ARTICLE

A Non-derivative MFCC Optimization Study of Cyclohexapeptide Monohydrate[†]

Xi-hua Chen[a],   John Z. Zhang[a,b*]

a. Department of Chemistry, New York University, New York, NY 10003, USA;
b. Institute of Theoretical and Computational Chemistry, Key Laboratory of Mesoscopic Chemistry of MOE, College of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210093, China

The MFCC-downhill simplex method is presented to study the binding structure of small ligands in large molecular complex systems. This method employs the Molecular Fractionation with Conjugated Caps (MFCC) approach to compute the interaction energy-structure relation of the system and implements the downhill simplex algorithm for structural optimization. The method is tested on a molecular system of $cyclo$-AAGAGG·H_2O to optimize the binding position of water molecule to the fixed cyclohexapeptide. The MFCC-downhill simplex optimization results are in good agreement with the crystal structure. An MFCC-Powell optimization method which uses the Powell's minimization algorithm is also described and tested on the same system. The MFCC-downhill simplex optimization is more efficient than the MFCC-Powell method.

Key words: MFCC-downhill simplex method, Binding structure, Optimization

I. INTRODUCTION

The determination of molecular structure by energy optimization is of great significance to both theoretical and applied chemistry. In particular, we are interested in predicting the binding structure of a small molecule in large biomolecular host, such as that of protein-ligand complex systems. Unveiling this sort of binding structure is crucial to understanding biochemical processes and to rational design of therapeutic inhibitors for specific protein targets.

Due to the considerable size of proteins which are usually of thousands of atoms, current structural optimization of the protein-ligand complexes is almost exclusively based on molecular mechanical (MM) energies derived from empirical force fields. The potential energy is given by some predefined empirical formulas and parameters, such as those in AMBER [1], CHARMM [2] and others [3-5], etc. Despite of some success of the force field methods in computational modelling of biological systems, there are intrinsic limitations. The MM methods ignore the electronic and polarization effects; the empirical potential functions and parameters must be developed from a collection of small molecules and transferred to larger biomolecules.

To provide more reliable potential energy for the molecular structural optimization, it is desirable to employ *ab initio* quantum mechanical methods. As a fact, in many cases, molecular quantum mechanics has

successfully reproduced and even challenged crystallographical experiments for accuracy. However, applications of common quantum chemistry methods [6-8] are limited to relatively small molecules due to the steep computational scaling with system size. Although optimization function is provided in many available quantum chemistry packages, it is only practical for a small system of typically less than two hundred atoms. Recent development of linear scaling methods has made significant advance in extending quantum calculation to large molecular systems, but their current applications to biomolecular systems are largely limited to semi-empirical methods [9-18].

Recently, we have developed a new approach for full quantum mechanical computation of interaction energies of small molecules binding to large polymers such as proteins and DNA segments [19-21]. In this Molecular Fractionation with Conjugated Caps (MFCC) approach, a protein, for example, is decomposed into amino acid-based fragments. Pairs of conjugated molecular caps are inserted at the cuts to satisfy the valency of the fragments. This approach differs from that by Kitaura and coworkers in that although the latter also divides the protein into amino acid fragments but it uses an electric static field instead of conjugated caps [22]. Another related method was recently proposed by Amovilli and coworkers in which the fragments are capped with a single hydrogen atom and its applicability is therefore limited [23]. In MFCC, the conjugated caps are adjustable and can (i) satisfy the valency requirement of the bonds being cut and preserve the property of the bonds; (ii) approximately represent the chemical environment of the fragment being capped. The interaction energy of the protein at a given structure with another molecule can be obtained by quan-
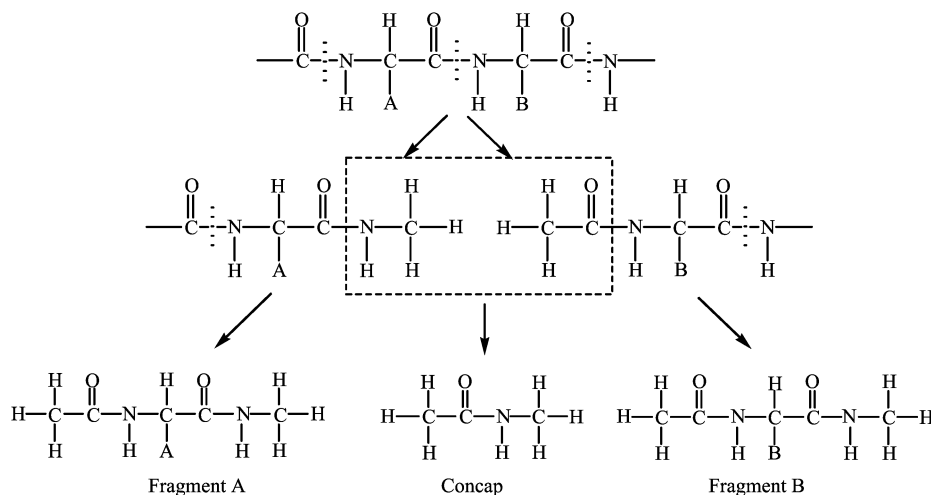
─────────

