Efficient Conformational Search Method for Peptides and Proteins: Monte Carlo Minimization with an Adaptive Bias

S. Banu Ozkan and Hagai Meirovitch*

Center for Computational Biology and Bioinformatics & Department of Molecular Genetics and Biochemistry, School of Medicine, University of Pittsburgh, BST W1058, Pittsburgh, Pennsylvania 15261

Received: March 14, 2003; In Final Form: June 1, 2003

The energy function of a protein consists of a tremendous number of minima. Locating the global energy minimum (GEM), which corresponds to the native structure, is a severe problem in global optimization. The commonly used Monte Carlo minimization (MCM) method is based on a random selection of torsional angle values. We suggest selecting these values with biased probabilities depending on the increased structure−energy correlations as the GEM is approached during the search. Our method applied to models of the 5-residue peptide Leu-enkephalin finds the GEM ∼2.7 faster than MCM.

Introduction

Many every day problems treated by computational methods (e.g., air traffic control, weather prediction) require *global* optimization of multivariable functions, an extremely difficult problem in applied mathematics. Global optimization is especially severe in protein folding, where the potential energy as a function of the 3D structure (conformation) of the polymer chain is highly "rugged", consisting of a tremendous number of local minima. An important theoretical goal is to locate the global energy minimum (GEM) structure, which ignoring entropic effects, is the most stable structure; therefore, it can be identified with the native structure of the protein.[1] Even the ability to predict the most stable *partial* structures of a protein, such as loops in homology modeling, is of practical importance in rational drug design. Therefore, a great deal of effort has been made in computational structural biology to develop efficient methods for global optimization [also called methods for conformational search (CS)], which has led to cross fertilization of ideas and exchange of techniques with the wider field of optimization theory in applied mathematics.[2,3]

On the molecular side, a branch of iterative CS methods (most of them stochastic) based on energy minimization, which allows efficient crossings of energy barriers, have been developed in the organic chemistry community[4−10] and for proteins, mainly by the Scheraga's group.[2,11−18] The common philosophy here is that a *significant* change (followed by minimization) of low energy structures leads with a high probability to a decrease in the energy as long as the change is not random over the entire conformational space that is populated predominately by high-energy structures. Thus, a relatively short pathway towards the global energy minimum is defined. In general, this approach has led to methods that are more efficient than simulated annealing[19−23] and the conventional Metroplolis Monte Carlo (MC)[24]and molecular dynamics,[25,26] which cross energy barriers very inefficiently at 300 K. A popular method in this category is the Monte Carlo minimization (MCM) of Li and Scheraga,[13] which has been implemented within the framework of a CS procedure for cyclic molecules and protein loops, the local torsional deformation method, suggested by our group.[27,28] With MCM, at each MC step a conformational change of the current structure i (with minimized energy E_i) is typically carried out by randomly selecting a *small* number of dihedral angles, defining their new values *at random* within the range [−180°,-180°], and minimizing the energy; the obtained trial structure j (with minimized energy E_j), is accepted with a Metropolis transition probability p_{ij},

$$p_{ij} = \min \{1, \exp[-(E_j - E_i)/k_B T]\} \quad (1)$$

where T is a temperature parameter that affects the efficiency significantly. Therefore, various temperature schedules were tested,[29,30] but the gain in efficiency (as compared to an optimal constant T) has been moderate at best. With another approach developed by Totrov and Abagyan,[31] the *random* selection of dihedral angle values is replaced by a *biased* selection based on the distribution of these angles in known protein structures, which has led to a significant increase in efficiency for α-helical peptides. Scheraga's group, on the other hand, has pursued a pure theoretical approach, seeking to gain efficiency not by relying on experimental data, but by organizing the structures in groups and selecting trial dihedral angles with some bias based on their distribution in low energy structures. Thus, with the conformational space annealing (CSA) method of Lee et al.,[12] where the MCM procedure is replaced by a genetic algorithm and a build-up procedure, the average number of energy minimizations required to reach the global minimum of Met-enkephalin was decreased by a factor of 2 as compared to MCM (using the ECEPP/3 potential).[32−34] The efficiency of a recently developed MCM-based procedure by Pillardy et al.,[15] the conformational family Monte Carlo (CFMC) is claimed to be comparable to that of CSA. The improved performance of these methods seems to stem mainly from the application of sophisticated structural clustering procedures. Other CS methods have been developed, which are not discussed here; see for example refs 35−38 and references therein.

