A Comparative Study of the Simulated-Annealing and Monte Carlo-with-Minimization Approaches to the Minimum-Energy Structures of Polypeptides: [Met]-Enkephalin

Akbar Nayeem, Jorge Vila,* and Harold A. Scheraga†

Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301

Received 14 August 1990; accepted 3 December 1990

A comparison of two methods for surmounting the multiple-minima problem, Simulated Annealing (SA) and Monte Carlo with Minimization (MCM), is presented with applications to [Met]-enkephalin in the absence and in the presence of water. SA explores a continuous space of internal variables, while MCM explores a discrete space consisting of the local energy minima on that space. Starting from random conformations chosen from the whole conformational space in both cases, it is found that, while SA converges to low-energy structures significantly faster than MCM, the former does *not* converge to a unique minimum whereas the latter does. Furthermore, the behavior of the RMS deviations with respect to the apparent global minimum (for enkephalin in the absence of water) shows no correlation with the observed overall energy decrease in the case of SA, whereas such a correlation is quite evident with MCM; this implies that, even though the potential energy decreases in the annealing process, the Monte Carlo SA trajectory does not proceed towards the global minimum. Possible reasons for these differences between the two methods are discussed. It is concluded that, while SA presents attractive prospects for possibly improving or refining given structures, it must be considered inferior to MCM, at least in problems where little or no structural information is available for the molecule of interest.

INTRODUCTION

The native conformation of a protein corresponds to a low-energy minimum of its potential energy, and in most cases this is probably the global minimum. Accordingly, prediction of protein conformation, a long-standing problem in molecular biophysics,[1-4] requires strategic and efficient methods that search for the global minimum of the intramolecular potential energy function in the conformational hyperspace of the macromolecule.[5] The difficulty with such a search arises from the fact that, because many local minima exist on the potential energy surface, minimization methods generally locate local rather than global minima, a problem referred to as "The Multiple-Minima Problem".[6] Thus, minimization searches usually end up in the minimum nearest to the starting conformation and, unless the latter accidentally happens to lie in the vicinity of the global minimum, such methods cannot determine the lowest-energy structure for the given force field.[5-7] Traditional Monte Carlo methods are very inefficient in searches for global minima, but can be improved by various techniques, e.g., by combining energy minimization with the Monte Carlo search.[8,9] The latter method [referred to as MCM (Monte Carlo with Minimization)] requires several consecutive minimization steps after each Monte Carlo search before converging to the lowest energy.

As an attractive alternative to the Monte Carlo search methods, the technique of simulated annealing (SA)[10,11] has recently been used in optimization problems involving the location of the global minimum-energy structure of polypeptides and proteins,[12-17] and for the crystallographic refinement of protein structures.[18,19] The essential feature of this method is that it combines a Monte Carlo search of conformational space at an initially high temperature with an appropriate "cooling schedule," i.e., a lowering of the temperature, which, if gentle enough, theoretically ensures that the system eventually gets trapped ("freezes") into the conformation of lowest energy.[20] This is clearly an improvement over a plain Monte Carlo search, since it biases the acceptance criterion (usually Metropolis[21]) in a way that apparently ensures that the system converges to an energy minimum. No minimization is carried out at any stage of simulated annealing, or at the final stage since, at the end of an SA run, the system has, in principle, reached a frozen state corresponding to a

*On leave from the National University of San Luis, Faculty of Science and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Instituto de Matemática Aplicada—San Luis. Ejercito de Andes 950, 5700 San Luis, Argentina, 1988–1991.

†To whom all correspondence should be addressed.

CCC 0192-8651/91/050594-12$04.00

local minimum.[10] The time-consuming minimization steps that are required in MCM are thus completely circumvented. The main drawback of this method, however, is that it depends heavily on the choice of the cooling schedule for its efficiency and success: a very slow cooling rate causes unnecessary wastage of computer time since the search does not spend much time in the minima, and conversely, a rapid cooling rate can cause the system to become supercooled, thus getting trapped in local minima from which it cannot escape. Furthermore, since the optimal cooling schedule is a sensitive function of the topography of the potential hypersurface, the cooling schedule is critically dependent upon the particular system whose energy is to be minimized, as well as on the nature of the potential energy function. The determination of the cooling schedule is a matter of trial and error.[20]

This article provides a comparison of the two methods, MCM and SA, and illustrates the application of these methods in locating the minimum-energy structure of [Met]-enkephalin, a pentapeptide molecule, whose lowest-energy conformation contains a type II′ β-turn involving Gly-Gly-Phe-Met.[8] This molecule, which has the amino-acid sequence Tyr-Gly-Gly-Phe-Met, was chosen because it has been used previously to assess other procedures for surmounting the multiple-minima problem,[8,9,22-25] and it has also been studied by using simulated annealing.[12-14] Consequently, its conformational space has been searched extensively, and the existing data enable us to compare the relative performance of the MCM and SA methods. The lowest-energy minimum of Met-enkephalin using ECEPP/2 (empirical conformational energy program for peptides) as the potential energy function[26-28] was located by using the MCM method,[8,9] and was subsequently verified by optimization in a space of higher dimensionality and then relaxing back to three dimensions.[24]

COMPUTATIONAL PROCEDURE

Potential Energy Function

The ECEPP/2 algorithm[26-28] used for modeling the polypeptide chain makes use of rigid covalent geometry, i.e., fixed bond lengths and bond angles, with the dihedral angles (i.e., ϕ, ψ, ω, χ) taken as the independent variables. The backbone of Met-enkephalin was terminated by a neutral NH_2- group at the α-amino end and a neutral –COOH group at the α-carboxyl end. All energy minimizations described here, including those for which the function to be minimized incorporated the hydration free energy (i.e., [Met]-enkephalin in the presence of water), were carried out with the algorithm SUMSL (secant unconstrained minimization solver) developed by Gay.[29]

