JCTC Journal of Chemical Theory and Computation

Divide and Conquer Hartree−Fock Calculations on Proteins

Xiao He and Kenneth M. Merz, Jr.*

Department of Chemistry and the Quantum Theory Project, 2328 New Physics Building, P.O. Box 118435, University of Florida, Gainesville, Florida 32611-8435

Received December 9, 2009

Abstract: The ability to perform ab initio electronic structure calculations that scale linearly with the system size is one of the central aims in theoretical chemistry. In this study, the implementation of the divide and conquer (DC) algorithm, an algorithm with the potential to aid the achievement of true linear scaling within Hartree−Fock (HF) theory, is revisited. Standard HF calculations solve the Roothaan−Hall equations for the whole system; in the DC-HF approach, the diagonalization of the Fock matrix is carried out on smaller subsystems. The DC algorithm for HF calculations was validated on polyglycines, polyalanines, and 11 real three-dimensional proteins of up to 608 atoms in this work. We also found that a fragment-based initial guess using the molecular fractionation with conjugated caps (MFCC) method significantly reduces the number of SCF cycles and even is capable of achieving convergence for some globular proteins where the simple superposition of atomic densities (SAD) initial guess fails.

Introduction

Ab initio quantum mechanical methods have been developed over the past several decades and successfully applied to the study of the chemical properties for small- to moderate-sized molecules. The routine application of these full quantum mechanical calculations on macromolecules (molecules containing greater than 500 atoms) continues to pose a great challenge for theoretical chemists. The major limitation of ab initio methods is the scaling problem, since the computational cost of ab initio methods increases considerably as the size of the molecule increases. For instance, Hartree−Fock (HF)[1] and density functional theory (DFT)[2] scale as $O(N^4)$, post-Hartree−Fock MP2[3] scales as $O(N^5)$, and the coupled cluster(CC)[4−9] method that includes single and double excitations (CCSD) scales as $O(N^6)$. In modern HF calculations, the computational cost for the 2-electron integrals can be reduced from $O(N^4)$ to $O(N^2)$ using a simple screening technique.[10] Hence, the dominant step for large molecules becomes the matrix diagonalization, which scales as $O(N^3)$. In this study, our goal was to reduce the computational cost of the diagonalization step in HF calculations to linear with system size.

* Correspondence author phone: 352-392-6973; fax: 352-392-8722, e-mail: merz@qtp.ufl.edu.

The state-of-the-art linear-scaling algorithms, which make the computational cost scale linearly $O(N)$ with the system size, have attracted the focus of many theorists during the past decade.[11−21] Much effort has been devoted to the development of linear-scaling methods in order to compute the total energy of large molecular systems at the Hartree−Fock (HF) or density functional theory (DFT) level.[12,15,18,22−26] One of the challenges is to speed up the calculation of long-range Coulomb interactions when assembling the Fock matrix elements. Fast multipole-based approaches have successfully reduced the scaling in system size to linear[14,16−18,25] and made HF and DFT calculations affordable for larger systems when small- to moderate-sized basis sets are utilized. The more recently developed Fourier transform Coulomb method of Fusti and Pulay[27,28] reduced the steep $O(N^4)$ scaling in basis set size to quadratic and makes the calculations much more affordable with larger basis sets.[29] There is also a class of fragment-based methods for quantum calculation of protein systems including the divide and conquer (DC) method of Yang,[22] Yang and Lee,[23] Dixon and Merz,[30] Gogonea et al.,[31] Shaw and St-Amant,[32] and Nakai and co-workers,[33−36] the adjustable density matrix assembler (ADMA) approach method of Exner and Mezey,[26,37−39] the fragment molecular orbital (FMO) method of Kitaura and co-workers,[13,40,41] and the molecular frac-

10.1021/ct9006635
Published on Web 01/07/2010

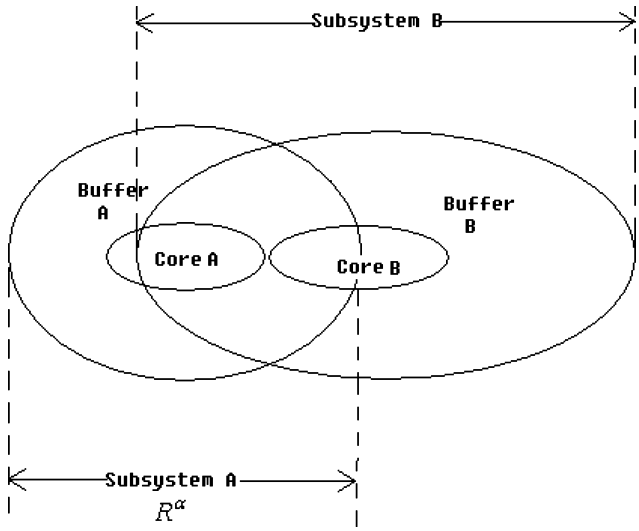

Figure 1. Graphical representation of the subsetting scheme used in divide and conquer calculations.

