Eur. Phys. J. B **30**, 577–580 (2002)
DOI: 10.1140/epjb/e2002-00416-2

THE EUROPEAN PHYSICAL JOURNAL B

Comparison of the ELP and multicanonical methods in simulation of the heptapeptide deltorphin

H. Arkın and T. Çelik[a]

Department of Physics, Hacettepe University, Beytepe 06532, Ankara, Turkey

Received 23 July 2002 / Received in final form 18 September 2002
Published online 31 December 2002 –

Abstract. To investigate the performance of the energy landscape paving (ELP) procedure for peptides, we apply it here to deltorphin, a linear heptapeptide with bulky side chains (H-Tyr[1]-D-Met[2]-Phe[3]-His[4]-Leu[5]-Met[6]-Asp[7]-NH_2) and compare the results with the Multicanonical method (MUCA) in regard of finding the low-energy structures. Deltorphin is modeled in vacuum by the potential energy function ECEPP.

PACS. 02.70.-c Computational techniques – 05.50.+q Lattice theory and statistics (Ising, Potts, etc.) – 82.20.Wt Computational modeling; simulation

The conformation space of proteins and peptides presents a complex energy profile consisting of a tremendous number of local minima separated by energy barriers. An ideal simulation scheme should freely visit the entire space and predominantly sample the significant conformations and also the transition states without generating many unimportant conformations. Because of energy barriers, conventional simulations in the canonical ensemble are of little use, they tend to get trapped in states of these energy local minima and the results thus will depend strongly on the initial conditions. One way to overcome this problem is to perform simulation in a generalized-ensemble where each state is weighted by non-Boltzman probability weight factor, so that a flat histogram in potential energy space may be realized. This allows the simulation to escape from any energy barrier and to sample much wider phase space than by conventional methods.

One of the most well-known, powerful generalized-ensemble methods is the Multicanonical algorithm (MUCA) [1]. But, in multicanonical procedure, the probability weight factors are not a priori known and have to be determined by iterations of trial simulations. This part of simulation can be non-trivial and very tedious for complex systems with many local minimum energy states. The other generalized-ensemble approaches, for example simulated tempering, $1/k$ sampling, replica-exchange methods are reviewed in references [2] and [3]. Application of the multicanonical approach to peptides was pioneered by Hansmann and Okamoto [4] and followed by others [5].

A considerably less attention has been paid to the fact that the system with Boltzmann statistics repeatedly visits the catchment basins that have previously been visited. If the goal is surveying the potential energy landscape, such events should be controlled because many of them bear no new information about the conformational space.

Recently, Wille and Hansmann [6] introduced a new global optimization method, energy landscape paving (ELP), which is designed to deform the energy surface to escape local minima as well as to direct the search towards the unexplored regions. ELP samples the significant local minima and the transition states without generating too many unimportant conformations. We would like to remind the reader that MUCA and other generalized-ensemble techniques provides one to obtain thermodynamic averages and fluctuations at different temperatures. However such feature is missing in ELP runs because it is essentially a Monte Carlo run at a pre-choosen low-temperature. Our aim in this work is to utilize the ELP method to find the global energy minimum (GEM) and the stable microstates pertaining to it and compare its efficiency to MUCA, rather than a complete thermodynamical investigation. Recently, an extensive comparison of ELP with simulated annealing has been made by Hsu *et al.* [7].

The central feature of ELP is to perform Monte Carlo (MC) simulation with a modified energy expression which enables to keep the search away from the already explored regions.

The weight for a state is taken as

$$w(\tilde{E}) = \mathrm{e}^{-\tilde{E}/k_B T}, \tag{1}$$

where T denotes temperature and $\tilde{E}$ is the following replacement of the energy E:

$$\tilde{E} = E + f(H(q,t)) \tag{2}$$

where $f(H(q,t))$ is a function of the histogram $H(q,t)$ in a chosen "order parameter" q. In order to test the

[a] e-mail: tcelik@hacettepe.edu.tr

efficiency of ELP, we adopted the simplest case and used the potential energy itself as an order parameter and the weight is generated by $\tilde{E} = E + H(E,t)$ where $H(E,t)$ is the histogram in energy. The histogram is updated at each MC step, hence the "time" dependence of $H(E,t)$, and normalized over the number of sweeps.

