

Prediction of Peptide Conformation by the Multicanonical Algorithm¹

Ulrich H.E. Hansmann² and Yuko Okamoto³

Abstract

We test the effectiveness of the multicanonical algorithm for the tertiary structure prediction of peptides and proteins. As a simple example we study Met-enkephalin. The lowest-energy conformation obtained agrees with that determined by other methods such as Monte Carlo simulated annealing. But unlike to simulated annealing the relationship to the canonical ensemble remains exactly controlled. Thermodynamic quantities at various temperature can be calculated from one run.

A protein or a peptide is a molecule that consists of a chain of N amino acid residues. There are 20 different amino acids known in nature. When N is large one calls the molecule a protein, otherwise a peptide. The prediction of tertiary structures of proteins, which determine their biological function, from their primary sequences remains one of the long-standing unsolved problems (for recent reviews, see, for example, Refs. [1]). It is widely believed that this structure corresponds to the global minimum in the energy. So the problem amounts to finding the global minimum energy out of a huge number of local minima separated by high tunneling barriers. Within the presently available computer resources, the traditional methods such as molecular dynamics and Monte Carlo simulations at relevant temperatures tend to get trapped in local minima. One of the methods which seem to alleviate this multiple-minima problem is simulated annealing.[2] However, a disadvantage of simulated annealing is that there is no established protocol for annealing and a certain number (which is not known *a priori*) of runs are necessary to evaluate the performance. Moreover, the relationship of the obtained conformations to the equilibrium canonical ensemble at a fixed temperature remains unclear.

These problems may be overcome by the multicanonical algorithm which was recently proposed by Berg *et al.*[3] Originally developed to overcome the supercritical slowing down of first-order phase transitions,[4] it has also been tested for systems with conflicting constraints such as spin glasses.[5, 6, 7] The latter systems suffer from a similar multiple-minima problem and it was claimed that the multicanonical algorithm outperforms simulated annealing in these cases.[6]

¹To appear in the *Proceedings of the Sixth Annual Workshop on Recent Developments in Computer Simulation Studies in Condensed Matter Physics, 22–26 Feb. 1993, Athens, Georgia.*

²Department of Physics and Supercomputer Computations Research Institute (SCRI), The Florida State University, Tallahassee, FL 32306, USA.

³Department of Physics, Nara Women's University, Nara 630, Japan.

The idea of this method is based on performing Monte Carlo simulations in a *multicanonical* ensemble[3, 8] instead of the usual (canonical) Gibbs-ensemble. In the canonical ensemble, configurations at an inverse temperature $\hat{\beta} \equiv 1/RT$ are weighted with the Boltzmann factor $\mathcal{P}_B(E) = \exp(-\hat{\beta}E)$. The resulting probability distribution is given by

$$P_B(E) \propto n(E)\mathcal{P}_B(E) , \quad (1)$$

where $n(E)$ is the spectral density. In the *multicanonical* ensemble,[3, 8] on the other hand, the probability distribution is defined in such a way that a configuration with any energy enters with equal probability:

$$P_{mu}(E) \propto n(E)\mathcal{P}_{mu}(E) = \text{const.} \quad (2)$$

Then it follows that the multicanonical weight factor should have the form

$$\mathcal{P}_{mu}(E) \propto n^{-1}(E) . \quad (3)$$

In order to define an explicit form of this weight factor, we introduce two parameters $\alpha(E)$ and $\beta(E)$ as follows:[3]

$$\mathcal{P}_{mu}(E) = \exp\{-(\hat{\beta} + \beta(E))E - \alpha(E)\} . \quad (4)$$

For any fixed $\beta(E)$ and $\alpha(E)$ this leads to the canonical weight factor with the inverse temperature $\beta = \hat{\beta} + \beta(E)$, hence the name ‘‘multicanonical’’. For a numerical simulation one needs estimators for the multicanonical parameters $\beta(E)$ and $\alpha(E)$. The iterative procedure by which one can get such estimators is described elsewhere.[9] Once the multicanonical parameters are determined, one multicanonical run is in principle enough to calculate all thermodynamic quantities by re-weighting.[10] Since in the multicanonical ensemble all energies enter with equal probability a simulation may overcome the barriers between local minima by connecting back to the high temperature states. In this way the global minimum can be explored.

In the present work we apply the multicanonical algorithm to the problem of protein folding, the tertiary structure prediction of peptides and proteins. The purpose of this work is primarily to test the effectiveness of the algorithm. For this reason we have studied one of the simplest peptide, Met-enkephalin. The lowest-energy conformation for the potential energy function ECEPP/2 [11] is known[12] and analyses with Monte Carlo simulated annealing with ECEPP/2 also exist.[13, 14]