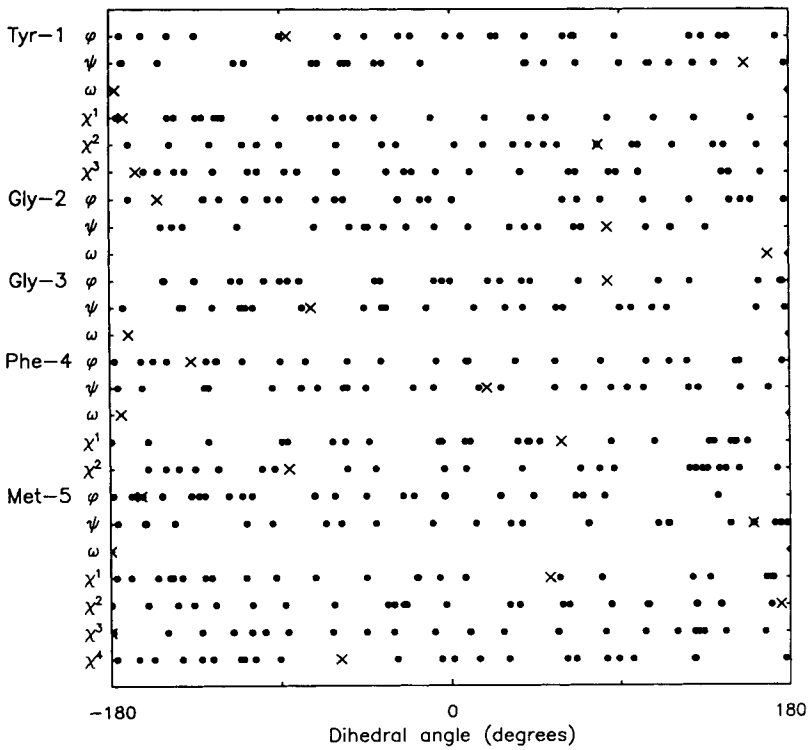

Figure 1. Distribution of dihedral angles for the 24 starting conformations of [Met]-enkephalin used in the MCM and SA procedures. In each case, "x" denotes the global minimum-energy structure.

Sampling Strategy

The search for the global minimum of the potential energy on the conformational energy surface started from either randomly selected starting points, or arbitrarily generated structures. The 24 independently generated starting structures thus chosen encompass a wide region of the whole conformational space, as can be seen from the distribution of the dihedral angles for the five residues in [Met]-enkephalin (Fig. 1).

During the Monte Carlo search, the conformations generated (starting from the initial ones) were perturbations of the current ones. With the exception of ω, which was varied only within $(180 \pm 15)°$, all backbone and side-chain dihedral angles were obtained by altering the values of one or more randomly chosen dihedral angles (from the whole range) by a random amount lying between -0.7π and 0.7π; this range was based on the concern for using an optimal choice that did not vary the dihedral angles over an unnecessarily large area of conformational space (and thus waste time), while at the same time not restricting the search to a small area. Following the suggestion of Li and Scheraga,[9] the sampling strategy used was chosen such that random changes involving k dihedral angles were generated with probabilities 2^{-k}, where $k = 1, 2, 3....$ In addition, since the overall conformation of a polypeptide depends more on its backbone than on its side chains, random changes in backbone dihedral angles were sampled more frequently than those of side-chain dihedral angles by a ratio of five to one.

Description of Algorithms

All calculations described here were carried out on an IBM-3090 supercomputer, with the codes vectorized and optimized for maximal efficiency.

In both the MCM and the SA approaches, a Monte Carlo sampling strategy was used, and newly generated conformations were either accepted or rejected in accordance with the Metropolis criterion.[21] To decide whether the Metropolis criterion was satisfied, the ECEPP/2 energy difference ΔE between the new and current conformations (E_{new} and $E_{current}$ respectively, and $\Delta E \equiv E_{new} - E_{current}$) was computed. At a given temperature T, the Metropolis criterion was considered satisfied, and the new conformation replaced the current one, if either $\Delta E < 0$, or (if $\Delta E > 0$) if a randomly generated number rnd lying between 0 and 1 fulfilled the condition: $rnd \leq \exp(-\Delta E/RT)$, where R is the gas constant. In order to preclude the high possibility of returning to the same current local minimum in the case of MCM, a minor modification was included in the Metropolis criterion—new conformations differing in energy from the current one were accepted only if $|(E_{new} - E_{current})/E_{new}| \geq 0.001$, i.e., minima differing in energy by 0.1% were considered identical for sampling purposes. If the Metropolis criterion was not satisfied, the new conformation was rejected and another one was considered.

While both methods use Monte Carlo sampling and proceed by using the Metropolis criterion, the differences in the two methods are: (1) MCM uses an energy minimizer (SUMSL) at each point along the search, so that the space of points sampled is a discrete one, whereas SA does not make use of minimization. Thus, the space sampled by the latter is continuous (and thus contains an infinite number of points); and (2) MCM is carried out at a fixed temperature, the temperature chosen being such as to yield acceptance ratios (being the ratio of the number of accepted structures to the total number of structures sampled) typically lying between 25–50%. In SA on the other hand, one starts at a typically high temperature (1000–5000 K) so that initially the acceptance ratio is high (ca. 80%) and the sampling can take place over a large area. The system is then allowed to "evolve" at this temperature, i.e., the Monte Carlo search is allowed to proceed for a finite number of steps. The temperature is then lowered by an amount determined by the particular annealing schedule,[10] defined as the sequence of temperatures and number of conformations sampled at each temperature during the search. The acceptance ratio is monitored at each temperature and, when this ratio falls below a certain amount (typically 10%) for five consecutive runs, the system is considered to be "frozen" and further decreases in temperature lead to no significant improvement (lowering) of the energy.

The relative progress of the two methods is measured not only in terms of the rate at which the energy decreases, but also by observing the change in the root-mean-square deviation (RMSD) of each newly accepted structure with respect to the minimum-energy structure as determined by Li and Scheraga,[8,9] this structure being assumed to correspond to the "global energy minimum." The RMSD was calculated over dihedral-angle space rather than cartesian coordinate space, and is defined as:

$$\mathrm{RMSD} = \sqrt{\sum_{i=1}^{N} \frac{(\theta_i - \theta_0)^2}{N}}$$

where θ_i and θ_0 denote the ith dihedral angles corresponding to the current conformation and the minimum-energy conformation, respectively, and N is the number of dihedral angles treated as variables (24 for Met-enkephalin). Whereas an overall lowering of the total energy (as the Monte Carlo search progresses) indicates that the search is drifting towards lower energy regions, the additional use of the RMSD tells us whether the search is indeed moving towards the "correct" minimum. Although the RMSD criterion is useful only when the minimum-

energy structure is known in the first place, it nonetheless provides a means of comparing the pathways followed by the two approaches with respect to a given structure. However, in the event that no structure uniquely describes the energy minimum (e.g., enkephalin in aqueous solution), the RMS deviation loses its significance.

RESULTS AND DISCUSSION

Twenty-four variable backbone and side-chain dihedral angles are required to specify the conformation of Met-enkephalin, and the conformational space is therefore expected to contain on the order of 3^{24}, or 10^{11} local minima. The number of low-energy structures for Met-enkephalin, however, estimated from the number of conformations of the individual residues that lie within 3 kcal/mol of the global minimum, is significantly smaller. Thus, using the numbers of such conformations for terminally-blocked residues from a previous study,[30] we estimate that there are $31 \times 7 \times 7 \times 14 \times 53$, or ca. 10^6 low-energy conformations of Met-enkephalin, with less than 1% of these lying within 3 kcal/mol of the global minimum of the pentapeptide.[22] Since MCM combines a Metropolis search with minimization, and the trajectory samples conformations that have low energies relative to those without minimization, e.g., SA (compare Figs. 2 and 3), the proper percentage of conformational space sampled by an MCM run is better determined not by the entire number of local minima (10^{11}), but the number of relatively low-energy minima (10^4).

