Ab Initio Stochastic Optimization of Conformational and Many-Body Degrees of Freedom

Scott A. McMillan, Ned C. Haubein, Randall Q. Snurr, and Linda J. Broadbelt*

Department of Chemical Engineering, 2145 Sheridan Road, Northwestern University, Evanston Illinois 60208-3120

Received March 1, 2003

Moderate to large size molecules in solution have complex energy surfaces due to intramolecular (conformational) and intermolecular (many-body) interactions. The first principles Monte Carlo (FPMC) method, previously shown to effectively locate minimum-energy structures for systems with only many-body complexity, has been extended to address conformational flexibility by adding three new Monte Carlo move types. The primary advantage of the FPMC method is the ability to efficiently locate minimum energy structures of molecules with conformational flexibility in the presence of explicit solvent molecules using highly accurate quantum chemical calculations. The additions to FPMC were validated by studying conformers of glycerol, glyceraldehyde, and a large humic acid monomer unit. The structure of glyceraldehyde in the presence of one and two water molecules was also explored to demonstrate the power of FPMC to study systems with both conformational and many-body degrees of freedom.

INTRODUCTION

Molecules consisting of more than a few atoms almost always have some appreciable degree of conformational flexibility. The conformers of a given molecule all possess the same connectivity of atoms, but each corresponds to a different local minimum on the molecular potential energy surface. Different conformations become important for even small systems. For example, ethane, containing only six atoms, has staggered and eclipsed conformations. Chair, twist-boat, and boat conformations of cyclohexane are other examples of conformers that arise due to molecular flexibility. The search for stable conformers of small to medium-sized molecules is an extremely active area of research.[2,3]

Conventional optimization algorithms for identifying molecular structures of minimum energy most often use the gradient of the energy to follow the energy surface downhill to the nearest minimum. Local minima are normally separated by peaks on the energy surface, so the conformation of the structure optimized using conventional techniques is usually very similar to the guess for the initial structure. For small molecules, initial structures corresponding to all possible conformations can be generated and then optimized using conventional techniques to find all of the local minima. A molecule such as glycerol, which only has 14 atoms, has 126 possible conformers to evaluate.[4] For larger molecules, such as enzymes and other proteins, the sheer number of possible conformations makes this approach impractical.

An alternative approach based upon the Monte Carlo algorithm biases the conformational search to low-energy regions of the energy surface (e.g., ref 3). Unlike conventional optimization techniques, Monte Carlo moves can be designed to overcome the energy barriers separating local energy minima and therefore are better suited to finding the conformer corresponding to the global energy minimum. However, the global energy minimum does not necessarily correspond to the most stable conformer due to entropic effects. Molecules undergo interconversion between conformations, and the relative population of each conformer depends on the free energy. Despite this limitation, a stochastic approach using the potential energy surface is usually the most practical way to explore the conformational freedom of molecules larger than a few dozen atoms.[2,3]

An additional complication of identifying stable conformers arises when the lowest-energy structure in a liquid solvent is sought. The lowest-energy conformer isolated in the gas phase may not correspond to the lowest-energy structure in a liquid solvent due to many-body interactions. Nonspecific solvent effects can be included via a continuum model with the appropriate dielectric constant. Specific solvent effects require the addition of a number of explicit solvent molecules to the system. The relative location of the explicit solvent molecules further complicates the energy surface. Many-body systems also have energy surfaces with multiple energy minima that are frequently shallow. Monte Carlo search algorithms, utilizing first principles quantum chemical energy evaluations, have been used successfully to locate the global energy structures of many-body systems.[1] In this work, we have extended the first principles Monte Carlo (FPMC) method to locate low-energy structures of molecular systems with both intramolecular (conformational) and intermolecular (many-body) degrees of freedom.

COMPUTATIONAL METHODS

FPMC is a modular Perl program that has been discussed in detail previously.[1] Perl is particularly adept at manipulating ASCII text files; this feature is significant since FPMC must control the operation of external commercial software through text input files and process the text output of each

* Corresponding author phone: (847)491-5351; e-mail: broadbelt@northwestern.edu.

10.1021/ci034038h CCC: $25.00
Published on Web 09/16/2003

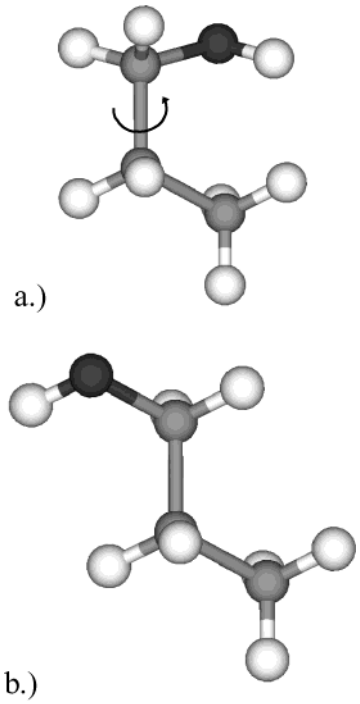

Figure 1. Illustration of the bond rotation move. The C−C−C−O dihedral angle is changed by 180°. Carbon atoms are shown in light gray, oxygen atoms in dark gray, and hydrogen atoms in white.

program. The additional computational overhead of Perl is negligible compared to the computational cost of the quantum chemical calculations. The primary functions of FPMC are to generate attempted molecular moves, set up the input and collect the output from the energy evaluation program, and accept or reject moves. Each of these functions will now be discussed in more detail.

Monte Carlo Moves. Seven types of moves have been implemented in FPMC. The type of move performed in a given iteration is selected randomly with user-specified probabilities for each type of move. Four of the moves, translation, rotation, partial optimization, and molecular interchange, have been discussed in the context of FPMC previously.[1] Briefly, the first two moves are the random translation of or rotation about the center of mass of a rigid molecule, respectively. Partial optimization moves consist of the conventional optimization of one molecule in a system, while the nuclear coordinates of all other molecules remain fixed. This move results in internal relaxation of the selected molecule, allowing it to respond to its current environment. Partial optimization very efficiently moves the system toward a local minimum. Molecular interchange swaps the centers of mass of two molecules without any molecular rotation. This move is always followed by a small number of full conventional optimization steps to remove the resulting strain in the system and increase the probability of accepting the interchange move.

Three additional types of moves have been added to FPMC to address conformational flexibility. The first new move, bond rotation, is illustrated in Figure 1. This type of move is common in conformational searches (e.g., ref 5). All bonds of the molecule for which a rotation would produce a change in the conformation of the molecule were considered for bond rotation. One of the torsion angles of the molecule is randomly selected and then rotated by a random amount.

