Efficiency of the Local Torsional Deformations Method for Identifying the Stable Structures of Cyclic Molecules

Canan Baysal and Hagai Meirovitch*

Supercomputer Computations Research Institute, Florida State University, Tallahassee, Florida 32306-4052

Received: September 9, 1996[⊗]

A new method for generating the low-energy structures of a chain molecule was proposed recently by us. This is a stochastic process where at each step an energy-minimized structure is changed by carrying out several *local* torsional deformations (LTDs) along the chain, which temporarily disrupt neighbors of the rotated bonds. The energy is then minimized and the disrupted bonds return to their usual geometry (in terms of bond lengths and angles) while the chain assumes a new conformation. This conformation is accepted (and then deformed) or rejected with the help of a "selection procedure" that gives preference to accepting the lower energy structures and, thus, directs the search toward the lowest energy regions, which include the global energy minimum (GEM) structure. The selection procedures tested are the Monte Carlo minimization (MCM) method of Li and Scheraga and the "usage directed" (UD) method of Still's group. LTD is a general method whose parameters can be optimized for any chain system. However, because of the local character of the conformational change, it is expected to be especially efficient for cyclic peptides, loops in proteins, and dense multichain systems. In this paper, LTD is applied to cycloheptadecane modeled by the MM2 force field, its parameters are optimized, and it is found to be more efficient than other methods. The results for this molecule and for an ECEPP model of the linear pentapeptide Leu-enkephalin show that MCM and UD are almost comparable in efficiency, with a slight advantage for MCM.

I. Introduction

The function of biomolecules, such as peptides and proteins, is determined to a large extent by their most stable three dimensional structures.[1,2] Therefore, it is important to predict these structures from theoretical considerations based solely on interatomic interactions. These interactions are usually modeled by an empirical potential energy function that defines the energy surface of the molecule, which even for a short peptide consists of a tremendous number of local energy wells (microstates). Identifying the most stable microstates, i.e., those of the lowest free energy, is not an easy task. Exact thermodynamic simulation methods, such as the Metropolis Monte Carlo (MC) method[3] and molecular dynamics[4] (MD), are very inefficient at room temperature since the molecule is likely to become trapped in a low-potential-energy well close to the conformation from which the simulation was started. Various techniques for a conformational search (CS) have been developed to surmount this problem;[5−12] however, most of them do the CS at the expense of replacing the search for microstates of low-free-energy with a search for low-minimized-energy structures. In practice, even the latter methods are limited to handle only relatively short peptides or flexible surface loops of a protein which are not well-determined experimentally by X-ray crystallography or multidimensional NMR. This approach has been applied, for example, to the complementarity-determining regions (CDRs) of antibodies,[9d,f] where the part of the structure that is well-defined is kept fixed, and to the "missing loops" in structure determination of a protein from the known structures of its homologous proteins[9b,g] (see also references cited in ref 1). Therefore, development of efficient methods for a conformational search is still a challenge in theoretical structural biology.

In a recent paper,[13] we proposed a new CS method, which is mainly designed for cyclic chain molecules, loops in proteins, and dense multichain systems. This is a stochastic process where at each of its steps an energy-minimized structure is changed by carrying out several *local* torsional deformations (LTDs) along the chain, which temporarily disrupt the neighboring bonds of the rotated ones; the energy is minimized and the disrupted bonds return to their usual geometry (in terms of bond lengths and angles) while the molecule assumes a new structure. This structure is then accepted or rejected with the help of a "selection procedure", where in the latter case another energy-minimized structure is selected. The selected structure is deformed and energy-minimized, and the process continues. The selection procedure gives preference to accepting the lower energy structures and, thus, directs the search toward the lowest energy regions in the conformational space, which include the global energy minimum (GEM) conformation. Several selection procedures have been studied in the literature, and we chose to use the MC minimization (MCM) method of Li and Scheraga;[5] the combined method is denoted LTD(MCM). It should be pointed out that LTD defines a significant deformation and therefore enables the system to cross energy barriers. Because of its local character, it is particularly suitable for treating dense systems. However, the extent of conformational change can be increased and optimized for any chain length and environment, in the same way that the efficiency of the Monte Carlo method for polymers is controlled by the the size of the repeated conformational move. Thus, the *global* conformational moves of the pivot algorithm lead to the most efficient sampling for a single chain in a good solvent,[14] while *local* moves are required for a polymer in a dense or geometrically confined environment.[15] Notice, however, that unlike these MC methods, LTD(MCM) does not lead to a Boltzmann distribution of conformations but rather constitutes a tool for generating low-energy structures.

LTD(MCM) was applied preliminarily to a relatively small cyclic molecule, cycloundecane described by the MM2 force field,[16] and was found to be more efficient than other tech-

[⊗] Abstract published in *Advance ACS Abstracts,* February 15, 1997.