Each run using MCM and SA for searching for the global minimum, which had previously been located by Monte Carlo minimization,[8] was therefore repeated by starting from 24 independently generated conformations (Fig. 1). The distribution of the dihedral angles for each of the five residues is shown in Figure 1 as solid circles, while the cross (×) de-

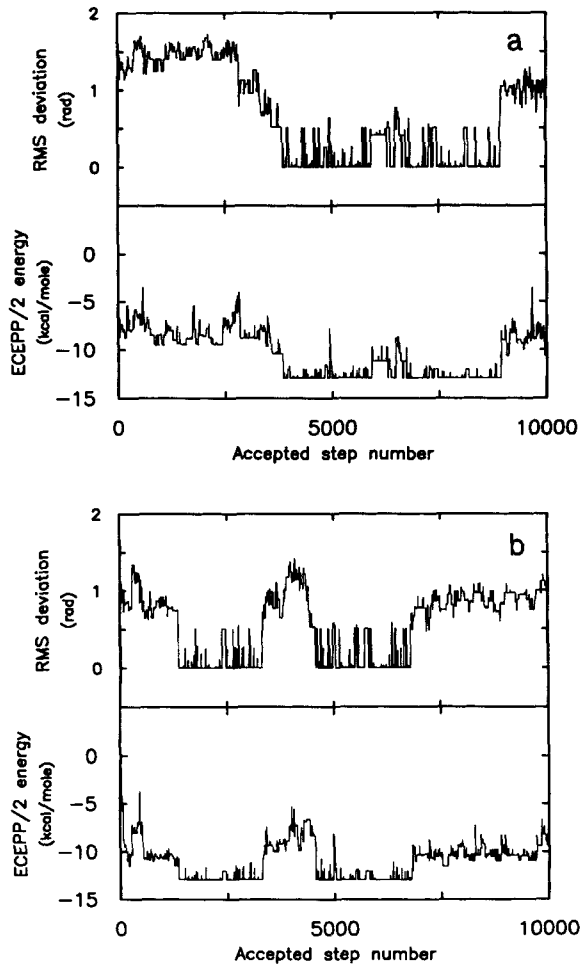

Figure 2. Progress of the MCM procedure starting from two different initial random conformations (a) and (b) whose dihedral angles are given in the first-row entries in Tables II and III, respectively. In each of the two cases, the ECEPP energy and the RMS deviation with respect to the global minimum-energy structure are shown.

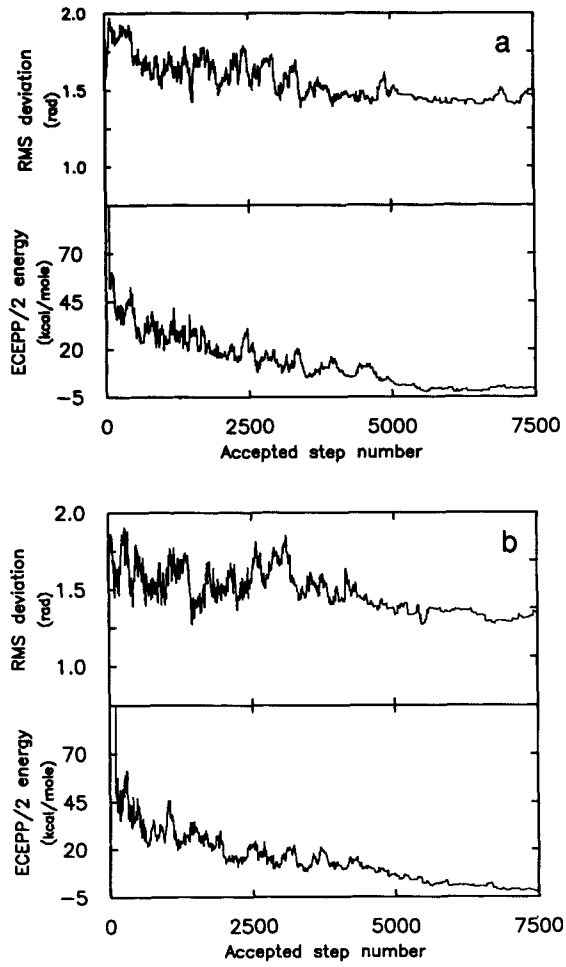

Figure 3. Progress of the SA procedure starting from two different initial random conformations (a) and (b) whose dihedral angles are given in the first-row entries in Tables II and III, respectively (the same as in Figure 2). In each of the two cases, the ECEPP energy and the RMS deviation with respect to the global minimum-energy structure are shown. The particular cooling schedule used was that of Wilson and Cui[29] (see text).

notes the apparent global minimum in agreement with that determined from previous studies.[9]

Using the sampling strategy involving random changes in two to four randomly chosen dihedral angles,[9] all the 24 structures were observed to converge to the same minimum-energy structure (within ± 0.7°) within 15,000 iterations, using the MCM method. The observed 50% acceptance rate implies that about 7,500 minima were sampled, i.e., ca. $7{,}500 \times 100/10^4$, or 75% of the low-energy minima in each such run, thus ensuring a sufficiently wide coverage of the relevant conformational space. A comparison between the minimum-energy structure obtained in the MCM runs reported here and those reported elsewhere is shown in Table I, together with the ECEPP/2 energies calculated for those conformations. The unique structure to which these MCM calculations converged is the same as that found previously.[8,9] The fact that all the randomly chosen initial starting conformations converged to the same structure also indicates that the probability of finding a structure lower in energy than the apparent global minimum obtained here is very low. On the other hand, when simulated annealing was attempted by using the same sampling strategy as for MCM, each of the 24 starting conformations converged (i.e., "froze") to a different structure, one that was often 5–15 kcal/mole higher in energy than the unique minimum found by using MCM. This was found to be the case in spite of attempting several different annealing strategies as proposed by other authors who reported success with SA (however, by using a rather simplified force field compared to ECEPP).[13–15] In our experience, the annealing schedule recently proposed by Wilson and Cui for polyalanines[17] has proven satisfactory for annealing purposes for [Met]-enkephalin.