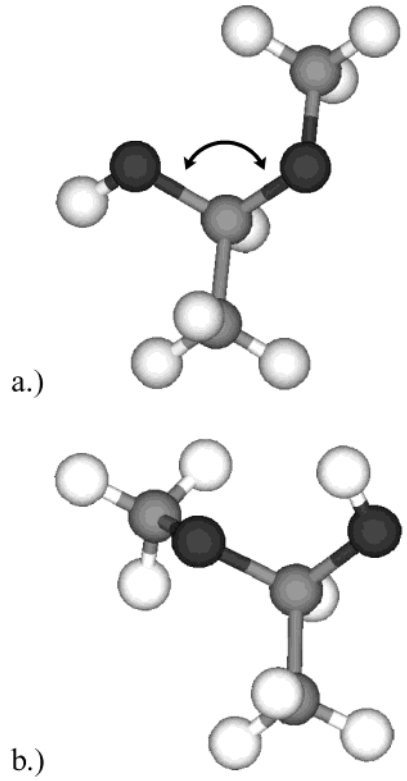

Figure 2. Illustration of the chiral inversion move. The hydroxy and methoxy groups are interchanged.

The resulting structure is then allowed to undergo a limited number of full conventional optimization steps. Simultaneous, multiple bond rotations were not permitted, although the method could be easily extended to allow for this possibility. Another new move, chiral inversion, operates on a randomly selected chiral atom in a molecule. Chiral atoms are automatically identified using the algorithm of Cieplak and Wisniewski.[6] A chiral center and two random groups attached to the chiral atom are then selected; the chiral center and the two atoms connected to it that are members of the selected groups define a plane. The two groups are rotated in this plane into each other's position (Figure 2). This type of move can lead to the overlap of atoms, particularly if the groups are bulky. Hence, the chiral inversion move is always followed by a small number of full conventional optimization steps. The third new move displaces all of the atoms in a molecule along random vectors. This move is known as the Random Incremental Pulse Search (RIPS).[7] Optimization using a potential force field with bond stretching terms is required to restore the molecule; geometry optimization using quantum chemistry cannot guarantee that the initial molecular connectivity will be recovered. The RIPS move can shift the system dramatically on the potential energy surface and change the chirality of the molecule. For simple molecules, the RIPS move was not required because the energy surface could be explored effectively with the other types of moves. In fact, application of the RIPS move to small molecules hindered convergence, as the required minimization using a force field often led to structures that were higher in energy on the quantum mechanical surface than the energy for the geometry prior to the RIPS move. In the present work, the RIPS move was applied to large molecules with many degrees of freedom.

Energy Program Interface. Energy evaluations on a molecular system are performed using external software that is controlled via an FPMC object. The energy object connects to the rest of FPMC through a standard interface. Setting up the input files, running the external software, and collecting the energy and new atomic positions are all localized in the energy object. Therefore, any external software may be used to evaluate the energy with FPMC once the FPMC energy object is created in Perl. FPMC is currently configured to operate with the Gaussian 98[8] and Tinker[9−14] software packages. Gaussian contains a comprehensive suite of quantum chemical methods including semiempirical, Hartree−Fock (HF), post Hartree−Fock, and density functional theory (DFT). Tinker is an open source molecular mechanics program that includes AMBER,[15] CHARMM,[16] and MM3[17] force field parameters. Unlike quantum chemical methods, molecular mechanics computes the energy on the same computational time scale that FPMC performs its functions. Since performing first principles, that is, quantum chemical, calculations is the primary focus of FPMC, no attempts were made to optimize the performance of FPMC for molecular mechanics energy evaluations. Both Gaussian and Tinker contain their own conventional optimization algorithms that can follow the energy gradient downhill to a local energy minimum.

Acceptance of Moves. The probability of accepting a move is based on the energy difference between the new and old configurations

$$P = \min\left[1, \exp\left(\frac{-\Delta E}{kT}\right)\right] \quad (1)$$

where P is the probability (Boltzmann factor) of accepting the move, ΔE is the difference between the new and old energies, k is Boltzmann's constant, and T is the absolute temperature. The energy difference is evaluated in terms of the electronic energies for quantum chemical calculations and the force field energy for molecular mechanics. Hence, FPMC will converge toward minima on these surfaces, not the free energy surface. The computational cost associated with the calculation of thermodynamic properties, particularly for the quantum chemical methods, is impractical at each Monte Carlo step. If a move lowers the energy, the move is always accepted. A move is automatically rejected if the intramolecular connectivity changes (e.g., an unsuccessful RIPS move).

FPMC also performs simulated annealing searches.[18] The temperature used for calculating the Boltzmann factor is gradually lowered during the Monte Carlo search. Moves which result in structures with higher energy are more likely to be accepted at higher temperatures. These unfavorable moves may be necessary to overcome otherwise insurmountable barriers on the energy surface. As the temperature decreases, the probability of accepting moves that increase the energy decreases. The FPMC cooling scheme and scaling of maximum displacement parameters are unchanged from the previous implementation.[1]

RESULTS AND DISCUSSION

FPMC has previously been applied to successfully locate low-energy structures of many-body systems regardless of the starting geometry.[1] Here, the ability of FPMC to locate energy minima in systems with both intramolecular (conformational) and intermolecular (many-body) degrees of freedom was validated by examining three test systems. First, FPMC was used to locate the lowest-energy conformer of glycerol. Second, the lowest-energy structures of D-glyceraldehyde were found in the absence of water and in the presence of one and two discrete solvent molecules. Finally, several low-energy structures of a humic acid monomer unit were determined with FPMC. The FPMC parameters used for all three test cases are listed in Table 1.

