Geometry Optimizations and Vibrational Spectra of Large Molecules from a Generalized Energy-Based Fragmentation Approach

Weijie Hua,[†] Tao Fang,[†] Wei Li,[†] Jian-Guo Yu,[‡] and Shuhua Li*,[†]

School of Chemistry and Chemical Engineering, Key Laboratory of Mesoscopic Chemistry of MOE, Institute of Theoretical and Computational Chemistry, Nanjing University, Nanjing 210093, P. R. China, and College of Chemistry, Beijing Normal University, Beijing 100875, P. R. China

The generalized energy-based fragmentation (GEBF) approach (Li, W.; Li, S.; Jiang, Y. *J. Phys. Chem. A* **2007,** *111,* 2193) is extended for geometry optimizations and vibrational spectra calculations of general large molecules or clusters. In this approach, the total energy and its derivatives, and some molecular properties, of a target system are obtained from conventional calculations on a series of subsystems derived from the target system. Each subsystem is electronically embedded in the background point charges generated by all other atoms outside the subsystem so that the long-range interactions and polarization effects between remote fragments are approximately taken into account. The approach computationally scales linearly with the system size and can be easily implemented for large-scale parallelization. By comparing the results from the conventional and GEBF calculations for several test molecules including a polypeptide and a water cluster, we demonstrate that the GEBF approach is able to provide quite reliable predictions for molecular geometries, vibrational frequencies, and thermochemistry data and satisfactory descriptions for vibrational intensities, for general molecules with polar or charged groups.

1. Introduction

Calculations of equilibrium geometries and vibrational frequencies for ab initio methods are still computationally difficult for quite large molecules. This is because such quantities require the calculations of the first and second derivatives (the gradients and the Hessian) of the molecular electronic energies with respect to the nuclear displacements, which are computationally more expensive than the single-point energy calculations. During the past two decades, lower or linear scaling algorithms[1–10] have been extensively pursued for electronic energies and energy derivative calculations especially at the Hartree–Fock (HF) and density functional theory (DFT) levels. These algorithms are mostly based on the near-sightedness of density matrices or the fast decay of electron correlation energies and thus are physically sound. Within these lower or linear scaling algorithms, corresponding schemes for calculating the analytic energy gradients or molecular properties have also been proposed.[1–10] However, these schemes have not been established as practical tools for geometry optimizations and frequency calculations of large molecules because the crossover between these schemes and the conventional methods occurs only at sufficiently large molecules (about 100~200 atoms).

On the other hand, various molecular fragmentation approaches,[11–35] although less robust than those standard linear scaling approaches, represent an appealing direction for approximately obtaining equilibrium geometries, vibrational frequencies, and molecular properties of very large molecules. The main idea of these approaches is to assemble the molecular energies or properties of a large molecule with the corresponding quantities from a series of subsystems, which are derived from the target molecule. Several typical fragmentation approaches include the density-matrix divide-and-conquer (D&C) approach,[11,12]

the fragment molecular orbital (FMO) method,[19–22] the energy-corrected molecular fractionation with conjugated caps (EC-MFCC) method,[27] the systematic molecular fragmentation method,[29,30] the cardinality-guided molecular tailoring approach (CG-MTA),[34,35] and the integrated multicenter molecular orbitals (IMiC MO) method.[24,25] Among these methods, the latter four can be classified as simple energy-based fragmentation approaches, which are computationally cheaper than the former two because in the latter four the calculation of each subsystem is independent or slightly dependent on the information of other subsystems. Another advantage of these energy-based fragmentation approaches is that they can be easily implemented at different theoretical levels with existing quantum chemistry packages. The IMiC MO method has been implemented to be applicable for the calculations of total energies, geometry optimizations, and vibrational frequencies of various molecular clusters (such as water clusters).[25] However, this method could not be used to treat covalently bonded macromolecules. Other energy-based approaches are designed to treat both macromolecules and molecular clusters, but their applications have been limited to neutral and nonpolar (or less charged) large molecules. For these systems, geometry optimizations,[25,29,34,35] Hessian and vibrational frequencies,[25,29,34] and molecular property[20,28,35,36] calculations with some fragmentation approaches have been reported.

Recently, we have proposed a generalized energy-based fragmentation (GEBF) approach,[28] which is applicable for general large molecules (macromolecules or clusters) with (or without) polar or charged groups. In this approach, a target system is first divided into small fragments, and then a series of subsystems are constructed from these fragments, as in other energy-based approaches.[24–27,29–34,37] Next, each subsystem is embedded in the field of point charges generated by other atoms of the whole system outside this subsystem. Then, the total energy of the target molecule can be obtained by simply

---

[†] Nanjing University.
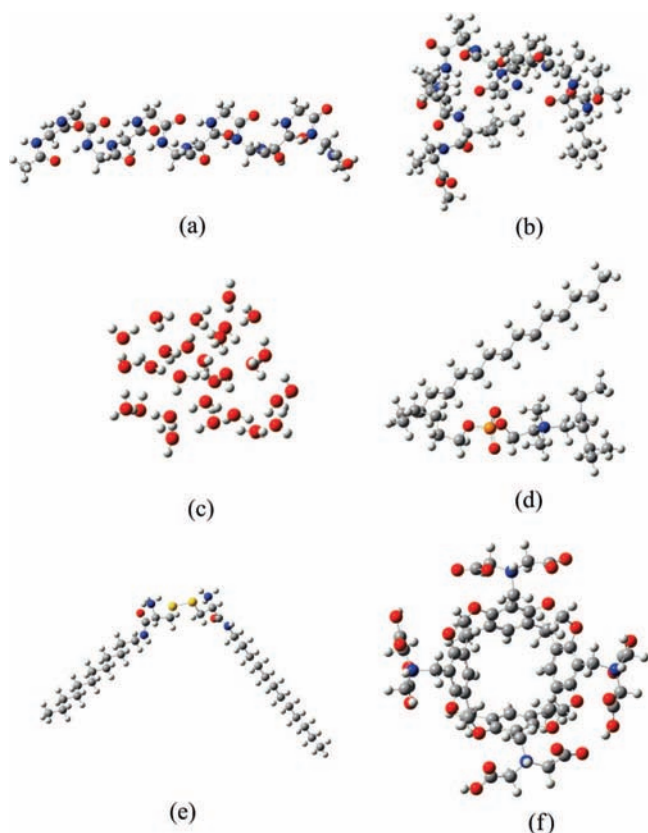[‡] Beijing Normal University.

