New Advance in Computational Chemistry: Full Quantum Mechanical ab Initio Computation of Streptavidin−Biotin Interaction Energy

Da W. Zhang, Yun Xiang, and John Z. H. Zhang*

Department of Chemistry, New York University, New York, New York 10003

Received: July 3, 2003; In Final Form: September 8, 2003

Benchmark full quantum mechanical Hartree−Fock calculation has been carried out to compute interaction energies for the streptavidin−biotin binding complex. In this report, the entire streptavidin−biotin interaction system with a total of 1775 atoms is treated by quantum mechanics. The full quantum energy calculation for this protein system is made possible by applying a recently developed MFCC approach in which the protein molecule is decomposed into amino-acid-based fragments that are properly capped. Ab initio calculations are performed at the Hartree−Fock level with a 3-21G basis set. The energies are computed for geometries of the binding complex near two configurations, corresponding to the crystal structure of the binding complex and a minimum energy geometry found from molecular force field, respectively. Comparisons are made of the computed ab initio energies with those from a force field. The present calculation shows that ab initio binding energies (at HF/3-21G level) are almost 30 kcal/mol larger than those given by a force field.

Fully quantum mechanical computation of protein interaction energy presents a grand challenge to computational chemists. The large size of proteins with thousands of atoms makes it impractical to attempt brute force ab initio calculations using standard quantum chemistry methods. Standard applications of ab initio methods such as HF (Hartree−Fock) and DFT (density functional method) are typically limited to systems of less than 100 atoms on standard workstations, due to steep scaling of computational time with the system size *N*. Efforts have been made to develop linear scaling methods to treat large biological systems over the past decade.[1−8] However, current applications of these methods are limited to semiempirical calculations for large biological systems such as proteins, due to huge computational cost. Recently, much effort has also been devoted to developing a hybrid quantum mechanical/classical mechanical (QM/MM) approach that combines quantum mechanics and molecular force field methods in energy calculations.[9−19]

We report benchmark ab initio computation of interaction energies for streptavidin−biotin binding using a new computational approach at ab initio HF level. The full ab initio computation is made possible by applying a new computational approach named molecular fractionation with conjugate caps (MFCC), which is recently developed for full quantum mechanical computation of interaction energies involving biological molecules such as proteins.[20,21] In the MFCC approach, the protein or peptide is decomposed into amino-acid-based fragments that are properly capped. Using this approach, computation of the interaction energy between a protein and another molecule can be carried out by separate ab initio calculations of interaction energies between protein fragments and the molecule of interest. The MFCC method is linear scaling, computationally efficient, and particularly suitable for computation on multi-processor computer systems.

The streptavidin−biotin system is one of the most tightly binding complexes for noncovalent binding of a protein and small ligand.[22] The high-affinity binding of biotin to streptavidin or closely related avidin has been exploited for decades. The highly specific and strong binding between these two molecules has remarkable technological utility. For example, the streptavidin−biotin system provides a means to examine a number of biochemical systems in biosensor applications, and it is one of the best characterized systems based on self-assembled monolayers.[23] This system has also served as an important focus for a wide variety of experimental and computational studies of ligand dissociation.[24−27] Theoretical calculations of free energies based on molecular dynamics simulation for streptavidin−biotin binding have also been reported.[31,32]

In the MFCC approach for computing protein/molecule interaction energy, the entire protein molecule is treated by quantum mechanics in a consistent fashion.[20] The MFCC method derives its great computational efficiency partly by foregoing some calculations of intra-protein interaction energy and focus on intermolecular interaction energy instead. The MFCC method has recently been successfully applied to calculating interaction energies for a protein−water system with a total of 985 atoms at HF, DFT, and MP2 levels.[21] In the present application of MFCC method to the streptavidin−biotin system, we first decompose the 121-amino acid streptavidin into 121 fragments by cutting all the backbone C_α−C bonds. At every position of cut, a pair of caps ($C_\alpha RH_2$- and $CONH_2$-) are inserted to cap the cutoff fragments. In addition, there are also 120 cap species formed by fusing pairs of conjugate caps inserted at all positions of cuts.