Table 1. Parameters Used in the FPMC Optimizations

		D-glyceraldehyde		Steelink	
parameter	glycerol	w/o water	w/ water	MM3	PM3
initial temp (K)	1000	1000	300	2000	2000
minimum temp (K)	1	1	50	300	300
cooling rate (K/move)	0.9	0.9	0.9	0.998	0.99
total number of moves	100	100	500	10000	1000
bond rotation %	90	90	20	45	47.5
chiral inversion %	0	0	0	45	47.5
partial optimization %	10	10	10	5	5
RIPS %	0	0	0	5	0
interchange %			10		
rotation %			30		
translation %			30		

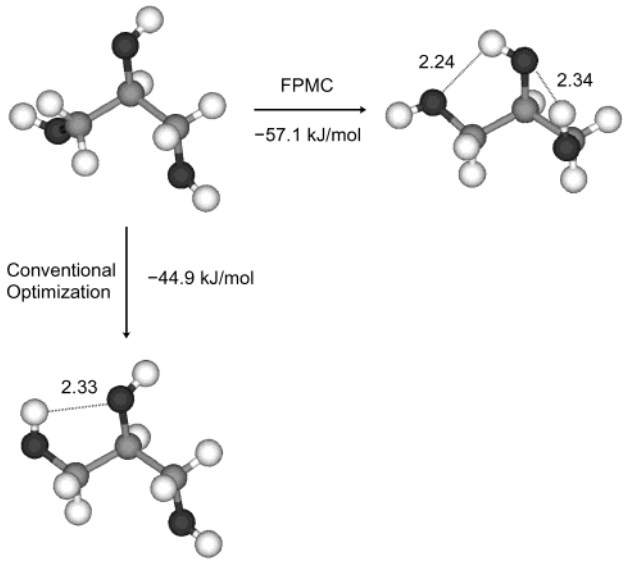

Figure 3. The structure of glycerol at the HF/6-31G* level of theory. Hydrogen bonds are indicated by dotted lines, and all distances are given in angstroms.

Glycerol. Glycerol (1,2,3-propanetriol) has 126 possible conformations. Callam et al.[4] systematically optimized the geometry of every conformer using the Hartree−Fock (HF) and B3LYP (DFT) methods with the 6-31G* basis set. The lowest-energy conformation was different for the two quantum chemical methods. While many FPMC searches could be used to generate a set of low-energy conformers, the emphasis here is to reproduce the lowest-energy conformations corresponding to the HF/6-31G* and B3LYP/6-31G* levels of theory.

Starting from the same initial structure, both FPMC and conventional optimizations were performed. The default parameters for conventional optimization in Gaussian were used. Figures 3 and 4 show the initial and final glycerol structures for the HF/6-31G* and B3LYP/6-31G* methods, respectively. Both conventionally optimized structures have

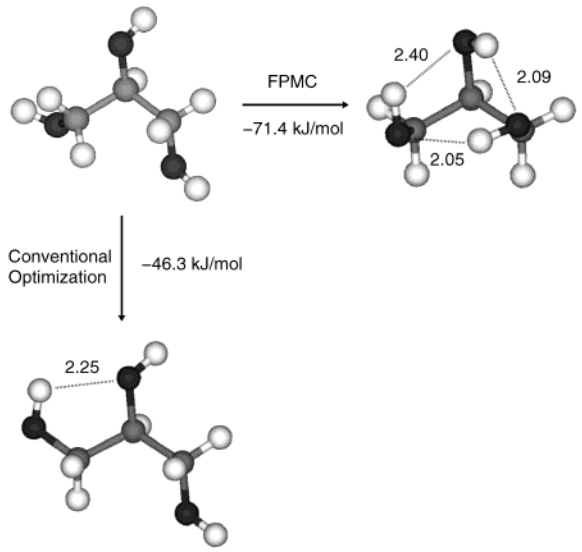

Figure 4. The structure of glycerol at the B3LYP/6-31G* level of theory. Hydrogen bonds are indicated by dotted lines, and all distances are given in angstroms.

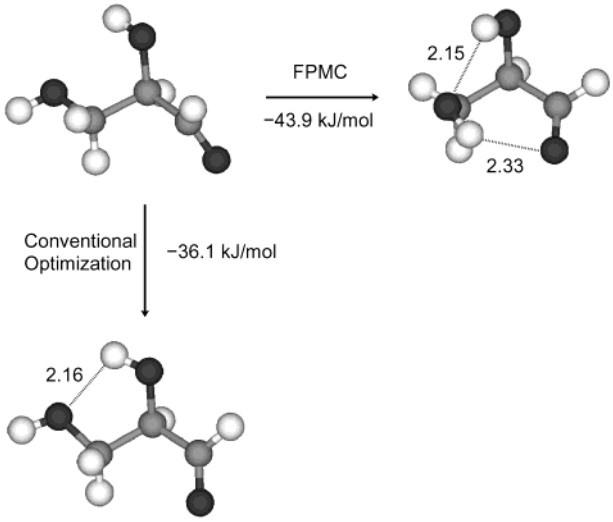

Figure 5. The structure of D-glyceraldehyde at the HF/3-21G level of theory. Hydrogen bonds are indicated by dotted lines, and all distances are given in angstroms.

only one intramolecular hydrogen bond formed by the alcohol groups. The HF/6-31G* FPMC geometry has two intramolecular hydrogen bonds. The electronic energy of the HF/6-31G* FPMC conformer is the same as conformer 95 in the labeling scheme of Callam et al.;[4] conformer 95 of the systematic search was the HF/6-31G* structure with the global minimum energy. The bond rotation moves implemented in FPMC lower the energy by 12.2 kJ/mol compared to the conventionally optimized result starting from the same initial structure. Note that the bond rotation moves were selected randomly from among all possible bond rotations as described in the methods section. Furthermore, simultaneous, multiple bond rotations were not allowed. The B3LYP/6-31G* FPMC geometry has three intramolecular hydrogen bonds, one more than the HF/6-31G* structure. Unlike any of the other structures, all three hydroxy groups form hydrogen bonds with each other. The structure identified by FPMC corresponds to the global energy minimum on the B3LYP/6-31G* energy surface (conformer 86 in Callam et al.[4]). The additional intramolecular interactions provide an additional 25.1 kJ/mol of stabilization compared to the single hydrogen bond in the conventionally optimized structure.

For both energy surfaces, the energy of the FPMC structure is lower than the conventionally optimized result starting from the same initial structure. Further, FPMC successfully converged to the global energy minimum identified in the exhaustive search by Callam et al.[4] for both the HF/6-31G* and B3LYP/6-31G* levels of theory. One of the key benefits of the FPMC method is that the guess for the initial structure is mostly arbitrary; FPMC, given a sufficient number of total moves, can locate the lowest-energy structure regardless of the initial geometry. Hence, the lowest-energy structure can be found knowing essentially only the molecular connectivity. Although 100 total moves were allowed for an FPMC glycerol run, the energy typically converged after about 20 total moves; usually 30−40% of the moves were accepted. At the HF/6-31G* level of theory, 16 conventional optimization cycles were necessary to optimize the glycerol molecule. If this value is representative of the glycerol conformers, then over 2000 conventional optimization cycles would be required to optimize all 126 possible conformers.

