PROTEINS: Structure, Function, and Genetics 14:102–109 (1992)

Monte Carlo Minimization With Thermalization for Global Optimization of Polypeptide Conformations in Cartesian Coordinate Space

Amedeo Caflisch, Peter Niederer, and Max Anliker
Institute of Biomedical Engineering and Medical Informatics of the Swiss Federal Institute of Technology, CH-8092 Zürich, Switzerland

ABSTRACT **A new minimization procedure for the global optimization in cartesian coordinate space of the conformational energy of a polypeptide chain is presented. The Metropolis Monte Carlo minimization is thereby supplemented by a thermalization process, which is initiated whenever a structure becomes trapped in an area containing closely located local minima in the conformational space. The method has been applied to the endogenous opioid pentapeptide methionine enkephalin. Five among 13 different starting conformations led to the same apparent global minimum of an in-house developed energy function, a type II′ reverse turn, the central residues of which are Gly-3-Phe-4. A comparison between the ECEPP/2 global minimum conformation of methionine enkephalin and the apparent one achieved by the present method shows that minimum-energy conformations having a certain similarity can be generated by relatively different force fields.**

Key words: force field, Metropolis criterion, methionine enkephalin, polypeptide folding

INTRODUCTION

Since Anfinsen's work in 1961,[1] increasing evidence has been presented which substantiates the hypothesis that the primary structure of a protein determines its tertiary structure, i.e., the amino acid sequence contains all the information required for unique folding in a proper solvent. Yet, there exists no algorithm to date which is able to generate the structure of a protein from its sequence. This is partly due to the multiple-minima problem[2] as well as to the inherent approximations in the formulation utilized to describe the energy and dynamics of the system consisting of the protein and the solvent. The former represents the major obstacle in finding the global minimum of the energy function, even in the case of an oligopeptide, which possesses a huge number of local minima due to the complexity and heterogeneity of systems characterized by a large number of degrees of freedom, many of which being strongly coupled.

The most common approximations to surmount the multiple-minima problem consists of the use of a reduced number of interaction centers (e.g., extended atom representation) and/or degrees of freedom (e.g., fixed bond distances and bond angles). Scheraga and his group developed several methods to locate the global minimum of the energy in the dihedral angle space of a small polypeptide chain (see [3] for a review). One of these methods, the Monte Carlo minimization (MCM) procedure,[4,5] yielded the global minimum of the ECEPP/2[6,7] energy for the pentapeptide Met-enkephalin.

We decided to develop a method for global optimization in the external coordinate space, because bond length and bond angle deviations from fixed ideal values, although relatively small, can be frequently found in experimentally determined structures of proteins. The protein description in cartesian coordinate space imply an increase in the number of variables by about a factor of three; in spite of that, a flexible geometry is necessary, as will be shown, to obtain folded conformations characterized by optimal nonbonded interactions at the expense of a small strain in the bond lengths and bond angles.

The present procedure combines the MCM method of Li and Scheraga[4,5] and a history dependent thermalization process, which allows to surmount high conformational energy barriers. The thermalization process is initiated whenever a conformation cannot escape during a given number of conventional MCM cycles from a set of local minima characterized by similar values of the conformational energy. Thereby, the Monte Carlo minimization with thermalization (MCMT), reported here, allows a given structure to escape from a local minimum by thermal perturbation at two different temperature

---

Received July 8, 1991; revision accepted September 20, 1991.
Address reprint requests to Dr. Amedeo Caflisch, Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, MA 02138.

TABLE Ia. Results of the Minimization Procedure*

	Starting conformation	EC	AH	EX	R2	R3	R6	R5
Step I	Heavy atoms DME[†]	1.39	0.45	0.02	0.33	0.60	0.13	1.34
	Backbone atoms DME[†]	0.41	0.31	0.01	0.06	0.29	0.06	0.59
	Total energy	−34.68	−35.34	−36.34	−35.38	−34.57	−35.71	−34.46
Step II	Heavy atoms DME[†]	1.39	0.01	0.00	0.01	0.01	0.01	1.34
	Backbone atoms DME[†]	0.41	0.00	0.00	0.00	0.00	0.00	0.60
	Total energy	−34.69	−36.38	−36.38	−36.38	−36.38	−36.38	−34.51
	Lowest energy conform.[§]	3–4 II′	3–4 II′	3–4 II′	3–4 II′	3–4 II′	3–4 II′	3–4 II