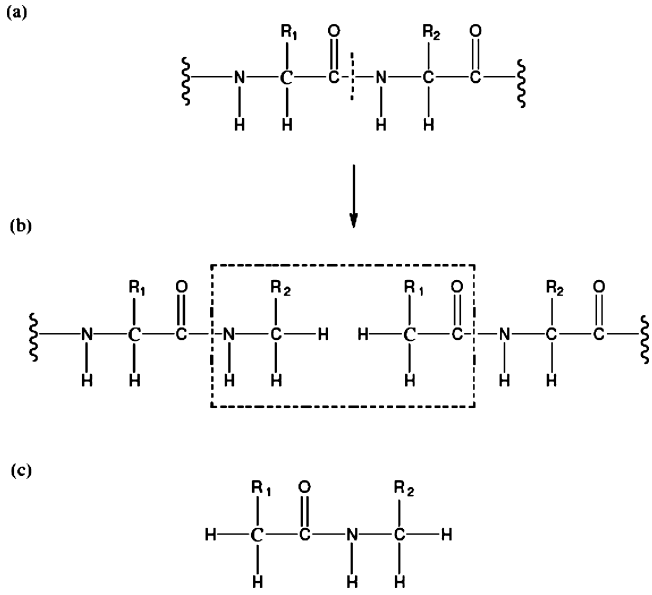

Figure 2. MFCC scheme in which the peptide bond is cut (a) and the fragments are capped with C_{cap} and its conjugate C_{cap}* (b). (c) Atomic structure of the concap. The concap is defined as the fused molecular species of C_{cap}*−C_{cap}.

tionation with conjugate caps (MFCC) approach developed by Zhang and co-workers.[42,43] Most applications of these methods to protein systems have been largely limited to semiempirical, HF, and DFT calculations. Among these approaches, FMO has been applied to higher level ab initio calculations such as second-order Møller−Plesset perturbation theory (MP2)[44] and coupled cluster theory (CC).[45] Nakai and co-workers recently proposed DC-MP2[33,36,46] and DC-CCSD[47] approaches; however, only systems of linear chains or near-linear chains have been tested so far for the divide and conquer algorithm for ab initio calculations.

In the DC algorithm, the total system is divided into small fragments. Atoms within adjustable buffer regions surrounding each fragment are included in the calculations to preserve the chemical environment of the divided subsystem. A set of local Roothaan−Hall equations is then solved for each

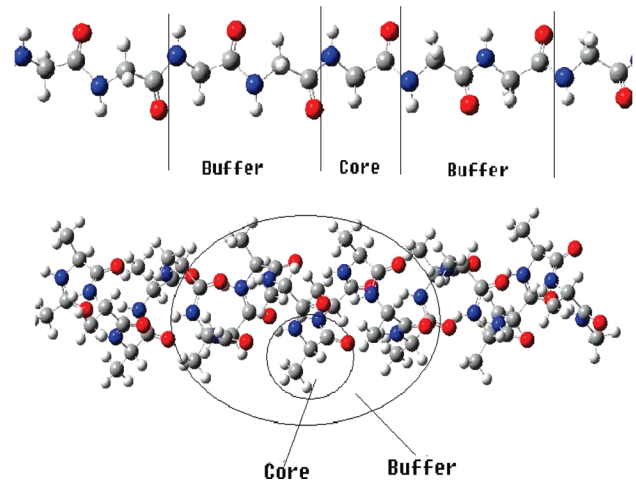

Figure 3. Subsetting schemes for divide and conquer calculations on the extended polyglycine $(Gly)_n$(upper) and polyalanine in an α-helical structure (α-$(Ala)_n$, bottom).