D-Glyceraldehyde. Without Water. Glyceraldehyde (2,3-dihydroxypropanal) is the simplest carbohydrate and one of the simplest molecules that exhibits chirality. Here, only the D-glyceraldehyde stereoisomer was investigated to mirror previous conformational investigations.[19,20] One of these investigations[19] explored the conformational structure with several quantum chemical methods, including the HF/3-21G and B3LYP/6-31G** methods. For the HF/3-21G energy surface, the electronic energy of the FPMC optimized result is 7.8 kJ/mol lower than the conventionally optimized structure, both starting from the same arbitrary initial structure (Figure 5). Conventional optimization required 13 cycles, while the FPMC energy typically converged after about 15 total moves. Once again, the bond rotation moves allow FPMC to cross barriers on the molecular potential energy surface unlike the conventional optimization scheme. Hence, the conventionally optimized geometry will generally resemble the initial structure. The conventionally optimized structure has a single intramolecular hydrogen bond between the two hydroxy groups; the carbonyl group does not interact with either hydroxy group. In contrast, all three functional groups interact in the lower energy FPMC structure. The terminal hydroxy group forms intramolecular hydrogen bonds with the carbonyl oxygen atom and the hydroxy group on the central carbon.

FPMC also results in a structure that is 17.7 kJ/mol lower in energy than the conventionally optimized result on the B3LYP/6-31G** energy surface (Figure 6). The FPMC energy typically converged after about 15 total moves, and conventional optimization required 16 cycles. Once again, the conventionally optimized result closely resembles the initial guess. Similar to the HF/3-21G FPMC structure, all three functional groups interact in the B3LYP/6-31G** FPMC geometry. However, in contrast to the HF/3-21G FPMC structure, the hydroxy group on the central carbon interacts with the carbonyl and terminal hydroxy groups as shown in Figure 6. The FPMC results for both of these quantum chemical methods agree with the theoretical gas-phase calculations reported previously in which several

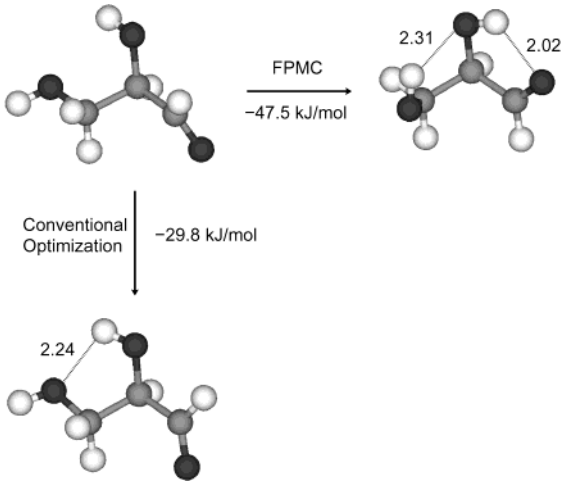

Figure 6. The structure of D-glyceraldehyde at the B3LYP/6-31G** level of theory. Hydrogen bonds are indicated by dotted lines, and all distances are given in angstroms.

individual conformers were evaluated.[19] Hence, FPMC successfully located the global minimum energy conformations of gas-phase D-glyceraldehyde, a molecule with a significant degree of intramolecular hydrogen bonding.

With Water. The structure of glyceraldehyde in the presence of one and two discrete water molecules was investigated with FPMC to examine its ability to perform simultaneous conformational and many-body optimizations. While the conformation of D-glyceraldehyde has been previously investigated with continuum solvent models,[20] the specific interaction of glyceraldehyde and water has not been previously examined with quantum chemistry. Only the HF/3-21G energy surface was examined directly with FPMC. The set of low-energy geometries produced by multiple HF/3-21G FPMC searches were then conventionally optimized with B3LYP/6-31G**. This two-step approach is attractive since the number of Monte Carlo steps required to fully explore a complex energy surface (i.e., one with intramolecular and intermolecular degrees of freedom) is usually large. The energy of the FPMC runs typically converged after 200−400 total moves, with 10−20% acceptance. Note in Table 1 that the array of the types of moves applied when water was present was much broader than for glyceraldehyde with no water present. Multiple entities require new moves to guide their relative placement. Also, while the single lowest-energy conformation frequently varies for different levels of theory, the *set* of low-energy conformers is usually similar regardless of the method.[4]

The structure of D-glyceraldehyde with one solvating water molecule is shown in Figure 7. The FPMC structure is 32 kJ/mol lower in energy than the conventionally optimized structure starting from the same arbitrary initial geometry. Conventional optimization required 82 cycles to converge. The conventional optimization significantly relocates the water molecule relative to the glyceraldehyde molecule. However, the glyceraldehyde conformation is largely unchanged as a result of the conventional optimization. The water molecule forms hydrogen bonds with both the terminal hydroxy group and the aldehyde hydrogen atom. The carbonyl oxygen atom and the central hydroxy group solely interact with each other. The water molecule in the FPMC geometry inserts itself into all of the intramolecular glyceraldehyde hydrogen bonds that were observed in the nonsolvated case; no intramolecular hydrogen bonds are formed with water present. Both hydroxy hydrogen atoms are directed toward the oxygen atom of the water molecule, while one of the hydrogen atoms of water is interacting with the carbonyl oxygen. This structure could not be easily guessed, and an exhaustive search would require the conventional optimization of hundreds or thousands of possible intra- and intermolecular configurations.

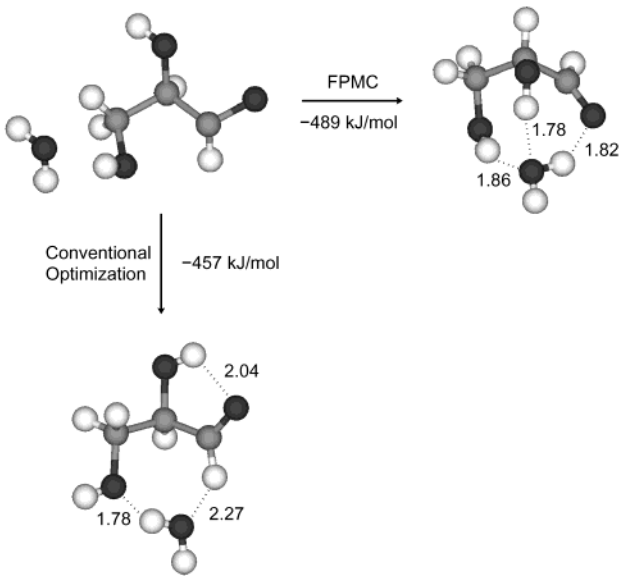

Figure 7. The structure of the D-glyceraldehyde and water system at the HF/3-21G level of theory. Hydrogen bonds are indicated by dotted lines, and all distances are given in angstroms.