TABLE Ib. Results of the Minimization Procedure

	Starting conformation	R1	R4	R0	R9	R8	R7
Step I	Heavy atoms DME[‡]	0.06	0.04	0.75	0.02	0.64	1.61
	Backbone atoms DME[‡]	0.01	0.01	0.15	0.01	0.04	0.33
	Total energy	−35.54	−35.72	−34.32	−35.64	−34.21	−34.36
Step II	Heavy atoms DME[‡]	0.00	0.01	0.65	0.01	0.01	1.61
	Backbone atoms DME[‡]	0.00	0.00	0.20	0.00	0.00	0.33
	Total energy	−35.75	−35.75	−34.95	−35.75	−35.75	−34.40
	Lowest energy conform.[§]	2–3 I′	2–3 I′	2–3 I′	2–3 I	2–3 I	2–3 I

*Step I corresponds to MCMT at a thermalization temperature of 10,000 K. The rms distance matrix error (DME) determines the degree of similarity in the pattern of intramolecular contacts between two conformations. DME values are in Angstroms; energy values are in kcal/mol.

[†] DME with respect to the apparent global minimum energy conformation (3–4 II′ reverse turn).

[§] In all runs, the type of the reverse turn generated by step I was not modified by step II.

[‡] DME with respect to the low lying local minimum energy conformation (2–3 I′ reverse turn).

levels. A temperature value which agrees with physiological conditions, as in the original MCM procedure, in which a new conformation with an energy higher than the old one is accepted on the basis of the Metropolis criterion with Boltzmann transition probability. If after a given number of cycles the structure has not escaped from a set of similar low energy conformations, a very high temperature during a short thermalization process is chosen.

The present method has been successfully applied on the Met-enkephalin structure with netural amino carboxyl termini (H-Tyr-Gly-Gly-Phe-Met-OH), which has been chosen to allow a comparison between our results and the ones obtained by the Scheraga group.

METHODS

An accurate description of the RES force field, which was utilized for all the cartesian space minimizations described in the present work, can be found in references 8 and 9; the former publication also contains a comparison between the RES energy terms and the ones of previously developed empirical energy functions. In the RES potential energy function,

$$E = E_b + E_a + E_t + E_{LJ} + E_C + E_H \qquad (1)$$

the first three addends on the right approximate individually the interactions between covalently bonded atoms, separated by 1, 2, and 3 covalent bonds, while the remaining three represent the interactions between nonbonded atoms, i.e., atoms separated by 3 or more covalent bonds: Lennard–Jones energy, electrostatic energy, and hydrogen bond energy. The parameters utilized by RES were derived from those of the CHARMM[10] empirical energy function. The heavy atoms and the polar hydrogen atoms (40 and 8 for Met-enkephalin) are considered as interaction centers. Their cartesian coordinates constitute the variables of the RES energy function. Furthermore, no cutoff for the nonbonded interactions is introduced and a distance-dependent dielectric constant is used. First derivatives with respect to the cartesian coordinates are evaluated analytically.

The minimization procedure consists of two steps.

In step I each starting structure is subjected to 50,000 MCMT cycles. Each cycle begins by a random change ($-180° \le \theta \le 180°$) involving n torsion angles (where n is chosen with a probability 2^{-n}, cf .5) randomly selected among all the variable dihedrals θ. Simultaneous random changes in more flexible torsion angles are required to escape from energetically favorable local minima in internal coordinate minimization with rigid geometry.[5] The randomly modified structure is conventionally minimized in cartesian coordinate space by 100 conjugate gradient iterations. The Metropolis criterion[11] is then applied at a temperature of 310 K. If after 100 MCM cycles the energy is not lowered by at least 1 kcal/mol, 5 thermalization cycles are performed by chosing a very high value of the