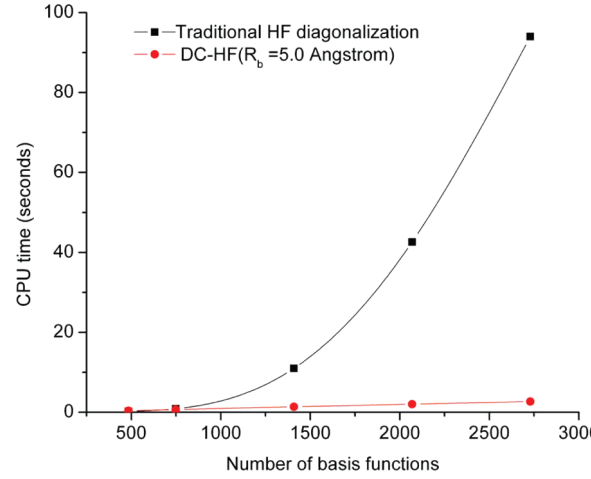

Figure 4. Average computational time to diagonalize the Fock matrix in each SCF cycle using traditional HF and DC-HF for a series of extended polyglycines at the HF/6-31G* level.

subsystem, and an approximate full density matrix of the entire molecular system is built up from subsystem contributions. By solving the HF self-consistent field (SCF) equation iteratively, the final converged full density matrix is used to obtain the total energy of the entire system. In this manner, linear scaling of the Fock matrix diagonalization step is achieved as a result of the fact that a set of smaller subsystem Fock matrices is diagonalized in the DC-HF approach rather than the global Fock matrix diagonalization for traditional HF calculations. Furthermore, divide and conquer calculations may be efficiently parallelized because the individual subsystem calculations are solved separately. In the DC-HF approach, the memory usage will increase linearly as the size of the system increases, which is also an important feature of this approach.

The aim of our current research is to further develop and validate the divide and conquer (DC)[22,23,30,32,46−48] methodology to aid in the application of ab initio methods to biomacromolecules. In this study, our goal is to validate the divide and conquer algorithm for Hartree−Fock calculations on globular proteins. Moreover, we propose a fragment-based

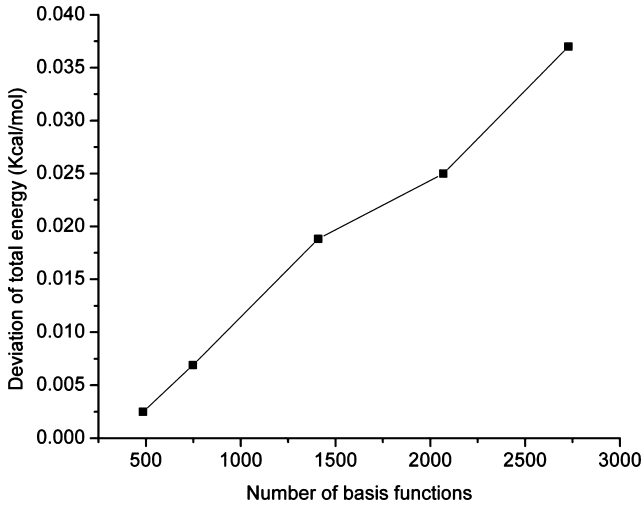

Figure 5. Accuracy of the total energy calculated by the DC-HF approach on extended polyglycine systems compared to full system calculations.

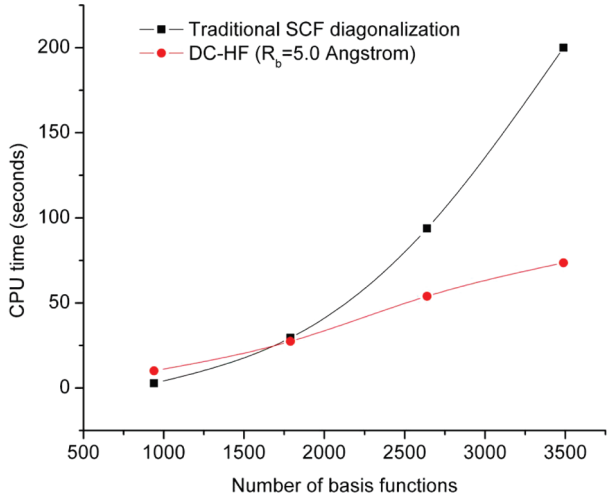

Figure 6. Similar to Figure 4 but for the polyalanine systems in an α-helical structure α-(Ala)$_n$.

initial guess using molecular fractionation with conjugated caps (MFCC) method to reduce the number of SCF cycles, and different division schemes are compared.

Computational Approaches

Divide and Conquer Approach on the Hartree−Fock Calculations. In protein systems, the divide and conquer approach is based on the chemical locality; this assumes that local regions of a protein are only weakly influenced by the atoms that are far away from the region of interest. The whole system is divided into fragments called core regions (Core$^\alpha$). A buffer region (Buffer$^\alpha$) is assigned for each core region to account for the environmental effects. The combination of every core region and its buffer region constitutes each individual subsystem (R^α) as illustrated in Figure 1. Local MOs of each subsystem are solved by the Roothaan−Hall equation

$$F^\alpha C^\alpha = S^\alpha C^\alpha E^\alpha \tag{1}$$

where F^α and S^α are local Fock matrix and local overlap matrix, respectively.