FIG. 1 The MFCC fragmentation of protein.

tum mechanical calculation of the interaction energies between the molecule and individual protein fragments and their conjugated caps (Concap molecules). Previous studies demonstrated that the MFCC approach is capable of giving excellent *ab initio* energies compared to the full system *ab initio* calculations [20,21]. It has also been applied recently to a number of real protein-ligand systems with thousands of atoms, such as HIV-I gp41-water [24] and adipocyte lipid-binding protein [25]. In addition, The MFCC approach is linear scaling and enables one to carry out practically first principle quantum mechanical calculations of interaction energies involving large biological molecules in a consistent quantum mechanical fashion on a variety of available *ab initio* methods such as HF, DFT, MP2, etc.

Since MFCC is able to simulate the interaction energy-structure relation [21], it is natural to apply MFCC to molecular structure optimization. There are many well-known optimization algorithms [26-28]. Although energy derivative-based optimization methods are generally more accurate and efficient for energy optimization, sometimes analytical derivatives of energies may be difficult to obtain and only energy values are available. The downhill simplex method [28-31] is one of those that require only function without derivative evaluations. By this algorithm, for a function of M variables, a set of $M+1$ vertices (points) are guessed to construct the initial simplex and the function is evaluated at each vertex. The vertex with the highest value is then replaced by anohter point which is obtained via one of the four proper operations: (a) a reflection, (b) a reflection and expansion away from the discarded point; (c) a contraction along one direction from the discarded point; and (d) a contraction along all dimensions towards the vertex with the lowest function value. With a series of such steps, the simplex adpapts the function landscape about the simplex and converges to the final minimum effectively. Powell's direction set method

is another gradient-free optimization algorithm [28,32]. By producing successive sets of mutually conjugated vector directions, the algorithm performs line searching until the minimum is located.

In the previous research [33], we have proposed an MFCC-downhill simplex method for the optimization of the binding structure of an ion to a larger molecules which can be treated by the MFCC approach [33]. The test on a 18-crown-6 adduct of potassium organometallic compound, [KCp(18-crown-6)] (Cp=cyclopentadienyl) shows the MFCC-downhill optimization results in good agreement both with the crystal structure and with the full-sysytem-downhill simplex optimizied structure but spends much less computation cost campared to that of full-system optimzation. It is the simplest case with only three dimensions (variables) because only the binding position of potassium ion is optimized.

In this work, we generalize the MFCC-downhill simplex method to optimize a small ligand molecule binding to a large host. The method is tested on a cyclic hexapeptide hydrate, Cyclo(L-ananyl-L-ananyl-glycyl-L-ananyl-glycyl-glycyl) hydrate [34] which is denoted as *cyclo*-AAGAGG·H$_2$O for convenience. An MFCC-Powell optimization method in which the interaction energy is computed by MFCC while Powell's minimization algorithm is adopted for optimization is also described and tested on this system. All the numerical energy calculations are obtained using Gaussian98 package [36].