GEBF Approach for General Large Molecules

J. Phys. Chem. A, Vol. 112, No. 43, 2008 **10865**

Figure 1. Systems under study: (a) $(Gly)_{12}$, (b) Deca, (c) $(H_2O)_{28}$, (d) Gel A, (e) Gel B, and (f) Gel C.
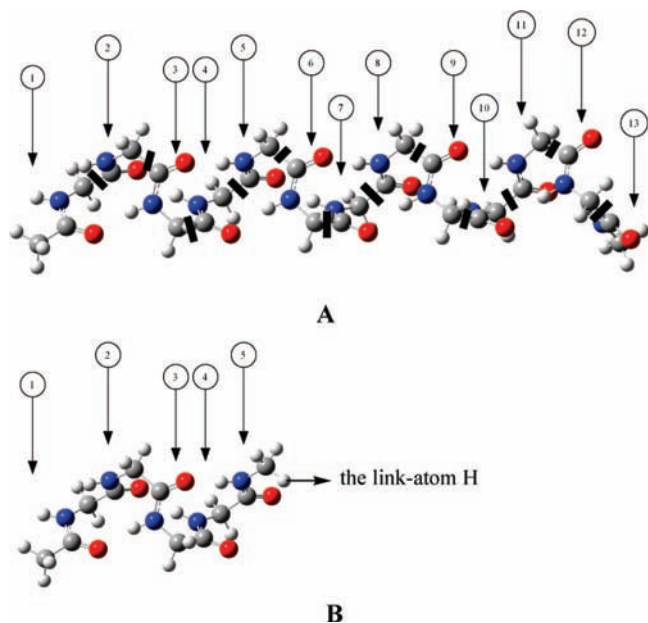

Figure 2. Fragmentation scheme for $(Gly)_{12}$: (a) All fragments. (b) The largest subsystem.

assembling the energies of all the subsystems. The dipole moment and static polarizability of the target system could also be computed in a similar way from the corresponding values of all constructed subsystems. It should be mentioned that an approach similar to our GEBF approach, named as the electrostatically embedded many-body expansion method, was also independently developed to calculate the electronic energies of water clusters.[37]

In this work, we further extend the applications of the GEBF approach to geometry optimizations, vibrational frequencies, and

vibrational intensities calculations of large molecules. Within the GEBF approach, the first and second derivatives of molecular electronic energies for a target molecule could be directly obtained by combining the corresponding derivatives of all subsystems, which are available with existing quantum chemistry packages. Then, the resulting energy gradients can be employed for geometry optimizations, and at the optimized geometry, the mass-weighted Hessian matrix can be diagonalized to obtain the vibrational frequencies and the normal coordinates. In a similar way, the derivatives of dipole moment and polarizability with respect to the normal coordinates can also be approximately computed for the target molecule, which allow the evaluation of infrared and Raman vibrational intensities to be possible.

This paper is organized as follows: In section 2, we describe the computational details for geometry optimizations, vibrational frequencies, and vibrational intensities calculations. In section 3, the GEBF approach is applied to perform geometry optimizations and vibrational spectra calculations for several neutral and charged molecules, and the results are compared with the corresponding values from the conventional approach. Finally, a brief summary is provided in section 4.

2. Methodology

2.1. Energy and Energy Derivatives. Within the GEBF approach,[28] the total energy of a target system can be derived from the total energies of a series of subsystems, each of which is placed in the presence of background charges generated by all atoms outside this subsystem. Assuming that the total energy of the subsystem m is denoted as $\tilde{E}_m$ or E_m (with or without the self-energy of all background point charges, respectively), then one has

$$\tilde{E}_m = E_m + \sum_{\substack{A<B \\ A,B \notin m}} \frac{Q_A Q_B}{R_{AB}} \quad (1)$$

Here, Q_A is the point charge on atom A outside the subsystem m, which could be taken as the natural population charge.[38,39] Then, the total energy of the target molecule can be approximately written as follows

$$E_{tot} = \sum_m C_m \tilde{E}_m - \left(\sum_m C_m - 1\right)\sum_{A<B} \frac{Q_A Q_B}{R_{AB}} \quad (2)$$

Here the summation over m runs over all of the constructed subsystems and C_m is the coefficient of the subsystem m. By differentiating eq 2 with respect to a certain Cartesian coordinate ($i = x, y, z$) of the Ith atom, one can obtain the approximate energy gradient as below (the derivation can be found in the Supporting Information)

$$\frac{\partial E_{tot}}{\partial q_{Ii}} \approx \sum_n C_n \frac{\partial E_n}{\partial q_{Ii}} - \sum_n C_n \sum_{A \notin n} f_{A,Ii} \quad (3)$$

Here, the summation over n is limited to those subsystems containing the atom I as a real atom, and $f_{A,Ii}$ denotes the i component of the Coulomb force between charge on A and charge on atom I

$$f_{A,Ii} \equiv Q_A Q_I\left(\frac{q_{Ai} - q_{Ii}}{R_{AI}^3}\right) \quad (4)$$

Furthermore, from eq 3, the second derivatives of the total energy can be approximately expressed as

$$\frac{\partial^2 E_{tot}}{\partial q_{Ii}\,\partial q_{Jj}} \approx \sum_n C_n \frac{\partial^2 E_n}{\partial q_{Ii}\,\partial q_{Jj}} - \sum_n C_n \sum_{A \notin n} g_{A,Ii,Jj} \qquad (5)$$

with

$$g_{A,Ii,Jj} \equiv Q_A Q_I \left[\frac{(-3)(q_{Ai} - q_{Ii})(q_{Aj} - q_{Ij})}{R_{AI}^2} + \delta_{ij} \right] R_{AI}^{-3}(\delta_{AJ} - \delta_{IJ}) \qquad (6)$$

2.2. Geometry Optimization. Once the gradients of all atoms in the target system are available from eq 3, we then proceed to find minimum structures starting from a given initial structure. In this work, we adopt the quasi-Newton algorithm[3,40,41] to locate the minimum structures. This method is computationally efficient because the approximate Hessian matrix (rather than the accurate Hessian matrix) is employed during the optimization process. The BFGS procedure is employed for updating the Hessian matrix.[42] The convergence criteria for the energy change and the maximum gradient are set to 0.01 milli-Hartree (mH) and 1mH/Bohr, respectively.

During the optimization, all subsystems are required to be reconstructed at each step (but the fragments remain the same as assigned in the first step). The way of constructing subsystems has been introduced elsewhere,[43] so here we will only give a brief description. The procedures include the following steps: (1) the target molecule is divided into a series of fragments (with approximately equal size) by cutting some single covalent bonds or hydrogen bonds; (2) for a given fragment, its environmental fragments are added sequentially to this fragment (according to their distances away from it) to form a derivative subsystem until the maximum number of fragments (η) is reached; (3) then the derivative subsystems with $m = \eta - 1, ..., 2, 1$ fragments are constructed sequentially, and the corresponding coefficients are determined according to the guiding rule that the net number of the m-fragment interaction energy term must be positive one in the total energy expression; (4) for a given subsystem, each of the atoms that are not included as real atoms is replaced by a background point charge calculated from natural population analysis[38,39] on all the primitive subsystems. In this fragmentation scheme, η is the only parameter and can be used to control the accuracy of the GEBF approach. It should be noticed that the positions of those link-atom hydrogens are not determined by the forces on them but determined by the coordinates and types of their bonded atoms, as discussed previously.[27]

2.3. Vibrational Frequencies and Thermochemistry Data Calculation. At the GEBF-optimized geometry, one can calculate the Hessian matrix of the whole system via eq 5, with the second derivatives of molecular electronic energies from all constructed subsystems. Then, vibrational frequencies and their corresponding normal modes can be obtained by diagonalizing the mass-weighted Hessian matrix of the whole system.