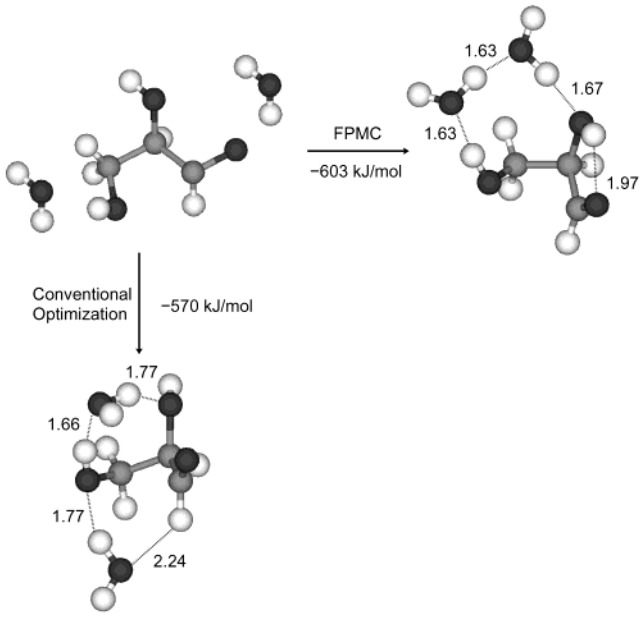

Figure 8. The structure of the D-glyceraldehyde and two-water system at the HF/3-21G level of theory. Hydrogen bonds are indicated by dotted lines, and all distances are given in angstroms.

Figure 8 shows the intermolecular geometry of D-glyceraldehyde and two water molecules. The FPMC result is 33 kJ/mol lower in energy than the conventionally optimized structure. The structure obtained by conventional optimization required 65 cycles. Once again the water molecules are significantly displaced from their initial positions in the

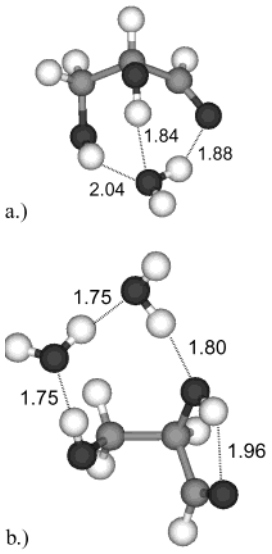

Figure 9. The structure of D-glyceraldehyde at the B3LYP/6-31G** level of theory with (a) one and (b) two waters. Hydrogen bonds are indicated by dotted lines, and all distances are given in angstroms.

conventionally optimized configuration. The glyceraldehyde conformation is similar to its initial geometry; however, the terminal carbon-alcohol group has rotated approximately 90° about the carbon−carbon bond. One of the water molecules interacts with both hydroxy groups. The second water molecule forms a hydrogen bond to the terminal hydroxy oxygen atom and the aldehyde hydrogen atom. The carbonyl oxygen atom does not interact with the water molecules nor the other glyceraldehyde functional groups. In contrast, the carbonyl oxygen is hydrogen bonded to the central hydroxy group in the FPMC-optimized geometry. The two water molecules form an extended hydrogen bonding structure connecting the two glyceraldehyde hydroxy groups.

Since the number of total moves is limited by the computational cost of ab initio calculations, the system may converge to a local energy minimum rather than the global energy minimum. Rather than increase the total number of moves, optimizations using FPMC were performed several times. FPMC moves that increase the energy are more likely to be accepted at the beginning of an FPMC run since the annealing temperature is high. Hence, configuration space should be more efficiently explored by multiple FPMC runs than by a single FPMC run with the same number of total moves. The FPMC structures shown in Figures 7 and 8 are the lowest-energy structures generated from several FPMC runs.

All of the unique FPMC structures from the multiple runs were conventionally optimized at the B3LYP/6-31G** level of theory. For both one and two water molecules, the lowest-energy structures obtained using B3LYP/6-31G** are very similar to the HF/3-21G FPMC structures (Figure 9). The bond lengths and angles vary slightly, but the relative position of water and the conformational structure of D-glyceraldehyde are the same. The lowest-energy structures of a system are not always the same for different levels of theory; for example, the HF/6-31G* and B3LYP/6-31G* structures of glycerol differ (Figures 3 and 4). However, by using a lower level of theory to generate a set of low-energy structures, which are typically the same regardless of the level of theory,[4] finding the lowest-energy structure at higher levels of theory becomes practical.

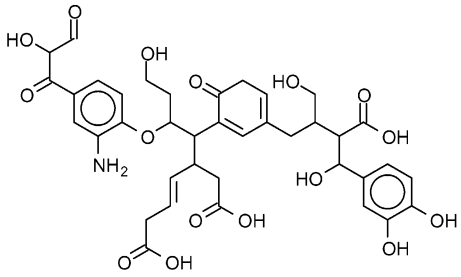

Figure 10. Molecular structure of the Steelink humic acid monomer unit.

For a given level of the theory, the lowest-energy conformation of D-glyceraldehyde depends on the number of solvent water molecules. The gas-phase conformation (Figure 5) is considerably different than either of the two water-solvated conformations (Figures 7 and 8). This is in direct contrast to the continuum approximation of D-glyceraldehyde in aqueous solution. The lowest-energy HF/6-31G** D-glyceraldehyde conformer in the continuum solvent closely resembles the gas-phase geometry. The FPMC results suggest that the specific solvent interactions alter the energy surface so that the lowest-energy conformation of D-glyceraldehyde changes significantly. Calculations of D-glyceraldehyde with many more explicit water molecules in addition to a continuum solvent are needed to fully resolve the conformation of D-glyceraldehyde in aqueous solution.

Humic Acid Monomer Unit. Humic substances are ubiquitous in the natural environment and are especially prominent in soil organic matter.[21] Adsorption of groundwater contaminants to soil organic matter is primarily responsible for limiting their bioavailability (ref 22 and references therein). A structural model of humic substances would greatly increase the understanding of how metal and organic contaminants interact with humic substances. The proposed humic acid monomer units of Steelink[23] and Davies et al.[24] are the logical starting points for building up the larger structure of humic substances.

