Conformational Search of Peptides and Proteins: Monte Carlo Minimization with an Adaptive Bias Method Applied to the Heptapeptide Deltorphin

S. BANU OZKAN, HAGAI MEIROVITCH
Center for Computational Biology and Bioinformatics & Department of Molecular Genetics and Biochemistry, University of Pittsburgh School of Medicine, BST W1058, Pittsburgh, Pennsylvania 15261

Received 3 June 2003; Accepted 8 October 2003

Abstract: The energy function of a protein consists of a tremendous number of minima. Locating the global energy minimum (GEM) structure, which corresponds approximately to the native structure, is a severe problem in global optimization. Recently we have proposed a conformational search technique based on the Monte Carlo minimization (MCM) method of Li and Scheraga, where trial dihedral angles are not selected at random within the range [−180°,180°] (as with MCM) but with biased probabilities depending on the increased structure-energy correlations as the GEM is approached during the search. This method, called the Monte Carlo minimization with an adaptive bias (MCMAB), was applied initially to the pentapeptide Leu-enkephalin. Here we study its properties further by applying it to the larger peptide with bulky side chains, deltorphin (H-Tyr-D-Met-Phe-His-Leu-Met-Asp-NH_2). We find that on average the number of energy minimizations required by MCMAB to locate the GEM for the first time is smaller by a factor of approximately three than the number required by MCM—in accord with results obtained for Leu-enkephalin.

 J Comput Chem 25: 565–572, 2004

Key words: conformational search; energy minimization; Monte Carlo; global energy minimum; protein folding

Introduction

Global optimization of multivariable functions is an unsolved problem in applied mathematics, which appears in a large number of every-day problems such as air traffic organization, weather prediction, economic cost/benefit analysis, graph theory, and so forth. Global optimization is especially severe in protein folding, where the potential energy as a function of the 3D structure of the protein chain is highly rugged, and locating the global energy minimum (GEM) among the tremendous number of local minima resembles a search for a needle in a haystack. Ignoring entropic effects, the GEM structure can be considered as the most stable and is therefore identified with the native structure of the protein.[1] Global optimization is also required for predicting the stability of partial structures of a protein, such as loops in homology modeling, or the binding of small ligands to a receptor, where such studies are 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), which has led to cross fertilization of ideas and exchange of techniques with the wider field of optimization theory in applied mathematics.[2–8]

On the molecular side, a branch of iterative conformational search methods (most of them stochastic) that is based on energy minimization has been developed in the organic chemistry community,[9–14] and for proteins, mainly by Scheraga's group.[15–23] A common feature of many of these methods is that a significant conformational change of the current structure is carried out at each step followed by energy minimization, which allows efficient crossings of energy barriers; the trial structure thus generated is then accepted or rejected by a certain *selection criterion*. The philosophy behind this approach is that a *significant* change of low energy structures (followed by minimization) leads on average to a decrease in their energy; however, the change should not be random over the entire conformational space, which is populated predominately by high-energy structures. Thus, a relatively short pathway towards the GEM is defined. The methods of this category differ by their conformational change procedures and selection criteria, and in general they have been found to be more efficient than simulated annealing,[24–28] the conventional Metropolis Monte Carlo (MC) method,[29] and molecular dynamics,[30,31]

Correspondence to: H. Meirovitch; e-mail: hagaim@pitt.edu

Contract/grant sponsor: NIH; contract/grant number: R01GM61916

where the last two methods cross energy barriers very inefficiently at 300 K.

In the past, we studied two of these methods, the Monte Carlo multiple minimum (MCMM) of Still's group using the "usage directed" selection criterion[11] and the Monte Carlo minimization (MCM) of Li and Scheraga.[15] While the performance of these methods, as applied to the pentapeptide Leu-enkephalin, was found to be comparable,[32] application of MCM is more straightforward, which has made it a popular technique; therefore, we have implemented MCM within the framework of our conformational search procedure for cyclic molecules and protein loops, the local torsional deformation method.[32,33]

With MCM, at each step a conformational change of the current structure i (with minimized energy E_i) is typically carried out by selecting at random 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}:

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

and the process is repeated many times. It should be pointed out that unlike a usual MC procedure where E_i is not minimized, MCM is based on *minimized* energies, and therefore the generated structures are not distributed according to the Boltzmann probability density, and T (which multiplies the Boltzmann constant k_B) is not a thermodynamic temperature but a parameter that can affect the efficiency significantly. Therefore, various temperature schedules were tested,[34] including a simulated annealing MCM procedure,[35] but the gain in efficiency (as compared to an optimal constant T) has been moderate at best. With a more substantial approach for improving MCM developed by Totrov and Abagyan,[36] the random selection of dihedral angle values has been 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.,[17] where the MCM procedure is replaced by a genetic algorithm and build-up procedures, the average number of energy minimizations required to reach the global minimum of Met-enkephalin has been decreased by a factor of two as compared to MCM (using the ECEPP/3 potential[37–39]). The efficiency of a recently developed MCM-based procedure by Pillardy et al.,[19] 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.