S1089-5639(96)02753-3 CCC: $14.00

niques.[13] In the present paper, we test it as applied to an MM2 model of a larger cyclic molecule, cycloheptadecane, which has become a benchmark for testing the performance of CS algorithms.[8e,g,9e] As in ref 13, the LTD results are compared to the results obtained with the Monte Carlo multiple minimum (MCMM) method of Still's group,[8f] which is based on a *global* torsional deformation and is one of the best methods available.[8g] In ref 8f, MCMM was tested with several different selection procedures, where the most efficient was found to be "usage directed" (UD). It is also of interest to compare the efficiencies of MCM and UD, and therefore, the calculations are carried out with MCMM(MCM), MCMM(UD), LTD(MCM), and LTD(UD). To check this point further, we apply MCM and UD to the linear pentapeptide Leu-enkephalin (H-Tyr-Gly-Gly-Phe-Leu-OH) described by the potential energy function ECEPP.[17]

LTD is an important ingredient of a new methodology, based on statistical mechanical considerations, for analyzing the multidimensional NMR data of peptides.[18] Short peptides or organic chain molecules in solution are in most cases random coils but under certain solvent conditions might produce medium- and long-range nuclear Overhauser effect intensities (NOEs). However, in many cases, the latter are only compatible with a molecule that populates *several* microstates in thermodynamic equilibrium, rather than the *single* microstate of a globular protein. Analysis of such NOE data is difficult because of the need to identify the most stable states and calculate their thermodynamic populations (see refs 18 and 19 and references cited therein). Our methodology has been developed and applied thus far to the pentapeptide Leu-enkephalin described by ECEPP, and it is being extended now to cyclic peptides. It consists of several stages.[18] First, the relative contribution of microstates to the partition function Z as their (minimum) energy is increased above the GEM are calculated. For Leu-enkephalin, it was found that at a temperature of 280 K, the localized microstates of energy within 2 and 3 kcal/mol above the GEM contribute 0.6 and 0.75 of Z, respectively.[18a] Such a study enables one to determine the significant energy range, i.e., that which provides the dominant contribution to the NOEs. The next step is to carry out an extensive conformational search (e.g., with LTD) for energy-minimized structures within this range. Since their number is relatively large, one selects a smaller set of structures that are *significantly different*, which become "seeds" for Monte Carlo (MC) simulations. The corresponding samples are called MC microstates, and their free energies, which lead to the populations, are obtained by the local states method.[20] The overall NOE intensity is an average over the individual contributions of the MC microstates, weighted by their populations. This methodology is also applicable to flexible loops or chain ends in proteins.

II. Methods

In this section, we discuss the difficulties in searching for the low-energy structures. We describe our LTD procedure together with some of the existing methods used in the CS of cyclic chain molecules, in particular MCMM, MCM, and UD.

II.1. Conformational Search Methods for Cyclic Molecules. The problem of conformational search can be gathered from Figure 5 of ref 21 or Figure 23 of ref 9e, where the number of energy-minimized structures is plotted vs the energy of the molecule whose conformational space is searched. It is shown that only a few minima exist near the GEM; their number increases strongly for higher energies, reaches a maximum for relatively high energy, and decreases again. This typical behavior (see also refs 18a and d) means that a CS method that is based on random selection of conformations followed by energy minimization will mostly lead to the highly populated high-energy structures. An efficient method, therefore, should give a strong preference for generating low-energy structures.

The existing CS methods for cyclic molecules can be divided into several categories. One approach includes the pioneering work of Gō and Scheraga,[9a,b] which provides a solution for the ring-closure problem of a peptide of N backbone dihedral angles. Thus, a linear conformation defined by $N-6$ angles is first determined at random or by a systematic search, and the values of the remaining six dihedral angles that lead to the ring closure are calculated (see also ref 1 and references cited therein). This method was originally developed for a force field that is based on rigid geometry, i.e., constant bond lengths and angles.[17] Bruccoleri and Karplus[9c,d] extended it to a protein described by flexible geometry. A related procedure was suggested by Weinberg and Wolfe; however, here the conformation of the first $N/2$ bonds is determined at random, while that of the last $N/2$ bonds is obtained under restrictions that guarantee the closure of the ring.[9e] With another method, "the tweak algorithm", a random conformation of the linear peptide, is first generated, and the dihedral angles are "tweaked" to close the ring.[9f] The common feature of these methods (and others; see refs 1 and 9e) is that a large set of ring conformations without severe atomic overlaps are generated first, and their energies are minimized at the next stage. Thus, while a large part of the conformational space is scanned evenly, it is likely that the small region of lowest energy will be missed, as discussed above.

With another group of methods based on MD and MC with or without simulated annealing,[7] the energy and the loop-closure condition are optimized simultaneously. In this category, we include variants of the multiple-copy procedure of Elber and Karplus[10a−c] and other methods.[10d−g] However, from a recent study of the properties of simulated annealing,[22] it seems that these methods would not lead to an efficient generation of low-energy minimized structures, which are mandatory for our NMR methodology. Note that in refs 7b and 8g, MD was also found to be inferior to other methods for generating such structures.

A different approach has been developed in the organic chemistry community in which low-energy conformations are obtained with the help of a stochastic process.[8a−g] Thus, at each step, an *existing* cyclic structure is first deformed and energy-minimized. Then, a decision is made whether to accept the new structure or another previously minimized structure that might be the original one; this is done with the help of a "selection procedure" (such as MCM or UD; see section II.3) that gives preference for accepting low-energy structures. The selected structure is then deformed and energy-minimized, and the process continues. The methods which pertain to this category (among them MCMM[8f] and LTD[13]) can be distinguished by their different deformation techniques and selection procedures. With the method of Saunders[8a,b] (see also ref 8c), the conformational change is carried out by applying small random "kicks" to the Cartesian coordinates of the atoms. The great advantage of this method stems from its simplicity; also, the small kicks make it suitable to handle dense systems. However, the method is expected to become relatively inefficient for a molecule with large conformational freedom. This problem is less severe with the method of Gotō and Ōsawa,[8d,e] which applies larger local *nonstochastic* deformations. The work of Braun[8h] can also be considered to pertain to this category. Extensive simulation studies of cycloalkanes modeled by MM2[8g] have shown that MCMM is more efficient than the random kicks approach[7b,8g] and is marginally less efficient than the Gotō−Ōsawa technique.[8d,e]