Furthermore, the resulting vibrational frequencies can be employed to calculate the vibrational partition function to evaluate zero-point vibrational energy and various thermochemistry data (together with the translational and rotational partition functions), such as enthalpy, entropy, Gibbs free energy, etc.

2.4. Vibrational Intensities. In the previous work,[28] we have calculated the dipole moment and polarizability with the formula below

$$\Omega \approx \sum_m^M C_m \Omega_m \qquad (\Omega = \mu_i, \alpha_{ij}) \qquad (7)$$

Differentiating eq 7 with respect to the Cartesian coordinate q_{Ii} ($i = x, y, z$) leads to

$$\frac{\partial \Omega}{\partial q_{Ii}} \approx \sum_m^M C_m \frac{\partial \Omega_m}{\partial q_{Ii}} \qquad (I \in \text{subsystem } m) \qquad (8)$$

According to this formula, the derivatives of dipole moment and polarizability are easily available from the corresponding values of all subsystems. By using the chain rule, one can also obtain the derivatives of dipole moment and polarizability with respect to the normal coordinates as follows

$$\frac{\partial \Omega}{\partial Q_k} = \sum_{a=1}^{3N} \frac{\partial \Omega}{\partial \xi_a} \frac{\partial \xi_a}{\partial Q_k} \qquad (k = 1, 2, ..., N_f) \qquad (9)$$

Here N_f is the number of vibrational degrees of freedom and ξ_a is the mass-weighted Cartesian coordinate defined as

$$\xi_1 = \sqrt{m_1}q_{1x}, \qquad \xi_2 = \sqrt{m_1}q_{1y}, \qquad \xi_3 = \sqrt{m_1}q_{1z},$$
$$\xi_4 = \sqrt{m_2}q_{2x}, ... \qquad (10)$$

With these derivatives, the infrared (IR) intensity I_k and Raman intensities R_k can be calculated with the expressions below, respectively[44]

$$I_k = \left| \frac{\partial \vec{\mu}}{\partial Q_k} \right|^2 \qquad (11)$$

$$R_k = (45\bar{\alpha}_k^2 + 7\gamma_k^2) \qquad (12)$$

where

$$\bar{\alpha}_k = \frac{1}{3} \sum_{a=x,y,z} \frac{\partial \alpha_{aa}}{\partial Q_k} \qquad (13)$$

$$\gamma_k^2 = \frac{1}{2}\left[3 \sum_{a,b=x,y,z} \left(\frac{\partial \alpha_{ab}}{\partial Q_k} \right)^2 - 9\bar{\alpha}_k^2 \right] \qquad (14)$$

2.5. Computational Procedures. Within the GEBF approach, a procedure of finding the minimum structures and computing vibrational frequencies and intensities includes the following steps:

(1) Read the coordinates of the target molecule and then manually divide the molecule into a series of fragments;

(2) Build subsystems (by setting η to be a given number) and obtain the point charges from natural population analysis (NPA charges in short) on each fragment;[27]

(3) Compute the energies and energy gradients for all subsystems in the presence of point charges;

(4) Obtain the total energy and gradients of the target molecule by eqs 2 and 3;

(5) Update the coordinates of the target system by using the quasi-Newton algorithm;

(6) Check the maximum force and the energy change. If the convergence criteria are not satisfied, go back to step (2) (with the NPA charges obtained at the latest geometry). Otherwise, output the coordinates of the optimized geometry;

(7) At the optimized geometry, construct the total Hessian matrix of the whole system from frequency calculations on all subsystems via eq 5;

(8) Diagonalize the mass-weighted Hessian matrix of the target system to calculate the vibrational frequencies and normal modes. Then vibrational frequencies are used to compute some thermochemistry quantities;

(9) Obtain the dipole moment and polarizability derivative tensors from calculations on all subsystems and construct the corresponding quantities for the target system via eq 8. Then the results as well as the normal coordinates are used to calculate the IR and Raman intensities.

GEBF Approach for General Large Molecules

J. Phys. Chem. A, Vol. 112, No. 43, 2008 **10867**

TABLE 1: Comparison of Total Energies and Optimized Geometrical Parameters Obtained from GEBF Calculations with Those from Conventional Calculations at the HF/6-31G* Level

molecule	basis functions[a]	total energy (a.u.)		rmsd[b]		
		conven.	GEBF	Δr (Å)	$\Delta\theta$ (deg)	$\Delta\varphi$ (deg)
(Gly)$_{12}$	881/353	−2728.848 06	−2728.846 56	0.014	0.83	3.20
(H$_2$O)$_{28}$	(532/133)	−2128.804 93	−2128.802 63	0.017	2.00	5.45
Deca	(1294/494)	−3502.877 85	−3502.876 44	0.027	1.63	5.75
Gel A	(710/349)	−1946.050 69	−1946.055 28	0.0003	0.14	0.88
Gel B	(706/326)	−2335.425 96	−2335.427 68	0.0005	0.17	5.41
Gel C	(1380/715)	−4169.926 26	−4169.910 40	0.004	0.86	8.25

[a] The total number of basis functions in the whole system and the largest subsystem, respectively. [b] Δr, $\Delta\theta$, and $\Delta\varphi$ represent the rmsd of bond lengths, bond angles, and dihedral angles, respectively.

TABLE 2: Comparison of Total Energies and Optimized Geometrical Parameters for (Gly)$_{12}$ Obtained from GEBF Calculations with Those from Conventional Calculations at Different Theoretical Levels

theoretical level	basis functions[a]	total energy (a.u.)		rmsd[b]		
		conven.	GEBF	Δr (Å)	$\Delta\theta$ (deg)	$\Delta\varphi$ (deg)
HF/3-21G	563/227	−2713.664 37	−2713.663 82	0.008	0.50	2.31
HF/6-31G*	881/353	−2728.848 06	−2728.846 56	0.014	0.83	3.20
HF/6-311G**	1212/492	−2729.549 26	−2729.547 80	0.013	0.84	3.38
HF/6-311+G**	1424/576	−2729.600 44	−2729.599 20	0.014	0.71	2.98
B3LYP/6-31G*	881/353	−2744.711 61	−2744.710 73	0.009	0.77	4.50

[a] The total number of basis functions in the whole system and the largest subsystem, respectively. [b] Δr, $\Delta\theta$, and $\Delta\varphi$ represent the rmsd of bond lengths, bond angles, and dihedral angles, respectively.