$$F^\alpha_{\mu\nu} = \begin{cases} F_{\mu\nu} \text{ if } \chi_\mu \in R^\alpha \text{ and } \chi_\nu \in R^\alpha \\ 0 \text{ elsewhere} \end{cases} \tag{2}$$

After the local MO coefficient matrices C^α are obtained, the total density matrix of the whole system is given by

$$P_{\mu\nu} = \sum_{\alpha=1}^{N_{\text{sub}}} P^\alpha_{\mu\nu} = \sum_{\alpha=1}^{N_{\text{sub}}} D^\alpha_{\mu\nu} p^\alpha_{\mu\nu} \tag{3}$$

where $D^\alpha_{\mu\nu}$ is the partition matrix

$$D^\alpha_{\mu\nu} = \begin{cases} 1 & \phi_\mu \in \text{Core}^\alpha \text{ and } \phi_\nu \in \text{Core}^\alpha \\ {}^1/_2 & \phi_\mu \in \text{Core}^\alpha \text{ and } \phi_\nu \in \text{Buffer}^\alpha \text{ or} \\ & \phi_\mu \in \text{Buffer}^\alpha \text{ and } \phi_\nu \in \text{Core}^\alpha \\ 0 & \phi_\mu \notin \text{Core}^\alpha \text{ and } \phi_\nu \notin \text{Core}^\alpha \end{cases} \tag{4}$$

and $p^\alpha_{\mu\nu}$ is the local density matrix defined by

$$p^\alpha_{\mu\nu} = \sum_i^{\text{LMOs}} n^\alpha_i C^\alpha_{\mu i} C^{\alpha *}_{\nu i} \tag{5}$$

where n^α_i is a smooth approximation to the Heaviside step function

$$n^\alpha_i = \frac{2}{1 + \exp[(\varepsilon^\alpha_i - \varepsilon_{\text{F}})/kT]} \tag{6}$$

ε_F is determined through the normalization of the total number of electrons of the whole system

$$N_{\text{elec}} = \sum_\alpha \sum_\mu (P^\alpha S^\alpha)_{\mu\mu} \tag{7}$$

After the density matrix is converged, the total HF energy is given as

$$E^{\text{DC}}_{\text{HF}} = \frac{1}{2} \sum_\alpha \sum_{\mu\nu} P^\alpha_{\mu\nu} (H^\alpha_{\mu\nu} + F^\alpha_{\mu\nu}) \tag{8}$$

where $H^\alpha_{\mu\nu}$ is the local one-electron core Hamiltonian matrix similar to the definition of local Fock matrix in eq 2.

For HF calculations on large systems, the construction of the Coulomb matrix and exchange matrix along with the diagonalization of the Fock matrix constitute the three most time-consuming steps. The Hamiltonian matrix diagonalization intrinsically scales as O(N^3). In the divide and conquer scheme the diagonalization calculation is performed on each submatrix, which will naturally make the SCF diagonalization step scale linearly with the number of subsystems. However, it is important to realize that the divide and conquer algorithm does not help to reduce the scale of computation of the Coulomb matrix and exchange matrix. The continuous fast multipole method (CFMM)[14,16−18,25,49−51] and the linear exchange K approach (LinK)[52,53] provide ways in which the Coulomb matrix and exchange matrix can be made linear scaling, respectively.

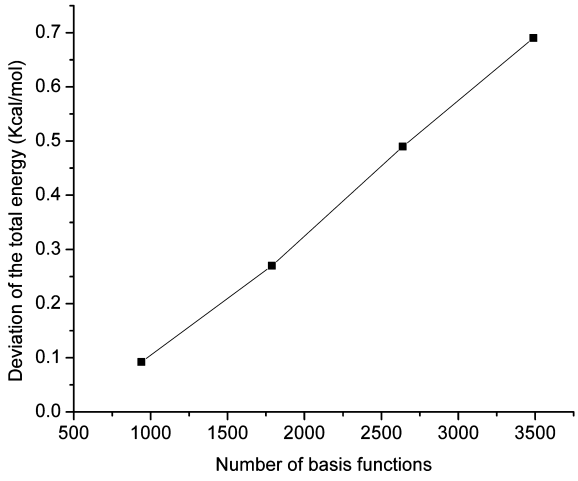

Figure 7. Similar to Figure 5 but for the polyalanine systems in an α-helix structure α-(Ala)$_n$.