TABLE II. Energy and Torsion Angles of the Minimum-Energy Structures

Potential	RES	RES	ECEPP/2
Structure	2–3 I′[*]	3–4 II′[†]	EC[‡]
Energy [kcal/mol]	−35.75	−36.38	
Tyr-1			
ϕ	−47	−44	−86
ψ	−114	16	156
ω	−179	−175	−177
χ_1	−177	−56	−173
χ_2	−50	−60	79
χ_3	1	−162	−166
Gly-2			
ϕ	61	−86	−154
ψ	22	54	83
ω	−176	179	169
Gly-3			
ϕ	76	75	84
ψ	8	−72	−74
ω	−176	168	−170
Phe-4			
ϕ	−136	−106	−137
ψ	−34	−20	19
ω	−173	−177	−174
χ_1	64	64	59
χ_2	93	101	95
Met-5			
ϕ	130	−77	−164
ψ	32	76	160
ω	−3	−11	−180
χ_1	60	−59	53
χ_2	−176	179	175
χ_3	175	−68	−180

[*] Runs R1 and R4 converged toward the RES low lying local minimum, a 2–3 I′ conformation, while R8 and R9 toward its mirror image.

[†] Runs AH, EX, R2, R3, and R6 converged toward the RES apparent global minimum, a 3–4 II′ conformation.

[‡] The ECEPP/2 global minimum energy conformation consists of a 3–4 II′ reverse turn.

temperature for the Metropolis criterion. The rationale for introducing a thermalization period derives from the consideration that, although the original MCM procedure reaches any local minimimum from any other one after a finite number of random sampling steps, disjointed sets of adjacent low energy conformations separated by very high barrier (for Met-enkephalin, turns with central residues 2–3 and turns with central residues 3–4) can not be accessed in a reasonable number of cycles.

The structure with the lowest energy is subjected to step II, which consists of 20,000 MCM cycles as in step I but now random changes involve only the side chain dihedrals and the torsion angles, which define the position of the hydrogens in the NH_2 terminus (ϕ_1) and the COOH terminal group (ψ_N and ω_N for a polypeptide chain of N residues). Furthermore, the temperature is kept constant at 310 K, i.e., no thermalization is allowed during step II.

The 13 following starting conformations of Met-enkephalin were utilized:

- Ten different structures having random values in all variable dihedrals (R0-R9);
- the ECEPP/2 global minimum structure[4,12] (EC);
- α-helical (AH; φ, ψ, and ω values of −57°, −47°, and 180°, respectively; random values of χs);
- totally extended structure (EX; all variable dihedrals set to 180°).

All calculations reported here were carried out on an IBM RS/6000-520 workstation, with the Pascal code optimized for maximal efficiency. Each conjugate gradient minimization (100 iterations) of the Met-enkephalin pentapeptide needed about 1.1 sec of CPU time thus requiring a total of about 21 hr for each run (70,000 cycles).

RESULTS AND DISCUSSION

Influence of the Thermalization Temperature

The 13 starting conformations were subjected to MCM at 310 K, characterized by more frequent sampling of backbone dihedrals as described.[5] The same starting structures were subjected to MCMT performed at a thermalization temperature (T_{therm}) of 2,000 K and 10,000 K. All runs consisted of 50,000 cycles; thereby, conformations with total energy below −33 kcal/mol were saved.

All minimum-energy conformations generated by MCM had an energy above −35 kcal/mol; furthermore, no jump between reverse turns involving residues 2–3 and 3–4 was observed during a run. The original MCM procedure did not allow escape from a set of adjacent local minima by a reasonable number of simulation cycles. The acceptance ratio (accepted/discarded structures), according to the Metropolis criterion at 310 K, was about 11% because of the frequent sampling of backbone dihedrals. Although the backbone conformations of the MCM minimum-energy structures consisted of reverse turns involving residues 2–3 or 3–4, both the backbone and side chain torsion angles had different values so that low energy structures generated from different starting conformations were not similar.