II. THEORETICAL METHOD AND COMPUTATION DETAILS

The MFCC approach is developed to compute the interaction energy between a rigid protein and its binding ligand [19]. The main idea of the MFCC approach is to divide a protein molecule into fragments (amino acid
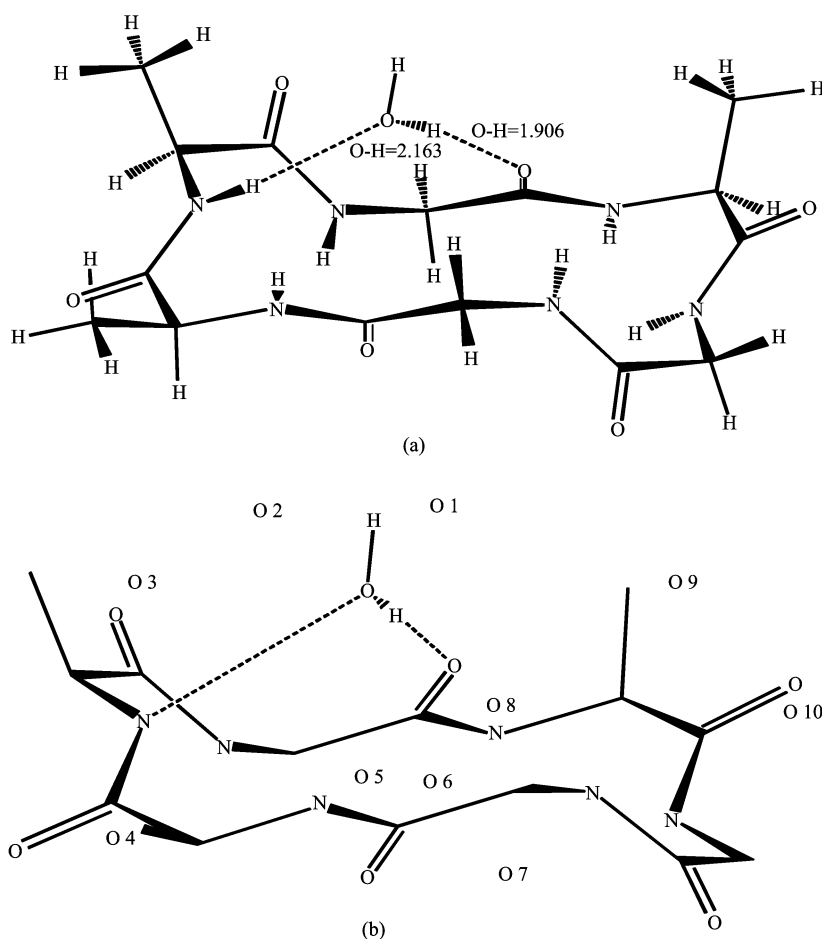
FIG. 2 (a) The crystal structure of *cyclo*-AAGAGG·H$_2$O. There are two H-bonds between H$_2$O and the cyclohexapeptide. (b) The full dimension (9-D) initial simplex of H$_2$O. For clarity, the cyclohexapeptide is shown with skeleton and H$_2$O molecules at the 10 vertices with only O atoms. H$_2$O in crystal structure remains for comparison with the initial vertices.

residues). Then both ends of each fragment are properly capped with a pair of chemical groups, or caps, to satisfy the valence requirement. The adjacent caps at each location of cut are conjugate in that they can fuse to form a proper molecular species, namely, concaps. Figure 1 illustrates how the MFCC scheme decomposes protein into fragments and concaps.

The interaction energy between the protein and its ligand, E_{P-L}, can therefore be approximated by the sum of individual fragments-ligand interaction substract the sum of concaps-ligand interaction [19,20]:

$$E_{P-L} = \sum_{i=1}^{n}(E_{F_i-L} - E_{F_i} - E_L)$$
$$- \sum_{i=1}^{n-1}(E_{CC_i-L} - E_{CC_i} - E_L) \qquad (1)$$

here, E_{F_i-L} is the ith fragment-ligand energy, E_{CC_i-L} the ith concap-ligand energy, E_{F_i} and E_{CC_i} are the energies of the ith fragment and ith concaps respectively, and E_L is the ligand's energy. For a protein of n amino

acids without crosslink, there are $n-1$ concaps needed. When the protein has crosslinks such as disulfide bonds, additional cutting of these bonds are needed [35].

In MFCC approach, the protein can also be cut at other locations as has been tested before [20], but cutting the peptide bond seems to be preferable in overall considerations. The concaps are also variable.