In this paper, we develop an MCM-based method that relies on the increasing correlation between structure and energy as the GEM is approached. Thus, a biased (rather than random) selection of dihedral angle values is imposed, which is adapted to the structural and energetic changes occurring continuously

* Corresponding author.

10.1021/jp0346615 CCC: $25.00
Published on Web 07/17/2003

during the search. This method, which relies on minimal structural organization, is called MCM with an adaptive bias (MCMAB). We demonstrate that the efficiency of MCMAB is enhanced by a factor of ~2.7 as compared to that of MCM for models of the linear penta-peptide Leu-enkephalin.

Models and Methods

Molecular Models. Leu-enkephalin (H-Tyr−Gly−Gly−Phe−Leu-OH) is modeled by the potential energy function ECEPP/2 that assumes rigid geometry (i.e., constant bond lengths and angles) and is based on Lennard-Jones, electrostatic, torsional, and hydrogen-bond potentials.[32,33] Two models are studied: In model I the peptide bond angles ω are kept fixed at 180°, and therefore a conformation is defined by 19 dihedral angles, the 10 backbone φ and ψ and the 9 side-chain dihedral angles, χ. In the second model (model II) the five backbone ω angles are allowed to vary and a conformation is defined by 24 variables. We adopt the standard dielectric constant $\epsilon = 2$ of ECEPP. This force field is implemented in the package Fantom,[39] which is used in the present calculations. We have chosen to use this force field because of its relative simplicity; thus, the limited number of variables makes it convenient to study the various aspects of the present new method.

The MCMAB Process. This process consists of three stages, the first $n < n_1$ MC steps, $n_1 \le n \le n_2$, and $n > n_2$, respectively, where for model I with constant ω, $n_1 = 50$, and $n_2 = 800$. During the first two stages, the usual MCM procedure is performed where the dihedral angle values are determined at random within the range [−180°, 180°], and $T = 400$ K (eq 1) is kept constant throughout the entire search.[40] However, in the second stage ($n_1 \le n \le n_2$) the program starts building the biased probabilities (over the range [−180°, 180°]), which are used only in the third stage. Thus, after n_1 MC steps the typical energy of the current conformation has been reduced significantly as compared to that of the starting structure ($n = 1$), meaning that the energy−structure correlations are strong enough to be taken into account. Therefore, subsequent accepted structures [by the MC criterion, eq 1] that differ significantly from each other are retained, where two structures are considered different if at least one dihedral angle differs by 30° or more. The acceptance rate of the retained structures is ~0.15. For the retained structures, the dihedral angle range [−180°, 180°] is divided into $m_{tot} = 3$ segments, [−120°, 0°], [0°, 120°], and the third segment that consists of the two ranges [−180°, −120°] and [120°, 180°]; these three segments denoted by m ($m = 1$, m_{tot}, where $m_{tot} = 3$) are centered at the three occurring rotamers, gauche-, gauche+, and trans, respectively. This division leads to a relatively smooth distribution, which is essential for the success of MCMAB (see below); the more "rugged " distributions obtained for $m_{tot} > 3$ have not led to an improvement in the efficiency of MCM.

We denote the retained structures by the index t; thus, the contribution of the tth retained structure (with energy E_t) to the selection probability of segment m ($m = 1,m_{tot}$) of dihedral angle k ($k = 1,19$ for constant ω) is proportional to the corresponding Boltzmann factor, $\exp[-E_t/k_B T^*]$, where T^* is a temperature parameter to be distinguished from T appears in the Metropolis criterion (eq 1). The *unnormalized* selection probability, $g^t_{m(k)}$ after the tth structure has been added to the group is

$$g^t_{m(k)} = g^{t-1}_{m(k)} + \exp(-E_t/k_B T^*) \quad (2)$$

and the normalized probability is

$$P^t_{m(k)} = \frac{g^t_{m(k)}}{\sum_{m=1}^{m_{tot}} g^t_{m(k)}} \quad (3)$$

The different structures are collected and $P^t_{m(k)}$ are updated during the entire MCMAB search (i.e., $n \ge n_1$). However, for each k, $P^t_{m(k)}$ is typically large (~0.9) for a specific segment m; hence, it is significantly smaller for the other two segments; this would lead to a very inefficient MCMAB process, where some regions in the conformational space become almost excluded, meaning that the bias should be much milder. Therefore, at the third stage ($n > n_2$), if $P^t_{m(k)}$ is smaller than p_{low} it is increased (becoming $p^t_{m(k)}$) such that $p^t_{m(k)} \sim p_{low}$,