Met-enkephalin in the Absence of Water

The relative progress of the MCM vs. SA methods, when applied to the same initial starting conformations (and using the same seeds for generating the random walk for the Monte Carlo runs) can be seen from Figures 2 and 3. Figures 2(a) and 2(b) show the change in the ECEPP/2 energy, together with the RMS deviation of the dihedral angles relative to the apparent global minimum (Table I), as the MCM proceeds, starting from the two different initial conformations. In each of these calculations, the system was allowed to evolve for 10,000 runs at a constant temperature (chosen as 300 K to correspond to room temperature). Each set of calculations shown here took four hours of CPU time on the IBM 3090. It should be noted, in both cases (which can be regarded as representative of the other cases attempted), that (1) by about 4000 iterations (1.5 CPU hours), convergence had been reached, and (2) the RMS deviation roughly follows the same course as the energy decrease. The latter implies that not only does the ECEPP/2 energy decrease as the MCM run proceeds, but also decreases *in the right direction.*

The same calculations, using the simulated annealing procedure described by Wilson and Cui[17] are shown in Figures 3(a) and 3(b) for the same initial starting conformations as in Figures 2(a) and 2(b).

Table I. Global energy-minimum structure of Met-enkephalin.

	Dihedral angle (degrees)[a]						
Residue	ϕ	ψ	ω	χ^1	χ^2	χ^3	χ^4
Tyr[1]	−86.2	156.2	−176.9	−172.6	78.7	−166.0	
	(−86)	(156)	(−177)	(−173)	(79)	(−166)	
	(−57)	(92)	(−177)	(−176)	(88)	(−153)	
	(95)	(133)	*****	(−174)	(−106)	(76)	
Gly[2]	−154.5	83.6	168.6				
	(−154)	(83)	(169)				
	(−81)	(59)	(−178)				
	(−146)	(81)	*****				
Gly[3]	83.7	−73.8	−170.2				
	(84)	(−74)	(−170)				
	(63)	(−70)	(172)				
	(100)	(−75)	*****				
Phe[4]	−137.0	19.3	−174.1	58.7	−85.5		
	(−137)	(19)	(−174)	(59)	(−85)		
	(−88)	(−37)	(−177)	(59)	(92)		
	(−157)	(147)	*****	(175)	(−96)		
Met[5]	−163.6	160.4	−179.7	52.7	175.2	−179.9	−58.5
	(−164)	(160)	(−180)	(53)	(175)	(−180)	(−59)
	(−78)	(66)	(−4)	(−176)	(58)	(−177)	*****
	(−129)	(−54)	*****	(−77)	(−60)	(144)	(170)

[a]The values in the three rows of parentheses for each residue are those taken from references 9, 13, and 14, respectively. The asterisks (*****) denote that the dihedral angle was not specified. The ECEPP/2 energies for the four conformations (assuming $\chi^4 = -59°$ for Met in the third set, and $\omega = 180°$ for all ω's in the fourth set) are −12.904, −12.904, 24.834, and 14.671 kcal/mol, respectively.

In each of these calculations, an initial temperature of 1400 K was selected. This choice for the optimum initial temperature was based on the fact that use of higher initial temperatures only wasted computer time and did not yield lower-energy structures upon freezing. On the other hand, use of temperatures lower than 1400 K prevented the system from surmounting potential barriers (2.8 kcal/mol) around local minima. Besides the choice of the initial temperature, the result of an SA run depends on: (1) The number of steps that the trajectory takes before the temperature is lowered; (2) the amount by which the temperature is lowered at the next cooling stage, and (3) the criterion used to determine when the system is considered to be frozen. In experimenting with various cooling strategies for Met-enkephalin, all these factors were considered in order to determine the optimal cooling schedule. This was found to be as follows: The Monte Carlo search was allowed to proceed for 250 steps, after which the temperature was decreased by an amount corresponding to an increase in 1/RT by a factor of 1.10. This will be referred to as the 250/1.10 schedule for convenience. At each step, a random number of dihedral angles, between 2 and 4, was varied within the range $\pm 0.7\pi$ from its current value. However, when the acceptance ratio fell below 0.25, this range was reduced to $\pm 0.35\pi$. Increasing the number of steps per temperature interval, and/or using a slower rate of cooling, did not improve the final result significantly (e.g., using 2,000 steps per temperature interval and increasing 1/RT by 1.01 resulted in a frozen structure that had an energy which was 1 kcal/mol lower than, and showed a comparable RMS deviation from, the 250/1.10 schedule; it also required much more computer time, viz., 85 CPU minutes compared to 1 CPU minute for the latter). The 250/1.10 cooling schedule was achieved in 30 temperature steps. At the end of each temperature run, the acceptance rate was calculated; if it fell below 0.1 for seven consecutive temperature runs, or if the 30 runs were over, the system was considered frozen and the annealing stopped. The final temperature thus reached was found to lie in the range 50–100 K at the end of 7500 total steps. These calculations required 1–1.5 CPU minutes on the IBM-3090. It should be noted that, while the CPU time required for the SA calculation (1 minute) is considerably less than the four CPU hours for the MCM calculation, the main point is that (1) the SA trajectory is frozen at 1 minute, and no further improvement in energy is possible; and (2) even when a considerably slower cooling rate is used (2000/1.01), no significant improvement in the energy was observed.

Figures 3(a) and 3(b) show that SA leads to energy decreases that are oscillatory, showing large amplitude oscillations at first due to the high temperatures but which gradually decrease in amplitude as the temperatures are lowered. At the end of the SA run, the final frozen structure has an ECEPP energy of -0.55 kcal/mol for Figure 3(a) and -2.23 kcal/mol for Figure 3(b). It should also be noted that the RMS deviations do not correlate in any way with the energy decreases, showing no tendency for the search to proceed along the path leading to the global minimum. A possible reason for this may be that the initial temperature is very high, and thus large barriers in energy may be overcome.

A relative assessment of the two methods may readily be made by comparing Figures 2 (MCM) and 3 (SA). In both examples, the final energy for the frozen SA structures are significantly higher than those calculated by using MCM (both structures in Figure 2 converged to a minimum-energy structure of -12.9 kcal/mol, whereas by using SA they converged at -0.55 and -2.23 kcal/mol, respectively).[31]

A possible criticism of the MCM method, however, is that it takes significantly longer (1.5–4 hours for convergence) than SA. Therefore, in additional tests, both calculations were run for the same length of CPU time. Thus, when MCM was allowed to run for 60–65 seconds, typically the time required for the SA calculations, the initial structure of Figure 2(a) reached a structure with an energy of -8.20 kcal/mol using MCM (compared with -0.55 kcal/mol by using SA), and that of Figure 2(b) reached a structure with an energy of -4.09 kcal/mol (compared with -2.23 kcal/mol by using SA). The initial structures, and those reached by using SA and MCM in the two cases are shown in Figures 4 and 5, and the corresponding dihedral angles are given in Tables II and III. Such behavior, showing the greater efficiency of the MCM method, was also observed for all the other cases examined.