In a canonical simulation, the probability to escape a local minima depends on the height of energy barriers. On the other hand, within ELP the probability to escape a local minimum increases with the increase in the histogram entries of that minimum, which in turn reflects the time the system stays in that minimum. During the simulation time, ELP smooths out the energy landscape locally in such way that the local minimum is no longer favored.

In this paper, we take an interest in the efficiencies of ELP and MUCA methods as applied to molecules of increasing size and examine the performance of the procedures in studying the low energy conformations. A step in this direction is carried out in this paper where ELP is applied and compared with MUCA to linear heptapeptide with bulky side chains, deltorphin (also known as dermenkephalin) (H-Tyr^1-D-Met^2-Phe^3-His^4-Leu^5-Met^6-Asp^7-NH_2). We have recently carried out a detailed study of deltorphin by multicanonical simulation [8]. This natural peptide found in frog skin, has high potency and receptor selectivity for δ opioid receptors. To understand the conformation-activity relationships, NMR studies of the solution structures of deltorphin in DMSO and cryoprotective solvents were carried out [9] and computational work based on these experiments was carried out as well [10].

Deltorphin is modeled by the ECEPP/2 potential [11], which assumes a rigid geometry, and is based on non-bonded, Lennard-Jones, torsional, hydrogen-bond, and electrostatic potential terms with the dielectric constant $\epsilon = 2$. This potential energy is implemented into the software package FANTOM [12]. We further fix peptide bond angles ω to their common value 180°, which leaves us with 36 dihedral angles as independent degrees of freedom ($n_F = 36$). At each update step, a trial conformation was obtained by changing one dihedral angle at random within the range [−180°; 180°], followed by the Metropolis test and an update of the suitable histogram. The dihedral angles were always visited in a predefined (sequential) order, going from Tyr to Asp; a cycle of N MC steps is called a sweep.

As pointed out in the Introduction, the characteristic behavior of ELP methods is shown in Figure 1 which is the time series of 5×10^5 sweeps for the ELP simulation of deltorphin at $T = 50$ K. For comparison, the standard Monte Carlo simulation at $T = 50$ K shows a time series confined to rather narrow range of energy -33 kcal/mol $\leq E \leq -28$ kcal/mol. The ELP time series has the typical time-dependent feature of continuously extending the covered range of energy. After long enough time elapsed, the time series becomes like the one achieved by multicanonical simulation. Another important feature one sees from the time series is that the simulation gets trapped in a local minima in the energy landscape,

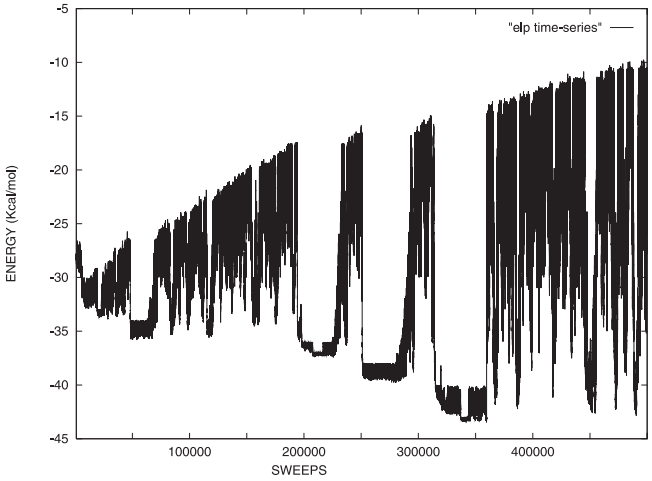

Fig. 1. Time series of the energy landscape paving simulation of deltorphin. For comparison, the standard Monte Carlo simulation at $T = 50$ K is performed and a time series fluctuating within a rather narrow range of energy -33 kcal/mol $\leq E \leq -28$ kcal/mol is obtained. MC time series is not plotted, otherwise the figure becomes confusing unless presented in color.