$$p^t_{m(k)} = (g^t_{m(k)} + d)/(\sum_m g^t_{m(k)} + d) \quad (4)$$

where

$$d = [p_{low} \sum_m g^t_{m(k)} - g^t_{m(k)}]/[1 - p_{low}] \quad (5)$$

and the other two probabilities are changed to $p^t_{m(k)}$ according to eq 4, where $d = 0$ in the numerator. We have found that for both models I and II, $p_{low} \sim 0.25$ is an optimal value meaning that the deviation from the random value, $1/m_{tot} = 1/3$ is not large. The energy of the structures added to the group decreases in the course of the search, and on average the structures added last have the strongest effect on the probabilities. However, the effect of the higher energy structures is not negligible because the optimal $T^* = T^*_1 = 750$ K is relatively high [we use the notation T^*_1 when single dihedral angles are treated; eqs 2 and 3 are also used (see below) to calculate probabilities for $\varphi-\psi$ pairs where the notation $T^*_{\phi\psi}$ is adopted]. Now, it should be pointed out that $p^t_{m(k)}$ is used only in the third stage, i.e., for MC steps $n > n_2$; then, if angle k is chosen to be changed, a segment m ($m = 1,m_{tot}$) is selected according to the probabilities $p^t_{m(k)}$ (rather than at random, i.e., with probability $1/ m_{tot} = 1/3$) and the value of the angle within the selected segment m is determined at random.

To increase efficiency, one can calculate biased probabilities also for pairs of backbone angles $\varphi-\psi$ and side chain angles $\chi_1-\chi_2$, and consider these probabilities in the MCMAB process. We did not attempt to apply the $\chi_1-\chi_2$ bias because Leu-enkephalin consists of two glycine residues without side chains. As discussed later, implementing the $\varphi-\psi$ bias for model I did not lead to a better efficiency, while for model II the $\varphi-\psi$ bias was found to be very effective.

To implement the $\varphi-\psi$ bias within the MCMAB procedure, the $\varphi-\psi$ region was divided into $m_{tot} = 4$ equal quadrants defined by partitioning the φ and ψ ranges ([−180°, 180°]) into two equal segments [−180°, 0°] and [0°, 180°]. The probabilities (preferences) of these $\varphi-\psi$ regions were obtained by eqs 2 and 3 using the same procedure and database of conformations collected for the single angle probabilities. However, the probabilities $P^t_{m(k)}$ (eq 3) of the four quadrants were found to be much more homogeneous than the $P^t_{m(k)}$ values of the single angles, and therefore the balancing process of eqs 4 and 5 was not applied, i.e., the $\varphi-\psi$ probabilities were defined by $P^t_{m(k)}$ (eq 3) rather than by $p^t_{m(k)}$ (eq 4). As mentioned above, for the $\varphi-\psi$ probabilities the temperature parameter T^* (eq 2) is denoted by $T^*_{\phi\psi}$.

TABLE 1: Number of MC Steps (energy minimizations) Required for Locating the Global Energy Minimum for the First Time[a]

model I constant $\omega = 180°$		model II variable ω	
MCM	MCMAB	MCM	MCMAB with $\varphi-\psi$ bias
460	352	36740	1667
697	697	882	882
6589	1811	2492	1514
2572	1407	9437	5598
334	463	3294	2301
2128	2292	18136	4319
315	315	4938	7126
3760	2082	9100	3674
17698	3762	585	585
295	295	9740	3565
5962	1917	1335	1335
2523	1792	1850	3955
8773	4730	3036	1855
8964	2853	16801	8070
15594	5548	29908	5920
5111	**2021**	**9885**	**3491**

[a] For each model the 15 MCM runs were started from different randomly selected structures (seeds), and at each row the MCMAB run was started from the corresponding MCM seed. The MCMAB results of model I are based on single angle probabilities only. The MCMAB results for model II are based on a procedure that combines single angle and $\varphi-\psi$ probabilities (see text). The average number of MC steps (no. of minimizations) appears in the bottom line bold-faced.

Thus, in the case where single angle and $\varphi-\psi$ probabilities are used together at each MC step [at the third stage ($n > n_2$)] the process has two options, to treat a $\varphi-\psi$ pair or a single angle. For model II, we have found an optimal probability, $p_{\varphi\psi}$, $= 0.25$ for treating a $\varphi-\psi$ pair, meaning that the probability for treating a single angle is 0.75. Thus, if a random number (within [0,1]) is smaller than 0.25, a $\varphi-\psi$ pair is selected at random, a quadrant is then chosen with $P^t_{m(k)}$ (eq 3) out of the $m_{\text{tot}} = 4$ quadrants, and φ and ψ values are selected at random from the corresponding ranges. In the other case, a dihedral angle is selected and treated according to the procedure described earlier for a single angle.