II.2. Structure Deformation: The MCMM and LTD Procedures. In contrast to these methods that consist of *local*, hence relatively small, conformational changes, MCMM,[8f] mentioned in the Introduction, is based on *global* conformational deformations induced by changes of torsional angles. Since this method is used in the present study, we shall describe it in detail as applied to a cycloalkane molecule of N carbon atoms labeled i, $1 \le i \le N$ (and the attached $2N$ hydrogen atoms). l_i denotes the bond connecting atoms i and $i + 1$, and l_N defines the bond between atoms N and 1. The four atoms $i - 1$, i, $i + 1$, and $i + 2$ define the torsion angle ϕ_i around l_i. With MCMM, the cyclic molecule is first treated as a linear chain by temporarily removing the bond potential of l_N, which is therefore called the *affected bond*. Next, q dihedral angles are selected at random and changed at random within the range $\phi_i^o \pm D$, where ϕ_i^o is the current angle value and D is a parameter, $D \le 180°$. If ϕ_i^o is changed, the locations of *all* the carbon atoms $i + 2$...N and the attached hydrogens are changed as well. If the length of the affected bond l_N remains within the *ring-closure range*, 0.5−3.5 Å, its bond potential is restored and the energy is minimized; otherwise, the conformation is discarded and a new one is generated. Chang *et al.* have concluded that for a symmetrical molecule of M independent torsions, the optimal values of q range between 2 and $M/3$. Thus, for cycloheptadecane, the best results were obtained when q was selected randomly at each step within the range $2 \le q \le 5$.[8f,g] This range is adopted in the present calculations as well. MCMM is used with a selection procedure as described in section II.3.

It should be pointed out that the global character of MCMM might become a disadvantage for a large N-bond ring. This is because the probability for ring closure (i.e., that carbons 1 and N will be located within the ring-closure range) decreases strongly with increasing N ($\sim N^{-1.5}$ for an ideal chain;[23] see also the results in refs 8f and 9e). Also, for a protein loop in a dense environment, MCMM might lead to undesired entanglements of the deformed loop with the other parts of the protein.

With LTD, we have sought to eliminate these expected problems of MCMM for large rings and loops in proteins. Thus, while LTD also consists of torsional deformations, its rotations are *local*, which enables one to control the extent of conformational change and adapt it to the particular molecular conditions. With a *local* rotation, ϕ_i around bond l_i, only carbon atom $i + 2$ and the hydrogens attached to carbons $i + 1$ and $i + 2$ are moved, while the rest of the atoms are kept fixed in their current positions. Figure 1a and 1b shows a local rotation applied to cyclodecane, where only the backbone atoms are shown for clarity. Notice that the affected bond is l_{i+2} (rather than l_N with MCMM), and it should remain within the ring-closure range of 0.5−3.5 Å. To achieve a larger conformational change, one can *locally* rotate m ($m \ge 1$) successive torsional angles, as depicted in Figure 1c, for $m = 2$; the corresponding method is denoted LTD$_m$. As with MCMM, the angles are determined at random within a range $\pm D$ around the current value. It should be pointed out that for a chain molecule, even a deformation that consists of several consecutive rotations is still local, and therefore, the length of the affected bond will mostly remain within the ring-closure range. Typically, b local rotations are applied simultaneously along the chain, where b increases with increasing the molecular size. Each of these segments must be separated by at least two unrotated bonds to guarantee independence. The case of $b = 2$ and $m = 1$ is depicted in Figure 1d. By changing the parameters m, D, and b, the efficiency of LTD can be adjusted to the specific molecular size and environment. Clearly, the local character of the LTDs make it suitable to handle dense macromolecular systems.

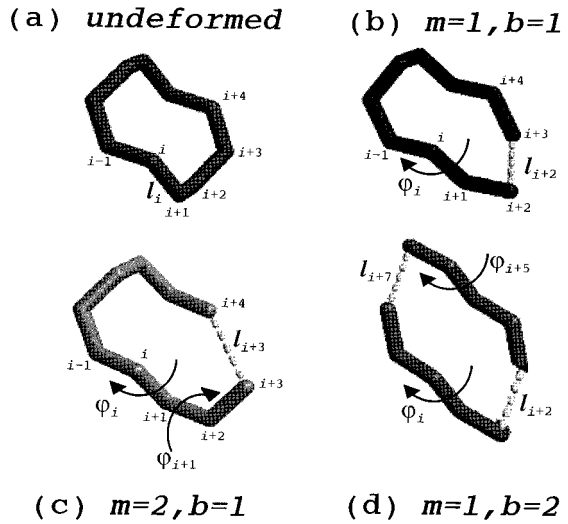

Figure 1. Local rotations of LTD illustrated for the carbon atoms of cyclodecane. (a) The undeformed GEM structure. (b) A local rotation ϕ_i around a single bond ($m = 1$) affects carbon $i + 2$, while the rest of the carbon atoms are kept fixed in their positions. The *affected bond*, l_{i+2}, is denoted by a dashed line. A typical deformation of LTD$_1$ consists of b simultaneous local rotations. Notice that a local rotation around bond l_i precludes such rotations around l_{i-1}, l_{i+1}, and l_{i+2} and that the initial coordinates of carbon $i + 2$ are used for defining a simultaneous rotation around l_{i+3}. (c) A local rotation (ϕ_i, ϕ_{i+1}) around the successive bonds l_i and l_{i+1} ($m = 2$). Only the positions of carbons $i + 2$ and $i + 3$ are changed, and the affected bond is l_{i+3}. A typical deformation of LTD$_2$ consists of b such double rotations. If $b > 1$, rotations around bonds l_{i-1}, l_{i+2}, and l_{i+3} are precluded. (d) Two simultaneous rotations ($b = 2$) of $m = 1$.