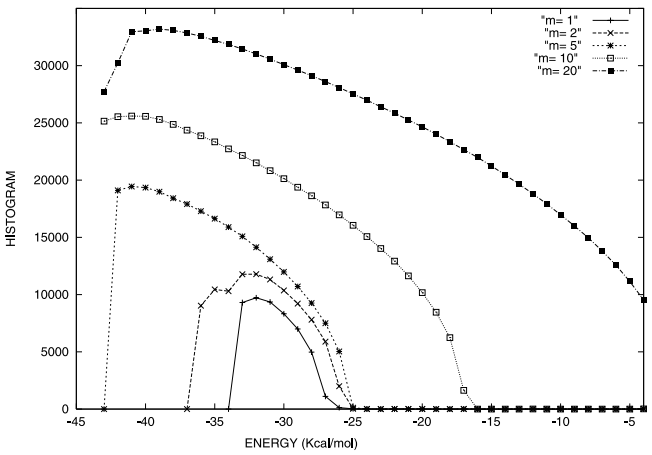

Fig. 2. The histogram of the energy landscape paving simulation at $T = 50$ K. The histograms are obtained following the mth step of simulation, where each step involves 50 000 sweeps. At the step $m = 10$, the GEM has been visited about 25 000 times while the simulation up to then have failed to find the GEM.

spends some time there to built histogram, then escapes to search other regions. But when the search hits the same pre-visited minima, it does not get trapped, almost immediately leaves and freely searches till gets trapped in another basin with lower energy. In a stepwise fashion, the search tries to reach the global minima and afterwards the stepwise entrapments disappear and the time series looks like the typical time series of multicanonical simulation (*e.g.* after 400 000 sweeps in Fig. 1).

In order to realize the dynamical feature of building the histogram in ELP simulation, we present in Figure 2, the histograms of the entries in each energy bin obtained following the mth step of the simulation, where each step involves 50 000 sweeps. The histogram for $m = 1$ is the

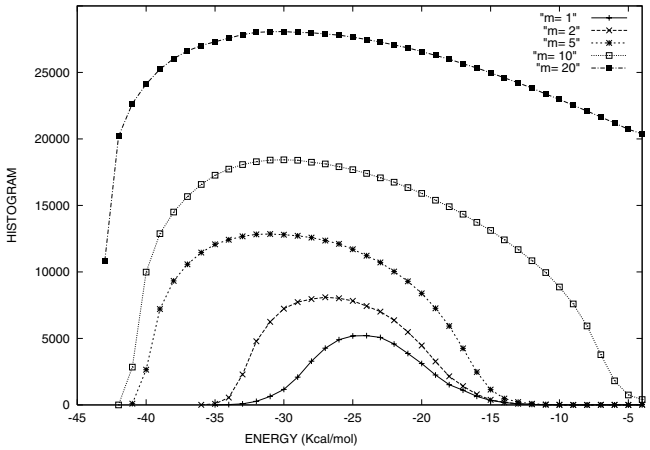

Fig. 3. The histogram same as of Figure 2 obtained by the ELP simulation at $T = 250$ K. Compared to the results of Figure 2, to find the GEM took much longer and at the $m = 20$th step, GEM had been visited about 10 000 times.

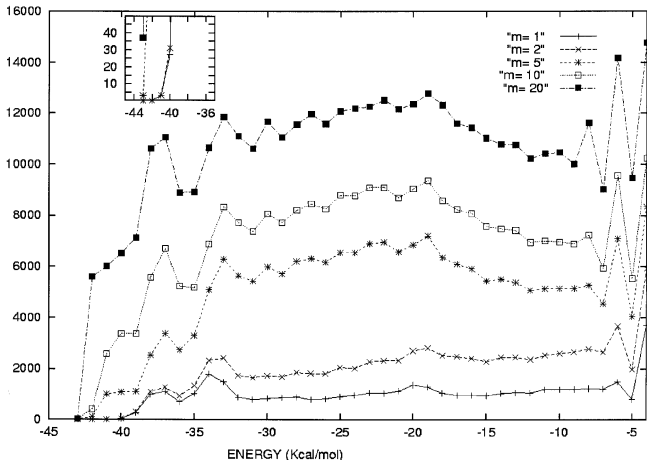

Fig. 4. The histogram of the multicanonical simulation. The multicanonical histogram is typically flat already for $m = 1$ step while in the ELP histograms the low-energy region is enhanced. Compared to Figures 1 and 2, number of visits to the lowest energy bin is much less for the multicanonical method.

typical standard Monte Carlo simulation peaked at the narrow range of -33 kcal/mol $\leq E \leq -28$ kcal/mol. As the time proceeds, the histogram get broader. At the step with $m = 10$, which reads as the number of sweeps 500 000, a fairly broad histogram is achieved with about 25 000 visits to the lowest bin in energy. In Figure 3, we show the histograms obtained in the ELP simulation of deltorphin at $T = 250$ K. The lowest energy bin has not been visited within the first 500 000 sweeps ($m = 10$) and visited about 10 000 times when the iteration time is extended up to one million sweeps ($m = 20$).