Results and Discussion

Our criterion of efficiency is defined by the number of energy minimizations required to reach the global energy minimum (GEM) for the first time; the smaller this number, the better the efficiency. Thus, the optimal values of n_1, n_2, T^* (eq 2), p_{low} (eq 5), and other parameters described below, have been determined according to this criterion by performing many MCM runs with different values of these parameters.

Model I. For model I the assumed GEM of -9.704 kcal/mol is known from previous studies. To compare the efficiencies of MCMAB and MCM for this model, we carried out 15 runs with each method at $T = 400$ K (eq 1) starting from the same randomly chosen conformations (the MCM procedure is described in detail in ref 40). In Table 1 we present for each run the number of energy minimizations required to reach the GEM for the first time, where the average number appears in the bottom line. The table reveals that for 13 runs MCMAB required less or equal number of minimizations than MCM and the average number of minimizations for MCM (5110) is 2.5 times larger than that of MCMAB (2021). As already mentioned, implementing the $\varphi-\psi$ bias together with the single angle bias did not improve the efficiency of MCMAB for model I further, probably due to the already fast search obtained with the single angles bias. It is important to point out that to a large extent the bias is not general but depends on the specific simulation run. Thus, if the biased probabilities obtained at the third stage of a certain run are used in the third stage of another run that started from a *different* structure these probabilities in general will be ineffective, i.e., the GEM will not be reached fast.

Model II. We also applied both MCMAB and MCM to model II (GEM $= -10.093$ kcal/mol), where the angles ω are allowed to vary during the energy minimization while they are not changed in the MC process. Because of the larger number of variables (24), the optimization is slower for model II than for constant ω, i.e., the number of MC steps required to reach the low energy region is increased; correspondingly, the optimal values, $n_1 = 500$, and $n_2 = 1800$ for model II are larger than those found optimal for model I ($n_1 = 50$; $n_2 = 800$). The temperature parameter $T_1^* = 800$ K (eq 2) optimized for model II is slightly higher than $T_1^* = 750$ K optimized for model I.

Fifteen randomly selected structures were generated for model II as "seeds" for MCM and MCMAB runs. The MCM results for the number of energy minimizations required to reach the GEM for the first time appear in the third column of Table 1. We performed 15 MCMAB runs based on single angle probabilities alone to find only a marginal decrease (by ~30%) in the average number of minimizations as compared to MCM (results are not shown). However, the MCMAB runs based on a combination of single angle probabilities and a $\varphi-\psi$ bias have led to results (appearing in the fourth column of Table 1) that in 12 cases are smaller than (i.e., better) or equal to those of MCM. The average number of minimizations required to reach the GEM by MCMAB (3491) is smaller than the MCM average (9885) by a factor of 2.8. This suggests that as the number of degrees of freedom of the backbone increases it is important to take into account also correlations between backbone dihedrals to achieve significant enhancement in the performance. It should be pointed out that the optimal temperature, $T\psi_\phi^* = 650$ K (eq 2) for the $\varphi-\psi$ probabilities is lower than 800 K used for the single angle probabilities, meaning that the effect of the lowest energy structures is more pronounced in the $\varphi-\psi$ case. Correspondingly, we have found that effective $\varphi-\psi$ probabilities should be based on lower energy structures than those required for the single angle probabilities. Thus, for the $\varphi-\psi$ calculations n_1 increases from 500 to1000 while $n_2 = 1800$ is unchanged.

Summary. We have introduced a new procedure, MCMAB, in which the selection of trial dihedral angles for a structural change is not random as in the usual MCM method but is biased toward values that have appeared frequently in previous MC steps. Applying MCMAB to models of Leu-enkephalin has shown that if the bias is moderate the global energy minimum can be located on average 2.7 times faster than with MCM; this factor is larger than factors obtained by other methods applied to similar molecules. Therefore, MCMAB is also much more efficient than simulated annealing that was found to be inferior to MCM.[19−23] Because MCMAB does not rely on structural organization, one would expect that tailoring its main features to available clustering techniques would enhance its efficiency even further. The present results suggest that MCMAB can also be improved by considering additional angular−energy correlations, such as those involving $\chi_1-\chi_2$, $\chi_2-\chi_3$, or $\varphi-\chi_1$. The parameters n_i are expected to increase for larger peptides; however, their relation to the molecular size and the sensitivity of all the parameters will be determined only in future studies of longer peptides. The present ideas can be used in conformation searches of any linear or cyclic macromolecule, loops in proteins, and other systems that can be expressed by internal coordinates.