In a recent article, here called article I,[40] we have developed an MCM-based method that relies on the increasing structure/energy correlation as the GEM is approached. Thus, a biased (rather than random) selection of dihedral angle values within the range [−180°,180°] is imposed, which is adapted to the structural and energetic changes occurring continuously during the search (and with minimal structural organization); this method is called MCM with an adaptive bias (MCMAB). We have demonstrated that for models of Leu-enkephalin (H-Tyr-Gly-Gly-Phe-Leu-OH) described by the ECEPP/2 force field, the number of energy minimizations required by MCMAB to locate the GEM for the first time is on average ≈2.7 smaller than that required by MCM.[40]

In this article we investigate the performance of MCMAB further as applied to a larger molecule, the linear heptapeptide deltorphin (also known as dermenkephalin), H-Tyr-D-Met-Phe-His-Leu-Met-Asp-NH_2), which unlike Leu-enkephalin consists of bulky side-chains. An interesting question is to examine the behavior of the parameters of MCMAB as the peptide size is increased. Deltorphin is a natural peptide, which is found in frog skin, and has high potency and receptor selectivity for δ opioid receptors. To understand the structure-activity relationships, NMR studies of the solution structures of deltorphin in DMSO and cryoprotective solvents were carried out[41] and computational work was performed based on these experiments.[42–44]

Molecular Model and Methods

Model

Deltorphin is modeled by the potential energy function ECEPP/2, which assumes rigid geometry (i.e., constant bond lengths and angles) and is based on Lennard-Jones, electrostatic, torsional, and hydrogen-bond potentials.[37–39] In addition to the rigid geometry, the peptide bond angles ω are kept fixed at 180°, and therefore a conformation is defined by a relatively small number, $K = 36$ of dihedral angles, the 14 backbone ϕ and ψ and the 22 side-chain dihedral angles, χ (a significantly larger number of variables, the $3N$ Cartesian coordinates of the N atoms, would be required for a force field with flexible geometry); therefore, using ECEPP facilitates the optimization of the various parameters of MCMAB. We adopt the standard dielectric constant $\varepsilon = 2$ of ECEPP, and use the software package FANTOM,[34,45] where this force field is implemented. Because ECEPP does not consider solvent effects, we do not attempt to compare our results to the experimental NMR data and only study the efficiencies of MCM and MCMAB.

Biased Probabilities for Single Dihedral Angles

The MCMAB procedure applied to deltorphin contains elements that have not been used for Leu-enkephalin; therefore, we describe here the entire procedure in detail, first as applied to single dihedral angles and later to pairs of ϕ-ψ and χ_1− χ_2 angles. For single dihedral angles the process consists of four stages, the first $n < n_1$ MC steps, $n_1 \le n \le n_2$, $n_2 < n \le n_3$, and $n > n_3$, where $n_1 = 4000$, $n_2 = 6500$, and $n_3 = 7000$. During the first two stages the usual MCM procedure is performed as described in ref. 46 (see also next section); in particular, 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. 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 in stages three and four. Thus, after n_1 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.

For the retained structures, the dihedral angle range [−180°,180°] is divided into $m_{\text{tot}} = 3$ 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 ($m = 1,m_{\text{tot}}$) 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). In article I we have found that the more "rugged" distributions obtained for $m_{\text{tot}} > 3$ do not improve the efficiency over that of MCM. We denote the retained structures by the index t; thus, the contribution of the t^{th} retained structure (with energy E_t), to the selection probability of segment m ($m = 1,m_{\text{tot}}$) of dihedral angle k ($k = 1,36$) is proportional to the corresponding Boltzmann factor, $\exp[-E_t/k_BT^*]$, where T^* is a temperature parameter that should be distinguished from T appears in the Metropolis criterion [eq. (1)]. The *unnormalized* selection probability, $g^t_{m(k)}$ after the t^{th} structure has been added to the group

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

and the normalized probability is

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

The different structures are collected and the $P^t_{m(k)}$ values are updated during the MCMAB search (i.e., for $n > n_1$). Eqs. (2) and (3) are used for defining single angle probabilities as well as probabilities for the pairs, ϕ-ψ and $\chi_1-\chi_2$, where the corresponding optimal values of T^* are different; therefore, for these cases T^* is replaced by T^*_1, $T^*_{\phi\psi}$, and $T^*_{\chi 1\chi 2}$, respectively, where $T^*_1 = 850$ K.