Two MCMT runs carried out at 310 K and T_{therm} of 2,000 K generated each one minimum-energy structure with a total energy below −35 kcal/mol. Seven MCMT runs carried out at 310 K and T_{therm} of 10,000 K generated each at least one minimum-energy structure with a total energy below −35 kcal/mol. The acceptance ratio was about 18 and 21% for the 2,000 K T_{therm} run and the 10,000 K T_{therm} run, respectively. To compare two conformations we utilize the rms distance matrix error (DME), which determines the degree of similarity in the pattern of intramolecular contacts between two

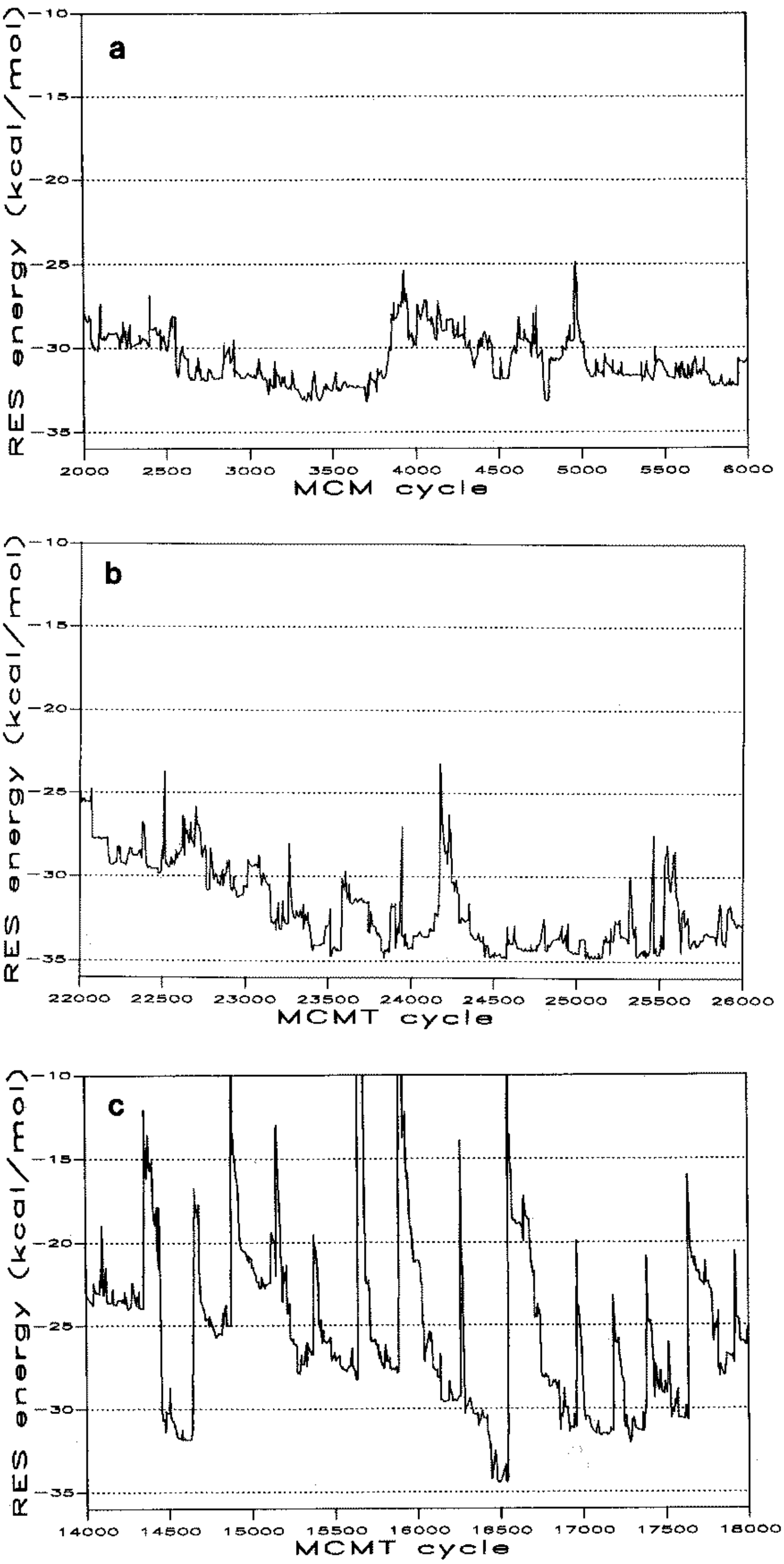

Fig. 1a–c. Legend appears on page 106.