Acknowledgment. This work was supported by NIH Grant R01GM61916 and by National Science Foundation Large Information Technology Research Grant, NSF 0225636.

References and Notes

(1) Vásquez, M.; Némethy, G.; Scheraga, H. A. *Chem. Rev.* **1994**, *94*, 2183−2239.
(2) Scheraga, H. A.; Pillardy, J.; Liwo, A.; Lee, J.; Czaplewski, C.; Ripoll, D. R.; Wedemeyer, W. J.; Arnautova, Y. A. *J. Comput. Chem.* **2002**, *23*, 28−34.
(3) Schlick, T. *Rev. Comput. Chem.* **1992**, *3*, 1−71.
(4) Saunders, M. *J. Comput. Chem.* **1991**, *12*, 645−663.
(5) Ferguson, D. M.; Raber, D. J. *J. Am. Chem. Soc.* **1989**, *111*, 4371−4378.
(6) Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379−4386.
(7) Gotō, H.; Ōsawa, E. *Tetrahedron Lett.* **1992**, *33*, 1343−1346.
(8) Kolossváry, I.; Keseru, G. M. *J. Comput. Chem.* **2001**, *22*, 21−30.
(9) Saunders, M.; Houk, K. N.; Wu, Y. D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. *J. Am. Chem. Soc.* **1990**, *112*, 1419−1427.
(10) Kolossváry, I.; Guida, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 5011−5019.
(11) Caflisch, A.; Niederer, P.; Anliker, M. *Proteins* **1992**, *14*, 102−109.
(12) Lee, J.; Scheraga, H. A.; Rackovsky, S. *J. Comput. Chem.* **1997**, *18*, 1222−1232.
(13) Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci.* **1987**, *84*, 6611−6615.
(14) Pillardy, J.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. A* **1999**, *103*, 9370−9377.
(15) Pillardy, J.; Czaplewski, C.; Wedemeyer, W. J.; Scheraga, H. A. *Helv. Chim. Acta* **2000**, *83*, 2214−2230.
(16) Pillardy, J.; Arnautova, Y. A.; Czaplewski, C.; Gibson, K. D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12351−12356.
(17) Ripoll, D. R.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 165−176.
(18) Sullivan, D. C.; Kuntz, I. D. *Proteins* **2001**, *42*, 495−511.
(19) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671−680.
(20) Wilson, S. R.; Cui, W.; Moskowitz, J. W.; Schmidt, K. E. *J. Comput. Chem.* **1991**, *12*, 342−349.
(21) Nayeem, A.; Vila, J.; Scheraga, H. A. *J. Comput. Chem.* **1991**, *12*, 594−605.
(22) Meirovitch, H.; Vásquez, M. *J. Mol. Struct.(THEOCHEM)* **1997**, *398*−*399*, 517−522.
(23) Baysal, C.; Meirovitch, H. *J. Comput. Chem.* **1999**, *20*, 1659−1670.
(24) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087−1092.
(25) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1959**, *31*, 459−466.
(26) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585−588.
(27) Baysal, C.; Meirovitch, H. *J. Phys. Chem.* **1997**, *101*, 2185−2191.
(28) Baysal, C.; Meirovitch, H. *J. Am. Chem. Soc.* **1998**, *120*, 800−812.
(29) Abagyan, R.; Argos, P. *J. Mol. Biol.* **1992**, *225*, 519−532.
(30) Von Freyberg, B.; Braun, W. *J. Comput. Chem.* **1991**, *12*, 1065−1076.
(31) Abagyan, R.; Totrov, M. *J. Mol. Biol.* **1994**, *235*, 983−1002.
(32) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361−2381.
(33) Sippl, M. J.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1984**, *88*, 6231−6233.
(34) Némethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 6472−6484.
(35) Amara, P.; Straub, J. E. *Phys. Rev. B* **1996**, *53*, 13857−13863.
(36) Andricioaei, I.; Straub, J. E. *J. Chem. Phys.* **1998**, *19*, 1445−1455.
(37) Church, B. W.; Shalloway, D. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 6098−6103.
(38) Pardalos, P. M.; Shalloway, D.; Xue, G. *J. Global Opt.* **1994**, *4*, 117−133.
(39) Von Freyberg, B.; Schaumann, T.; Braun, W. *FANTOM, User Manual and Instructions*; ETH Zürich: Zürich, 1993.
(40) Meirovitch, H.; Meirovitch, E. *J. Comput. Chem.* **1997**, *18*, 240−253.