We also carried out an SA run by choosing the initial point as the global minimum itself (in the absence of water). The Monte Carlo SA trajectory actually caused the system to "walk out" of the minimum, finally reaching a point of convergence 3 kcals/mol *higher* in energy than the starting point. This seems to be a direct consequence of employing high initial temperatures that cause the system to surmount large energy barriers. On the other hand, since MCM uses a minimizer at each stage of the trajectory and is normally carried out at room temperature (300 K), the system oscillates around the global minimum.

Met-enkephalin in the Presence of Water

Because of the computational complexity involved in using explicit water molecules for the protein folding problem, models that consider proteins in solution often treat the solvent as a continuum, the solvation free energy depending on the surface area of the protein that is accessible to solvent (see reference 32 and references therein). In spite of this simplified treatment, these calculations, since they

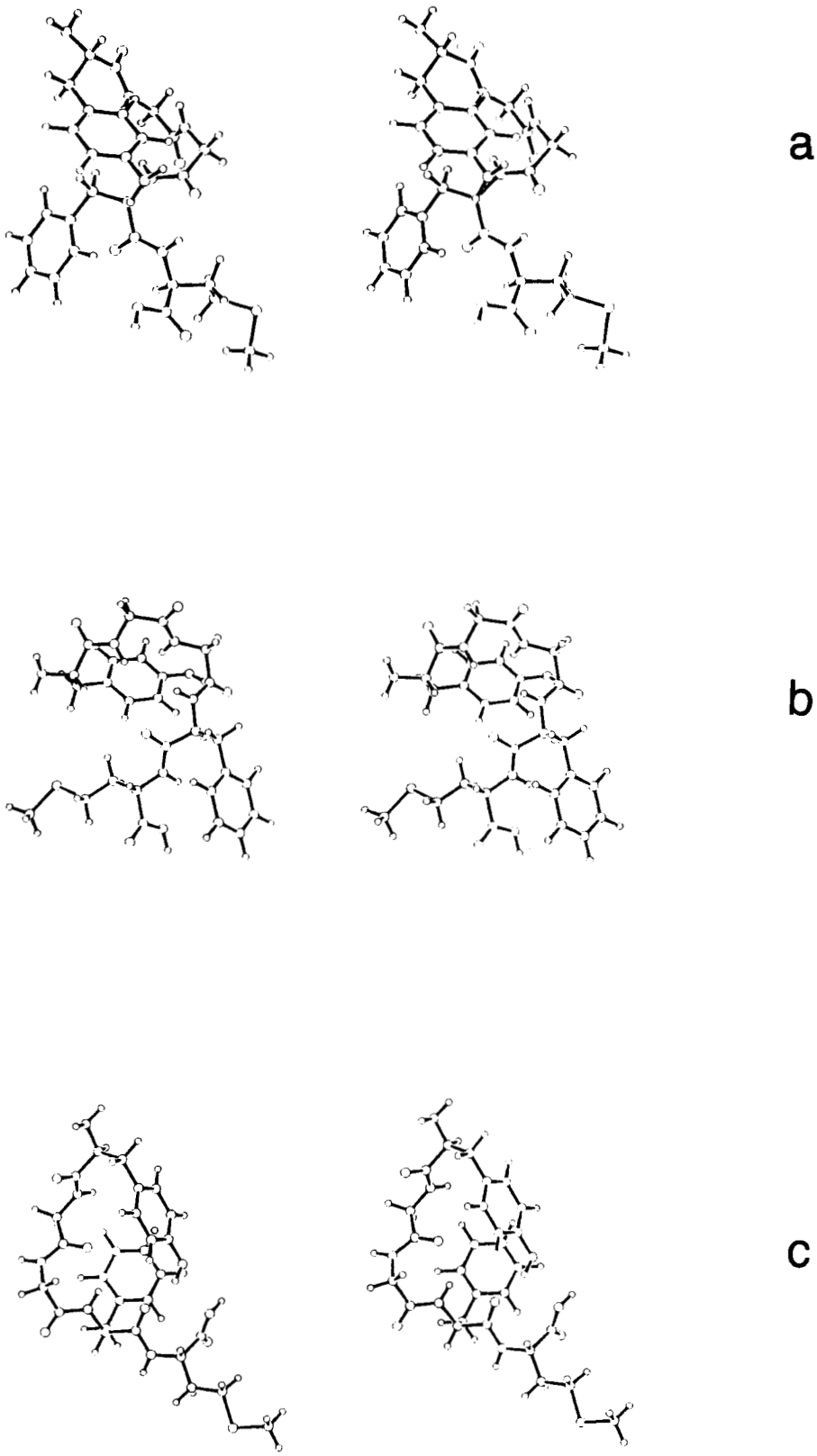

Figure 4. Stereo views of structures of [Met]-enkephalin showing (a) the initial conformation in Figures 2(a) and 3(a); (b) the converged structure by using SA; and (c) the structure reached in the same length of CPU time as (b), but by using MCM instead. The ECEPP/2 energies in these three cases are 2.1×10^6, -0.55, and -8.20 kcal/mole, respectively.