Met-enkephalin has the amino-acid sequence Tyr-Gly-Gly-Phe-Met. For our simulations the backbone was terminated by a neutral NH_2^- group at the N-terminus and a neutral $-\text{COOH}$ group at the C-terminus as in the previous works of Met-enkephalin.[12] The potential energy function that we used is given by the sum of the electrostatic term E_{es} , the van der Waals energy E_{vdW} , and hydrogen-bond term E_{hb} for all pairs of atoms in the peptide together with the torsion term E_{tors} for all torsion angles:

$$E_{tot} = E_{es} + E_{vdW} + E_{hb} + E_{tors} \quad (5)$$

$$E_{es} = \sum_{(i,j)} \frac{332q_iq_j}{\epsilon r_{ij}} , \quad (6)$$

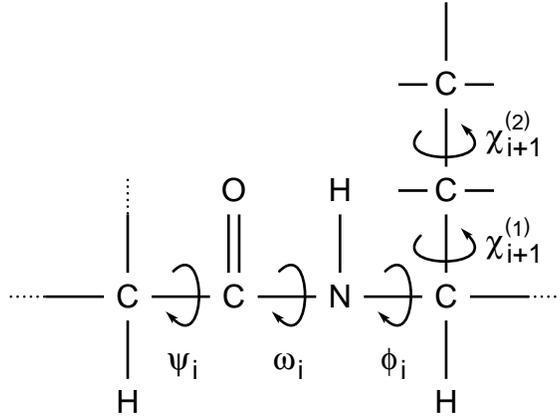


Figure 1: Definition of dihedral angles

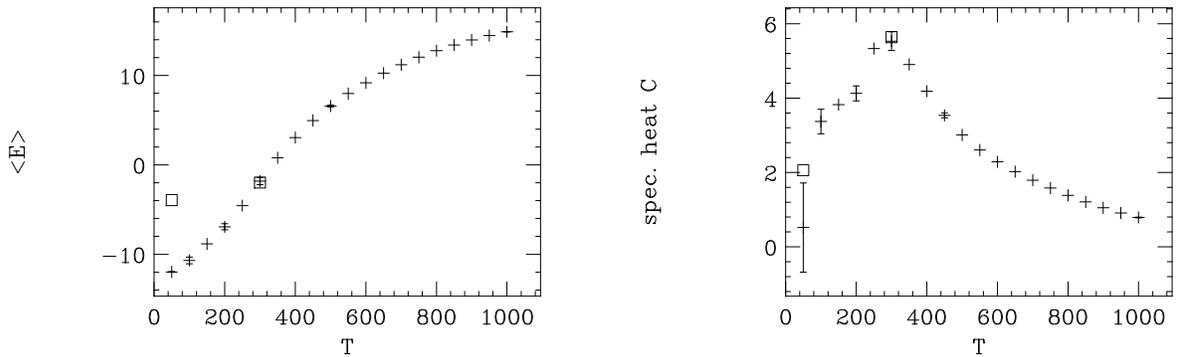


Figure 2: Average energy $\langle E \rangle$ and specific heat C of Met-enkephalin as a function of temperature evaluated by multicanonical algorithms. The results of canonical simulations at fixed temperatures (50 K and 300 K) are also plotted (\square).

$$E_{vdW} = \sum_{(i,j)} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right), \quad (7)$$

$$E_{hb} = \sum_{(i,j)} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right), \quad (8)$$

$$E_{tors} = \sum_l U_l (1 \pm \cos(n_l \alpha_l)). \quad (9)$$

r_{ij} is the distance between the atoms i and j , and α_l is the torsion angle for the chemical bond l . For a definition of these angles which represent the true degrees of freedom see Fig. 1. The parameters ($q_i, A_{ij}, B_{ij}, C_{ij}, D_{ij}, U_l$ and n_l) for the energy function were adopted from ECEPP/2,[11]. The effect of surrounding atoms of water is neglected and the dielectric constant c is set equal to 2. The computer code KONF90,[15] was modified to accommodate the multicanonical method. The peptide-bond dihedral angles ω were fixed at the value 180° for simplicity, which leaves 19 angles ϕ_i, Ψ_i and χ_i as independent variables.

In Fig. 2a we show the average energy, obtained by our method in a run with 10^5

Table 1: Table I. Energy and dihedral angles of the lowest-energy conformations of Met-enkephalin obtained by multicanonical runs. Conformation A is the lowest-energy conformation obtained by Monte Carlo simulated annealing (taken from Ref. 13).

Conformation	A	1	2	3	4	5	6
E [kcal/mol]	-11.9	-11.9	-12.0	-12.0	-12.1	-12.0	-11.9
ϕ_1	98	90	91	90	97	96	98
ψ_1	154	153	152	154	151	153	156
ϕ_2	-161	-160	-157	-161	-158	-161	-163
ψ_2	69	72	64	71	71	68	65
ϕ_3	65	64	66	63	64	64	66
ψ_3	-93	-95	-92	-95	-94	-89	-92
ϕ_4	-85	-82	-80	-77	-83	-85	-80
ψ_4	-27	-26	-29	-32	-30	-31	-29
ϕ_5	-83	-81	-82	-78	-80	-82	-86
ψ_5	142	142	138	137	145	151	147
χ_1^1	-179	179	-177	179	179	-178	-176
χ_1^2	-112	-110	-117	-109	-111	-115	-114
χ_1^3	149	144	146	143	149	145	142
χ_4^1	180	-176	178	177	180	-178	180
χ_4^2	73	79	81	86	79	78	78
χ_5^1	-65	-64	-67	-67	-66	-67	-66
χ_5^2	180	-179	180	180	-176	180	176
χ_5^3	179	178	179	-179	-179	-178	-178
χ_5^4	-55	-66	-59	-62	-61	-60	-57

sweeps, as a function of temperature. The value ≈ -12 kcal/mol at $T = 50$ K is very close to the global-minimum energy obtained by other methods.[12, 13, 14] In Fig. 2b we likewise present the “specific heat” (per residue), which is defined by

$$C = \beta^2 \frac{\langle E^2 \rangle - \langle E \rangle^2}{5} . \quad (10)$$

It has a peak around $T = 300$ K, which indicates that this temperature is important for peptide folding.