Table 1. Number of SCF Cycles Needed To Reach Convergence for the SAD and MFCC Initial Guess at the HF/6-31G* Level

	DC		non-DC[a]	
system	SAD initial guess	MFCC initial guess	SAD initial guess	MFCC initial guess
Gly_6	18	10	12	7
Gly_{10}	18	11	12	7
Gly_{20}	18	10	12	6
Gly_{30}	18	10	12	6
Gly_{40}	18	8	12	7
α-$(Ala)_{10}$	18	15	12	9
α-$(Ala)_{20}$	16	12	12	9
α-$(Ala)_{30}$	16	12	12	8
α-$(Ala)_{40}$	15	12	12	8

[a] In the SCF procedure of the non-DC case every 10 previous Fock matrices were stored in the DIIS calculations, while for the DC case every 2 previous Fock matrices were stored in the DIIS calculations until the root-mean-squared change of the density matrix elements reaches 10^{-4} au, after which the DIIS technique was turned off.

MFCC Initial Guess. Here we introduce a fragment-based initial guess for ab initio calculations using the molecular fractionation with conjugate caps (MFCC) algorithm as described elsewhere.[42,54,55] In the spirit of the MFCC approach, the full density matrix of the protein can be assembled by a linear combination of fragment density matrices

$$P_{\mu\nu} = \sum_{i=1}^{N_f} P^f_{\mu\nu}(i) - \sum_{j=1}^{N_c} P^{cc}_{\mu\nu}(j) \quad (9)$$

where $P^f_{\mu\nu}(i)$ is the density matrix element of the ith protein fragment and $P^{cc}_{\mu\nu}(j)$ is the density matrix element of the jth conjugate cap. N_f and N_c are the total number of the protein fragments and conjugate caps, respectively. The MFCC partition scheme to cut a protein into amino acid fragments and conjugate caps is shown in Figure 2. First, a series of single-point HF calculations on all the fragments and conjugate caps are performed; then the full density matrix of the protein obtained using the converged fragment density matrices based on eq 9 is taken as the initial guess for the subsequent divide and conquer HF calculations. All ab initio calculations were implemented in an in-house-developed quantum chemistry package QUICK.[56]

Table 2. Converged Total Energies (au) (at the HF/6-31G* level) Using Two Different Subsetting Schemes: Residue Based with a Buffer of 5 Å and Atom Based with a Buffer of 7 Å[a]

system	residue-centric core region	atom-centric core region	deviation (kcal mol^{-1})
Gly_{10}	−2314.783296	−2314.783272	−0.015
Gly_{20}	−4382.595749	−4382.595726	−0.014
Gly_{30}	−6450.407962	−6450.407938	−0.015
Gly_{40}	−8518.221662	−8518.221679	0.011
α-$(Ala)_{20}$	−5164.086850	−5164.086911	0.038
α-$(Ala)_{30}$	−7622.660188	−7622.660373	0.116
α-$(Ala)_{40}$	−10081.238571	−10081.238839	0.168
MUD			0.054

[a] MUD: mean unsigned deviation.

Results and Discussion

Accuracy and Timing Comparisons. In this section we assess the DC-HF approach performance by performing calculations on two types of simple systems: extended polyglycine $(gly)_n$ and an α-helix of polyalanine (α -$(ala)_n$, see Figure 3). All calculations discussed here use the 6-31G* basis set. A buffer radius of $R_b = 5.0$ Å was adopted for all DC-HF calculations. Within this definition we include all the residues that contain any atoms within 5 Å from the core region as part of the buffer region. A comparison of the CPU time required to solve the SCF equations on the extended polyglycine $(gly)_n$ ($n = 6-40$) using the standard HF and DC-HF approaches is shown in Figure 4. As expected, the computational scale for the DC-HF diagonalization calculation is O(N), in contrast to $O(N^{2.9})$ for the traditional HF SCF diagonalization on the full Fock matrix of the entire system. Moreover, since the polyglycine is extended, the basis set cross-over point is between 485 and 749. Figure 5 shows the deviation of DC-HF energy compared to the full system calculation on extended polyglycine systems. The error becomes larger as the size of the system increases; however, all of the deviations are within 0.04 kcal mol^{-1}. This good accuracy suggests that we can employ the divide and conquer scheme to study large, 3-dimensional systems. The computational cost and accuracy of DC-HF for α-$(ala)_n$ ($n = 10-40$) systems are illustrated in Figures 6 and 7, respectively. Because of the helix structure, each subsystem contains a larger number of residues than in the extended system using the same buffer size. As illustrated in Figure 6, the cross-over point is around 1789, which is over 2 times larger than for the polyglycine example. Each DC-HF diagonalization SCF cycle in this example scales as $O(N^{1.1})$, in contrast to $O(N^{2.7})$ for the traditional HF diagonalization cost. Furthermore, the total energy errors for the α-helical polyalanines are slightly larger than those for the extended polyglycine systems (see Figure 7), but they are still in a good agreement with the full system calculations since the largest error is less than 0.7 kcal mol^{-1} for α-$(ala)_{40}$.