II.3. Selection Procedures: The UD and MCM Methods. With both MCMM and LTD, two selection procedures are used, UD[8f] and MCM.[5] UD was found by Chang *et al.* to be the most efficient among several methods tested. Thus, during the process, the computer retains all the *different* energy-minimized conformations with energy within E_{cut} above the lowest energy structure found. To each member of the set, an index is assigned, which keeps track of the number of times this member was used. If a deformation and minimization does not lead to a new structure of this set, the conformer of the set with the lowest index is selected as a new candidate for deformation; if two or more conformers have the same index, priority is given to that with the lowest energy.

With MCM, at each step of the process, the minimized deformed conformation k (obtained from j) is accepted with a Metropolis MC probability p_{jk},

$$p_{jk} = \min(1, \exp[-(E_k - E_j)/k_B T^*]) \qquad (1)$$

while j is accepted again (i.e., k is rejected) with the probability $1 - p_{jk}$. Here, E_k and E_j are the corresponding *minimized* energies, k_B is the Boltzmann constant, and T^* is a temperature parameter that affects the efficiency of MCM. The accepted conformation is then deformed, and the process continues. The effect of changing T^* and using thermalization procedures has been previously discussed.[24] With MCM, as with UD, the coordinates and energies of *all* the energy-minimized structures, including those which were rejected, are stored in a file for further analysis.

The computer programs for both methods are based on the random incremental pulse search (RIPS) program of Ferguson and Raber,[8c,25] which includes the MM2 force field.[16] With

TABLE 1: Number of Energy Minimizations Required for Locating the GEM and the Number of Detected Minima within 1, 2, and 3 kcal/mol above the GEM for Cycloheptadecane[a]

	2000 minimizations[c]				
method and parameters[b]	GEM[d]	0−1 kcal/mol[e]	0−2 kcal/mol[e]	0−3 kcal/mol	10000 minimizations[f]
LTD_1 ($b = 4$, $D = 180$)	**499 ± 159**	**10**	56	**172 ± 3**	**11**, **69**, 245
LTD_1 ($b = 4$, $D = 180$)[g]	864 ± 258	9	**57**	166 ± 2	**11**, **69**, **247**
LTD_2 ($b = 3$, $D = 90$)	1038 ± 186	9	54	163 ± 2	10, 68, 244
LTD_1 ($b = 4$, $D = 90$)	584 ± 175	**10**	55	142 ± 2	**11**, 68, 218
LTD_2 ($b = 3$, $D = 180$)	571 ± 197	8	43	121 ± 3	10, 65, 229
MCMM ($q = 3-5$, $D = 180$)	1055 ± 250	9	47	133 ± 4	**11**, 67, 233
exact[h]		11	69	262	11, 69, 262

[a] Best results are boldfaced. [b] Methods are defined in the text and the parameters are given in parentheses. [c] Averages from 5 runs of 2000 minimizations each. [d] Average number of minimizations required to locate the GEM. [e] Statistical error is less than 1 in all cases. [f] Overall number of minima located within 1, 2, and 3 kcal/mol of the GEM in a *single* run of 10^4 minimizations. [g] The temperature parameter, T^*, of MCM is varied between 200 and 600 K. $T^* = 200$ K in all other cases. [h] The number of known energy minima.

RIPS, the stability of an energy minimum is verified by applying small random kicks to each of the Cartesian coordinates (their maximal size is 0.05 Å such that the atoms move inside a sphere of 0.09 Å), minimizing the energy again and checking that the structure returned to its original position. If the minimum is found to be unstable, the structure is discarded and a new deformation is carried out according to the selection procedure used. This procedure was successfully applied previously.[8a−c,h] To check if two structures are identical, the RMS deviation between the corresponding torsional angles was calculated. Because of the symmetry of the molecule, this calculation should be carried out for 2×17 sets of angles for each pair. If in one of these calculations the RMS deviation is smaller than 10°, the two structures are considered to be identical.[9c] Such calculations enable one to obtain a set of different structures from the larger set of structures generated in the simulation.

III. Results and Discussion

The performance of LTD was previously checked on the relatively small molecule cycloundecane described by MM2.[13] However, to optimize its parameters and test its efficiency, we now apply it to cycloheptadecane described by MM2, which is characterized by a much denser spectrum of energy minima than cycloundecane. Cycloheptadecane is a convenient model for our studies because it has been investigated extensively by various methods and all its energy-minimized structures within 3 kcal/mol above the GEM are anticipated to be known.[8g] For comparison we also present calculations carried out with MCMM, adopting the optimized parameters of Chang et al.;[8f] i.e., $q = 3-5$, and $D = \pm 180°$. Both methods are used with MCM and UD as selection procedures, and the efficiency of MCM and UD is further examined as applied to the pentapeptide Leu-enkephalin described by ECEPP.[17]

Before comparing the performance of these methods, one has to define the criteria of efficiency. The average number of steps required for identifying the GEM structure is an indispensable criterion in this respect. In addition, we have observed a number of different energy-minimized structures within 0−1, 0−2, and 0−3 kcal/mol above the GEM generated in a given number of minimizations; for cycloheptadecane, the exact numbers for these ranges are known to be 11, 69, and 262, respectively (the GEM is 19.09 kcal/mol).