Figure 4 displays the energy histograms for the multicanonical simulation of deltorphin. Comparing to the previous two figures, one can follow how the energy histograms are filled at mth step of iterations. Notice that the lowest energy bin has been visited only 37 times within the first one million iterations and has not been sampled for the first 500 000.

For peptides it is not only of interest to obtain thermodynamic averages and fluctuations at different temperatures but also to find the most stable regions in conformational space populated by the molecule, which allows to identify the most stable wide microstates.

In Table 1 we have shown the number of conformations found in energy bins of 1 kcal/mol above the GEM with ELP and MUCA methods. First part of the table (part A) shows the result of the first 5×10^5 steps and the second part (part B) is for total of 10^6 steps for both methods. Because the efficiency of ELP strongly depends on the temperature, we carried out two different ELP simulations each of 10^6 steps at temperatures $T = 50$ K and 250 K. From the Table 1 it is obvious that ELP sampled more low temperature conformations compared to MUCA, except for the second and the third lowest energy bins where the MUCA simulation encountered an entrapment in a wide macrostate. In searching the low-energy conformations, ELP search at $T = 50$ K is clearly more effective then the one at higher temperature.

Next, it is of interest to investigate the conformational coverage provided by these methods, in particular in the low energy region. In a recent work [8], we have investigated the low energy microstates of deltorphin pertaining to GEM by MUCA and the Monte Carlo Minimization (MCM) method developed by Li and Scheraga [13]. While MCM method includes minimization at each update step and the simulated annealing (SA) works as a local optimizer in its final phase, MUCA and ELP do not include any minimization steps. In order to get better insight of these algorithms, it would be fair to make another comparison of the results of simulations after they are subjected to further refinement. In order to classify the microstates according to the potential wells they belong around thermodynamically stable different structures, each conformation of our simulation data was subjected to energy minimization. Following the methods proposed by Meirovitch *et al.* [14], we have adopted a variance criterion whereby two structures are considered different if at least two corresponding dihedral angles differ by 2° or more. We minimized by Newton-Raphson method the energy conformations generated in 10^6 sweeps of MUCA and ELP runs. The lowest energy conformation (our suspected GEM) is [8]

$$E = -44.1058\,\text{kcal/mol}. \tag{3}$$

The number of *different* structures found in energy bins of 1 kcal/mol above $E = -44.11$ kcal/mol appear in Table 2. We compare the conformational coverage of the low energy region obtained with MUCA and two different ELP simulations at $T = 50$ K and $T = 250$ K. The number of different structures lying within the lowest energy bin obtained after the minimization is greater for MUCA, while ELP samples more structures at all energy ranges. This is important due to the fact that MUCA covers a large range of energies in an approximately homogeneous way, while with ELP a strong preference is given for simulating the

Table 1. Number of conformations in energy bins of 1.0 kcal/mol above $E = -44.11$ kcal/mol as obtained by the MUCA and the ELP Methods. Part A shows the result of the first 5×10^5 sweeps and part B is for total of 10^6 sweeps.

	Energy(Kcal/mol)	MUCA	ELP $T = 50$ K	ELP $T = 250$ K
A	−44.11 to −43.11	1 207	10 770	-
	−43.11 to −42.11	20 116	13 825	3
	−42.11 to −41.11	25 225	14 023	2 845
	−41.11 to −40.11	7 506	13 863	99 86
	−40.11 to -39.11	3 838	20 362	12 880
	−39.11 to −38.11	4 729	20 449	14 503
B	−44.11 to −43.11	2 674	27 729	10 843
	−43.11 to −42.11	45 350	30 238	20 189
	−42.11 to −41.11	46 091	32 962	22 656
	−41.11 to −40.11	12 188	33 044	24 125
	−40.11 to −39.11	8 485	33 192	25 245
	−39.11 to −38.11	10 983	33 081	26 039

Table 2. Number of *significantly different* energy minimized structures in energy bins of 1 kcal/mol above $E = -44.11$ kcal/mol as obtained by the MUCA and the ELP methods. While Table 1 compares the number of conformations obtained in MUCA and ELP simulations, here we present the number of different conformations which are classified according to the criterion mentioned in the text, obtained following the further minimization of both data. The results of only 10^5 sweeps are presented in the table.