In this work, the system tested is a cyclic hexapeptide hydrate, (*cyclo*-AAGAGG·H$_2$O). The crystal structure of *cyclo*-(L-Ala-L-Ala-Gly-L-Ala-Gly-Gly)·2H$_2$O was determined and analyzed in detail by Hossain and Helm [34]. Two water molecules are almost symmetrically above and below the cyclic hexapeptide ring, each hydrogen-bonding to a closest amide and a closest carbonyl group. For simplicity, we choose one of the water molecule to optimize (Fig.2(a)). The peptide ring is treated with the general MFCC scheme and is fragmentated into six residues that are then capped. The resulted fragments are three 2-acetylamino-*N*-methyl-propionamine (CH$_3$CONHCH(CH$_3$)CONHCH$_3$) for the alanine residues and three 2-acetylamino-*N*-methyl-acetamine (CH$_3$CONHCH$_2$CONHCH$_3$) for the glycine

residues. The six concaps are uniformly N-methyl-acetamide ($CH_3CONHCH_3$). The interaction energy between the water molecule and the cyclic hexapeptide is now simulated by a sum of the interaction energies between the water and the fragments subtract those between the water and the concaps. In this illustrative system, the binding structure of water to the peptide is dominated by the interaction, in particular, the two hydrogen bonds between water and the peptide. The available crystal structure provides a good reference for theoretical structure optimization.

The MFCC-downhill simplex method is thus applied to optimize water position in the fixed cyclic hexopep-tide. For a ligand of N atoms, there are $M=3N$ co-ordinate variables and the initial simplex must contain $M+1$ vertices. But if the ligand can be approximated by a rigid body, then the dimensions drops to $M=6$, including three translational and three rotational co-ordinates. In our test, the water is firstly treated in full dimension (9-D). The initial simplex is shown in Fig.2(b). The 10 vertices are arbitrarily choosen. They form an irregular decahedron surrounding the crystal position of water molecule. The vertices' RMSDs (root mean square deviation) from the crystal water structure range from 1.0 Å to 3.0 Å.

An MFCC-downhill simplex optimization is then carried out until the simplex converges. One should note that in this 9-D case, energy refers to a summation of two parts: the interaction energy of water-cyclohexopeptide which is computated by MFCC approach and the relative energy of the variational water with respect to that of an isolated water molecule. The quantum chemical energy calculations are performed with HF/6-31+G* and HF/3-21G. Secondly, the water molecule is treated as a rigid body (6-D). Quaternion is used to represent the rotation coordinates. The seven vertices of the initial simplex are arbitrarily chosen. They form an irregular heptahedron surrounding the crystal position of the water molecule. The picture is similar to that of the 9-D simplex and is thus omitted. The quantum chemical calculations are performed with HF/3-21G, HF/6-31+G* and a higher level of MP2/6-31+G*. Finally, a full dimension (9-D) optimization is performed by MFCC-Powell approach to compare its efficiency with that by MFCC-downhill simplex method. The quantum chemical calculations are performed with HF/3-21G. For all optimization, the energy convergence criterion is set as 10^{-8} a.u..

III. RESULTS AND DISCUSSION

The first MFCC-downhill simplex optimization performed on *cyclo*-AAGAGG·H_2O is to optimize H_2O full dimensionally. Since H_2O is strongly binding to the cyclohexapeptide via two hydrogen bonds, it is expected that the internal structure of H_2O will be slightly different from that of a free H_2O molecule. In this case,

the driving force of optimization is a summary energy of the the interaction energy of water-cyclohexopeptide and the relative energy of the variational water with respect to that of an isolated water molecule(optimized by HF/6-31+G*). Starting from the 10 vertices of the initial simplex, a total number of 282 vertices are visited before the optimization converges at an energy criterion of 10^{-8} a.u. at the last 9 vertices. The overall profile of energy evolution is shown in Fig.3(a). The energies at some first vertices are very high due to the close contact of H_2O with the cyclohexapeptide. But after about the 110th vertices, the energy is lowered to a fine tune stage, as shown more clearly in Fig.3(b). Almost half of the time of optimization is spent to achieve the high energy convergence criterion. Further research might use MFCC-downhill simplex optimization at the beginning and some derivative optimization method at the end to speedup convergence.