III.1. Optimization of the LTD Parameters. The parameter T^* of MCM was studied first. We found that for $T^* < 200$ K and $T^* > 400$ K, the regions of 1−3 and 0−1 kcal/mol above the GEM, respectively, were not searched effectively, while $T^* = 200$ and 300 K have led to the best results. Therefore, in one run, T^* was varied within the range 200−600 K, while in all the other MCM calculations we used $T^* = 200$ K. The LTD parameters (i.e., m, b, and D) were optimized using MCM. Their effectiveness was preliminarily checked for relatively short runs of 1000 minimizations, and a subgroup of them was further tested in longer simulations of 2000 and 10^4 minimizations. In order to obtain an efficient search, a limited conformational change should be applied; i.e., the deformed conformation should bear some resemblance to its predecessor. The effect of a substantial change is similar to that of a random search, which leads to the high-energy structures. Therefore, the parameters m, b, and D have to be optimized simultaneously. For the present 17-membered ring, we used $m = 1$ and 2 only, for which the *maximum* number of independent simultaneous rotations is $b = 5$ and 4, respectively. We have therefore tested the values of b within the ranges 3−5 and 2−4 for LTD_1 and LTD_2, respectively, where $D = 90°$, 120°, and 180°.

Our preliminary runs revealed that LTD is more sensitive to the values of D than to those of b. With LTD_1, the average number of minima (over the results for $b = 3$, 4, or 5) after 1000 minimizations in the 0−3 kcal/mol range is 99, 119, and 132 for $D = 90°$, 120°, and 180°, respectively, which means that $D = 180°$ leads to the highest efficiency. On the other hand, for LTD_2, $D = 90°$ is the best choice, where the average number of minima (over $b = 2$, 3, or 4) is 112, 80, and 77 for $D = 90°$, 120°, and 180°, respectively.

Averages for selected sets of parameters obtained from additional 1000 minimizations (i.e., altogether 2000) support these findings, where the number of minima generated with the "better" and "worse" sets of parameters are on the order of 170 and 130, respectively. This is also demonstrated by the results of Table 1. We carried out long LTD(MCM) simulations of 10^4 energy minimizations each for the better and worse parameter sets found for LTD_1 and LTD_2 in the preliminary calculations. Since all the generated structures were retained, we could calculate the averages of various properties based on 5 samples of 2000 minimizations each. The results for the better parameter sets appear in the first three rows. The results in the sixth row were obtained with MCMM(MCM) using the optimal MCMM parameters of Chang et al.[8f] In all these calculations, $T^* = 200$ K except for the second run where T^* was varied within the range 200−600 K, by increments of 100 K after every 200 minimizations. The average step at which the GEM was found for the first time varies between 499 and 1055; these results are not always the smallest for the better models. However, to assess such correlations more reliably, the relatively large statistical errors should be decreased by increasing the number of runs. The average number of minima found within the 0−1 kcal/mol range is comparable for all the runs; this means that in 2000 minimizations, all the methods are able to generate most of the 11 structures that pertain to this range.

TABLE 2: Comparison between the Efficiencies of the Selection Procedures MCM and UD Used with LTD and MCMM and Applied to Cycloheptadecane[a]

method and parameters[b]	selection procedure	0−1 kcal/mol	0−2 kcal/mol	0−3 kcal/mol
LTD_2 ($b = 3$, $D = 180$)	MCM ($T^* = 200$ K)	10	65	229
LTD_2 ($b = 3$, $D = 180$)	UD ($E_{cut} = 3$ kcal/mol)	10	65	224
LTD_2 ($b = 3$, $D = 180$)	UD ($E_{cut} = 5$ kcal/mol)	11	63	218
MCMM ($q = 3-5$, $D = 180$)	MCM ($T^* = 200$ K)	11	67	233
MCMM ($q = 3-5$, $D = 180$)	UD ($E_{cut} = 3$ kcal/mol)	11	67	234
exact[c]		11	69	262

[a] Number of located minima within 1, 2, and 3 kcal/mol above the GEM obtained in single runs of 10^4 minimizations. [b] The methods and parameters are defined in the text. [c] The number of known energy minima.

For the 0−2 kcal/mol range (69 structures), the difference in efficiency starts to show up, between the comparable results of the first 4 runs and the significantly lower results of the last 2 runs.

For the 0−3 kcal/mol range (262 structures), the first 3 runs become the most efficient, in accord with the preliminary calculations discussed above. In the last column, we present the total number of different structures found in the whole run of 10^4 minimizations for the three energy ranges. For the 0−3 kcal/mol range, these values are fully correlated with the averages obtained from the 2000 minimizations. It should be noted that using variable T^* had an almost negligible effect on the efficiency. Also, all the structures within the 0−2 kcal/mol range (69) can be generated in 10^4 minimizations. However, the much more populated 0−3 kcal/mol range could not be searched completely in 10^4 minimizations by any method. In fact, generating the 16 missing structures of the second run requires a much larger sample size than that (10^4) used for generating the first 247 structures. Indeed, we could recover the 262 different structures of the 0−3 kcal/mol range only by analyzing all the 6×10^4 structures obtained in the runs of Table 1.