Using the MFCC approach, the expression of interaction energy for the streptavidin−biotin system (V(B−S)) is given by summations over individual interaction energies[20]

$$V(B-S) = \sum_{i=1}^{121} V(B-F_i) - \sum_{i=1}^{120} V(B-C_i)$$

where $V(B - F_i)$ and $V(B - C_i)$ denote the interaction energy of biotin with the *i* th fragment of streptavidin and with the *i* th cap fragment, respectively. Both streptavidin and biotin are kept rigid with structures given by PDB (protein data bank) id 1stp.

Figure 1 shows the crystal structure of streptavidin−biotin complex as obtained from PDB. A more detailed structure near

10.1021/jp0359081 CCC: $25.00
Published on Web 10/11/2003

Figure 1. Structure of streptavidin−biotin binding complex from protein data bank (PDB id 1stp).

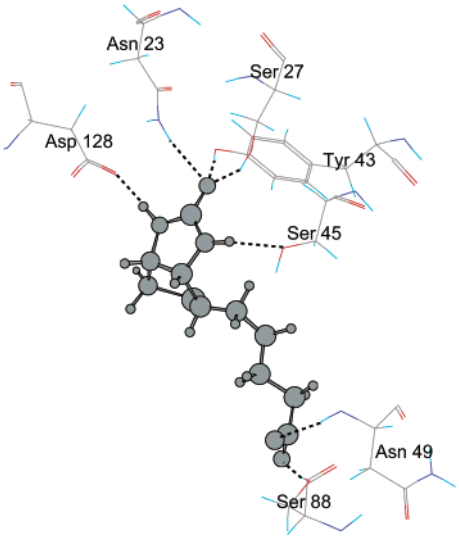

Figure 2. Hydrogen bonding network between biotin and related residuals of streptavidin. The broken lines denote intermolecular hydrogen bonding

the binding site where hydrogen bonding between biotin and residues of streptavidin are explicitly shown in Figure 2. In this study, the streptavidin and biotin are kept rigid throughout the calculation. We first employed molecular force field to perform a local energy optimization of geometries near the crystal structure by using the conjugate gradient method. A slightly shifted geometry with a somewhat lower energy is obtained from the optimization procedure. The MFCC ab initio calculation at HF/3-21G level was carried out to generate interaction energies near the vicinity of both structures. Due to computational cost, however, energy minimization using HF method was not performed. Instead, one-dimensional potential curves along each direction of the three-dimensional space are produced.

In our calculation, we fix the protein in space and move the rigid biotin along one of the three coordinate axes at a time. This generates three one-dimensional potential curves with the starting position as the origin of the coordinate. Figure 3a shows calculated streptavidin/biotin potential energy curves along three directions in which the origin of the coordinate represents the crystal structure of the complex. The three potential curves correspond to energy profiles for individual biotin movement along the (*x*,*y*,*z*) axes. Here the energy is normalized such that

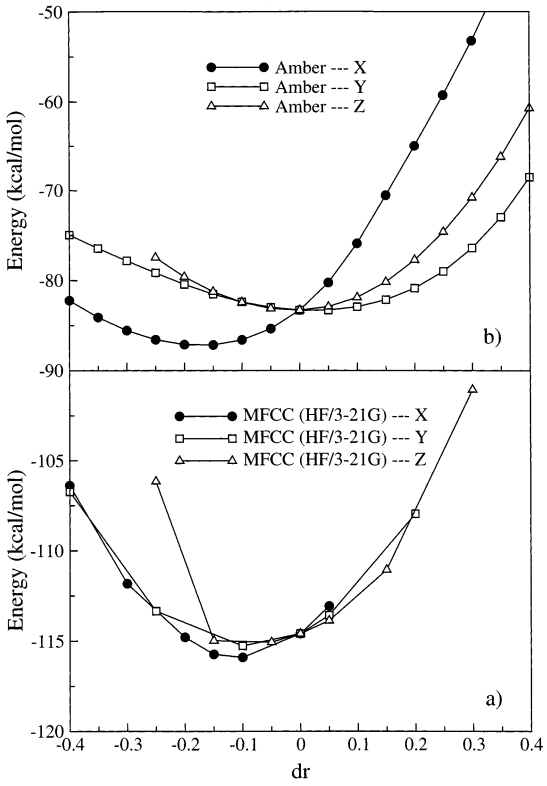

Figure 3. Streptavidin−biotin potential energy curves along (*x*,*y*,*z*) directions. The origin of the *x*-coordinate corresponds to the crystal structure from PDB (a) Results from MFCC HF/3-21G calculation. (b) Results from AMBER force field.