Because the MCM search is biased towards the low energy region some angle values appear with higher probability than others. Thus, it turns out that 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, from the third stage on ($n_2 < n$), if $P^t_{m(k)}$ is smaller than p_{low} it is increased [becoming $p^t_{m(k)}$] such that $p^t_{m(k)} \approx p_{\text{low}}$

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

where

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

and the other two probabilities $p^t_{m(k)}$ are changed according to eq. (4), where $d = 0$ in the numerator; if two segments have low probability, their probabilities are increased to the value p_{low}. We have found that $p_{\text{low}} = 0.28$ is an optimal value, meaning that the deviation from the random value, $1/m_{\text{tot}} = 1/3$ is not large. The energy of the structures added to the group decreases in the course of the search, and therefore, on average, those added last have the strongest effect on the probabilities. However, the effect of the higher energy structures is not negligible because the optimal value, $T^*_1 = 850$ K is relatively high.

Notice that for $n \le n_2$ the usual MCM is applied, while $p^t_{m(k)}$ is used only in the fourth stage, that is, for MC steps $n > n_3$; thus, if angle k is chosen to be changed, a segment $m(k)$ ($m = 1,m_{\text{tot}}$) is selected according to the probabilities $p^t_{m(k)}$ (rather than at random, i.e., with probability $1/m_{\text{tot}} = 1/3$) and the value of the angle within the range of the selected m is determined at random. We have found that the efficiency of the process increases by "bridging" these two regions by an intermediate region, $n_2 < n \le n_3$, where slightly biased probabilities, denoted $p^{*t}_{m(k)}$ are used. $p^{*t}_{m(k)}$ are calculated from the *unnormalized* probabilities $f^t_{m(k)}$:

$$f^t_{m(k)} = (1/m_{\text{tot}} + C_R p^t_{m(k)} t) \tag{6}$$

$$p^{*t}_{m(k)} = f^t_{m(k)} / \sum_m f^t_{m(k)} \tag{7}$$

where $C_R = 0.0025/n_2$ is a small constant. Thus, the bias in $p^{*t}_{m(k)}$ is always small, increasing slightly as t increases. It should be emphasized again that to enable the search process finding short pathways to the GEM it is necessary to allow the molecule to visit large portions of conformational space and therefore application of a relatively small bias is crucial for the success of MCMAB.

Biased Probabilities for Pairs of Dihedral Angles

As shown later, for MCMAB applied to deltorphin to be successful one should take into account also the correlations between pairs of dihedral angles, such as the backbone ϕ-ψ or the side-chain angles $\chi_1-\chi_2$. To implement probabilities for pairs of angles, the regions of the backbone ϕ-ψ and side-chain $\chi_1-\chi_2$ were divided into the $m_{\text{tot}} = 4$ equal quadrants ([$\Delta\phi$] = [−180°,0°], [$\Delta\psi$] = [−180°,0°]), ([$\Delta\phi$] = [−180°,0°], [$\Delta\psi$] = [0°,180°]), and so forth. The probabilities (preferences) of these regions were obtained by eqs. (2) and (3) using the same procedure and database of conformations collected for the single angle probabilities. The optimal temperatures found, $T^*_{\phi\psi} = T^*_{\chi_1\chi_2} = 700$ K [eq. (2)] are lower than 850 K optimized for the single probabilities, meaning that the effect of the lowest energy structures is more pronounced in this case. Correspondingly, we have found that effective ϕ-ψ and $\chi_1-\chi_2$ probabilities should consist of lower energy structures than those required for the single angle probabilities. Thus, for ϕ-ψ and $\chi_1-\chi_2$, n_1 was increased from 4000 to 5000, while $n_2 = 6500$ and $n_3 = 7000$ are unchanged; eqs. (4) and (5) are used with p_{low} 0.2 (as compared to a random selection of $1/m_{\text{tot}} = ¼ = 0.25$).

For $n > n_2$ the program can select to change a pair of ϕ-ψ, a pair of $\chi_1-\chi_2$, or a single angle; we have found the optimal probabilities, $p_{\phi\psi} = 0.25$, $p_{\chi_1\chi_2} = 0.1$, hence $p_1 = 0.65$ for these

cases, respectively. Thus, if, for example, a random number is smaller than 0.25, a ϕ-ψ pair is chosen at random out of the seven available pairs and a quadrant (denoted m_c) is than selected according to the $p^t_{m(k)}$ [eq. (4)] or $p^{*t}_{m(k)}$ [eq. (7)], $m = 1,4$. Finally, new ϕ and ψ values are determined randomly within the chosen range m_c, the structure is changed accordingly, its energy is minimized and compared to that of the previous structure; the minimized structure is accepted or rejected according to the MC criterion [eq. (1)].