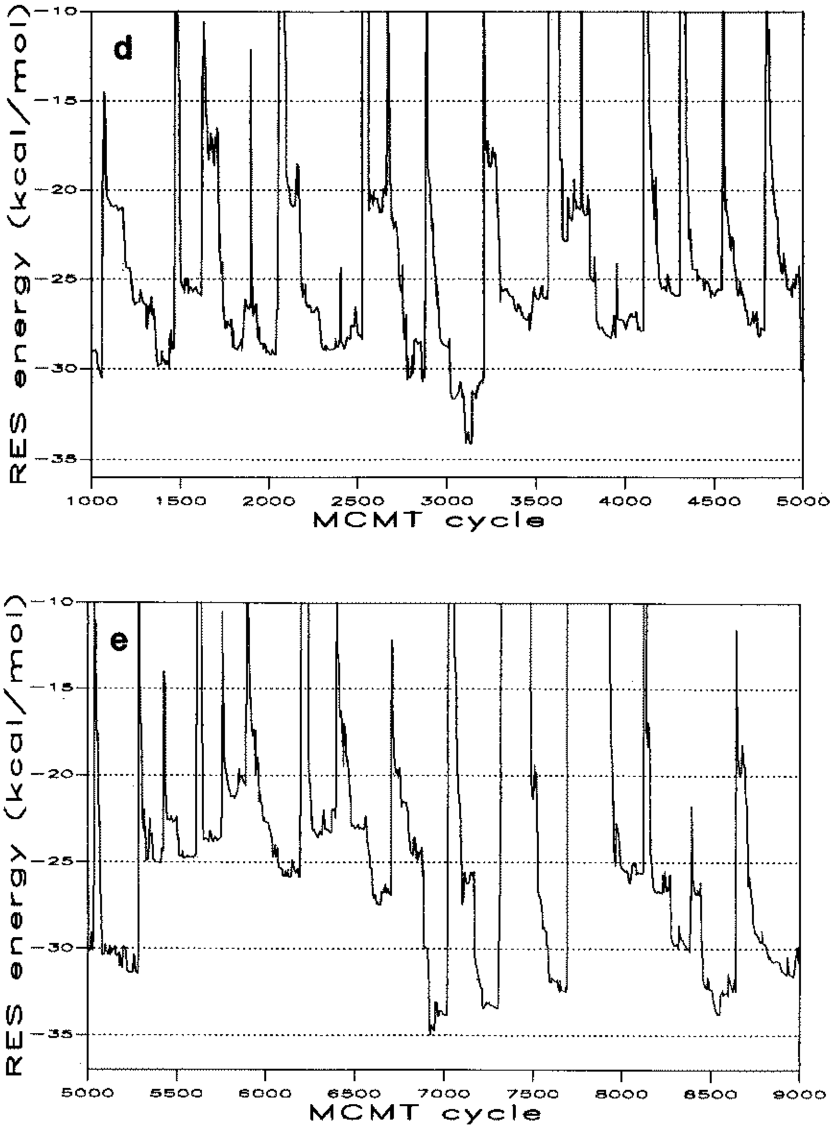

Fig. 1. The simulation interval containing the lowest-energy structure is shown for 5 different runs starting from conformation R5. Only the accepted steps are plotted. (**a**) Progress of the MCM procedure at 310 K. (**b—e**) Progress of the MCMT carried out at thermalization temperatures of 2,000, 10,000, 50,000, and 90,000 K, respectively.

structures. At the lower T_{therm} value only R1 led to a conformation similar (heavy atoms DME within 0.10 Å and backbone atoms DME within 0.01 Å) to the 3–4 II′ apparent global minimum (see further) with a total energy of −36.37 kcal/mol. At T_{therm} of 10,000 K the run starting from the totally extended structure (EX) led to a conformation similar to the 3–4 II′ apparent global minimum, R1 and R4 generated structures similar to the 2–3 I′ low lying local minimum (see further) while R9 yielded a structure similar to the 2–3 I, which is the mirror image of 2–3 I′ (cf. Table I). Furthermore, the AH, R2, and R6 runs led to a minimum energy structure resembling the 3–4 II′ apparent global minimum (all heavy atoms DME within 0.50 Å).

An analysis of the conformations characterized by an energy below −33 kcal/mol showed that the MCMT runs performed at T_{therm} of 2,000 K could not induce a jump between the 2–3 and 3–4 reverse turns associated with a minimum energy conformation. In contrast, the MCMT runs performed at T_{therm} of 10,000 K allowed the conformation to jump between these two disjoined sets of minimum energy conformations 2.5 times on the average during 50,000 cycles, with a maximum of 5 times (run R2) and a minimum of once (run R5).