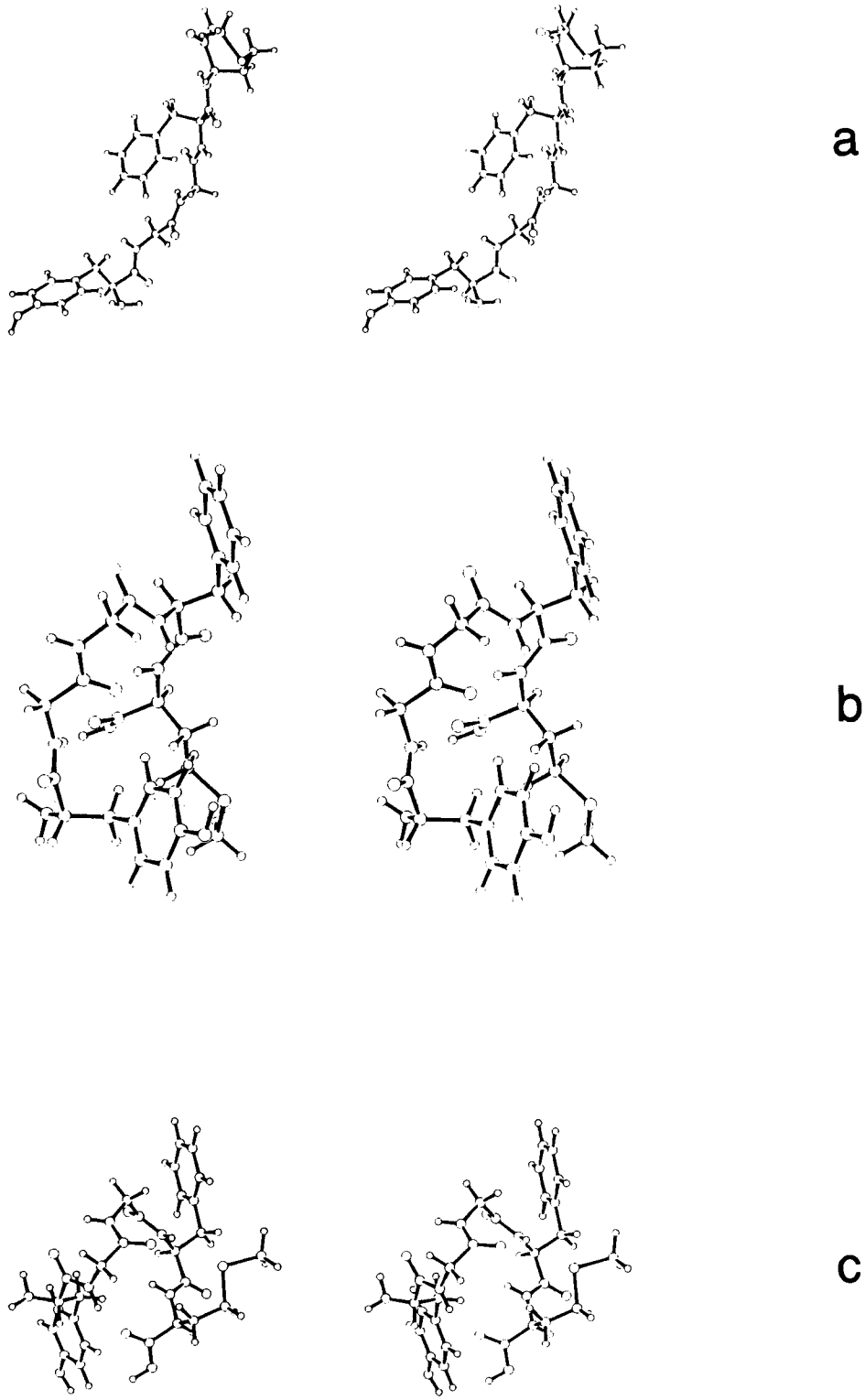

Figure 5. Stereo views of structures of [Met]-enkephalin showing (a) the initial conformation in Figures 2(b) and 3(b); (b) the converged structure by using SA; and (c) the structure reached in the same length of CPU time as (b), but by using MCM instead. The ECEPP/2 energies in these three cases are 337.4, −2.23, and −4.09 kcal/mole, respectively.

involve a computation of the surface areas exposed to solvent, often require extensive CPU time in order for an efficient search of the conformational space to be carried out.[32] Since SA, on the other hand, does not require energy minimization (an *additional* time-consuming step in the MCM search besides the calculation of surface areas), it seemed initially that the SA method might be an attractive alternative to MCM for calculating low-energy structures in solution. The relative performance of SA versus MCM was there-

Table II. Structures of the different conformations of Met-enkephalin shown in Figures 4(a), 4(b), and 4(c).[a]

Residue	Dihedral angle (degrees)						
	ϕ	ψ	ω	χ^1	χ^2	χ^3	χ^4
Tyr[1]	−174.6	−154.2	180.0	−176.1	−28.4	63.5	
	−147.6	137.5	−177.6	167.4	−105.3	154.7	
	54.0	141.5	−177.2	−179.1	67.0	−167.6	
Gly[2]	159.6	−45.2	180.0				
	−156.1	41.0	−177.5				
	−78.0	158.7	−176.1				
Gly[3]	−79.9	12.8	180.0				
	138.3	−49.8	−175.4				
	−79.8	77.1	173.0				
Phe[4]	−76.8	71.0	180.0	137.0	−150.7		
	−147.8	150.7	176.1	−175.8	82.8		
	−153.6	162.3	−178.0	62.6	−90.4		
Met[5]	−105.1	37.5	180.0	−126.5	−160.3	129.2	30.7
	−153.4	−34.0	171.8	179.6	174.8	170.5	167.9
	−66.1	107.0	−176.6	−173.1	175.8	179.9	−60.0

[a]The three rows of entries for each residue correspond to Figures 4(a) (initial), 4(b) (SA), and 4(c) (MCM), respectively.

fore carried out on [Met]-enkephalin in aqueous solution.

In the model for hydration used here, the free energy of solvation G_i for a given amino-acid i with N atoms is given by:[32–34]

$$G_i = \sum_{k=1}^{N} \sigma_k A_{i,k}$$

where σ_k is the free energy of solvation per unit area for the kth atom, and $A_{i,k}$ is the corresponding solvent accessible surface area for the kth atom. σ_k was obtained from NMR measurements of the coupling constants for peptides in water,[32] while $A_{i,k}$ was obtained from an analytical surface area calculation following the algorithm of Connolly.[35] (Several other hydration models discussed elsewhere[32] are not included in our discussion since they fall outside the scope of this work.)

Three runs each of MCM and SA were carried out in order to compare the relative performances of the two methods. With the SA method, each run on [Met]-enkephalin in water started at 1400 K as before, allowing the Monte Carlo trajectory to proceed for 250 steps, and then lowering the temperature by an amount corresponding to a 10% rise in $1/RT$. Because of the additional term for hydration free energy in the potential energy, these calculations typically required 0.5–1 CPU hour on the IBM-3090 (as opposed to 1–2 CPU minutes in the absence of water) because the solvent exposed surface area had to be calculated at each point of the trajectory. The MCM calculations were carried out at two temperatures, 100 K and 300 K. Rather than allow MCM to proceed for 10,000 steps as before, however, the MCM calculations were allowed to proceed for a maximum of 40 CPU minutes, in order to monitor the progress of the MCM trajectory in times comparable to SA. In this time, 150 steps of MCM were taken. It was observed, however, that, in the cases studied, within 15 accepted steps the MCM trajectory reached lower-energy structures than that at which

Table III. Structures of the different conformations of Met-enkephalin shown in Figures 5(a), 5(b), and 5(c).[a]

Residue	Dihedral angle (degrees)						
	ϕ	ψ	ω	χ^1	χ^2	χ^3	χ^4
Tyr[1]	6.2	−108.5	180.0	−50.8	−101.7	67.0	
	−148.6	−41.4	178.8	−176.7	−96.0	3.5	
	−85.2	157.0	173.7	−61.9	100.9	173.9	
Gly[2]	−121.7	−146.9	180.0				
	106.6	−167.5	178.0				
	−76.6	72.1	−173.8				
Gly[3]	−151.2	162.6	180.0				
	63.3	−93.3	−178.4				
	76.1	−75.2	178.8				
Phe[4]	176.1	−175.8	180.0	−5.0	37.6		
	−132.5	21.4	179.9	−56.8	−71.4		
	−144.4	29.9	−180.0	−56.4	−79.8		
Met[5]	−166.9	−58.3	180.0	167.3	62.8	−115.2	82.6
	−145.5	139.1	−178.2	−176.6	171.3	174.3	−70.4
	−132.6	133.2	179.2	−70.0	92.4	−172.7	−59.1

[a]The three rows of entries for each residue correspond to Figures 5(a) (initial), 5(b) (SA), and 5(c) (MCM), respectively.