TABLE 3: Highest 30 Vibrational Frequencies (cm^{-1}) at the Respectively Optimized Structures Calculated with the Conventional HF and GEBF-HF Calculations

normal modes	(Gly)$_{12}$		(H$_2$O)$_{28}$		Gel A		Gel B	
	conven.	GEBF[a]	conven.	GEBF[a]	conven.	GEBF[a]	conven.	GEBF[a]
1	3268.71	1.96	3986.54	1.48	3223.68	1.11	3223.58	0.77
2	3281.97	−2.11	3990.08	0.86	3227.36	0.43	3228.15	0.33
3	3287.11	0.26	3991.44	3.41	3229.06	−0.15	3233.11	0.01
4	3308.83	2.16	3993.16	11.46	3229.57	0.67	3237.75	0.35
5	3311.42	1.04	3997.86	7.28	3231.56	1.50	3241.96	0.65
6	3320.84	−1.51	4001.49	5.84	3235.65	0.42	3245.45	0.63
7	3321.39	−1.63	4004.69	11.46	3236.98	−0.36	3249.70	0.79
8	3321.69	−1.34	4008.01	9.42	3242.03	0.91	3252.86	1.12
9	3321.75	−1.20	4008.45	10.69	3249.50	0.78	3256.47	0.60
10	3322.00	−0.48	4010.38	9.64	3253.33	−1.12	3261.08	−0.51
11	3322.36	−0.67	4024.30	−0.38	3255.71	2.20	3262.08	0.46
12	3322.43	−0.25	4026.68	7.59	3257.97	1.48	3262.89	−0.10
13	3322.57	0.00	4027.99	12.27	3258.43	1.15	3266.70	1.18
14	3323.59	1.08	4041.07	16.30	3259.96	−0.31	3267.19	1.23
15	3324.80	0.40	4050.33	12.16	3264.04	−0.38	3268.24	1.56
16	3333.62	−2.14	4051.88	11.82	3266.54	−1.32	3269.48	1.66
17	3337.62	3.34	4060.11	3.98	3267.84	0.37	3299.04	0.56
18	3837.56	2.64	4061.55	4.33	3269.03	1.23	3301.48	−0.46
19	3838.54	3.56	4063.24	10.45	3274.18	−0.54	3317.33	−1.39
20	3841.90	4.28	4068.23	7.96	3274.38	−0.67	3325.60	0.54
21	3844.48	3.31	4075.92	1.67	3275.05	−0.99	3334.42	−2.09
22	3846.04	3.10	4079.06	1.23	3279.98	−5.10	3335.48	−1.80
23	3847.59	2.62	4082.40	19.94	3285.00	−1.56	3531.19	2.89
24	3849.38	4.70	4131.41	0.66	3311.22	−1.25	3542.60	−1.42
25	3851.56	3.28	4132.47	−0.02	3319.54	−1.03	3673.35	0.15
26	3853.21	2.39	4134.23	−1.19	3329.47	−1.27	3678.56	2.33
27	3855.25	4.96	4134.44	0.36	3353.20	3.04	3721.61	0.28
28	3870.99	4.27	4135.05	0.60	3374.14	0.13	3738.56	0.81
29	3879.69	4.10	4135.22	1.42	3388.69	−1.24	3854.02	1.62
30	3906.44	5.36	4135.97	1.12	3404.82	4.58	3859.06	1.25

[a] The deviation of the GEBF frequency relative to the conventional value.

In the GEBF approach, all the subsystems are calculated independently, thus the program can be readily parallelized so that calculations on different subsystems can be done on different compute nodes or processors. Both the message-passing interface (MPI)[45] and the open multiprocessing (OpenMP)[46] parallel techniques have been implemented.

3. Results and Discussions

To illustrate the applicability of the GEBF method, we carry out geometry optimizations and vibration spectra calculations for several moderate systems and compare the results with the corresponding values from the standard treatment. The selected

TABLE 4: Thermochemistry Data Obtained from the Conventional and GEBF Calculations[a]

	$(Gly)_{12}$		$(H_2O)_{28}$	
thermochemistry data	conven.	GEBF[b]	conven.	GEBF[b]
ZPE (a.u.)	0.842 16	−0.000 47	0.776 66	−0.001 88
E_0 (a.u.)	−2728.005 90	0.001 03	−2128.028 27	0.000 42
U (a.u.)	−2727.952 85	0.001 26	−2127.960 53	0.001 41
H (a.u.)	−2727.951 90	0.001 25	−2127.959 59	0.001 42
G (a.u.)	−2728.103 43	0.000 21	−2128.130 09	−0.001 21
C_v (cal·mol^{-1}·K^{-1})	190.636	0.34	235.445	2.38
S (cal·mol^{-1}·K^{-1})	318.920	2.21	358.862	5.50

[a] E_0 represents the sum of electronic and zero-point energies (ZPE). U, H, G, C_v, and S denote the internal energy, enthalpy, Gibbs free energy, thermal capacity, and entropy, respectively. [b] The deviation of the GEBF value relative to the conventional value.
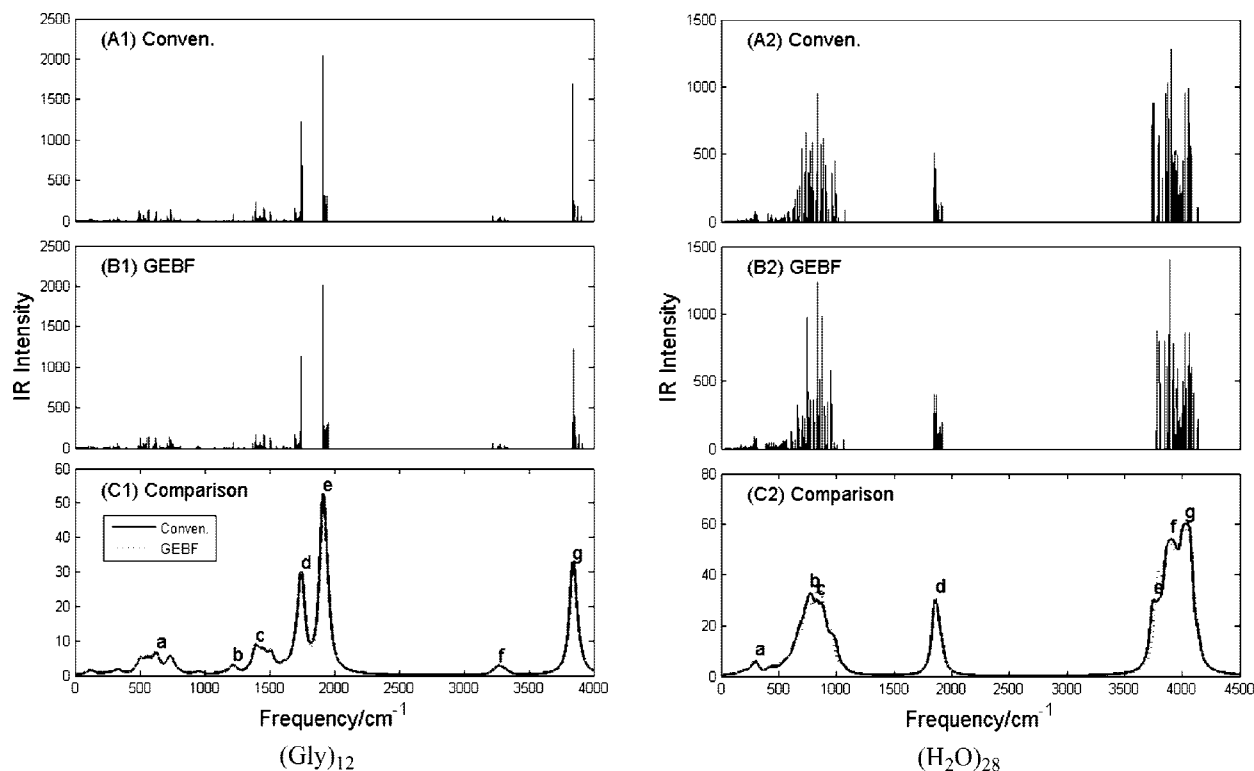