the infinite separation of the rigid protein and biotin has zero energy. As shown in Figure 3a, a slight shift ($dr = -0.1$ Å) gives lower energy than that at the origin (crystal structure). The calculated ab initio binding energy in Figure 3a is about 116 kcal/mol. Such strong noncovalent binding is largely the result of strong hydrogen bonding network formed between biotin and streptavidin. As illustrated in Figure 2, the complex can form a total of seven intermolecular hydrogen bonds between biotin and streptavidin, in addition to the weaker van der Waals attractions.

For comparison, we also show energies given by AMBER[33] force field in Figure 3b. We see from Figure 3 that the minimum energy geometries given by ab initio calculation and the force field are in generally good agreement with each other. This is not surprising, because the interaction energy of the complex is dominated by strong hydrogen bonding. The main difference between the ab initio and molecular force field calculation, however, is in calculated interaction energies. For instance, the binding energy (lowest energy) of the complex from force field calculation in Figure 3b is about 88 kcal/mol, while ab initio HF/3-21G calculation gives about 116 kcal/mol in Figure 3a, a difference of 28 kcal/mol!

One explanation for the large difference in binding energy between classical force field and quantum mechanical calculation is that the current force field does not account for electron polarization effect. This can be especially important when the molecular species is electrically charged such as in biotin. The failure to account for electron polarization is one of the major drawback of the current force field, and could seriously underestimate molecular interaction energies when significant charge redistribution is involved.

We also picked a slightly different geometry as a new origin with lower energy by shifting the biotin along the negative

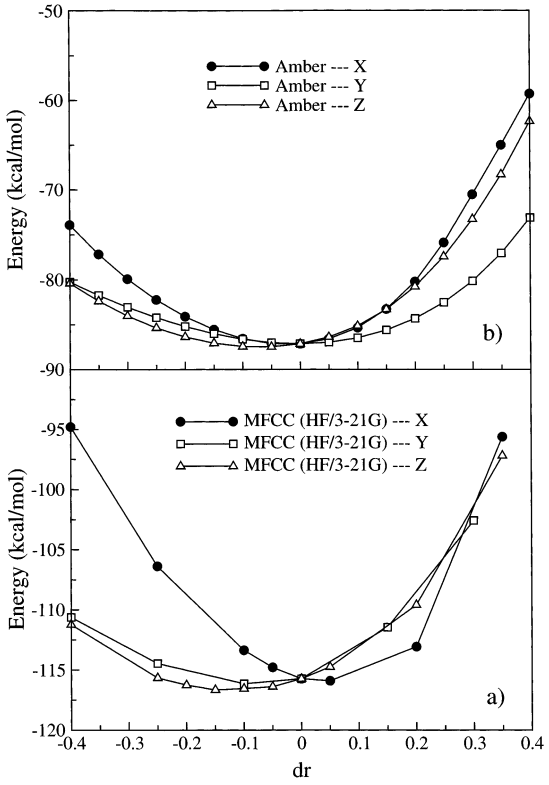

Figure 4. Streptavidin−biotin potential energy curves along (*x*,*y*,*z*) directions. The origin of the *x*-coordinate is slightly shifted away from the crystal structure obtained by energy minimization using molecular force field. (a) Results from MFCC HF/3-21G calculation. (b) Results from AMBER force field.

x-direction to $\Delta x = -0.15$ Å as shown in Figure 3b. We then computed potential energy curves again along all three directions at the same HF/3-21G level. The ab initio calculated potential curves are shown in Figure 4a together with the corresponding force field energies in Figure 4b. Here again, we see that the ab initio and force field calculations generally agree with each other on minimum energy geometries of the complex. The binding energies differ again by about 30 kcal/mol with the force field result giving energies too small. However, one should keep in mind that the calculated ab initio energies at HF/3-21G level can have large errors. Thus, we should not put too much emphasis on this energy difference until calculations at higher level ab initio calculations with larger basis set are performed.