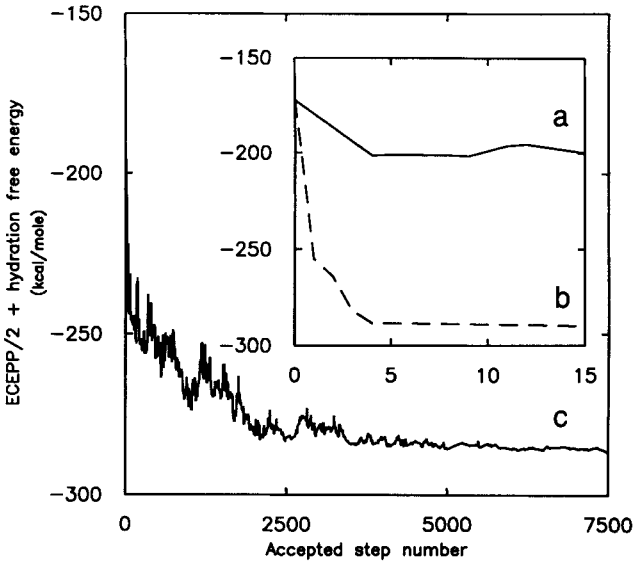

Figure 6. Comparison of performance of the (a) SA and (b) MCM methods for the first 15 steps in the case of [Met]-enkephalin in the presence of water (inset). In both cases, the starting structure was the ECEPP/2 global minimum in the absence of water (Table I). Curve (c) corresponds to 7500 steps of an SA run, requiring 34 CPU minutes. It should be noted that, in spite of the significantly shorter time, 7 CPU minutes (15 steps), (b) reached a lower-energy structure (−289.9 kcal/mol) compared to −286.4 kcal/mol for SA in 34 CPU minutes.

the SA trajectory froze after 7500 steps; the time required for this was about 7 CPU minutes. [In order to enhance the efficiency of these methods, the codes were vectorized and parallelized, a step particularly desirable for MCM which evaluates analytical derivatives of the surface areas as a function of the internal coordinates of the molecule (in the presence of solvent, each local minimization was observed to require at least 10 times as much time as that in the absence of solvent).]

The progress of the trajectory using the two methods is shown in Figure 6. In each case, the initial starting conformation was chosen to be the global minimum in the absence of water (see Table I for internal coordinates), and the same seed was used to generate the random walk in both cases. The conformations of the final structures thus obtained are given in Table IV. The CPU times in the two cases were 34 minutes for SA, and 7 minutes for MCM. It should be noted that, in spite of running for considerably less time than SA, the MCM trajectory reached a structure of lower energy (−289.9 kcal/mol) than that with SA (−286.4 kcal/mol). This was also observed when the initial conformations were either chosen randomly or fully extended (not shown).

As discussed above, a possible reason for the greater efficiency of MCM may be that, owing to the high initial temperature used in SA, the SA trajectory can surmount high energy barriers. As the temperature is gradually lowered, the trajectory settles down in the vicinity of that point on the conformational surface where the temperature was lowered (250 steps). (In practice, using a higher initial temperature and/or a larger number of steps for equilibration at a given temperature was not found to improve the results significantly. An initial temperature of 1400 K and 250 steps per temperature were found to be adequate.) On the other hand, the choice of (1) a lower temperature (300 K), and (2) the walk along the local minima for MCM ensures that the MCM trajectory proceeds along local minima of decreasing energy and at the same time has a low probability for accepting high-energy structures.

In these cases, however, since [Met]-enkephalin in solution does not appear to show a unique structure,[9] no reference structure was available for calculating an RMS deviation. The real point, however, is that MCM approaches a lower-energy structure more *efficiently* than SA, providing further justification for the conclusion that MCM is more efficient than SA, for this problem.

CONCLUSIONS

Our studies of [Met]-enkephalin in the absence of water showed that, while SA converges to a lower-energy structure significantly faster than MCM, the former does *not* converge to a unique minimum in the case of the twenty-four randomly chosen initial

Table IV. Structures of the two conformations of Met-enkephalin in water.[a]

	Dihedral angle (degrees)						
Residue	ϕ	ψ	ω	χ^1	χ^2	χ^3	χ^4
Tyr[1]	−76.3	154.8	179.1	−68.7	123.3	160.2	
	−75.1	173.3	168.5	−75.7	115.5	179.7	
Gly[2]	−160.9	108.0	−165.7				
	−162.1	149.0	−170.5				
Gly[3]	−178.8	−150.4	179.0				
	175.5	−165.9	168.5				
Phe[4]	−90.8	128.1	−166.0	−83.0	121.7		
	−81.7	76.5	−168.5	−73.1	111.6		
Met[5]	−143.4	98.9	−172.1	−63.1	−161.9	−176.5	56.3
	−131.7	99.5	−179.9	−63.9	−158.5	−73.3	58.5

[a]The two rows of entries for each residue correspond to Figures 6a and 6b, respectively.