Figure 3. IR spectra of $(Gly)_{12}$ and $(H_2O)_{28}$ at the HF/6-31G* level: (a) Conventional results. (b) GEBF results. (c) Comparison of conventional and GEBF results using Lorentz broadening with HWHM = 30 cm^{-1}.

systems include three neutral molecules, a peptide 3_{10}-helix-CH$_3$CO-(Gly)$_{12}$-NHCH$_3$ [denoted as $(Gly)_{12}$ for convenience], a decapeptide Boc-Leu-Aib-Val-Gly-Leu-Aib-Val-^{D}Ala-^{D}Leu-Aib-OMe (denoted as Deca), and a water cluster $(H_2O)_{28}$ and three charged molecules, Gel A, Gel B, and Gel C, which are usually termed as hydrogelators.[47] For vibrational spectra calculations, we take four molecules, $(Gly)_{12}$, $(H_2O)_{28}$, Gel A, and Gel B, as examples. The Cartesian coordinates of the optimized structures obtained from both approaches are available in the Supporting Information. The structures of all selected molecules are schematically shown in Figure 1. The GEBF approach has been implemented in the LSQC package.[48] All conventional calculations on subsystems and the full systems are carried out with the GAUSSIAN 03 package.[49]

3.1. Geometry Optimization. First, we perform geometry optimizations for six molecules shown in Figure 1. Calculations are done at the HF level with the 6-31G* basis set. The initial geometries of these molecules are usually obtained with the force-field methods or the semiempirical AM1 method, and their coordinates are provided in the Supporting Information. As described above, the first step in the GEBF calculation is to divide the studied system into fragments of roughly equal size,

and all of the other steps can be automatically done. For illustration, the fragmentation scheme for $(Gly)_{12}$ is shown in Figure 2. As shown in Figure 2, $(Gly)_{12}$ is divided into 13 fragments by cutting the C−C bond within each residue. The largest subsystem ($\eta = 5$) is also displayed in Figure 2. As shown in this figure, the link-atom hydrogen atom is added to one terminal carbon atom for valence saturation in the subsystem, and the C−H bond length is set to 1.07 Å.[27,28] For the water cluster, each water is chosen as a fragment, and for the other four molecules, their fragmentation schemes are given in Figure S1 of the Supporting Information. It is worthwhile to mention some geometrical features of the selected systems. The species $(Gly)_{12}$ has the 3_{10}-helix structure, being a highly polar molecule.[27] Deca is a folded polypeptide with an intricate structure.[50] $(H_2O)_{28}$ is a hydrogen-bonded cluster with significant polarization interactions between any two monomers. The other three species, Gel A, Gel B, and Gel C, are all charged. Gel A has two positively charged amines −N$^+$H$_3$, and Gel B has one positively charged amine −N$^+$(CH$_3$)$_2$− and one negatively charged phosphate −PO$_4^-$. In Gel C, there are four positive amines −N$^+$H− and four negative carboxylic acid groups −COO$^-$.

GEBF Approach for General Large Molecules

J. Phys. Chem. A, Vol. 112, No. 43, 2008 **10869**

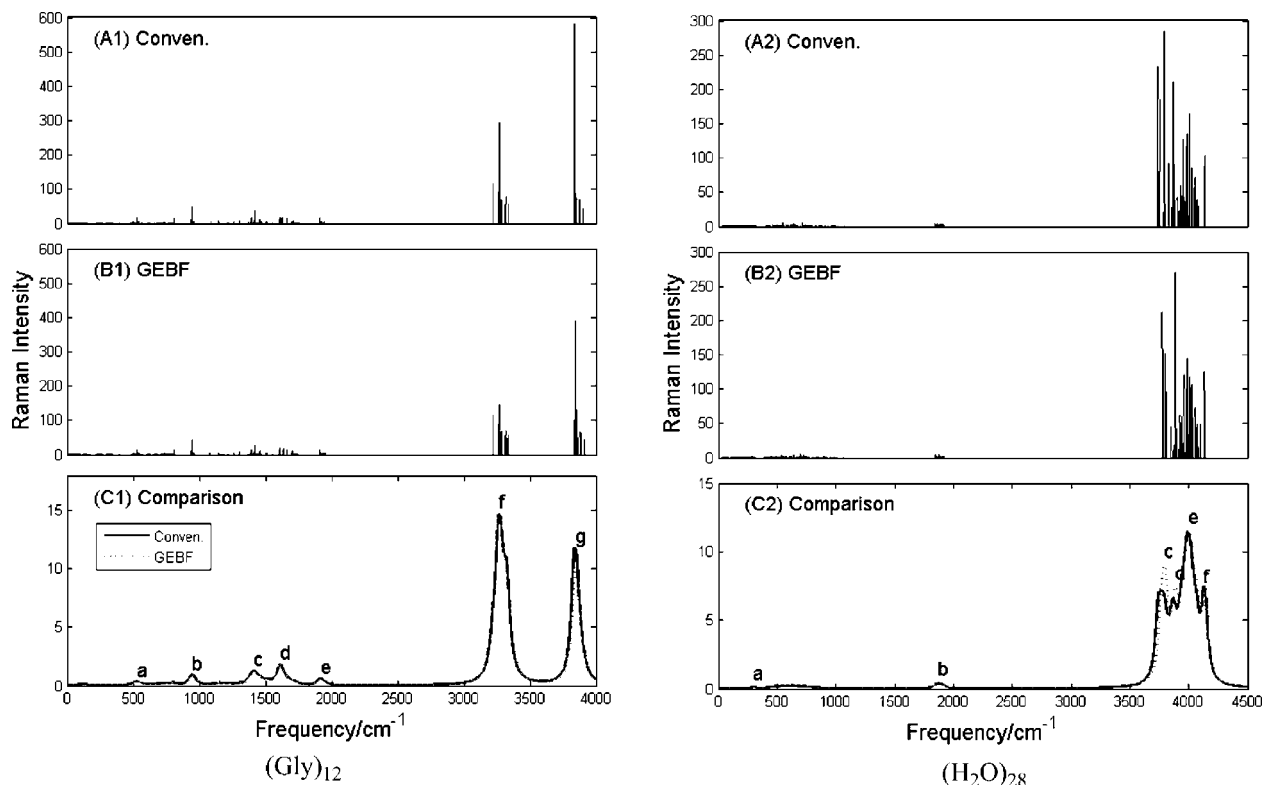**Figure 4.** Raman spectra of $(Gly)_{12}$ and $(H_2O)_{28}$ at the HF/6-31G* level: (a) Conventional results. (b) GEBF results. (c) Comparison of conventional and GEBF results using Lorentz broadening with HWHM = 30 cm^{-1}.

