chem. Phys.* **2004**, *120*, 11386−11391.

(8) Gao, A. M.; Zhang, D. W.; Zhang, J. Z. H.; Zhang, Y. An efficient linear scaling method for ab initio calculation of electron density of proteins. *Chem. Phys. Lett.* **2004**, *394*, 293−297.

(9) Madura, J. D.; Davis, M. E.; Gilson, M. K.; Wade, R. C.; Luty, B. A.; McCammon, J. A. Biological Applications of Electrostatic Calculations and Brownian Dynamics Simulations. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1994; Vol. 5, pp 229−267.

(10) Politzer, P.; Murray, J. S. In *Molecular Electrostatic Potentials: Concepts and Applications*; Murray, J. S., Sen, K., Eds.; Elsevier: Amsterdam, 1996; pp 649−660.

(11) Naray-Szabo, G. Electrostatic Catalysis. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; Wiley & Sons: Chichester, U.K., 1998; pp 905−912.

(12) Gilson, M. K.; Honig, B. *Proteins* **1988**, *4*, 7.

(13) Tomasi, J.; Mennucci, B.; Cammi, R. In *Molecular Electrostatic Potentials: Concepts and Applications*; Murray, J. S., Sen, K., Eds.; Elsevier: Amsterdam, 1996; pp 1−85.

(14) Hummer, G.; Pratt, L. R.; Garcia, A. E.; Berne, B. J.; Rick, S. W. *J. Phys. Chem. B* **1997**, *101*, 30 17.

(15) Tobias, D. J. *Curr. Opin. Struct. Biol.* **2001**, *11*, 253.

(16) Matrin, M. E.; Sanchez, M. L.; Olivares del Valle, F. J.; Aguilar, M. A. *J. Chem. Phys.* **2002**, *116*, 1613.

(17) Orozco, M.; Canela, E. I.; Franco, R. *Eur. J. Biochem.* **1990**, *188*, 155.

(18) Rogriguez, J.; Manault, F.; Sanz, F. *J. Comput. Chem.* **1993**, *14*, 922.

(19) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.

(20) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620.

(21) Dykstra, C. E. *Chem. Rev.* **1993**, *93*, 3.

(22) Wang, W.; Donini, O.; Reges, C. M.; Kollman, P. A. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211.

(23) Halgren, T. A.; Damm, W. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236.

(24) Naray-Szabo, G.; Surjan, P. R. In *Theoretical Chemistry of Biological Systems*; Naray-Szabo, G., Ed.; Elsevier: Amsterdam, 1986; pp 1−30.

(25) Warshel, A.; Aqvist, J. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 267.

(26) Yang, W. T. *Phys. Rev. Lett.* **2004**, *66*, 1438.

(27) Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2000**, *318*, 614.

(28) Exner, T. E.; Mezey, P. *J. Phys. Chem. A* **2002**, *106*, 11791−11800.