Harmonic Free Energy

The relative stability of different energy minimized structures is determined *correctly* by the free energy rather than the energy. Therefore, we have also obtained the harmonic free energy F_i^{har} around selected minimized structures i from the harmonic entropy, S_i^{har}:

$$S_i^{\text{har}} = -\frac{k_B}{2}\ln[\text{Det(Hessian)}] + S^{\text{har}}(T) \quad (8)$$

where Det stands for determinant and Hessian is the matrix of second derivatives of the energy with respect to the dihedral angles calculated at their minimized values.[47] $S^{\text{har}}(T)$ is an additive term that only depends on T and the units of the angles used; to calculate the difference ΔS_{ij} between two harmonic potential wells i and j of the same molecule at a given temperature, $S^{\text{har}}(T)$ cancels out and can be ignored. The free energy is

$$F_i^{\text{har}} = E_i - TS_i^{\text{har}} \quad (9)$$

Results and Discussion

The lowest minimized energy obtained in many MCM and MCMAB runs is -44.105 kcal/mol, which is considered to be the GEM (also denoted GEM1). Our criterion for efficiency is defined as the number of energy minimizations required to reach the GEM for the first time: the smaller this number is, the better the efficiency. Thus, the optimal values of n_1, n_2, n_3,T^* [eq. (2)], p_{low} [eq. (5)], and so forth, have been determined according to this criterion by performing many MCMAB runs with different values of these parameters.

To compare the efficiencies of MCMAB and MCM, we carried out 12 runs with both methods at $T = 400$ K [eq. (1)] starting from the same randomly generated conformations. The details of the MCM procedure used are as described in ref. 46. Thus, the number of dihedral angles l to be changed at each MC step is

$$l = \min\{K, \text{int}[1. - \ln(r + 0.00001)]\} \quad (10)$$

where $K = 36$ is the number of dihedral angles, r is the random number distributed uniformly within [0,1], and int(a) is the integer value closest to a from below. The particular l dihedral angles are determined at random where the side-chain angles with symmetry, such as χ_2 of Phe, χ_3 and χ_4 of Leu, and so forth, are chosen with a probability of 0.06, and the other angles with a probability of 0.94. Then, each selected angle is changed randomly within the range $\pm 180°$ and the energy of the resulting structure is minimized; the minimized trial structure is accepted or rejected according to the MC criterion [eq. (1)]. For MCMAB two series of runs were performed, one based on single angle probabilities only, and the other on single angle probabilities and ϕ-ψ and χ_1-χ_2 probabilities.

Table 1. Number of Energy Minimizations Required for Locating the Global Energy Minimum for the First Time.[a]

MCM	MCMAB with single angle, φ–ψ, and χ_1–χ_2 Probabilities	MCMAB with single angle probabilities only
2754	2754	2754
15933	8971	7284
23485	20092	37063
96925	6534	6657
1387	1387	1387
45146	18295	200000
140331	24174	11844
29901	23113	93477
6433	6433	6433
108507	58000	192254
45607	7909	16553
87266	20465	15897
50306	**16510**	**49300**

[a]The bold-faced numbers in the bottom row are the average values.

Efficiency of MCM and MCMAB to Locate the Gem

Table 1 demonstrates the efficiency of MCM and MCMAB for finding the GEM. Each of the 12 rows presents the number of energy minimizations required to reach the GEM structure of -44.1 kcal/mol for the first time, where all the results that appear in the same row are based on runs started from the same randomly selected "seed" structure. For each column, the average of these numbers appears in the bottom line. The results in the second column were obtained by MCMAB based on single angle as well as ϕ-ψ and χ_1-χ_2 probabilities; they are better than or equal to the MCM results (first column) where the corresponding averages are 16,510 and 50,306 minimizations, respectively; that is, MCMAB is more efficient than MCM by a factor of 3.1. The results in the third column, on the other hand, were obtained with an MCMAB procedure that consists of single angle probabilities only. While eight of these MCMAB results are better (smaller) than the corresponding MCM values, the averages, 49,300 (MCMAB) and 50,306 (MCM) are comparable. However, neglecting the results of a single run where the GEM could not be located by MCMAB after 200,000 MC steps (minimizations) does not change the MCM average but leads to a decrease of the MCMAB average by a factor of $\approx$1.4 to 35,600.