TABLE 5: Peaks and Intensities of IR Spectra for $(Gly)_{12}$ and $(H_2O)_{28}$ Calculated with the GEBF and Conventional Approaches at the HF/6-31G* Level

| | $(Gly)_{12}$ | | | | $(H_2O)_{28}$ | | | |
| | peak (cm^{-1}) | | intensity | | peak (cm^{-1}) | | intensity | |
normal modes	conven.	GEBF[a]	conven.	GEBF	conven.	GEBF[a]	conven.	GEBF
a	626	−6	6.8	7.2	299	−6	5.9	6.6
b	1220	0	3.0	3.0	776	−18	32.8	31.2
c	1398	−1	8.8	8.8	834	8	29.5	36.4
d	1746	−3	29.9	30.2	1860	−1	30.9	30.5
e	1915	4	52.3	49.4	3754	45	30.2	42.6
f	3271	0	2.9	2.9	3909	−14	54.6	54.4
g	3842	6	32.7	29.4	4027	7	59.9	58.5

[a] The deviation of the GEBF frequency relative to the conventional value.

It is necessary to mention how to select the optimal parameter η (the maximum number of fragments) for GEBF calculations since this parameter is expected to correlate with the accuracy of the GEBF approach. The guiding rule for selecting this parameter is that for each fragment in a given system it is covalently bonded or hydrogen-bonded fragments must be included in the corresponding primitive subsystem. Once the hydrogen bond criteria are given,[51] the parameter η could be determined from the fragmentation procedure before the GEBF calculations. A more practical way to choose η is to perform test calculations for a given system at some typical geometries with several η values. If further increase of the parameter brings little change to the total energy, then this parameter could be used for subsequent geometry optimizations and other calculations. For example, for $(Gly)_{12}$ and $(H_2O)_{28}$, the GEBF-HF energies with different η values are given in Table S1 (Supporting Information). It can be seen that the accuracy of the GEBF approach increases with increasing value of η (when η is relatively small), but then it stays almost unchanged when the parameter is further increased. Thus, η can be selected to

be 5 for $(Gly)_{12}$ and 7 for $(H_2O)_{28}$, respectively. For other molecules, the optimal η values from our test calculations are: 6 for Deca, 4 for Gel A and Gel C, and 5 for Gel B.

For all these species, the geometrical differences between the optimized geometries obtained with the GEBF and conventional approaches are listed in Table 1, altogether with the total energies at the respectively optimized geometries. For comparison, the geometry optimization with the standard approach and the GEBF approach starts from the same initial structure. One can see that the root-mean-square deviations (rmsd's) between the two optimized structures are quite small for each molecule. The largest rmsd is about 0.027 Å for bond lengths, 2.18° for bond angles, and 8.25° for dihedral angles. The total energies at the, respectively, optimized structures from conventional and GEBF approaches are also quite close to each other, being less than 5.0 mH (except for Gel C). For Gel C, a relatively larger energy error (15.9 mH) reflects that our present fragmentation procedure may not be quite appropriate for structures with many cyclic ring structures.
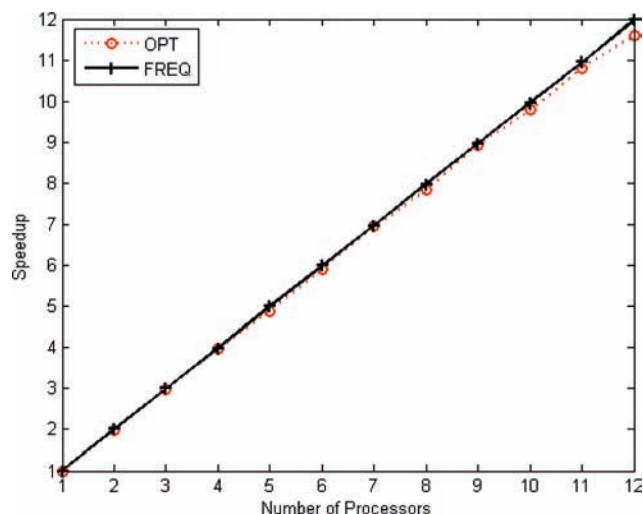
Figure 5. Speedup of the GEBF method for geometry optimization and frequency calculation of $(H_2O)_{28}$ at the HF/6-31G* level with the MPI parallel technique. The speedup is calculated as the sequential time divided by the parallel wall time of N processors.

It may be also worthwhile to evaluate the performance of the GEBF approach at different theoretical levels. For this purpose, we perform the geometry optimizations for $(Gly)_{12}$ at the HF level with the 3-21G, 6-31G*, 6-311G**, and 6-311+G** basis sets and the B3LYP[52] level with the 6-31G* basis set. The results from the GEBF and conventional approaches are collected in Table 2. It can be seen that the deviation between the optimized structures from these two approaches is very small, being almost independent of the basis sets or theoretical methods. At the respectively optimized geometries, the difference of the total energies from both approaches is always less than 2.0 mH, at all theoretical levels used.

From the discussions above, one can see that the GEBF approach is capable of producing optimized molecular geometries, which are very good approximations to the optimized geometries from the conventional approach, for both neutral and charged molecules. Since all the subsystems are kept under a given size in the GEBF approach (by the parameter η), this approach is expected to be quite useful in computing the optimized structures for very large molecules. It should be noticed that GEBF works fairly well for geometry optimizations of some molecules with intricate structures. For example, drastic geometry changes occur during the optimization of Deca, as the energy lowering from the initial geometry to the optimized geometry is about 1.4 a.u.

3.2. Vibrational Frequencies and Thermochemistry Data. Before vibrational frequencies are computed, it may be interesting to see how accurate the GEBF-HF Hessian matrix elements are compared to the Hessian matrix elements from conventional HF calculations. For $(Gly)_{12}$, $(H_2O)_{28}$, Gel A, and Gel B at their HF-optimized geometries, our calculations showed that the rmsd between the GEBF-HF and conventional HF Hessian matrix elements is no more than 1.9×10^{-4} a.u., and the maximum deviation is no more than 2.2×10^{-3} a.u for all of these systems (the detailed information is provided in Table S2, Supporting Information). Thus, the Hessian from GEBF calculations can be used to obtain reasonably accurate vibrational frequencies for large systems, in which conventional calculations are not available.