The structure of the Steelink humic acid monomer unit is shown in Figure 10. The structure has a high degree of conformational flexibility as well as several stereocenters. Unlike glycerol or glyceraldehyde, the lowest-energy structure of the Steelink monomer unit cannot be found with absolute certainty; the sheer number of possible conformations presents an insurmountable obstacle. However, FPMC can still be applied to find a set of low-energy conformations. We sought to validate the FPMC method for a system with many degrees of freedom by reproducing Steelink results from the literature.[25,26] Since the Steelink unit contains almost 100 atoms, even moderate-level quantum chemical calculations are not practical for the numerous energy evaluations required by the FPMC method. Instead, the MM3 molecular mechanics (Tinker) and PM3 semiempirical quantum chemical (Gaussian) methods were employed. While the PM3 method is arguably "first principles", the MM3 method uses force field parameters fitted to experimental data and high-level quantum chemical calculations. However, despite the name of the FPMC method, FPMC is not limited to first principles energy evaluations; a key feature is that any

Table 2. Lowest Five Steelink Energies (kJ/mol) Found during MM3 FPMC Optimizations

conformer	MM3	PM3[a]	B3LYP/6-31G**[a]
M8	137.7	−2147.7	−7 173 304
M5	139.3	−2145.1	−7 173 330
M1	145.2	−2142.5	−7 173 264
M9	146.8	−2133.6	−7 173 301
M17	150.0	−2160.7	−7 173 295
conventional[b]	290.0	−2085.3	−7 173 152
ref 25	n/a	−515.5[c]	n/a
ref 26	n/a	−2558.1[d]	n/a

[a] Single point calculation on MM3 structure. [b] Conventional optimization of initial structure. [c] PM3 single point on force field-generated conformers. [d] PM3/SM5.4 conventional optimization on force field-generated conformers.

Table 3. Lowest Five Steelink Energies (kJ/mol) Found during PM3 FPMC Optimizations

conformer	PM3	B3LYP/6-31G**[a]
P12	−2560.1	−7 173 391
P20	−2552.3	−7 173 378
P10	−2551.7	−7 173 356
P6	−2549.2	−7 173 309
P4	−2548.8	−7 173 310
conventional[b]	−2487.1	−7 173 196
ref 25	−515.5[c]	n/a
ref 26	−2558.1[d]	n/a

[a] Single point calculation on PM3 structure. [b] Conventional optimization of initial structure. [c] PM3 single point on force field-generated conformers. [d] PM3/SM5.4 conventional optimization on force field-generated conformers.

external energy evaluation software can be employed. Hence, the suitability of using more computationally efficient molecular mechanics methods can be directly compared to not only the more accurate but also the more computationally expensive, quantum chemical calculations.

Sein et al.[25] employed a multistep algorithm to locate low-energy Steelink conformers. Briefly, they first conventionally optimized one arbitrary conformer for each of the 64 possible Steelink stereoisomers using the Tripos/Sybyl force field.[27] A stochastic RIPS search, also with the Tripos/Sybyl force field, was then performed on the five lowest-energy structures from the first step. A single point PM3 calculation was then performed on the set of lowest-energy conformations from the RIPS step; only the PM3 energies were reported. Bruccoleri et al.[26] performed Monte Carlo searches on an arbitrary initial structure also using the Tripos/Sybyl force field. The 60 lowest-energy conformers generated from the Monte Carlo search were then conventionally optimized using PM3, with and without the SM5.4 continuum solvation method;[28] only the solvated PM3/SM5.4 energies were reported.

Twenty conformers of the Steelink unit were generated with FPMC for the MM3 energy surface starting from an arbitrary initial structure. The FPMC parameters are listed in Table 1. Table 2 shows the energies corresponding to the five lowest-energy Steelink structures obtained by FPMC with MM3. The results of PM3 and B3LYP/6-31G** single point calculations are also reported in the same table. Not surprisingly, the FPMC results are clearly superior to a direct conventional optimization on the arbitrary initial structure. For a molecule with as many internal degrees of freedom as the Steelink monomer unit, the probability of choosing an initial structure with no energy barrier preventing the direct conventional optimization to the global minimum is extremely small. The MM3, PM3, and B3LYP energies of the M8 conformer are 152, 62, and 152 kJ/mol lower than the conventionally optimized structure, respectively. The energetic order of the MM3-generated conformers varies; for example, the M8 conformer has the lowest MM3 energy, while the M5 conformer has the lowest B3LYP energy.

Comparing the PM3 energies from the literature to the values in Table 2 reveals that FPMC produces conformers that are dramatically lower in energy compared to those found by Sein et al.[25] This is especially surprising for the MM3 FPMC search since it also utilized the RIPS move. It is therefore likely that the addition of the bond rotation, chiral inversion, and partial optimization moves in FPMC can account for the difference in the PM3 energies. The FPMC PM3 energies are comparable, although higher in energy, to that found by Bruccoleri et al.[26] However, the PM3 single point energies on the FPMC/MM3 generated conformers are not directly comparable to the PM3 energies reported by Bruccoleri et al.[26] First, their investigation includes solvation effects through the SM5.4 continuum solvation model. Second, they conventionally optimized their conformers on the PM3/SM5.4 energy surface, while our values in Table 2 are only single point PM3 calculations on the MM3 FPMC generated structures.

FPMC conformers were also generated on the PM3 energy surface to provide a better comparison between the FPMC method and the energies reported by Bruccoleri et al.[26] The same initial starting geometry was used for both the FPMC/MM3 and FPMC/PM3 investigations. The five lowest-energy conformers from the PM3 FPMC search are listed in Table 3. The FPMC results are once again superior to the conventional optimization of the arbitrary initial structure. The PM3 energies of the FPMC/PM3 structures are considerably lower than the single point PM3 energies corresponding to configurations identified by the MM3 FPMC search (Table 2). This is not surprising considering that the PM3 FPMC search is locating minima on the PM3 energy surface, while the MM3 FPMC is not. The PM3 energies from the FPMC/PM3 search are very similar to the PM3/SM5.4 energy reported by Bruccoleri et al.[26] and include a value that is lower in energy (P12) than theirs. The similarity between these energies shows that the FPMC method can successfully locate low-energy structures, even on very complicated energy surfaces.