In the current DC-HF approach, the scale for the computation of the Coulomb matrix is still $O(N^2)$ after prescreening

Table 3. Total Energies (au) of Three-Dimensional Globular Proteins Obtained Using Standard Full System HF/6-31G* Calculations and Divide and Conquer HF/6-31G* Calculations Using the MFCC Initial Guess[a]

system	number of atoms	number of basis functions	standard full system calculation	DC using MFCC initial guess	deviation (kcal mol^{-1})
Trp-cage	304	2610	−7439.721780	−7439.722124	−0.22
1VTP	396	3418	−10014.756053	−10014.755741[b]	0.20
1BBA	582	5033	−15103.299186	−15103.302595	−2.14
1AML	598	5178	−15140.895905	−15140.897305[b]	−0.88
1BHI	591	5124	−15989.697592	−15989.696544	0.66
1BZG	573	4851	−13680.602670	−13680.602916[b]	−0.15
2JPK	589	5000	−13854.809422	−13854.810188[b]	−0.48
2KCF	576	4991	−14599.178617	−14599.180118	−0.94
2PPZ	608	5111	−14957.602116	−14957.605696	−2.25
2RLK	588	5089	−14589.701015	−14589.702771[b]	−1.10
2YSC	578	5108	−14634.254517	−14634.257181	−1.67
MUD					0.97

[a] MUD: mean unsigned deviation. [b] Did not converge using the SAD initial guess.

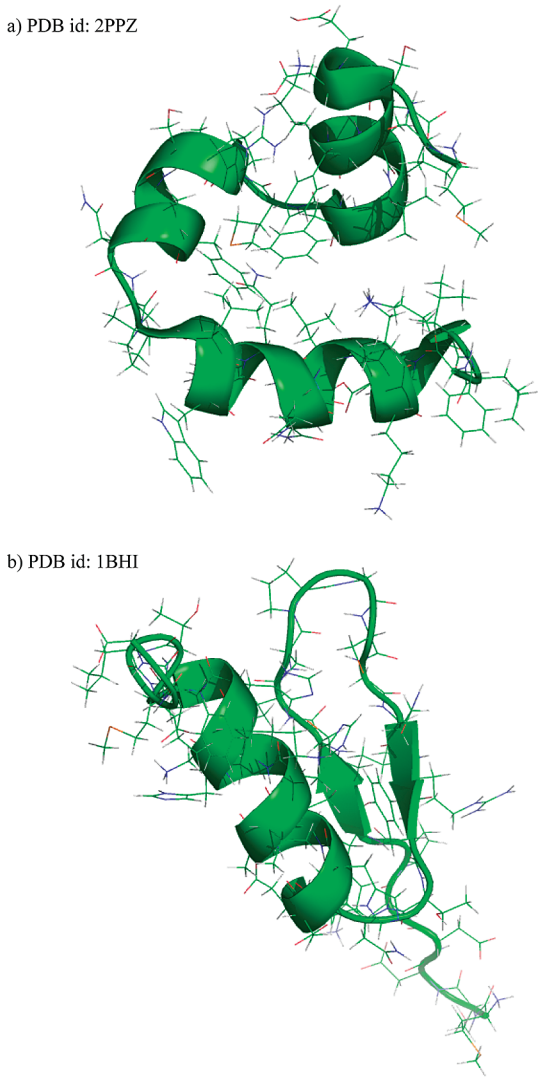

Figure 8. Two representative three-dimensional protein structures studied in this work.

the two-electron integrals.[10] When we apply eq 2 to construct the subsystem Fock matrix, the long-range Coulomb interactions between the local subsystem and distant atoms cannot be circumvented; thus, it should be emphasized that the DC algorithm itself does not reduce the scale of Coulomb and exchange matrix evaluations, and other approaches are necessary to achieve this result (e.g., CFMM).[14,16,17,49]

MFCC Initial Guess for DC-HF Calculations. Next, we compare the number of SCF cycles necessary to reach convergence when the SAD (superposition of atomic densities) and MFCC initial guesses are used in the divide and conquer scheme using the 6-31G* basis set (see Table 1). The convergence criterion in all examples was set to 10^{-6} au on the root-mean-squared change of the density matrix elements and 10^{-4} au for the maximum change of the density matrix elements. Nakai and co-workers[35] and Shaw and St-Amant[32] noted that DIIS causes SCF calculations to oscillate at the final stage of the SCF convergence process due to the slight errors introduced by the DC approximation for assembling the density matrix (see eq 3). In our HF DC calculations, the DIIS technique was turned off when the root-mean-squared change of the density matrix elements reaches 10^{-4} au. We also found that although DIIS works in the early stages of the SCF procedure, we get the best performance when only two previous Fock matrices were stored in the DIIS calculations. One can see from Table 1 that the HF DC calculations usually require more SCF cycles than the non-DC runs; however, for the polyglycine and polyalanine systems, the MFCC initial guess helps to reduce the number of SCF cycles in both DC and non-DC cases.