The root mean square deviation (RMSD) of H_2O at each vertices is shown in Fig.3(c). At about the 110th vertices, the simplex has been found in an area very close to the crystal structure of H_2O. According to the algorithm, the final 9 meet the energy convergence criterion vertices. Thus we take the average of the 9 structures as the final result which is in good agreement with the crystal structure. The RMSD of this structure is 0.0592 Å.

Figure 3(d) gives the profile of the H_2O internal energy relative to a free H_2O_i energy. The free H_2O is one optimized by HF/6-31+G*. The relative internal energy of the final structure is 0.120 kJ/mol higher than the free H_2O, compared to 6.659 kJ/mol of the experimental H_2O in the crystal. It will be interesting to compare the final H_2O structure with its crystal structure and theoretical prediction of a free H_2O molecule (Table I). After MFCC-downhill simplex optimization, the internal structure of H_2O is relaxed for lower internal energy while the binding hydrogen bonds are reproduced.

TABLE I Comparison of the structures of H_2O optimized in this work (Final H_2O), free H_2O optimized by HF/6-31+G*, and H_2O in *cyclo*-AAGAGG · H_2O crystal (Exp.). The two hydrogen bonds, $O-H_1 \cdots O$ and $O-H(N)$, are shown in Fig.1(a). E is the energy relative to the free H_2O.

	Final H_2O	Exp.	Free H_2O
$H-O-H/(°)$	107.00	107.68	106.54
$O-H_1/Å$	0.942	0.910	0.948
$O-H_2/Å$	0.950	0.920	0.948
$H_1 \cdots O/Å$	1.942	1.906	N/A
$O \cdots H(N)/Å$	2.199	2.163	N/A
$E/(kJ/mol)$	0.120	6.659	0.0