These results demonstrate that for a molecule that is not short enough, the single angle probabilities do not provide the search process with a strong guidance towards the GEM, while adding the ϕ-ψ and χ_1-χ_2 correlations constitutes a more effective bias. It

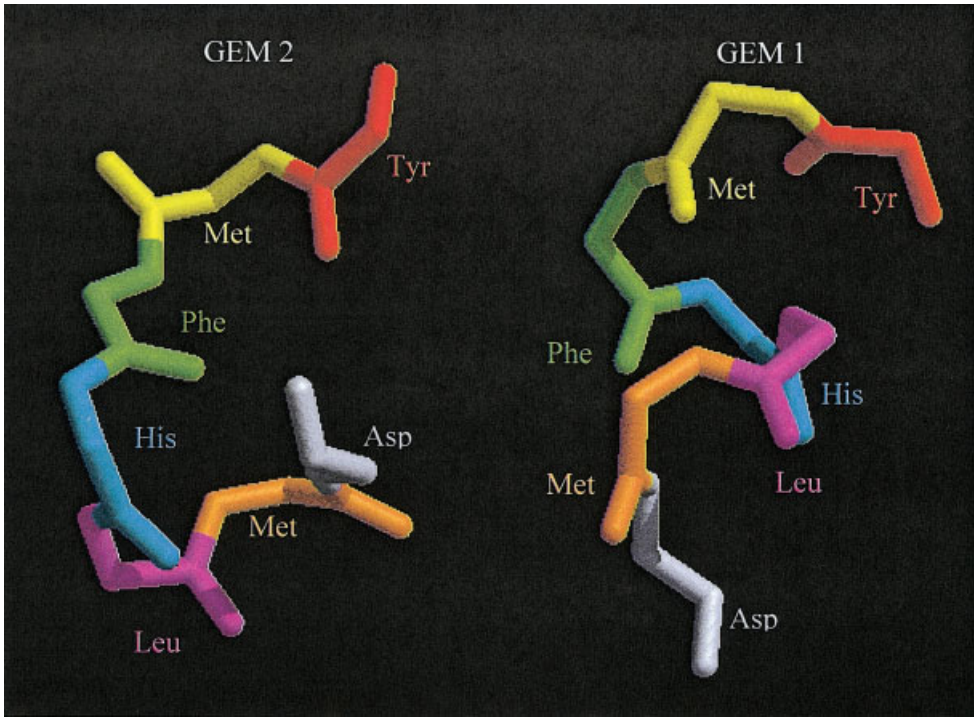

Figure 1. The two lowest energy minimized structures, GEM1 and GEM2, of deltorphin.

should be pointed out that for the pentapeptide Leu-enkephalin modeled by ECEPP/2 with constant $\omega = 180°$, single angle probabilities were found to be effective, where on average the GEM was located 2.5 times faster by MCMAB than by MCM (see article I). However, when the number of backbone degrees of freedom was increased by allowing the peptide bond ω to vary, the MCMAB procedure based on single angle probabilities has led to the smaller factor ≈1.5. Again, adding the effect of the ϕ-ψ probabilities has increased the latter factor to ≈2.8. The importance of applying such correlations for conformational search has been recognized by others (e.g., ref. 48).

Coverage of the Low Energy Region

A peptide is not expected to reside in the GEM structure but typically will exhibit *intermediate flexibility,* where several low energy potential wells are populated in thermodynamic equilibrium. Therefore, it is of interest not only to locate the GEM but also other low energy minima.[9,11,12,14,33,35,49,50,51] We have developed a methodology for treating intermediate flexibility based on an extensive conformational search that generates a large group of low energy minimized structures, from which a smaller set of structures that are significantly different are sorted.[33,51,52] These structures become "seeds" for MC simulations that span their vicinities. The free energies, hence the relative populations of these samples, are calculated by the local states method[53–55] and the contributions of these samples to the physical quantities of interest are weighted by the populations and compared with the experiment.

In this context it should be pointed out that while the GEM (also denoted GEM1) is −44.105 kcal/mol, the second lowest minimum of −43.95 kcal/mol, denoted GEM2, is only $\Delta E = 0.15$ kcal/mol higher than GEM1 and the corresponding harmonic free energies at 300 K, −3.842 and −3.663 kcal/mol [see eqs. (8) and (9)] also differ only slightly by $\Delta F^{\text{har}} = 0.18$ kcal/mol. While the stability of these structures is comparable they are significantly different, as can be judged from their ribbon graphs illustrated in Figure 1, their relatively large root mean square deviation (RMSD) of 5.18 Å, and the existence of significantly different backbone ϕ-ψ pairs for Tyr[1], Phe[3], and Asp[7] as presented in Table 2. In fact, the GEM1 structure creates a type II′ turn around Met[2] and Phe,[3]

Table 2. Dihedral Angles (in Degrees) of the Two Lowest Energy Structures, GEM1 and GEM2.

	GEM1 (−44.11 kcal/mol)		GEM2 (−43.95 kcal/mol)	
Sequence	ϕ	ψ	ϕ	ψ
Tyr	−174	138	−169	−36
Met	62	−122	87	−104
Phe	−61	−33	−150	−169
His	−67	−35	−66	−29
Leu	−69	−41	−64	−40
Met	−63	−37	−74	−41
Asp	−72	−39	−162	95

Table 3. Number of Low Energy Conformations in Energy Bins of 0.5 kcal/mol above GEM of −44.1 kcal/mol as Generated by MCM and MCMAB.