After obtaining the optimized geometries for $(Gly)_{12}$, $(H_2O)_{28}$, Gel A, and Gel B at the HF/6-31G* level, we then proceed to calculate their vibrational frequencies at their respective opti-

mized geometries with both the GEBF-HF and conventional HF approaches. For simplicity, only the highest 30 calculated frequencies are listed in Table 3 for comparison. It can be seen from Table 3 that the maximum difference between the GEBF frequencies and those from the conventional approach is 5.36 cm^{-1} for $(Gly)_{12}$, 19.94 cm^{-1} for $(H_2O)_{28}$, 5.10 cm^{-1} for Gel A, and 2.89 cm^{-1} for Gel B, respectively. For all calculated frequencies (lower frequencies are not shown), we find that the maximum difference is 7.6, 36.9, 5.47, and 8.02 cm^{-1} for $(Gly)_{12}$, $(H_2O)_{28}$, Gel A, and Gel B, respectively, and the rmsd's of frequencies in the whole range are 2.3, 10.6, 1.10, and 1.34 cm^{-1} for these four molecules, respectively. This result indicates that the vibrational frequencies from the GEBF approach are in reasonable agreement with those from the conventional approach. It should be mentioned that the results described above are obtained with the convergence criteria of the maximum gradient being 1.0 mH/Bohr. In this case, the highest frequency among the six "zero" frequencies (corresponding to translation and rotational motions of the whole system) from the GEBF approach is 6.08 cm^{-1} for $(Gly)_{12}$ and 13.10 cm^{-1} for $(H_2O)_{28}$, respectively. However, when the converge criteria for the maximum gradient is tightened to be 0.3mH/Bohr for geometry optimizations, these zero frequencies now have very small values (8.43×10^{-7} and 6.50×10^{-6} cm^{-1} for $(Gly)_{12}$ and $(H_2O)_{28}$, respectively). However, those other (3N-6) vibrational frequencies are only slightly improved for both molecules.

With the obtained vibrational frequencies, the calculated thermochemistry data for two molecules, $(Gly)_{12}$ and $(H_2O)_{28}$, are listed in Table 4. The maximum differences between zero-point energies and various terms obtained from conventional and GEBF approaches are less than 1.3 and 1.4 mH for $(Gly)_{12}$ and $(H_2O)_{28}$, respectively, while for the thermal capacity and the entropy, the values from the GEBF approach deviate from the conventional values by less than 1.5%. Thus, the thermochemistry quantities from the GEBF approach also show good agreement with those from the conventional approach.

3.3. IR and Raman Spectra. With the calculated vibrational frequencies and IR or Raman intensities, we present the IR and Raman spectra for $(Gly)_{12}$ and $(H_2O)_{28}$ in Figures 3 and 4, respectively. Lorentz broadenings of the spectra lines calculated from the conventional and GEBF approaches are also made with the same half-width at half-maximum (HWHM), 30 cm^{-1} (Figures 3c and 4). From Figures 3 and 4, one can see that the spectral shapes from the conventional and GEBF approaches are nearly consistent. For better comparison, the main peaks and their IR or Raman intensities are summarized in Tables 5 and 6, respectively. In the IR spectra, the rmsd's of the peaks are 3.7 cm^{-1} for $(Gly)_{12}$ and 19.6 cm^{-1} for $(H_2O)_{28}$, and the rmsd's of the intensities are 5% for $(Gly)_{12}$ and 17% for $(H_2O)_{28}$. While in the Raman spectra, the rmsd's of the peaks are 2.7 cm^{-1} for $(Gly)_{12}$ and 16.1 cm^{-1} for $(H_2O)_{28}$, and the rmsd's of the intensities are 7% for $(Gly)_{12}$ and 12% for $(H_2O)_{28}$. From these results, one can conclude that the IR or Raman intensities from the GEBF approach are still fairly satisfactory for $(Gly)_{12}$ and acceptable for $(H_2O)_{28}$. The reason why the GEBF approach gives less accurate IR or Raman intensities for $(H_2O)_{28}$ than for $(Gly)_{12}$ may be possibly due to the fact that the minimum structure $(H_2O)_{28}$ from the GEBF-HF optimization may be slightly different from the one from the conventional HF optimization (because many local minima exist for $(H_2O)_{28}$).

3.4. Computational Timing. To evaluate the parallel efficiency of the GEBF approach, we also carry out the geometry optimization and vibrational frequency calculation for $(H_2O)_{28}$ at the HF/6-31G* level by using up to 12 processors for

TABLE 6: Peaks and Intensities of Raman Spectra for $(Gly)_{12}$ and $(H_2O)_{28}$ Calculated with the GEBF and Conventional Approaches at the HF/6-31G* Level

	$(Gly)_{12}$				$(H_2O)_{28}$			
	peak (cm^{-1})		intensity		peak (cm^{-1})		intensity	
normal modes	conven.	GEBF[a]	conven.	GEBF	conven.	GEBF[a]	conven.	GEBF
a	530	1	0.4	0.4	300	−7	0.1	0.1
b	948	−1	1.0	0.9	1874	−2	0.4	0.4
c	1414	0	1.3	1.2	3758	33	7.3	9.0
d	1614	−1	1.9	1.8	3870	20	6.6	7.5
e	1917	3	0.6	0.6	3997	−3	11.4	11.6
f	3270	2	14.8	14.7	4132	−1	7.5	7.2
g	3842	6	11.8	10.0				

[a] The deviation of the GEBF frequency relative to the conventional value.

parallelization. The MPI[45] parallel technique is used in the platform with three nodes, each of which contains four Intel Xeon 2.66 GHz processors. All of the calculations in this work are carried out with this platform. The implementation of a parallel code for energy-based fragmentation methods has been reported in detail earlier.[34,35] Here a similar parallelization strategy, as described previously,[34,35] is adopted. The computing network consists of one managing processor and more computing processors. The missions of the managing processor include the construction of all subsystems for the target system at a given geometry, the distribution of the jobs on available compute nodes, and the assembly of the energy gradients or the Hessian for the target system if needed. On each compute processor, one subsystem is calculated at a time. For the GEBF approach, the communication time between the managing processor and compute processors is negligible, compared to the computation time. The speedup is obtained by dividing the sequential time by the parallel wall time, as shown in Figure 5. One can see that the speedup is nearly linear with respect to the numbers of processors. Therefore, the GEBF approach can be efficiently parallelized and is expected to be applicable for large molecules with hundreds or thousands of atoms as long as a lot of compute nodes are available. On the other hand, when one node (four processors) is used for both GEBF-HF and conventional HF optimizations, we find that the computation time is 8 min/step on average for GEBF-HF, which is longer than 3 min/step for the conventional HF optimization. This result indicates that the crossover between GEBF-HF and conventional HF optimizations occurs at a water cluster larger than $(H_2O)_{28}$. For single-point energy calculations, our previous study[28] showed that the crossover point appears at $(H_2O)_{48}$ at the HF/6-311G** level. Thus, when the target system becomes larger and larger, the GEBF optimization will eventually outperform the conventional optimization method. For vibrational frequency calculation, the advantage of the GEBF approach is already apparent even for the relatively small molecule $(H_2O)_{28}$. For $(H_2O)_{28}$, the computational time is 128 min for the conventional HF method but only 45 min for the GEBF-HF approach, which indicates that a time saving of about 65% is achieved.