In the practical application to study ligand-protein binding structure, it is often desirable to reduce the number of coordinate variables. The ligand might be treated as a rigid body so that there are only six coordi-
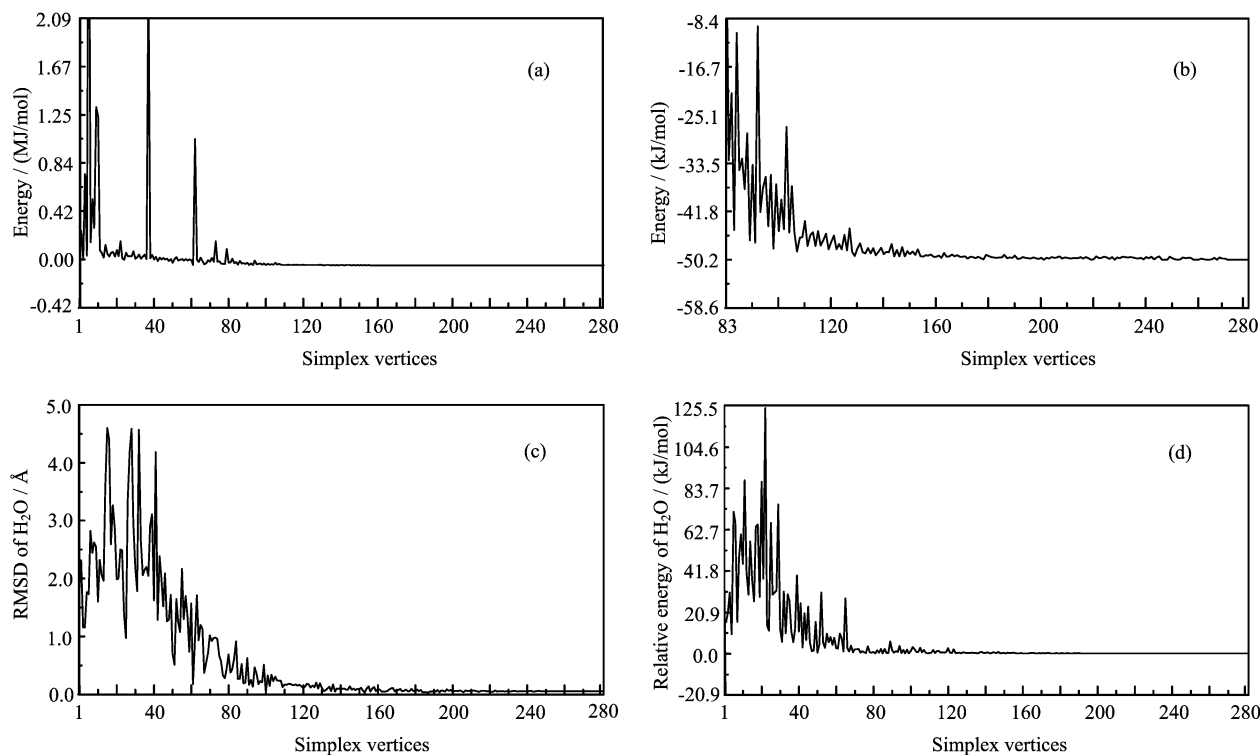
FIG. 3 Full dimension(9-D) MFCC-downhill simplex optimization of *cyclo*-AAGAGG·H_2O. level: HF/6-31+G*. (a) The overall profile of energy evolution. (b) Energy evolution at the last 200 vertices. (c) RMSD evolution of H_2O from its crystal structure in *cyclo*-AAGAGG·H_2O. (d) The profile of H_2O energy evolution during optimization.
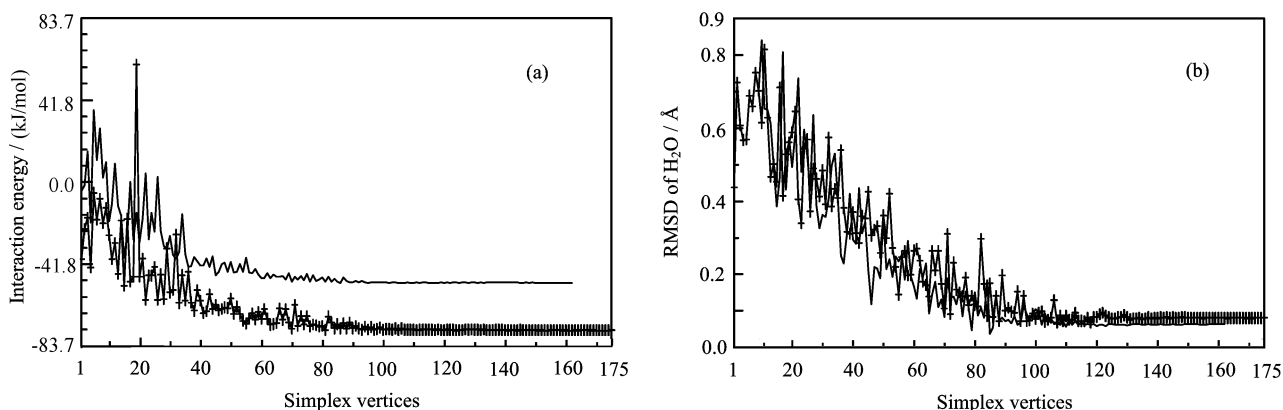

FIG. 4 Rigid body (6-D) MFCC-downhill simplex optimization of *cyclo*-AAGAGG·H_2O. Ligand H_2O is treated as a rigid body. HF/6-31+G* results are in solid line and MP2/6-31+G* results in dashes line with pluses. (a) Profile of interaction energy evolution. (b) RMSD evolution of H_2O from its crystal structure in *cyclo*-AAGAGG·H_2O.

nate variables, or a semi-rigid body that allows some interested coordinates to vary. The starting ligand structure can be taken either from experimental measurement or from *ab initio* optimization. In the illustrative system of *cyclo*-AAGAGG·H_2O, the binding structure of H_2O to the cyclohexapeptide is dominated by the two hydrogen bonds. This allows one to treat the H_2O as a rigid body when looking for its binding position. We therefore test the rigid-body MFCC-downhill simplex optimization on the *cyclo*-AAGAGG·H_2O system with

the structure of a free H_2O molecule which is previously optimized with HF/6-31+G*. Two optimizations are carried out with *ab initio* levels of HF/6-31+G* and MP2/6-31+G*, respectively, both starting from the same initial 7-vertex simplex.