Bin (kcal/mol)	Energy (kcal/mol)	MCM (2°)	MCMAB (2°)	MCM (65°)	MCMAB (65°)
0.0–0.5	−44.1 to −43.6	16	14	12	11
0.5–1.0	−43.6 to −43.1	92	60	88	58
1.0–1.5	−43.1 to −42.6	281	214	254	205
1.5–2.0	−42.6 to −42.1	616	413	570	392
2.0–2.5	−42.1 to −41.6	1469	950	1333	876
2.5–3.0	−41.6 to −41.1	2491	1649	2300	1510
3.0–3.5	−41.1 to −40.6	3554	2354	3304	2205
3.5–4.0	−40.6 to −40.1	4240	2783	3988	2604
4.0–4.5	−40.1 to −39.6	4375	3044	4153	2851
4.5–5.0	−39.6 to −39.1	4224	2903	4077	2742
5.0–5.5	−39.1 to −38.6	3831	2588	3684	2465
5.5–6.0	−38.6 to −38.1	3478	2257	3355	2168

where residues 4–7 are helical. Thus, the low energy region of this molecule consists of at least two "funnels", meaning that it is more complex than the low energy region of Leu-enkephalin with constant ω, where the two lowest energy structures differ only by side-chain dihedrals.

In view of the above discussion, we have also investigated the conformational coverage of the low energy region obtained by MCM and MCMAB. Thus, we computed the number of different energy minimized structures within a certain energy range above the GEM, where two structures are considered different according to two criteria, if at least one dihedral angle differs by 2° or by 65°. The 2° (or 1°) criterion has been used by others;[34] the 65° criterion enables one to identify minima representing nonlocal (i.e., relatively large) potential wells, as those related to side-chain rotamers and defined over a 120°-range. Such *significantly different* energy minimized structures are used as seeds for MC simulations within the framework of a methodology for treating intermediate flexibility discussed above.[33,51–55]

For each method the data are based on three runs of 10^5 MC steps (minimizations), where only the structures accepted by the MC procedure [eq. (1)] were analyzed. It should be pointed out that in a similar analysis carried out recently for deltorphin,[56] the side-chain symmetry was not taken into account and a large number of structures were found even in the lowest energy bin. In the present analysis, however, the symmetry of the side-chains is considered. For example, structures of the benzene ring of Phe defined by χ_2 and χ_2 +180° are counted only once; therefore, the number of structures here is significantly smaller.

Table 3 presents the number of different structures found in energy bins of 0.5 kcal/mol above the GEM of −44.1 kcal/mol. The table reveals that for the first (lowest energy) bin MCM is only slightly more efficient than MCMAB, where the number of structures found is 16 versus 14 for the 2° criterion, and 12 versus 11 for the 65° criterion, respectively. For the other bins, significantly more minima are found with MCM than with MCMAB, which might seem surprising in view of the much better efficiency of MCMAB in locating the GEM. However, the fact that the bias introduced by MCMAB shortens the pathways towards the GEM also means that smaller parts of the low energy regions of conformational space are visited by MCMAB than by MCM, leading to the results of Table 3. This picture is in accord with the relatively low acceptance rate obtained in the MC process [eq. (1)] with MCMAB, 0.096 as compared to 0.18 obtained with MCM, meaning that the number of MCM structures analyzed, ≈54,000, is two times larger than the number of structures used in the analysis of the MCMAB runs. We have carried out a similar analysis for MCM and MCMAB runs each of 2×10^5 minimizations where any structure generated (i.e., accepted or rejected) was considered (data not shown). MCM still was found to generate more structures than MCMAB, but the difference between the corresponding numbers decreased, as compared to those presented in Table 3, especially for the higher energy bins.

It is of interest to compare the optimal parameters obtained for MCMAB as applied to the ECEPP-modeled peptides studied thus far, Leu-enkephalin with constant and variable ω (article I) and deltorphin with constant ω. Clearly, the parameters n_i are expected to increase with increasing the molecular size and the number of degrees of freedom. Indeed, we have obtained n_1 = 50, 500, and 4000 and n_2 (or n_3 for deltorphin) equals 800, 1800, and 7000, for Leu-enkephalin with constant ω, variable ω, and deltorphin with constant ω, respectively. However, these data still do not allow extrapolation to larger peptides. Other parameters are within a close range to each other as can be seen from Table 4. Thus, T_1^* for the single angle probabilities [eq. (2)] is 750, 850, and 850 K for Leu-enkephalin with constant ω, variable ω, and deltorphin with constant ω, respectively. $T^*_{\phi\psi}$ = 650 K for Leu-enkephalin with variable ω and $T^*_{\phi\psi} = T^*_{\chi_1\chi^2}$ = 700 K for deltorphin. The optimized probability, $p_{\phi\psi}$ for selecting a ϕ-ψ pair to be changed in an MC step (in the last stage) is the same, 0.25, for Leu-enkephalin with variable ω and deltorphin; also, p_{low} is equal to 0.25 for both models of Leu-enkephalin and is 0.28 for deltorphin. These results provide an initial picture of the sensitivity of the MCMAB parameters, suggesting that their values for larger peptides might be close to the above values, thus facilitating the final parameters' optimization.