4. Conclusions

In this work, we have extended the applications of the GEBF approach to the calculations of the optimized geometries and vibrational spectra for general molecules with charged or polar groups. By comparing the GEBF results with those from the conventional calculations, we demonstrate that the GEBF approach is capable of giving quite reliable predictions for molecular geometries, vibrational frequencies, and thermochemistry data and fairly satisfactory predictions for vibrational

intensities of general molecules. These results are quite encouraging for full ab initio quantum chemistry calculations of various closed-shell large systems. Because the GEBF calculations on various subsystems are almost independent, a highly efficient parallel code could allow ab initio geometry optimizations of very large systems to become routinely possible as long as tens or hundreds of compute nodes are available.

It is worth mentioning that the present implementation of the GEBF approach could be further improved. First, the fragmentation of the target molecule is done manually in the present implementation (by simply modifying its connectivity), and this can be done automatically with a computer program in the future. We noticed that an automatic fragmentation procedure was successively implemented in FMO[19−22] and CG-MTA[34,35] methods. Second, in the present GEBF-based optimizations, the initial fragments are kept frozen during the geometry optimizations (subsystems may be different from one geometry to another). This procedure works well as long as no chemical bonds are breaking within each fragment during optimization. However, this procedure would break down when a proton is transferred from one fragment to another during optimization. In this case, a proton itself may be chosen to be a fragment. We are extending the GEBF approach to such difficult cases. Another area we are exploring is the implementation of the GEBF approach for ab initio level molecular dynamics or Monte Carlo simulations of complex systems.

Supporting Information Available: The detailed derivation of eq 3, the fragmentation schemes for some systems under study, the Cartesian coordinates of the initial structures, and optimized geometries of all studied systems for the conventional and GEBF methods. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Azhary, A. E.; Rauhut, G.; Pulay, P.; Werner, H.-J. *J. Chem. Phys.* **1998**, *108*, 5185.
(2) Burant, J. C.; Strain, M. C.; Scuseria, G. E.; Frisch, M. J. *Chem. Phys. Lett.* **1996**, *248*, 43.
(3) Farkas, O.; Schlegel, H. B. *J. Chem. Phys.* **1999**, *111*, 10806.
(4) Larsen, H.; Helgaker, T.; Olsen, J.; Jørgensen, P. *J. Chem. Phys.* **2001**, *115*, 10344.
(5) Niklasson, A. M. N.; Challacombe, M. *Phys. Rev. Lett.* **2004**, *92*, 193001.

(6) Ochsenfeld, C.; Head-Gordon, M. *Chem. Phys. Lett.* **1997**, *270*, 399.

(7) Rauhut, G.; Werner, H.-J. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4853.

(8) Saebø, S.; Baker, J.; Wolinski, K.; Pulay, P. *J. Chem. Phys.* **2004**, *120*, 11423.

(9) Schütz, M.; Werner, H.-J.; Lindh, R.; Manby, F. R. *J. Chem. Phys.* **2004**, *121*, 737.

(10) Weber, V.; Niklasson, A. M. N.; Challacombe, M. *Phys. Rev. Lett.* **2004**, *92*, 193002.

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

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

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

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

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

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

(17) Li, W.; Li, S. *J. Chem. Phys.* **2005**, *122*, 194109.

(18) Gu, F. L.; Aoki, Y.; Korchowiec, J.; Imamura, A.; Kirtman, B. *J. Chem. Phys.* **2004**, *121*, 10385.

(19) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.

(20) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832.

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

(22) Fedorov, D. G.; Ishida, T.; Uebayasi, M.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 2722.

(23) Hirata, S.; Valiev, M.; Dupuis, M.; Xantheas, S. S.; Sugiki, S.; Sekino, H. *Mol. Phys.* **2005**, *103*, 2255.

(24) Morita, S.; Sakai, S. *J. Comput. Chem.* **2001**, *22*, 1107.

(25) Sakai, S.; Morita, S. *J. Phys. Chem. A* **2005**, *109*, 8424.

(26) Li, W.; Li, S. *J. Chem. Phys.* **2004**, *121*, 6649.

(27) Li, S.; Li, W.; Fang, T. *J. Am. Chem. Soc.* **2005**, *127*, 7215.

(28) Li, W.; Li, S.; Jiang, Y. *J. Phys. Chem. A* **2007**, *111*, 2193.

(29) Deev, V.; Collins, M. A. *J. Chem. Phys.* **2005**, *122*, 154102.

(30) Collins, M. A.; Deev, V. A. *J. Chem. Phys.* **2006**, *125*, 104104.

(31) Bettens, R. P. A.; Lee, A. M. *J. Phys. Chem. A* **2006**, *110*, 8777.

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

(33) Li, W.; Fang, T.; Li, S. *J. Chem. Phys.* **2006**, *124*, 154102.

(34) Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R. *J. Chem. Phys.* **2006**, *125*, 104109.

(35) Gadre, S. R.; Ganesh, V. *J. Theor. Comput. Chem* **2006**, *5*, 835.

(36) Lee, A. M.; Bettens, R. P. A. *J. Phys. Chem. A* **2007**, *111*, 5111.

(37) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46.

(38) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211.

(39) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735.

(40) Schlegel, H. B. *J. Comput. Chem.* **1982**, *3*, 214.

(41) Pulay, P.; Fogarasi, G. *J. Chem. Phys.* **1992**, *96*, 2856.

(42) Leach, A. R. *Molecular Modelling: Principles and Applications*; Addison Wesley Longman: London, 1996.

(43) Li, W.; Dong, H.; Li, S. In *Frontiers in Quantum Systems in Chemistry and Physics*; Wilson, S., Grout, P. J., Piecuch, P., Maruani, J., Delgado-Barrio,G., Eds.,in press.

(44) Person, W. B.; Zerbi, G. *Vibrational intensities in infrared and Raman spectroscopy*; Elsevier scientific publishing company: Amsterdam-Oxford-New York, 1982.

(45) http://www-unix.mcs.anl.gov/mpi/.

(46) http://www.openmp.org.

(47) Estroff, L. A.; Hamilton, A. D. *Chem. Rev.* **2004**, *104*, 1201.

(48) Li, S.; Li, W.; Fang, T.; Ma, J.; Jiang, Y. *LSQC*, Version 1.1; Nanjing University: Nanjing, 2006.

(49) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y. ; Kitao,O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth,G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, revision D.01*; Gaussian, Inc.: Wallingford, CT, 2004.

(50) Aravinda, S.; Shamala, N.; Bandyopadhyay, A.; Balaram, P. *J. Am. Chem. Soc.* **2003**, *125*, 15065.

(51) The criteria for hydrogen bonds X−H···Y in our calculations is $r_{H···Y} \leq 2.9$ Å, $r_{X···Y} \leq 3.5$ Å, and $\angle X−H···Y \geq 120°$.

(52) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.