During the production run the system reached the global-energy minimum region six times. The lowest-energy conformation within each visit is listed in Table I together with the global-minimum energy conformation (Conformation A in Table I) obtained by simulated annealing.[13] Conformations 1–6 are the results at Monte Carlo steps 20128, 39521, 44462, 65412, 89413, and 95143. Hence, the system reached the lowest-energy region in every 5000 to 20000 Monte Carlo steps. The energies are almost all equal, and the lowest-energy value in the present work (-12.1 kcal/mol) is slightly less than the previous result (-11.9 kcal/mol) by simulated annealing.[13] Most of the dihedral angles of the six conformations also agree with the corresponding ones of Conformation A within $\approx 5^\circ$.

We have applied the recently developed multicanonical algorithm to the problem of predicting the peptide conformation. This method avoids getting trapped in a local min-

imum of energy function by connecting back to high temperature states and enhances in this way the probability to find the global minimum. We have demonstrated the effectiveness of the algorithm by reproducing the lowest-energy conformation of Met-enkephalin. Furthermore, the multicanonical algorithm can yield various thermodynamic quantities as a function of temperature from only one production run.

Acknowledgements:

Our simulations were performed on the SCRI cluster of fast RISC workstations. This work is supported, in part, by the Department of Energy, contract DE-AC03-76SF00515, DE-FG05-87ER40319, DE-FC05-85ER250000 and by the Deutsche Forschungsgemeinschaft under contract H180411-1. Y.O likes to thank SCRI for the kind hospitality extended to him during a visit.

References

- [1] M. Levitt, *Curr. Opin. Struct. Biol.*, **1**, 224 (1991).
- [2] S. Kirkpatrick, C.D. Gelatt, Jr., and M.P. Vecchi, *Science*, **220**, 671 (1983).
- [3] B.A. Berg and T. Neuhaus, *Phys. Lett.*, **B267**, 249 (1991).
- [4] B.A. Berg and T. Neuhaus, *Phys. Rev. Lett.*, **68**, 9 (1992); B. Berg, U. Hansmann and T. Neuhaus, *Phys. Rev.* **B 47** (1993) 497; B. Berg, U. Hansmann and T. Neuhaus, *Z. Phys.* **B 90** (1993) 229; W. Janke, B. Berg and M. Katoot, *Nucl. Phys.*, **B382** 649 (1992);
- [5] B. Berg and T. Celik, *Phys. Rev. Lett.*, **69** 2292 (1992).
- [6] B. Berg and T. Celik, *Int. J. Mod. Phys.* **C 3** (1992) 1251.
- [7] B. Berg, T. Celik and U. Hansmann, to appear in *Europhysics Letters*.
- [8] G.M. Torrie and J.P. Valleu, *J. Comp. Phys.*, **23**, 187 (1977).
- [9] U. Hansmann and Y. Okamoto, FSU-SCRI-93-12, submitted to *J. Comp. Chem.*
- [10] A.M. Ferrenberg and R.H. Swendsen, *Phys. Rev. Lett.*, **61**, 2635 (1988); **63**, 1658(E) (1989), and references given in the erratum.
- [11] F.A. Momany, R.F. McGuire, A.W. Burgess, and H.A. Scheraga, *J. Phys. Chem.*, **79**, 2361 (1975); G. Némethy, M.S. Pottle, and H.A. Scheraga, *J. Phys. Chem.*, **87**, 1883 (1983); M.J. Sippl, G. Némethy, and H.A. Scheraga, *J. Phys. Chem.*, **88**, 6231 (1984).
- [12] Z. Li and H.A. Scheraga, *Proc. Natl. Aca. Sci., U.S.A.*, **84**, 6611 (1987); A. Nayeem, J. Vila, and H.A. Scheraga, *J. Comp. Chem.*, **12**, 594 (1991).
- [13] Y. Okamoto, T. Kikuchi, and H. Kawai, *Chem. Lett.*, **1992**, 1275.
- [14] B. von Freyberg and W. Braun, *J. Comp. Chem.*, **12**, 1065 (1991).
- [15] H. Kawai, Y. Okamoto, M. Fukugita, T. Nakazawa, and T. Kikuchi, *Chem. Lett.*, **1991**, 213; Y. Okamoto, M. Fukugita, T. Nakazawa, and H. Kawai, *Protein Eng.*, **4**, 639 (1991).