Table 4. Comparison of the MCMAB Parameters Used for Leu-Enkephalin and Deltorphin.[a]

Parameters	Leu-enkaphalin (constant ω)	Leu-enkaphalin (variable ω)	Deltorphin (constant ω)
T_1^*—single angle	750 K	850 K	850 K
$T_{\phi\psi}^*$ or $T_{\chi_1\chi_2}^*$	—	650 K	700 K
p_{low} for single angles	0.25	0.25	0.28
$p_{\phi\psi} : p_{\chi_1\chi_2}$	—	0.25:—	0.25:0.1

[a]The parameters for Leu-enkephalin are taken from article I. T_1^*, $T_{\phi\psi}^*$, and $T_{\chi_1\chi_2}^*$ are defined in eq. (2) and explained in the text following eq. (3). p_{low} is defined in eqs. (4) and (5). $p_{\phi\psi}$ and $p_{\chi_1\chi_2}$ are the probabilities for treating a ϕ–ψ or a χ_1–χ_2 pair in the last stage of MCMAB, respectively.

Summary

In this work the MCMAB method has been tested further by applying it to an ECEPP/2 model of the heptapeptide deltorphin that is significantly larger than Leu-enkephalin studied in our previous work, article I. It has been demonstrated that on average the number of energy minimizations required by MCMAB to locate the GEM is 3.1 times smaller than that required by MCM, a factor that is slightly larger than ≈2.7 obtained for Leu-enkephalin. It should be emphasized that using single angle probabilities has been found insufficient and the above efficiency has been gained by taking into account also correlations of the backbone pairs ϕ-ψ and side-chain pairs $\chi_1-\chi_2$, as has also been found for Leu-enkephalin with variable ω. This suggests that including correlations between χ_2 and χ_3, ϕ and χ_1, and correlations among three and four angles might be important for larger peptides—a point that has been emphasized by others as well.[48] The fact that the optimal parameters for the three models studied thus far are comparable gives reason to believe that they will not change significantly for other peptides as well. On the other hand, the parameters n_i are expected to increase with the molecular size but this correlation might be established only in the future when larger peptides than deltorphin will be studied. However, MCM has been found more efficient than MCMAB to locate low energy minima. It should be pointed out that calculation of the probabilities with MCMAB does not increase computer time significantly (20,000 MC steps require ≈5 h CPU on a PC equipped with a 2 GHz processor) because most of the calculation time is spent on the energy minimizations. Because MCMAB does not rely on a structural organization, one would expect that tailoring its main features to available clustering techniques[17,19] would enhance its efficiency even further. The present ideas can be incorporated in conformational search procedures of linear or cyclic macromolecules, loops in proteins, and other systems that can be expressed by internal coordinates.