Figure 4 (a) and (b) show the energy and RMSD profiles of both optimizations. Optimization using MP2 visits 175 vertices before convergence while that using HF takes 162 vertices. Both of the final H_2O structures are in good agreement with its crystal structure. The
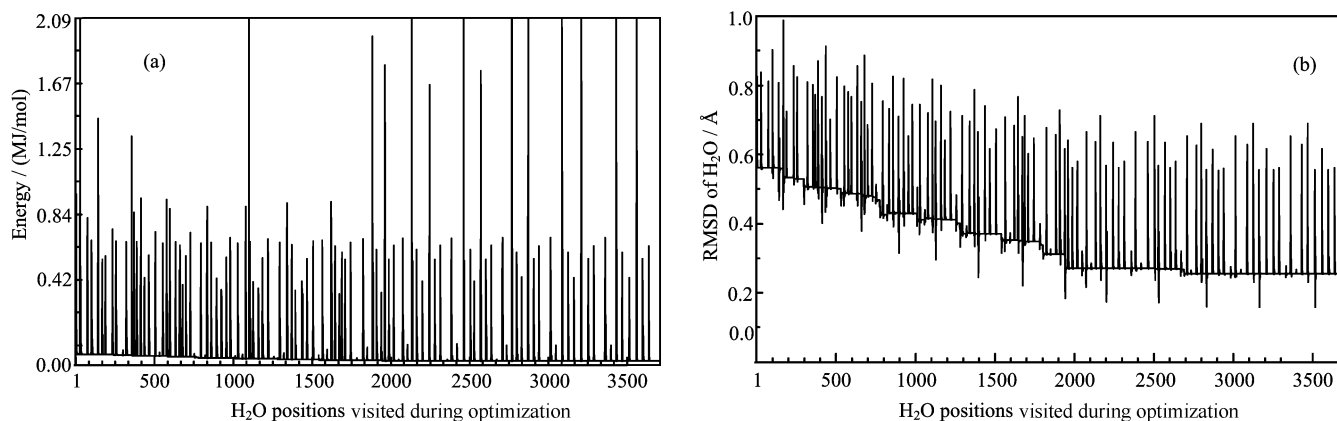
FIG. 5 Full dimension (9-D) MFCC-Powell optimization of *cyclo*-AAGAGG·H_2O. level: HF/3-21. (a) Profile of interaction energy evolution. (b) RMSD evolution of H_2O from its crystal structure in *cyclo*-AAGAGG·H_2O.

RMSD is 0.080 Å in the case of MP2 and 0.062 Å in the case with HF. MFCC-downhill simplex optimization using HF can achieve comparable result with that using more expensive MP2.

It is interesting to compare the MFCC optimization approach with the downhill simplex optimization and that with another non-derivative algorithm, Powell's method. Analogous to MFCC-downhill simplex optimization approach, MFCC-Powell method uses the MFCC approach to compute the interaction energy-structure relation of the system but implements the Powell's algorithm [28,33] for structural optimization. It is tested on the same *cyclo*-AAGAGG·H_2O to optimize H_2O full dimensionlly(9-D) with an *ab initio* level of HF/3-21G. The energy and RMSD profiles are shown in Fig.5. It computes the energies at 3711 points before convergence. Clearly, the MFCC-downhill simplex optimization is much more efficient than the MFCC-Powell method.

IV. CONCLUSION

In this work, we generalize the MFCC-downhill simplex optimization method to study the binding structure of samll ligands in large molecular systems. By this method, the large host molecule is fixed and treated by an appropriate MFCC (Molecular Fractionation with Conjugated Caps) scheme into small fragment molecules and concap molecules. Therefore the interaction energy-structure relation of the small molecule and the large host can be calculated by quantum chemical methods. The ligand can either be treated with full dimension to optimize both its own structure and binding position to the host, or be treated as a rigid body that allows only the binding position is optimized. The downhill simplex algorithm is employed to carry out structural optimization. Then the MFCC-downhill simplex method is tested on a

cyclo-AAGAGG·H_2O to optimize the structure of H_2O molecule. The MFCC-downhill optimization results in good agreement with the crystal structure. The positive results can be extended to larger ligand-protein systems and imply the potential application of the MFCC-downhill simplex method to a variety of areas such as drug design, drug encapsulation, enzyme catalysis, ect. Finally, an MFCC-Powell optimization method which uses the Powell's minimization algorithm is also described and tested on the same system. The MFCC-downhill simplex optimization is much more efficient than the MFCC-Powell method.