Two further runs, each consisting of 30,000 MCMT cycles, were started from conformation R5, with a T_{therm} value of 50,000 K respectively 90,000 K to investigate the effect of very high values of the

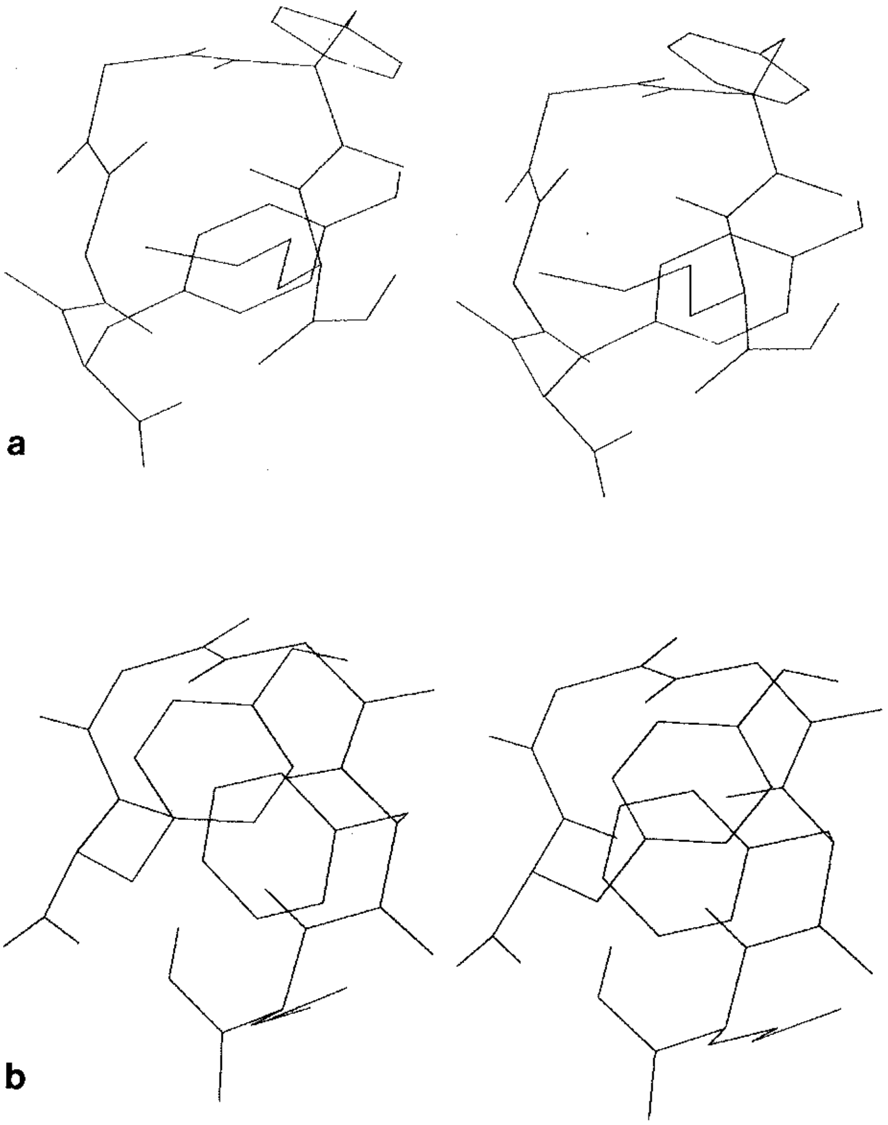

Fig. 2. Stereoscopic view of the heavy atoms and polar hydrogens of the RES[8] minimum-energy structures of Met-enkephalin (H-Tyr-Gly-Gly-Phe-Met-OH) obtained by the MCMT procedure. (**a**) Type II′ reverse turn involving Gly-Gly-Phe-Met. (**b**) Type I′ reverse turn involving Tyr-Gly-Gly-Phe.