In summary, if one seeks to find the GEM or to search the 0−1 kcal/mol range, all the methods studied are almost comparable in efficiency. However, for the 0−2 and 0−3 kcal/mol ranges, the first four and the first three LTD models, respectively, yield better results than the other methods including MCMM. This conclusion, however, should be taken with some caution since it is based on single runs of 10^4 minimizations.

III.2. The UD and MCM Procedures. All the results presented in Table 1 were obtained with MCM. It is of interest to study the effect of MCM on the efficiency by comparing its results to those obtained with UD. The only comparison between these methods was carried out by Chang *et al.*,[8f] who found that they generate the GEM structure with the same efficiency. However, comparative studies for the efficiency of these methods in detecting other minima were not previously made. We decided to investigate this question further and carried out additional runs of 10^4 minimizations for MCMM(UD) (E_{cut} of 3 kcal/mol as in ref 8f) and for LTD_2(UD) (with $b = 3$, $D = 180°$) using E_{cut} of 3 and 5 kcal/mol. Even though these LTD parameters are not the optimal, improvements that may be caused by UD are expected to be more pronounced for this LTD model. The UD results and their MCM counterparts that appear in Table 2 are almost identical; the only exception is the UD value obtained with $E_{cut} = 5$ for the 0−3 kcal/mol range, which is slightly smaller than the corresponding MC value (229 vs 218). Whether this difference is significant or stems from a statistical fluctuation is impossible to determine from the present data that are based on single runs. We decided to check if this comparable efficiency of MCM and UD is general or only holds for the present highly symmetrical cyclic molecule. Thus, both methods were applied to a more complex molecule, the linear pentapeptide Leu-enkephalin (H-Tyr-Gly-Gly-Phe-Leu-OH) modeled by ECEPP/2.[17] ECEPP is based on Lennard-Jones, electrostatic, torsional, and hydrogen-bond potentials; however, in contrast to MM2, it assumes rigid geometry (i.e., constant bond lengths and angles), and the angles ω are kept fixed at 180°. Therefore, for Leu-enkephalin, a conformation is defined solely by 19 variables, the 10 backbone dihedral angles ϕ and ψ, and the 9 side-chain dihedral angles χ.

TABLE 3: Efficiency of MCM and UD as Applied to Leu-enkephalin[a]

energy range above GEM, kcal/mol	MCM[b]	UD $E_{cut} = 3$ kcal/mol	UD $E_{cut} = 5$ kcal/mol
0.0−0.5	**8**	**8**	**8**
0.5−1.0	**12**	**12**	**12**
1.0−1.5	**20**	**20**	**20**
1.5−2.0	**73**	68	69
2.0−2.5	113	**119**	108
2.5−3.0	227	**244**	219
GEM[c]	**2330 ± 470**	3100 ± 1300	2920 ± 470
n_u/n	6/7	4/7	6/7

[a] Total number of different energy-minimized structures within energy bins of 0.5 kcal/mol above the GEM. For each method the results are based on a set of $n = 7$ independent runs of 10^4 minimizations each. Best results are boldfaced. [b] Temperature parameter T^* is 300 or 600 K. [c] Average number of minimizations at which the GEM is generated for the first time. These averages are based on the n_u runs (out of n) at which the GEM structure was generated.

With both UD and MCM, an energy-minimized structure is deformed with a procedure that can be considered as MCMM for a linear chain. First one determines m, the number of trial angles to be changed,

$$m = \text{int}[1 - \ln(p)] \qquad (2)$$

where p is a random number within the range [0,1] and int(a) is the truncated integer value of a. Thus, the probability for $m = 1$ and 2 is ∼0.63 and 0.23, respectively. The specific m angles are then determined at random, where side-chain angles are selected with a relatively low probability of 0.06. Each selected angle is then changed at random within the range $\pm 0.85\pi$ around its current value,[26] and this deformed structure is energy-minimized. The decision whether to accept the latter is made, of course, differently by UD or MCM. All these calculations were carried out with the program FANTOM.[21,24,27]

With UD as well as MCM, each run is based on 10^4 minimization steps starting from a randomly selected conformation. With UD, two sets of simulations were carried out, for $E_{cut} = 3$ and 5 kcal/mol (denoted (UD(3) and UD(5), respectively), where each set consists of $n = 7$ such runs. The different structures of each set were found, and their distribution in energy bins of 0.5 kcal/mol above the GEM was calculated (Table 3). The criterion for variance of two conformations is that at least one angle differs by more than than 2°. The MCM

results are based on the first seven MCM simulations of 10^4 minimization steps ($T^* = 300$ and 600 K) carried out in ref 18d for this model of Leu-enkephalin. The data of these simulations were analyzed in the same way described above for the UD runs.

Table 3 reveals that with MCM, the GEM structure is generated in six out of the seven runs and the average step at which this structure is obtained for the first time is the lowest among the corresponding results of the three methods. Therefore, as far as this criterion of efficiency is concerned, MCM is significantly better than UD(3), where the GEM structure is found only in four runs and is slightly better than UD(5). For the first three bins, all the methods find the same number of different energy-minimized structures. For the fourth bin, the MCM value is the best (i.e., the largest), and for the last two bins, the results of UD(3) are the best followed by those of MCM and UD(5), respectively.

Altogether MCM appears to be slightly (but not significantly) more efficient than UD. We find MCM more convenient to use than UD, which requires defining E_{cut} and comparing *all* the accumulated unique structures in the course of the simulation. MCM depends only on T^*, which does not affect the efficiency significantly over a large range of values and can easily be incorporated into thermalization procedures.[24] The fact that MCM and UD have comparable efficiency is important since UD can be incorporated in CS methods that are less suitable to use with MCM. Such a CS method is SADA,[28] which we plan to improve and combine with UD.