References

1. Vásquez, M.; Némethy, G.; Scheraga, H.A. Chem Rev 1994, 94, 2183.
2. Schlick, T. Rev Comput Chem 1992, 3, 1.
3. Tancredi, T.; Temussi, P.A.; Picone, D.; Amodeo, P.; Tomatis, R.; Salvadori, S.; Marastoni, M.; Santagada, V.; Balboni, G. Biopolymers 1991, 31, 751.
4. Andricioaei, I.; Straub, J.E. J Chem Phys 1998, 19, 1445.
5. Amara, P.; Straub, J.E. Phys Rev B 1996, 53, 13857.
6. Pardalos, P.M.; Shalloway, D.; Xue, G. J Glob Opt 1994, 4, 117.
7. Church, B.W.; Shalloway, D. Proc Natl Acad Sci USA 2001, 98, 6098.
8. Wales, D.J.; Doye, P.K. J Phys Chem 1997, 101, 5111.
9. Saunders, M. J Comput Chem 1991, 12, 645.
10. Ferguson, D.M.; Raber, D.J. J Am Chem Soc 1989, 111, 4371.
11. Chang, G.; Guida, W.C.; Still, W.C. J Am Chem Soc 1989, 111, 4379.
12. 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.
13. Kolossváry, I.; Guida, W.C. J Am Chem Soc 1996, 118, 5011.
14. Kolossváry, I.; Keseru, G.M. J Comput Chem 2001, 22, 21.
15. Li, Z.; Scheraga, H.A. Proc Natl Acad Sci USA 1987, 84, 6611.
16. Ripoll, D.R.; Scheraga, H.A. Biopolymers 1990, 30, 165.
17. Lee, J.; Scheraga, H.A.; Rackovsky, S. J Comput Chem 1997, 18, 1222.
18. Pillardy, J.; Liwo, A.; Scheraga, H.A. J Phys Chem A 1999, 103, 9370.
19. Pillardy, J.; Czaplewski, C.; Wedemeyer, W.J.; Scheraga, H.A. Helv Chim Acta 2000, 83, 2214.
20. Pillardy, J.; Arnautova, Y.A.; Czaplewski, C.; Gibson, K.D.; Scheraga, H.A. Proc Natl Acad Sci USA 2001, 98, 12351.
21. Caflisch, A.; Niederer, P.; Anliker, M. Proteins 1992, 14, 102.
22. Sullivan, D.C.; Kuntz, I.D. Proteins 2001, 42, 495.
23. 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.
24. Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. Science 1983, 220, 671.
25. Wilson, S.R.; Cui, W.; Moskowitz, J.W.; Schmidt, K.E. J Comput Chem 1991, 12, 342.
26. Nayeem, A.; Vila, J.; Scheraga, H.A. J Comput Chem 1991, 12, 594.
27. Meirovitch, H.; Vásquez, M. J Molec Struct (Theochem) 1997, 398–399, 517.
28. Baysal, C.; Meirovitch, H. J Comput Chem 1999, 20, 1659.
29. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. J Chem Phys 1953, 21, 1087.
30. Alder, B.J.; Wainwright, T.E. J Chem Phys 1959, 31, 459.
31. McCammon, J.A.; Gelin, B.R.; Karplus, M. Nature 1977, 267, 585.
32. Baysal, C.; Meirovitch, H. J Phys Chem 1997, 101, 2185.
33. Baysal, C.; Meirovitch, H. J Am Chem Soc 1998, 120, 800.
34. Von Freyberg, B.; Braun, W. J Comput Chem 1991, 12, 1065.
35. Abagyan, R.; Argos, P. J Mol Biol 1992, 225, 519.
36. Abagyan, R.; Totrov, M. J Mol Biol 1994, 235, 983.
37. Momany, F.A.; McGuire, R.F.; Burgess, A.W.; Scheraga, H.A. J Phys Chem 1975, 79, 2361.
38. Sippl, M.J.; Neméthy, G.; Scheraga, H.A. J Phys Chem 1984, 88, 6231.
39. 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.
40. Ozkan, S.B.; Meirovitch, H. J Phys Chem B 2003, 107, 9128.
41. Temussi, P.A.; Picone, D.; Tancredi, T.; Tomatis, R.; Salvadori, S.; Marastoni, M.; Balboni, G. FEBS Lett 1989, 247, 283.
42. Nikiforovich, G.V.; Hruby, V.J. Biochem Biophys Res Commun 1990, 174, 1053.
43. Nikiforovich, G.V.; Hruby, V.J.; Prakash, O.; Gehrig, C.A. Biopolymers 1991, 941, 55.
44. Nikiforovich, G.V.; Prakash, O.; Gehrig, C.A.; Hruby, V.J. J Am Chem Soc 1993, 115, 3399.
45. Von Freyberg, B.; Schaumann, T.; Braun, W. FANTOM, User Manual and Instructions; ETH Zurich: Zurich, 1993.

46. Meirovitch, H.; Meirovitch, E. J Comput Chem 1997, 18, 240.
47. Gō, N.; Scheraga, H.A. J Chem Phys 1969, 51, 4751.
48. Ripoll, D.R.; Liwo, A.; Scheraga, H.A. Biopolymers 2003, 46, 117.
49. Gotō, H.; Ōsawa, E. Tetrahedron Lett 1992, 33, 1343.
50. Head, M.S.; Given, J.A.; Gilson, M.K. J Phys Chem A 1997, 101, 1609.
51. Meirovitch, E.; Meirovitch, H. Biopolymers 1996, 38, 69.
52. Meirovitch, H.; Meirovitch, E. J Phys Chem 1996, 100, 5123.
53. Meirovitch, H. Chem Phys Lett 1977, 45, 389.
54. Meirovitch, H.; Koerber, S.C.; Rivier, J.E.; Hagler, A.T. Biopolymers 1994, 34, 815.
55. Meirovitch, H.; Lipkowitz, K.B. Rev Comp Chem 1998, 12, 1.
56. Yasar, F.; Arkin, H.; Celik, T.; Berg, B.A.; Meirovitch, H. J Comput Chem 2002, 23, 1127.