conformations whereas the latter does. The energy difference between the minima reached by the two methods is typically 5–15 kcal/mol in favor of MCM. In the presence of water, MCM reaches a lower-energy structure in *less* time than SA. Furthermore, the RMS deviations with respect to the apparent global minimum (for [Met]-enkephalin in the absence of water) appear to show no correlation with the observed overall energy decrease in the case of SA, while such a correlation is evident with MCM. This implies that, even though the potential energy decreases in the annealing process, the Monte Carlo trajectory does not proceed towards the global minimum. A possible reason for this may be that SA starts at an initially high temperature, allowing the system to surmount barriers that are much higher than the barriers separating local minima, so that even the global minimum (if reached during the course of a trajectory) can be abandoned. Such is not the case with MCM, which is normally carried out at room temperature (300 K); thus, large energy barriers (ca. >1 kcal/mol) cannot be overcome as easily. Therefore, while, on the one hand, starting at a high temperature serves the purpose of exploring the gross features of the space, it suffers from the drawback that one can escape relatively easily from minima that may lie close to the global minimum. Secondly, the lowering of temperature (in SA) at the end of a specified number of runs (at a constant temperature) gradually biases the search in favor of those regions that lie within the vicinity of the point reached at the end of a given constant temperature run. In addition to these two reasons, the success of SA is critically dependent on the nature of the energy surface. As shown by Wilson and Cui,[17] this includes both the number of local minima on, as well as the shape of, the conformational energy surface. The failure of SA to converge in the present case (Met-enkephalin) must, therefore, be due to the large number of local minima (3^{24}) that needs to be sampled, as well as the fact that the surface of Met-enkephalin is considerably complex, as stated by Wilson and Cui: "The energy surface for Met-enkephalin (18 dihedrals) is much more complex than the symmetrical one for Ala_{10} (20 dihedrals)."[36] Thus, an SA run on a 10-residue polyalanine chain,[17] which has about 3^{20} local minima but, unlike enkephalin, has a *"symmetrical"* energy surface, locates the global minimum using the cooling schedule described above (i.e., 250/1.10, and 30 temperature steps). Thus, although SA seemed to work for polyalanine,[17] its success in this case must be attributed to the relative symmetry of the energy surface. This is indeed a serious limitation. MCM, on the other hand, uses minimization, and therefore samples a discrete space of variables. Thus, the actual shape of the energy surface is much less relevant. Therefore, while SA presents attractive prospects for possibly improving or refining given structures, it must be considered inferior to MCM, at least in cases where little or no structural information is available for the molecule of interest.

This work was supported by grant No. GM-14312 from the National Institutes of Health and grant No. DMB84-01811 from the National Science Foundation. J. V. was supported by a fellowship from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) of Argentina (1988–1990). The computations were carried out at the Cornell National Supercomputer Facility, a resource at the Cornell Center for Theory and Simulation in Science and Engineering, which receives major funding from the National Science Foundation and IBM Corporation, with additional support from New York State and members of the Corporate Research Institute.

References

1. H.A. Scheraga, *Adv. Phys. Org. Chem.* **6,** 103 (1968).
2. C.B. Anfinsen, *Science,* **181,** 223 (1973).
3. R. Jaenicke, Ed. *Protein Folding,* Elsevier/North Holland Biomedical Press, Amsterdam, 1980.
4. C. Ghelis and J. Yon, *Protein Folding,* Academic Press, New York, 1982.
5. H.A. Scheraga, *Chemica Scripta,* **29A,** 3 (1989).
6. K.D. Gibson and H.A. Scheraga, in *Structure and Expression Volume 1: From Proteins to Ribosomes,* R.H. Sarma and M.H. Sarma, Eds., Adenine Press, Guilderland, New York, 1988, pp. 67–94.
7. C. Tosi, R. Fusco, L. Caccianotti, and V. Parisi, *J. Math. Chem.,* **3,** 311 (1989).
8. Z. Li and H.A. Scheraga, *Proc. Natl. Acad. Sci., U.S.A.,* **84,** 6611 (1987).
9. Z. Li and H.A. Scheraga, *J. Mol. Struc. (Theochem),* **179,** 333 (1988).
10. S. Kirkpatrick, C.D. Gelatt, Jr., and M.P. Vecchi, *Science,* **220,** 671 (1983).
11. D. Vanderbilt and S.G. Louie, *J. Comput. Phys.* **56,** 259 (1984).
12. S.R. Wilson, W. Cui, J.W. Moskowitz, and K.E. Schmidt, *Tetrahedron Lett.* **29,** 4373 (1988).
13. J.W. Moskowitz, K.E. Schmidt, S.R. Wilson, and W. Cui, *Int. J. Quant. Chem: Quant. Chem. Symp.,* **22,** 611 (1988).
14. H. Kawai, T. Kikuchi, and Y. Okamoto, *Protein Eng.,* **3,** 85 (1989).
15. C. Wilson and S. Doniach, *Proteins,* **6,** 193 (1989).
16. P. Auffinger and G. Wipff, *J. Comp. Chem.* **11,** 19 (1990).
17. S.R. Wilson and W. Cui, *Biopolymers,* **29,** 225 (1990).
18. A.T. Brunger, *J. Mol. Biol.,* **203,** 803 (1988).
19. M. Nilges, A.M. Gronenborn, A.T. Brunger, and G.M. Clore, *Protein Eng.* **2,** 27 (1988).
20. W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes: The Art of Scientific Computing,* Cambridge University Press, Cambridge, 1986, 326–334.
21. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, *J. Chem. Phys.,* **21,** 1087 (1953).
22. M. Vásquez and H.A. Scheraga, *Biopolymers,* **24,** 1437 (1985).
23. a. G.H. Paine and H.A. Scheraga, *Biopolymers,* **24,** 1391 (1985); b. *ibid,* **25,** 1547 (1986); c. *ibid,* **26,** 1125 (1987).
24. E.O. Purisima and H.A. Scheraga, *J. Mol. Biol.,* **196,** 697 (1987).
25. D.R. Ripoll and H.A. Scheraga, *J. Prot. Chem.* **8,** 263 (1989).

26. F.A. Momany, R.F. McGuire, A.W. Burgess, and H.A. Scheraga, *J. Phys. Chem.*, **79,** 2361 (1975).
27. G. Némethy, M.S. Pottle, and H.A. Scheraga, *J. Phys. Chem.* **87,** 1883 (1983).
28. M.J. Sippl, G. Némethy, and H.A. Scheraga, *J. Phys. Chem.* **88,** 6231 (1984).
29. D.M. Gay, *ACM Trans. Math. Software*, **9,** 503 (1983).
30. M. Vásquez, G. Némethy, and H.A. Scheraga, *Macromolecules*, **16,** 1043 (1983).
31. Following the suggestions of the two referees of this article, two modifications to traditional simulated annealing were also attempted: (1) At the final stage of the SA run, the frozen conformation was subjected to local energy minimization. In both cases shown in Figures 3(a) and 3(b), this led to an energy decrease of only about 2.3 kcal/mol (with the energy still being much higher than that of the global minimum), and the RMS deviation was not improved. (2) Instead of using the final conformation reached at the end of a given temperature interval, the minimum energy conformation encountered during that interval was used as the starting point for the next (lower) temperature interval. In this case, the trajectory took a different course and resulted in a frozen structure that was not very different in energy from the previous one (lower by about 2 kcal/mol, but still about 6 kcal/mol higher than the global minimum).
32. J. Vila, R.L. Williams, M. Vasquez, and H.A. Scheraga, *Proteins* in press.
33. D. Eisenberg and A.D. McLachlan, *Nature*, **319,** 199 (1986).
34. T. Ooi, M. Oobatake, G. Némethy, and H.A. Scheraga, *Proc. Natl. Acad. Sci., USA*, **84,** 3086 (1987).
35. M.L. Connolly, *J. Appl. Cryst.*, **16,** 548 (1983).
36. For example, see footnote on page 228 of reference 17.