Energy (Kcal/mol)	MUCA	ELP $T = 50$ K	ELP $T = 250$ K
−44.11 to −43.11	1 566	764	1 171
−43.11 to −42.11	2 438	1 500	2 990
−42.11 to −41.11	2 670	2 983	4 306
−41.11 to −40.11	3 188	3 933	5 178
−40.11 to −39.11	2 855	4 638	6 139
−39.11 to −38.11	2 454	5 627	6 338

low energy region. Here we would like to remind the reader that in MUCA simulation, one needs to carry out first a tedious work to build the multicanonical parameters by at least one million sweeps and then the production run is performed, while with ELP one reaches to an even better statistics in studying the low-energy structures by shorter simulation with a much simpler form of weight function. Therefore ELP is simpler to perform and saves computer time for reaching the global energy minimum and also the significant low energy conformations near the global energy minimum.

The objective of this work has been to investigate the performance of the MUCA and ELP methods for peptides in regard of investigating the most stable microstates pertaining to the global energy minimum. Here, we simulated the heptapeptide deltorphin with bulky side chains. A very good coverage of the lowest energy bins is provided by the two methods. However, extensively long computer time is needed in MUCA simulation and the probability weight factors are not *a priori* known and have to be determined by iterations of trial simulations, while ELP simulation is much simpler to implant and more effective in sampling the lowest energy region of the conformational space.

References

1. B.A. Berg, Fields Insti. Commun. **26**, 1 (2000)
2. U.H.E. Hansmann, Y. Okamoto, Ann. Rev. Comp. Physics **5**, 129 (1999)
3. A. Mitsutake, Y. Sugita, Y. Okamoto, Biopolymers (Peptide Science) **60**, 96 (2001)
4. U.H.E. Hansmann, Y. Okamoto, J. Comput. Chem. **14**, 1333 (1993)
5. M.-H. Hao, H.A. Scheraga, J. Phys. Chem. **98**, 4940 (1994); J. Phys. Chem. **98**, 9882 (1994); A. Kolinski, W. Galazka, Skolnick, J. Proteins **26**, 271 (1996); J. Higo, N. Nakajima, H. Shirai, A. Kidera, H. Nakamura, J. Comput. Chem. **18**, 2086 (1997)
6. U.H.E. Hansmann, L.T. Wille, Phys. Rev. Lett. **88**, 068105 (2002)
7. H.P. Hsu, S.C. Lin, U.H.E. Hansmann, Acta Cryst. A **58**, 259 (2002)
8. F. Yaşar, H. Arkın, T. Çelik, B.A. Berg, H. Meirovitch, J. Comput. Chem. **23** (12) 1127 (2002)
9. (a) P.A. Temussi, D. Picone, T. Tancredi, R. Tomatis, S. Salvadori, M. Marastoni, G. Balboni, FEBS Lett **247**, 283 (1989). (b) T. Tancredi, P.A. Temussi, D. Picone, P. Amodeo, R. Tomatis, S. Salvadori, M. Marastoni, V. Santagada, G. Balboni, Biopolymers **31**, 751 (1991)
10. (a) G.V. Nikiforovich, V.J. Hruby, J. Biochim. Biophys. Res. Commun. **173**, 521 (1990); (b) G.V. Nikiforovich, V.J. Hruby, O. Prakash, C.A. Gehrig, Biopolymers **31**, 941 (1991). (c) G.V. Nikiforovich, O. Prakash, C.A. Gehrig, V.J. Hruby, J. Am. Chem. Soc. **115**, 3399 (1993)
11. F.A. Momany, R.F. McGuire, A.W. Burgess, H.A. Scheraga, J. Phys. Chem. **79**, 2361 (1975); M.J. Sippl, G. Némethy, H.A. Scheraga, J. Phys. Chem. **88**, 6231 (1984)
12. (a) B. von Freyberg, T. Schaumann, W. Braun, *FANTOM User's Manual and Instructions* (ETH Zürich, Zürich, 1993). (b) B. von Freyberg, W. Braun, J. Comput. Chem. **14**, 510 (1993)
13. (a) Z. Li, H.A. Scheraga, Proc. natl. Acad. Sci. USA **84**, 6611 (1987); (b) J. Mol. Struct. (Thechem.) **179**, 333 (1988)
14. (a) H. Meirovitch, E. Meirovitch, J. Lee, J. Phys. Chem. **99**, 4847 (1995); (b) H. Meirovitch, E. Meirovitch, Biopolymers **38**, 69 (1996)