IV. Summary

LTD defines a prescription for inducing conformational change on a chain molecule by carrying out *local* random rotations around small groups of consecutive bonds along the chain. This temporarily disrupts bonds that are neighbors to the rotated ones, which return to their usual geometry (in terms of bond lengths and angles) by energy minimization. The extent of deformation is controlled by the number of consecutive rotated bonds in a group (m), the number of groups (b), and the maximal angular change (D) and thus can be optimally adjusted to any molecular size and environment. However, this type of conformational change is expected to be especially useful for cyclic peptides, loops in proteins, and dense multichain systems, such as polymers in the glassy state.[15d,29] For the latter systems, the conventional methods are not effective in inducing substantial configurational rearrangements on the chains. In contrast, the repeated local deformations of LTD enable the system to cross high-energy barriers and relax to configurations that are energetically favorable. The search is directed toward the low-energy regions by selection procedures such as MCM or UD that filter out the higher energy structures. LTD has been developed as an important ingredient of our methodology for analyzing the multidimensional NMR data of peptides with intermediate flexibility. Such peptides populate significantly several microstates in thermodynamic equilibrium, and identifying them from theoretical considerations requires generating a complete set of low-energy structures within some cutoff energy above the GEM.

The parameters of LTD(MCM) were optimized as applied to the MM2 model of cycloheptadecane. The efficiency was found to be more sensitive to the choice of m, b, and D than of T^*. The optimal sets of parameters are those which lead to the maximal number of different structures within the 0−3 kcal/mol range above the GEM. These numbers are significantly larger for LTD(MCM) than those obtained with MCMM (based on MCM or UD), which is one of the best methods available. These results are in accord with those obtained previously for cycloundecane in ref 13. This justifies the general use of LTD, in particular for large cyclic peptides and loops in proteins, where it is expected to be the most suitable method. Our calculations for both cycloheptadecane and the linear pentapeptide Leu-enkephalin modeled by ECEPP have shown that MCM and UD are almost comparable in efficiency with only slight advantage for MCM. Thus, UD can be incorporated in CS procedures that are not suitable for using MCM.

The conclusions of this work are based on a relatively small and highly symmetrical cyclic molecule without side chains; therefore, the properties of LTD should further be studied for more complex systems. The method is now being applied to *cyclo* (D-Pro-Phe-Ala-Ser-Phe-Phe) described by the GROMOS[30] force field, in order to decipher its dynamic structure from NMR data.

Acknowledgment. We thank Professor M. Saunders for providing us with the structures of cycloheptadecane and Dr. M. Vásquez for valuable discussions. We acknowledge support from the Florida State University Supercomputer Computations Research Institute, which is partially funded by the U.S. Department of Energy (DOE) under Contract Number DE-FC05-85ER250000. This work was also supported by DOE Grant DE-FG05-95ER62070.

References and Notes

(1) Vásquez, M.; Némethy, G.; Scheraga, H. A. *Chem. Rev.* **1994**, *94*, 2183.

(2) Karplus, M.; Petsko, G. A. *Nature* **1990**, *347*, 631.

(3) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.

(4) (a) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1959**, *31*, 459. (b) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585.

(5) Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611.

(6) Caflisch, A.; Niederer, P.; Anliker, M. *Proteins: Struct. Funct. Genet.* **1992**, *14*, 102−109.

(7) (a) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671. (b) Wilson, S. R.; Cui, W.; Moskowitz, J. W.; Schmidt, K. E. *J. Comput. Chem.* **1991**, *12*, 342. (c) Okamoto, Y.; Kikuchi, T.; Kawai, H. *Chem. Lett.* **1992**, *3*, 1275.

(8) (a) Saunders, M. *J. Am. Chem. Soc.* **1987**, *109*, 3150. (b) Saunders, M. *J. Comput. Chem.* **1991**, *12*, 645. (c) Ferguson, D. M.; Raber, D. J. *J. Am. Chem. Soc.* **1989**, *111*, 4371. (d) Gotō, H.; Osawa, E. *J. Am. Chem. Soc.* **1989**, *111*, 8950. (e) Gotō, H.; Osawa, E. *Tetrahedron Lett.* **1992**, *33*, 1343. (f) Chang, G.; Guida, W. C.; Still W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379. (g) Saunders, M.; Houk, K. N.; Wu, Y.-D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. *J. Am. Chem. Soc.* **1990**, *112*,1419. (h) Braun, W. *Biopolymers* **1987**, *26*, 1691.

(9) (a) Gō, N.; Scheraga, H. A. *Macromolecules* **1970**, *3*, 188. (b) Dudek, M. J.; Scheraga, H. A. *J. Comput. Chem.* **1990**, *11*, 121. (c) Bruccoleri, R. E.; Karplus, M. *Biopolymers* **1987**, *26*, 137. (d) Bruccoleri, R. E.; Haber, E.; Novotný, J. *Nature* **1988**, *335*, 564. (e) Weinberg, N.; Wolfe, S. *J. Am. Chem. Soc.* **1994**, *116*, 9860. (f) Shenkin, P. S.; Yarmush, D. L.; Fine, R. M.; Wang, H.; Levinthal, C. *Biopolymers* **1987**, *26*, 2053. (g) Moult, J.; James, M. N. G. *Proteins* **1986**, *1*, 146.