Residue-Centric Core Region versus Atom-Centric Core Region. Previously, all calculations used a residue-based definition for the core region. We also examined an atom-based subsetting strategy for the core region in polyglycines and polyalanines. One can see from Table 2 that the converged total energies using the atom-centric core region were almost identical to those using a residue-based cutoff. Indeed, the overall mean unsigned deviation is as low as 0.054 kcal mol^{-1}. This is an attractive aspect of the divide and conquer approach since it allows for parallelization at the atom level rather than at the much larger reside-based cutoff scheme.

Validation on Three-Dimensional Protein Systems. No previous studies have utilized the divide and conquer HF approach on three-dimensional globular proteins. In order to address this point, we validated the accuracy of divide and conquer HF/6-31G* calculations on 11 real proteins. The systems ranged from 304 to 608 atoms and are listed in Table 3. The proteins consisted of α-helical structures (see Figure 8a) or are mixed $\alpha-\beta$ structures (see Figure 8b). As shown

in Table 3, the largest deviation is 2.25 kcal mol^{-1} and the overall mean unsigned deviation is only 0.97 kcal mol^{-1} compared to standard full system calculations. Importantly, the observed error is larger than what was observed for the one-dimensional examples but is still within acceptable limits. This study sets the stage for the wide application of divide and conquer calculations on real protein systems. Furthermore, we found that for five proteins the divide and conquer HF calculations are not able to reach convergence using the simple SAD initial guess while all cases converged using the MFCC initial guess. Therefore, we conclude that the quality of the initial guess plays an important role in ensuring the convergence of divide and conquer calculations. Fragment-based electron density provides a much better quality initial guess with linear-scaling computational cost and, ultimately, much less computational time when compared to full system calculations.

Conclusions

In this study, divide and conquer HF theory was revisited in order to examine its potential to study three-dimensional constructs and a new and effective initial guess scheme was introduced. We first validated the accuracy of the divide and conquer HF/6-31G* calculations on 11 three-dimensional globular proteins. The overall mean unsigned error was within 1 kcal mol^{-1} when compared to standard full system calculations. Furthermore, we found that the fragment-based initial guess using the MFCC approach reduces the number of SCF cycles for polyglycine and polyalanine systems. Moreover, the MFCC initial guess facilitated SCF convergence for several of the globular proteins, where the SAD initial guess was unable to yield a converged wave function.

Acknowledgment. We thank the NIH (GM GM044974) for financial support of this research. Computing support from the University of Florida High Performance Computing Center is gratefully acknowledged.

References

(1) Szabo, A.; Ostlund, N. S. *Modern quantum chemistry: introduction to advanced electronic structure theory*, 1st ed.; McGraw-Hill: New York, 1989.

(2) Parr, R. G.; Yang, W. T. *Annu. Rev. Phys. Chem.* **1995**, *46*, 701.

(3) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 0618.

(4) Bartlett, R. J.; Musial, M. *Rev. Mod. Phys.* **2007**, *79*, 291.

(5) Čížek, J. *J. Chem. Phys.* **1966**, *45*, 4256.

(6) Crawford, T. D.; Schaefer, H. F. *Rev. Comput. Chem.* **2000**, *14*, 33.

(7) Kállay, M.; Gauss, J. *J. Chem. Phys.* **2005**, *123*, 214105.

(8) Kállay, M.; Surján, P. R. *J. Chem. Phys.* **2001**, *115*, 2945.

(9) Bomble, Y. J.; Stanton, J. F.; Kállay, M.; Gauss, J. *J. Chem. Phys.* **2005**, *123*, 054101.

(10) Strout, D. L.; Scuseria, G. E. *J. Chem. Phys.* **1995**, *102*, 8448.

(11) Schwegler, E.; Challacombe, M. *J. Chem. Phys.* **1996**, *105*, 2726.

(12) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085.

(13) Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 6904.

(14) Challacombe, M.; Schwegler, E. *J. Chem. Phys.* **1997**, *106*, 5526.

(15) Friesner, R. A.; Murphy, R. B.; Beachy, M. D.; Ringnalda, M. N.; Pollard, W. T.; Dunietz, B. D.; Cao, Y. X. *J. Phys. Chem. A* **1999**, *103*, 1913.