In addition to the PM3 heat of formation, the B3LYP/6-31G** electronic energy was calculated for each of the conformers (Tables 2 and 3). The B3LYP quantum chemical approximation is much less severe than the PM3 method, which replaces many of the two-electron integrals with fitted parameters;[29] the relative B3LYP energies should provide the best estimate of the true energetic stability of the conformers. On this basis, the FPMC search on the PM3 energy surface generated the best set of conformers. The lowest-energy PM3 conformer (P12) is 60.8 kJ/mol lower on the B3LYP energy surface than the best MM3 conformer (M5). The M5 and P12 monomers are illustrated in Figure 11. The MM3 conformers generally favor a compact structure with some degree of ring stacking, while the PM3 conformers prefer an elongated structure without any ring stacking.[26,30] The MM3 force field method generally overestimates the

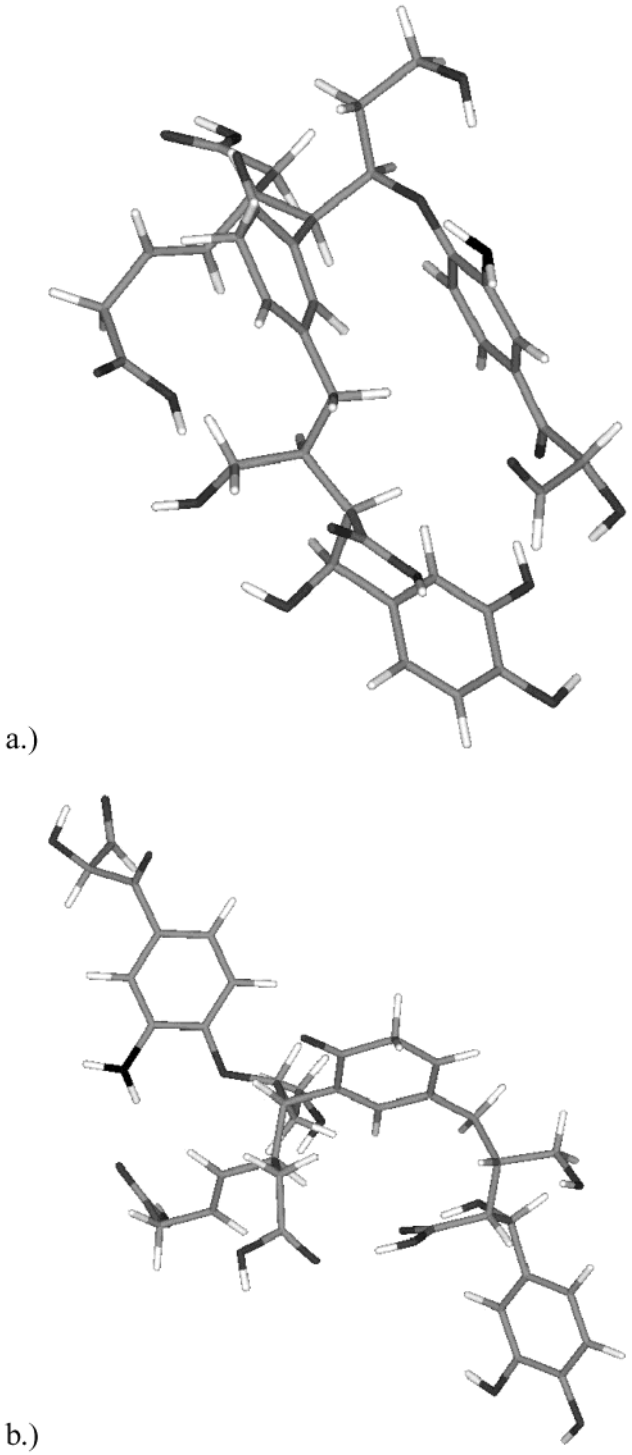

Figure 11. Three-dimensional structure of the Steelink humic acid monomer unit (a) conformer M5 and (b) conformer P12. Carbon atoms are shown in light gray, oxygen atoms in dark gray, nitrogen atoms in black, and hydrogen atoms in white.

additional stabilization due to aromatic ring stacking. The MM3 method was parametrized to reproduce the experimental structure of small molecules and may not accurately predict the structure of larger molecules such as the Steelink monomer unit. First principles calculations, even in the semiempirical limit of the PM3 method, are generally applicable regardless of the molecular system. The B3LYP results suggest that the semiempirical PM3 method more accurately reflects the true Steelink energy surface than the MM3 force field. One of the key features of the FPMC method is the ability to use any energy evaluation method, from high-level quantum chemical calculations to semiempirical methods to force field models, to explore complicated energy surfaces.

CONCLUSIONS

The structure of molecules with conformational flexibility in solution is difficult to determine using computational chemistry due to the complexity of the potential energy surface. The first principles Monte Carlo (FPMC) technique was successfully extended to address this issue. First principles calculations are considerably more expensive than molecular mechanics, limiting the application of FPMC to relatively small systems. However, quantum chemical methods are not limited by the availability or applicability of force fields. FPMC calculations successfully found the minimum energy structure of glycerol in the gas phase at two levels of theory. The structure of D-glyceraldehyde was determined in the gas phase as well as in the presence of explicit solvent molecules. Finally, a set of low-energy Steelink humic acid monomer geometries was determined that significantly improves upon some of the data found in the literature. In all of these cases, the energy of the FPMC-optimized geometry was lower than the energy of the corresponding conventionally optimized geometry. The ability of FPMC to locate minimum energy structures of systems with conformational and many-body degrees of freedom starting from arbitrary initial geometries is a demonstrated strength of the method.

ACKNOWLEDGMENT

This work was supported by the EMSI program of the National Science Foundation and the Department of Energy (CHE-9810378) at the Northwestern University Institute for Environmental Catalysis.

Supporting Information Available: The energies and Cartesian coordinates of all of the structures. This material is available free of charge via the Internet at http://pubs.acs.org. The PERL scripts comprising FPMC may be obtained free of charge by contacting the corresponding author.

REFERENCES AND NOTES

(1) Haubein, N. C.; McMillan, S. A.; Broadbelt, L. J. Many-Body Optimization Using an Ab Initio Monte Carlo Methodol. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 68−74.

(2) Leach, A. R. A Survey of Methods for Searching the Conformational Space of Small and Medium-Sized Molecules. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley: New York, 1991; Vol. 2, pp 1−55.

(3) Treasurywala, A. M.; Jaeger, E. P.; Peterson, M. L. Conformational Searching Methods for Small Molecules. III. Study of Stochastic Methods Available in SYBYL and MACROMODEL. *J. Comput. Chem.* **1996**, *17*, 1171−1182.

(4) Callam, C. S.; Singer, S. J.; Lowary, T. L.; Hadad, C. M. Computational Analysis of the Potential Energy Surfaces of Glycerol in the Gas and Aqueous Phases: Effects of Level of Theory, Basis Set, and Solvation on Strongly Intramolecularly Hydrogen-Bonded Systems. *J. Am. Chem. Soc.* **2001**, *123*, 11743−11754.