(10) (a) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161. (b) Simmerling, C.; Elber, R. *J. Am. Chem. Soc.* **1994**, *116*, 2534. (c) Zheng, Q.; Kyle, D. J. *Proteins* **1996**, *24*, 209. (d) Carlacci, L.; Englander, S. W. *J. Comput. Chem.* **1996**, *17*, 1002. (e) Carlacci, L.; Englander, S. W. *Biopolymers* **1993**, *33*, 1271. (f) van Schaik, R. C.; van Gunsteren, W. F.; Berendsen, H. J. C. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 97. (g) van Schaik, R. C.; Berendsen, H. J. C.; Torda, A. E.; van Gunsteren, W. F. *J. Mol. Biol.* **1993**, *234*, 751.

(11) (a) Berg, B. A.; Neuhaus, T. *Phys. Lett. B* **1991**, *267*, 249. (b) Lee, J. *Phys. Rev. Lett.* **1993**, *71*, 211. (c) Hansman, U. H. E.; Okamoto, Y. *J. Comput. Chem.* **1993**, *14*, 1333.

(12) (a) Meirovitch, H. *J. Phys. A* **1982**, *15*, L735. (b) Meirovitch, H.; Vásquez, M.; Scheraga, H. A. *J. Chem. Phys.* **1990**, *92*, 1248.

(13) Baysal, C.; Meirovitch, H. *J. Chem. Phys.* **1996**, *105*, 7868.

(14) (a) Lal, M. *Mol. Phys.* **1969**, *17*, 57. (b) Madras, N.; Sokal, A. D. *J. Stat. Phys.* **1987**, *47*, 573.

(15) (a) Verdier, P. H.; Stockmayer, W. H. *J. Chem. Phys.* **1962**, *36*, 227. (b) Kremer, K.; Baumgärtner, A.; Binder, K. *J. Phys. A* **1982**, *15*, 2879. (c) Wall, F. T.; Mandel, F. *J. Chem. Phys.* **1975**, *63*, 4592. (d) Dodd,

L. R.; Boone, T. D.; Theodorou, D. N. *Mol. Phys.* **1993**, *78*, 961. (e) Kolinski, A.; Skolnick, J.; Yaris, R. *J. Chem. Phys.* **1987**, *86*, 7164. (f) Meirovitch, H. *Macromolecule*s **1984**, *17*, 2038.

(16) Burket, U.; Allinger, N. L. *Molecular Mechanics*; American Chemical Society: Washington, DC, 1982.

(17) (a) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361. (b) Sippl, M. J.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1984**, *88*, 6231.

(18) (a) Meirovitch, H.; Meirovitch, E.; Lee, J. *J. Phys. Chem.* **1995**, *99*, 4847. (b) Meirovitch, E.; Meirovitch, H. *Biopolymers* **1996**, *38*, 69. (c) Meirovitch, H.; Meirovitch, E. *J. Phys. Chem.* **1996**, *100*, 5123. (d) Meirovitch, H.; Meirovitch, E. *J. Comp. Chem.*, in press.

(19) (a) Kessler, H.; Griesinger, C.; Lautz, J.; Muller, A.; van Gunsteren, W. F.; Berendsen, H. J. C. *J. Am. Chem. Soc.* **1988**, *110*, 3393. (b) Bruschweiler, R.; Blackledge, M.; Ernst, R. R. *J. Biomol. NMR* **1991** *1*, 3. (c) Cicero, D. O.; Barbato, G.; Bazzo, R. *J. Am. Chem. Soc.* **1995**, *117*, 1027.

(20) (a) Meirovitch, H. *Chem. Phys. Lett.* **1977**, *45*, 389. (b) Meirovitch, H. *Phys. Rev. A* **1985**, *32*, 3709. (c) Meirovitch, H.; Vásquez, M.; Scheraga, H. A. *Biopolymers* **1987**, *26*, 651. (d) Meirovitch, H.; Koerber, S. C.; Rivier, J.; Hagler, A. T. *Biopolymers* **1994**, *34*, 815.

(21) Shaumann, T. S.; Braun, W.; Wütrich, K. *Biopolymers* **1990**, *29*, 679.

(22) Meirovitch, H.; Vásquez, M. *THEOCHEM*, in press.

(23) de Gennes, P. G. *Scaling Concepts in Polymer Physics*; Cornell University: Ithaca, NY, 1985.

(24) von Freyberg, B.; Braun, W. *J. Comput. Chem.* **1991**, *12*, 1065.

(25) Ferguson, D. M.; Raber, D. J. QCPE 588, Quantum Chemistry Program Exchange, Department of Chemistry, Indiana University, Bloomington, IN 47401.

(26) Nayeem, A.; Vila, J.; Scheraga, H. A. *J. Comput. Chem.* **1991**, *12*, 594.

(27) von Freyberg, B.; Schaumann, T.; Braun, W. *FANTOM User's Manual and Instructions*; ETH Zürich: Zürich, 1993.

(28) (a) Meirovitch, H.; Scheraga, H. A. *Macromolecules* **1981**, *14*, 1250. (b) Meirovitch, H.; Meirovitch, E.; Michel, A. G.; Vásquez, M. *J. Phys. Chem.* **1994**, *98*, 6241.

(29) (a) Theodorou, D. N.; Suter, U. W. *Macromolecules* **1985**, *18*, 1467. (b) Roe, R. J. *J. Chem. Phys.* **1994**, *100*, 1610.

(30) van Gunsteren, W. F.; Berendsen, H. J. C. *Groningen Molecular Simulation (GROMOS) Library Manual*; Biomos b.v., Groningen: The Netherlands, 1987.