thermalization temperature. Both runs could not generate structures with a total energy below −35 kcal/mol. Among the accepted structures of the runs starting from the R5 conformation, the following rates had a total energy below −25 kcal/mol: 99% for MCM, 87, 42, 39, and 37% for MCMT at T_{therm} of 2,000, 10,000, 50,000, and 90,000 K, respectively. Since at very high values of the thermalization temperature high energy conformations were sampled too frequently, we could conclude that a T_{therm} value of 10,000 K represents an optimal choice for the Met-enkephalin structure. With this value of T_{therm} the need for an exhaustive search of conformational space is balanced by a sparse sampling of high energy structures. Figure 1a contains a part of the history of the R5 MCM while Figure 1b–e exhibits part of the MCMT at the four thermalization temperatures discussed above. It should be noted that larger samples of conformations with energy above −25 kcal/mol correspond to higher values of T_{therm}.

Minimum-Energy Structures

The conformation with the lowest energy generated by each of the 13 MCMT runs performed at a T_{therm} value of 10,000 K was subjected to step II consisting of 20,000 MCM cycles (no thermalization, Metropolis criterion applied at 310 K) of random changes involving only side chain dihedrals. Five conformations (lowest energy structures generated by MCMT starting from AH, EX, R2, R3, R6) converged (all heavy atoms DME within 0.01 Å) toward the apparent global minimum of the RES energy

TABLE IIIa. Energy[*] Values in kcal/mol of the RES Minimum-Energy Structures

Turn	Bond	Angle	Torsion	van der Waals	Coulomb	Hydrogen bonds	Total
3–4 II′	0.32	2.75	3.53	−14.94	−11.23	−16.80	−36.38
2–3 I′	0.19	2.13	1.39	−18.60	−8.86	−11.99	−35.75

[*] The RES force field has been described in detail.[8]

TABLE IIIb. Hydrogen Bonds of the RES Apparent Global Minimum, 3–4 II′[*]

Donor	Acceptor
Y1	M5
G2	M5[†]
G3	Y1
F4	G2[†]
M5	G2
M5 OH	F4
Y1 S	F4[‡]

[*] All hydrogen bonds between backbone atoms with exception of the last one where S means side chain. M5 OH means the hydroxyl group of the C terminus.

[†] Hydrogen bond of both the RES and the ECEPP/2 global minimum structure (EC).

[‡] The LC conformation has Y1 S as donor and G3 as acceptor.

TABLE IIIc. Hydrogen Bonds of the RES Low Lying Local Minimum, 2–3 I′[*]

Donor	Acceptor
Y1	M5 OH
F4	Y1
M5	Y1
M5 OH	Y1

[*] All hydrogen bonds between backbone atoms. M5 OH means the hydroxyl group of the C terminus.

function at −36.38 kcal/mol, which constitutes a type II′ reverse turn, the central residues of which are Gly-3-Phe-4 (3–4 II′, see Tables Ia and II). R1 and R4 converged toward a local minimum at −35.75 kcal/mol, which is a type I′ reverse turn, the central residues of which are Gly-2-Gly-3 (2–3 I′, see Tables Ib and II), while R8 and R9 converged towards its mirror image (2–3 I, see Table Ib). Although EC, R5, R0, and R7 did not reach neither the apparent global minimum nor the low lying local minimum, their lowest energy structure was a 3–4 II′, 3–4 II, 2–3 I′, and 2–3 I reverse turn, respectively. The acceptance ratio of step II was about 31%, which is relatively large, because the random perturbations affected only the side chains dihedrals.

In the RES apparent global minimum of Met-enkephalin, a 3–4 II′ turn, the Phe-4 and Met-5 side chains are above the plane formed by the backbone while Tyr-2 is located below it (see Fig. 2a). The local minimum, 2–3 I′, shows a different arrangement of the side chains; Phe-4 lies above while Tyr-1 and Met-5 are positioned below the backbone plane (see Fig. 2b). As a consequence, the RES total energy of 3–4 II′ is lower than the one of 2–3 I′ because of the more favorable Coulomb and hydrogen bonds interactions (cf. Table IIIa). In both minimum-energy structures the largest strain is located in the Tyr-1 atoms, especially in the N-terminus; in the apparent global minimum conformation there is also some strain in the torsion angles defining the position of the hydrogen atom of the hydroxyl group. The latter is compensated by the additional hydrogen bond between the hydroxyl group of the Tyr-1 side chain and the backbone oxygen of Phe-4. Although there is more strain in the apparent global minimum structure than in the 2–3 I′ local minimum (higher values of the bonded interactions, cf. Table IIIa), the former is characterized by a lower total energy because of the more favorable Coulomb and hydrogen bond interactions. Tables IIIb–c list the hydrogen bonds of the minimum-energy conformations, 3–4 II′ and 2–3 I′, respectively; a hydrogen bond was assumed if both its donor–acceptor distance was below 3.1 Å and its donor–hydrogen–acceptor angle was above 135°.