(16) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1994**, *230*, 8.

(17) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1996**, *253*, 268.

(18) Scuseria, G. E. *J. Phys. Chem. A* **1999**, *103*, 4782.

(19) Korchowiec, J.; Lewandowski, J.; Makowski, M.; Gu, F. L.; Aoki, Y. *J. Comput. Chem.* **2009**, *30*, 2515.

(20) Jiang, N.; Ma, J.; Jiang, Y. S. *J. Chem. Phys.* **2006**, *124*, 114112.

(21) Daniels, A. D.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 1321.

(22) Yang, W. T. *Phys. Rev. Lett.* **1991**, *66*, 1438.

(23) Yang, W. T.; Lee, T. S. *J. Chem. Phys.* **1995**, *103*, 5674.

(24) Kohn, W. *Phys. Rev. Lett.* **1996**, *76*, 3168.

(25) Strain, M. C.; Scuseria, G. E.; Frisch, M. J. *Science* **1996**, *271*, 51.

(26) Exner, T. E.; Mezey, P. G. *J. Phys. Chem. A* **2002**, *106*, 11791.

(27) Fusti-Molnar, L. *J. Chem. Phys.* **2003**, *119*, 11080.

(28) Fusti-Molnar, L.; Pulay, P. *J. Chem. Phys.* **2002**, *117*, 7827.

(29) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; DiStasio, R. A.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C. P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F.; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172.

(30) Dixon, S. L.; Merz, K. M. *J. Chem. Phys.* **1996**, *104*, 6643.

(31) Gogonea, V.; Westerhoff, L. M.; Merz, K. M. *J. Chem. Phys.* **2000**, *113*, 5604.

(32) Shaw, D. M.; St-Amant, A. *J. Theor. Comput. Chem.* **2004**, *3*, 419.

(33) Kobayashi, M.; Nakai, H. *Int. J. Quantum Chem.* **2009**, *109*, 2227.

(34) Akama, T.; Fujii, A.; Kobayashi, M.; Nakai, H. *Mol. Phys.* **2007**, *105*, 2799.

(35) Akama, T.; Kobayashi, M.; Nakai, H. *J. Comput. Chem.* **2007**, *28*, 2003.

(36) Kobayashi, M.; Akama, T.; Nakai, H. *J. Chem. Phys.* **2006**, *125*, 204106.

(37) Exner, T. E.; Mezey, P. G. *J. Comput. Chem.* **2003**, *24*, 1980.

(38) Exner, T. E.; Mezey, P. G. *J. Phys. Chem. A* **2004**, *108*, 4301.

(39) Exner, T. E.; Mezey, P. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 4061.

(40) Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2002**, *351*, 475.

(41) Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2006**, *433*, 182.

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

(43) He, X.; Zhang, J. Z. H. *J. Chem. Phys.* **2005**, *122*, 031103.

(44) Fedorov, D. G.; Ishimura, K.; Ishida, T.; Kitaura, K.; Pulay, P.; Nagase, S. *J. Comput. Chem.* **2007**, *28*, 1476.

(45) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2005**, *123*, 134103.

(46) Kobayashi, M.; Imamura, Y.; Nakai, H. *J. Chem. Phys.* **2007**, *127*, 074103.

(47) Kobayashi, M.; Nakai, H. *J. Chem. Phys.* **2008**, *129*, 044103.

(48) Dixon, S. L.; Merz, K. M. *J. Chem. Phys.* **1997**, *107*, 879.

(49) Schwegler, E.; Challacombe, M. *J. Chem. Phys.* **1999**, *111*, 6223.

(50) Burant, J. C.; Strain, M. C.; Scuseria, G. E.; Frisch, M. J. *Chem. Phys. Lett.* **1996**, *248*, 43.

(51) Shao, Y. H.; White, C. A.; Head-Gordon, M. *J. Chem. Phys.* **2001**, *114*, 6572.

(52) Ochsenfeld, C. *Chem. Phys. Lett.* **2000**, *327*, 216.

(53) Ochsenfeld, C.; White, C. A.; Head-Gordon, M. *J. Chem. Phys.* **1998**, *109*, 1663.

(54) Chen, X. H.; Zhang, J. Z. H. *J. Chem. Phys.* **2006**, *125*, 044903.

(55) Chen, X. H.; Zhang, Y. K.; Zhang, J. Z. H. *J. Chem. Phys.* **2005**, *122*, 184105.

(56) He, X.; Ayers, K.; Brothers, E.; Merz, K. M. *QUICK*; University of Florida: Gainesville, FL, 2008.

CT9006635