(5) Li, Z.; Scheraga, H. A. Monte Carlo-Minimization Approach to the Multiple-Minima Problem in Protein Folding. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 661−6615.

(6) Cieplak, T.; Wisniewski, J. L. A New Effective Algorithm for the Unambiguous Identification of the Stereochemical Characteristics of Compounds During Their Registration in Databases. *Molecules* **2001**, *6*, 915−926.

(7) Ferguson, D. M.; Raber, D. J. A New Approach to Probing Conformational Space with Molecular Mechanics: Random Incremental Pulse Search. *J. Am. Chem. Soc.* **1989**, *111*, 4371−4378.

(8) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; J. A. Montgomery, J.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Rega, N.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; 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.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98, Revision A.11.3*; Gaussian, Inc.: Pittsburgh, PA, 2002.

(9) Ponder, J. W.; Richards, F. M. An Efficient Newton-like Method for Molecular Mechanics Energy Minimization of Large Molecules. *J. Comput. Chem.* **1987**, *8*, 1016−1024.

(10) Kundrot, C. E.; Ponder, J. W.; Richards, F. M. Algorithms for Calculating Excluded Volume and Its Derivatives as a Function of Molecular Conformation and Their Use in Energy Minimization. *J. Comput. Chem.* **1991**, *12*, 402−409.

(11) Dudek, M. J.; Ponder, J. W. Accurate Modeling of the Intramolecular Electrostatic Energy of Proteins. *J. Comput. Chem.* **1995**, *16*, 791−816.

(12) Kong, Y.; Ponder, J. W. Reaction Field Methods for Off-Center Multipoles. *J. Chem. Phys.* **1997**, *107*, 481−492.

(13) Dudek, M.; Ramnarayan, K.; Ponder, J. W. Protein Structure Prediction Using a Combination of Sequence Homology and Global Energy Minimization II. Energy Functions. *J. Comput. Chem.* **1998**, *19*, 548−573.

(14) Pappu, R. V.; Hart, R. K.; Ponder, J. W. Analysis and Application of Potential Energy Smoothing for Global Optimization. *J. Phys. Chem. B* **1998**, *102*, 9725−9742.

(15) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollmann, P. AMBER: A Package of Computer Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995**, *91*, 1−41.

(16) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(17) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. Molecular Mechanics − The MM3 Force-Field for Hydrocarbons. *J. Am. Chem. Soc.* **1989**, *111*, 8551−8566.

(18) Kirkpatrick, S.; Gelatt, C. D., Jr.; Vecchi, M. P. Optimization by Simulated Annealing. *Science* **1983**, *220* (4598), 671−680.

(19) Lozynski, M.; Rusinska-Roszak, D.; Mack, H.-G. MP2 and Density Functional Studies of Hydrogen Bonding in Model Trioses: D-(+)-Glyceraldehyde and Dihydroxyacetone. *J. Phys. Chem. A* **1997**, *101*, 1542−1548.

(20) Kikuchi, O. Ab initio GB Study of Conformational Space of D-glyceraldehyde in Aqueous Solution. *J. Mol. Struct.-Theochem* **2000**, *496*, 145−152.

(21) Davies, G.; Ghabbour, E. A.; Steelink, C. Humic Acids: Marvelous Products of Soil Chemistry. *J. Chem. Educ.* **2001**, *78*, 1609−1614.

(22) Luthy, R. G.; Aiken, G. R.; Brusseau, M. L.; Cunningham, S. D.; Gschwend, P. M.; Pignatello, J. J.; Reinhard, M.; Traina, S. J.; Weber, W. J., Jr.; Westall, J. C. Sequestration of Hydrophobic Organic Contaminants by Geosorbents. *Environ. Sci. Technol.* **1997**, *31*, 3341−3347.

(23) Steelink, C. Implications of Elemental Characteristics of Humic Substances. In *Humic Substances in Soil, Sediment, and Water*; Aiken, G. R., McKnight, D. M., Wershaw, R. L., MacCarthy, P., Eds.; Wiley: New York, 1985; pp 457−476.

(24) Davies, G.; Fataftah, A.; Cherkasskiy, A.; Ghabbour, E. A.; Radwan, A.; Jansen, S.; Kolla, S.; Paciolla, M. D.; Sein, L. T., Jr.; Buermann, W.; Balasubramanian, M.; Budnick, J.; Xing, B. Tight Metal Binding by Humic Acids and Its Role in Biomineralization. *J. Chem. Soc., Dalton T.* **1997**, 4047−4060.

(25) Sein, L. T., Jr.; Varnum, J. M.; Jansen, S. A. Conformational Modeling of a New Building Block of Humic Acid: Approaches to the Lowest Energy Conformer. *Environ. Sci. Technol.* **1999**, *33*, 546−552.

(26) Bruccoleri, A. G.; Sorenson, B. T.; Langford, C. H. Molecular Modeling of Humic Structures. In *Humic Substances: Structures, Models and Functions*; Ghabbour, E. A., Davies, G., Eds.; Royal Society of Chemistry: Cambridge, 2001; pp 193−208.

(27) Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. Validation of the General-Purpose Tripos 5.2 Force-Field. *J. Comput. Chem.* **1989**, *10*, 982−1012.

(28) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Model for Aqueous Solvation Based on Class IV Atomic Charges and First Solvation Shell Effects. *J. Phys. Chem.* **1996**, *100*, 16385−16398.

(29) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209−220.

(30) Kamishima, M.; Kojima, M.; Yoshikawa, Y. Calculation of Stacking Geometries by a Modified MM2 Method. *J. Comput. Chem.* **2001**, *22*, 835−845.

(31) Sein, L. T., Jr. Personal communication, 2002. Sein et al.[25] report the PM3 binding energy, −41 518.7 kJ/mol, the only value shown in the Hyperchem graphical user interface, not the heat of formation. The PM3 energy reported by most software packages is not the binding energy but rather the heat of formation. The Hyperchem PM3 binding energy can be converted to the heat of formation by subtracting the sum of the PM3 heats of formation of the constituent atoms, −41 003.2 kJ/mol. This conversion was confirmed by calculating the binding energy as well as the heat of formation for several of the Steelink monomer structures. Hence, the heat of formation corresponding to the lowest-energy conformer found by Sein et al.[25] is −515.5 kJ/mol. This corrected value is listed in Tables 2 and 3.

CI034038H