The RES apparent global minimum conformation of Met-enkephalin and the ECEPP/2 one consist both of a 3–4 II′ reverse turn. Their heavy atoms and backbone DME are 0.90 and 0.71 Å, respectively. The main difference between them consists of the Tyr-1 side chain orientation (see Table II and compare Fig. 2a of this paper to Fig. 1 of reference 4); in the former (RES) it points toward Phe-4 (hydrogen bond) and Met-5 while in the latter (ECEPP/2) it is oriented toward Gly-3 (hydrogen bond) and Phe-4. Both backbone hydrogen bonds of the ECEPP/2 global minimum are also present in the RES lowest energy conformation. The latter conformation has some additional hydrogen bonds because of the different N-terminus position allowed by some strain in the bond angles and torsions.

CONCLUSIONS

Monte Carlo minimization supplemented by thermalization applied to the RES energy function of Met-enkephalin showed better optimization performances in cartesian coordinate space than the original MCM procedure. The former method, applied at a thermalization temperature of 10,000 K, allowed a jump between reverse turns with central residues Gly-2-Gly-3 and Gly-3-Phe-4, thereby reaching low energy conformations. This was not accomplished by the original MCM at 310 K even when backbone dihedrals were more frequently sampled than side chain torsion angles.

Ideally, minimization procedures should always reach the global minimum conformation starting from any random structure. The method proposed here led to the same apparent global minimum of the energy function for 5 among 13 different Met-enkephalin starting conformations, while 4 starting structures got trapped in the same low lying local minimum and the remaining 4 in further local minima. Failure to reach the global minimum in all runs has been a consequence of insufficient conformational sampling. New sampling techniques for the stochastic perturbations of the conformation should improve the performance of MCMT.

ACKNOWLEDGMENTS

We thank Drs. Hein Osenberg, Markus Muser, and Thomas Leemann for computer assistance.

REFERENCES

1. Anfinsen, C.B., Haber, E., Sela, M., White, F.H., Jr. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc. Natl. Acad. Sci. U.S.A. 47:1309–1314, 1961.
2. Némethy, G., Scheraga, H.A. Theoretical studies of protein conformation by means of energy computations. FASEB 4:3189–3197, 1990.
3. Scheraga, H.A. Calculations of stable conformations of polypeptides, proteins, and protein complexes. Chem. Scripta 29A:3–13, 1989.
4. Li, Z., Scheraga, H.A. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. Proc. Natl. Acad. Sci. U.S.A. 84:6611–6615, 1987.
5. Li, Z., Scheraga, H.A. Structure and free energy of complex thermodynamic systems. J. Mol. Struct. (Theochem.) 179:333–352, 1988.
6. Momany, F.A., McGuire, R.F., Burgess, A.W., Scheraga, H.A. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. J. Phys. Chem. 79:2361–2381, 1975.
7. Némethy, G., Pottle, M.S., Scheraga, H.A. Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids. J. Phys. Chem. 87:1883–1887, 1983.
8. Caflisch, A. Monte Carlo-docking of polypeptide chains; a general model with applications to oligopeptides bonded to class I MHC proteins. Ph.D. thesis, ETH Zürich, 1991. ETH diss 9400.
9. Caflisch, A., Niederer, P., Anliker, M. Monte carlo-docking of oligopeptides to proteins. Proteins: Struct. Funct. Genet. 13:223–230, 1992.
10. Brooks, BR, Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 4:187–217, 1983.
11. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. Equation of state calculations by fast computing machines. J Chem. Phys. 21:1087–1092, 1953.
12. Purisima, E.O., Scheraga, H.A. An approach to the multiple-minima problem in protein folding by relaxing dimensionality. Tests on enkephalin. J. Mol. Biol. 196:697–709, 1987.