[1] D. A. Case, D. A. Pearlman, J. W. Caldwell, T. E. Cheatham III, J. Wang, W. S. Ross, C. Simmerling, T. Darden, K. M. Merz, R. V. Stanton, A. Cheng, J. J. Vincent, M. Crowley, V. Tsui, H. Gohlke, R. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. Weiner, and P. A. Kollman, *AMBER 7*, San Francisco: University of California, (2002).
[2] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[3] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. diNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[4] A. T. Hagler, E. Huler, and S. Lifson, J. Am. Chem. Soc. **96**, 5319 (1977).
[5] G. Nemethy, M. S. Pottle, and H. A. Scheraga, J. Phys. Chem. **87**, 1883 (1983).
[6] W. J. Hehre, L. Radom, P. Schleyer, and J. Pople, *Ab initio Molecular Orbital Theory*, New York: Wiley, (1986).
[7] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry*, New York: McGraw-Hill, (1989).
[8] R. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, New York: Oxford University Press, (1989).

[9] W. Yang, Phys. Rev. Lett. **66**, 1438 (1991).

[10] T. S. Lee, D. M. York, and W. Yang, J. Chem. Phys. **105**, 2744 (1996).

[11] W. Yang and T. S. Lee, J. Chem. Phys. **103**, 5674 (1995).

[12] W. Kohn, Phys. Rev. Lett. **76**, 3168 (1996).

[13] S. L. Dixon and K. M. Merz, Jr., J. Chem. Phys. **104**, 6643 (1996).

[14] V. Gogonea, L. M. Westerhoff, and K. M. Jr. Merz, J. Chem. Phys. **113**, 5604 (2000).

[15] J. J. P. Stewart, Int. J. Quantum Chem. **58**, 133 (1996).

[16] A. D. Daniels, J. M. William, and G. E. Scuseria, J. Chem. Phys. **107**, 425 (1997).

[17] A. D. Daniels and G. E. Scuseria, J. Chem. Phys. **110**, 1321 (1999).

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

[19] D. W. Zhang, and J. Z. H. Zhang, J. Chem. Phys. **119**, 3599 (2003).

[20] D. W. Zhang, X. H. Chen, and J. Z. H. Zhang, J. Comput. Chem. **24**, 1846 (2003).

[21] T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayasi, and K. Kitaura, Chem. Phys. Lett. **318**, 614 (2000).

[22] X. H. Chen, and J. Z. H. Zhang, J. Chem. Phys. **120**, 11386 (2004).

[23] C. Amovilli, I. Cacelli, S. Campanile, and G. Prampolini, J. Chem. Phys. **117**, 3003 (2002).

[24] D. W. Zhang and J. Z. H. Zhang, J. Theo. Comput. Chem. **1**, 43 (2004).

[25] Y. Xiang, D. W. Zhang, and J. Z. H. Zhang, J. Comput. Chem. **25**, 1431 (2004).

[26] R. P. Brent, *Algorithms for Minimization Without Drivatives*, Englewood Cliffs, NJ: Prentice-Hall, (1973).

[27] J. E. Dennis and R. B. Schnabel, *Numberical Methods for Unconstrained Optimization and Nonlinear Equa-*

tions, Englewood Cliffs, NJ: Prentice-Hall, (1983).

[28] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd Ed., Cambridge: Cambridge University Press, (1992).

[29] W. Spendley, G. R. Hext, and F. R. Himsworth, Technometrics **4**, 441 (1962).

[30] J. A. Nelder and R. Mead, Computer J. **7**, 308 (1965).

[31] L. A. Yarbro and S. N. Deming, Anal. Chem. Acta **73**, 391 (1974).

[32] X. H. Chen, D. W. Zhang, and J. Z. H. Zhang, J. Chem. Phys. **120**, 839 (2004).

[33] M. J. D. Powell, Computer J. **2**, 155 (1964).

[34] X. H. Chen and J. Z. H. Zhang, J. Theo. Comput. Chem. **3**, 277 (2004).

[35] M. B. Hossain and D. van der Helm, J. Am. Chem. Soc. **100**, 5191 (1978).

[36] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, *GAUSSIAN 98, Revision A.6*, Pittsburgh, PA: Gaussian Inc., (1998).