This study presents the first full ab initio HF calculation of interaction energy for streptavidin−biotin complex, which is of significant technological importance. Applications of the existing quantum chemistry methods to such a large biological system with 1775 atoms are beyond the current limit and the present ab initio calculation is made possible by using the newly developed MFCC approach. To provide an idea of the computational efficiency of the MFCC approach, calculation of a single point energy at HF/3-21G level for the streptavidin−biotin complex (1775 atoms) takes about 15 h on a Dell 2.66 GHz Pentium 4 workstation running linux. This makes the MFCC approach extremely promising for practical applications in calculating protein interaction energies ab initio, especially when large multi-processor machines are utilized.

Some comments about the validity of MFCC method is in order here. For small peptides, the results of MFCC calculation have been compared with full system ab initio calculations. These comparisons show excellent agreement between the MFCC and full system calculations, and the results are indepdnent of the level of ab initio methods and the size of basis set used.[20,28] For large protein systems, no such comparison with full system ab initio calculation can be made. However, we expect the fragmentation approach will continue to work well, as evidenced by related works that also employed fragmentation scheme in a different context.[29,30] We hope to carry out more extensive studies to access the accuracy of the MFCC method in application to large biological systems.

References and Notes

(1) Yang, W. *Phys. Rev. Lett.* **1991**, *66*, 1438.
(2) Lee, T.-S.; York, D. M.; Yang, W. *J. Chem. Phys.* **1996**, *105*, 2744.
(3) Kohn, W. *Phys. Rev. Lett.* **1996**, *76*, 3168.
(4) Dixon, S. L.; Merz, K. M., Jr. *J. Chem. Phys.* **1996**, *104*, 6643.
(5) Gogonea, V.; Westerhoff, L. M.; Merz, K. M., Jr. *J. Chem. Phys.* **2000**, *113*, 5604.
(6) Stewart, J. J. P. *Int. J. Quantum Chem.* **1996**, *58*, 133.
(7) Daniels, A. D.; William, J. M.; Scuseria, G. E. *J. Chem. Phys.* **1997**, *107*, 425.
(8) Daniels, A. D.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 1321.
(9) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(10) Singh, U. C.; Kollman, P. *J. Comput. Chem.* **1986**, *7*, 718.
(11) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
(12) Gao, J.; Xia, X. *Science* **1992**, *258*, 631.
(13) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; Wiley: New York, 1991.
(14) Stanton, R. V.; Hartsough, D. S.; Merz, K. M., Jr. *J. Phys. Chem.* **1993**, *97*, 11868.
(15) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.
(16) Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46.
(17) Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290.
(18) Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **1999**, *20*, 1468.
(19) Castillo, R.; Andres, J.; Moliner, V. *J. Am. Chem. Soc.* **1999**, *121*, 12140.
(20) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599.
(21) Zhang, D. W.; Zhang, J. Z. H., to be submitted.
(22) Weber, P. C.; Ohlendorf, D. H.; Wendoloski, J. J.; Salemme, F. R. *Science* **1989**, *243*, 85−88.
(23) Wilcheck, M., Bayer, E. A., Eds.; *AvidinCbiotin Technology, Methods in Enzymology*; Academic Press: San Diego, 1990; Vol. 184.
(24) Moy, V. T.; Florin, E. L.; Gaub, H. E. *Science* **1994**, *266*, 257−259.
(25) Grubmuller, H.; Heymann, B.; Tavan, P. *Science* **1996**, *271*, 997−999.
(26) Freitag, S.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 8384−8389.
(27) Hyre, D. E.; Amon, L. M.; Penzotti, J. E.; Le Trong, I.; Stenkamp, R. E.; Lybrand, T. P.; Stayton, P. S. *Nature Struct. Bio.* **2002**, *9*, 582−585.
(28) Zhang, D. W.; Chen, X. H.; Zhang, J. Z. H. *J. Comput. Chem.*, in press.
(29) Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2000**, *318*, 614.
(30) Amovilli, C.; Cacelli, I.; Campanile, S.; Prampolini, G. *J. Chem. Phys.* **2002**, *117z*, 3003.
(31) Dixon, R. W.; Kollman, P. *Proteins* **1999**, *36*, 471−473.
(32) Dixit, S. B.; Chipot, C. *J. Phys. Chem. A* **2001**, *105*, 9795−9799.
(33) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III.; DelBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1.