Fizika Nizkikh Temperatur, 2007, v. 33, No. 9, p. 1052–1054

Effective conformational search methods for biological macromolecules

T. Çelik

Turkish Academy of Sciences, Piyade Sokak No:27, Çankaya, Ankara 06550, Turkey
E-mail: tcelik@tuba.gov.tr

Received November, 24 2006

The configuration space of many complex physical systems presents a rough energy landscape consisting of tremendous number of local minima separated by high energy barriers. One way to overcome these barriers is to perform the simulation in a generalized-ensemble where each state is weighted by a non-Boltzmann probability weight factor. Multicanonical Ensemble Approach overcomes this difficulty by performing a random walk in one-dimensional energy space. Our attempts to design hybrid generalized ensemble algorithms will be reported. The folding of a protein into its native structure involves one or more transitions between distinct phases. The representation of the energy landscape would be useful for the determination of the conformational transition temperatures. Such a study would lead to clear indications of the equilibrium conformations of proteins and provide a detailed picture of the folding pathway. The topographic structure of energy landscape of short peptides will be presented.

PACS: 02.70.Uu Applications of Monte Carlo methods;
82.20.Wt Computational modeling; simulation.

Keywords: Generalized ensembles, protein folding.

The problem of protein folding entails the study of a non-trivial dynamics along pathways embedded in a rugged energy landscape. The conventional simulation methods are not effective because the system becomes trapped for long simulation time in a potential well. The trapping problem of the Monte Carlo and Molecular Dynamics methods can be alleviated to a large extent by the Multicanonical MC method (MUCA) [1,2], which was applied initially to lattice spin models and its relevance for complex systems was first noticed in [1]. Applications of the multicanonical approach to peptides was pioneered by Hansmann and Okamoto [3] and followed by others [4]; simulations of protein folding with MUCA and related generalized ensemble methods are reviewed in Ref. 5.

The multicanonical ensemble based on a probability function in which the different energies are equally probable.

$$P^{MU}(E) \approx n(E)w(E) = \text{const}\,.$$

However, implementation of MUCA is not straightforward because the density of states $n(E)$ is unknown. The weights $w(E) \approx 1/n(E)$ are calculated in the first stage of simulation process. First, energy range is divided into $L - 2$ equal segments i defined by $[E_{i-1}, E_i]$ where;

$$E_i - E_{i-1} = \Delta E_i > 0 \text{ for } i = 2, ..., L\text{–}1 \text{ with}$$
$$E_0 = -\infty \quad \text{and } E_L = +\infty\,.$$

The definition of the parametrization of the Muca weights read:

$$w_i(x) = \exp(-b_i E_x + a_i) \quad \text{for} \quad E_{i-1} < E_x \le E_i$$

where a_i and b_i are multicanonical parameters which are calculated by an iterative procedure. The iterative procedure is followed by a long production run based on the fixed w's where equilibrium configurations are sampled. Re-weighting techniques (see Ferrenberg and Swendsen and literature given in their second reference) enable one to obtain Boltzmann averages of various thermodynamic properties over a large range of temperatures [6].

By setting up a one-dimensional random walk in energy space, the multicanonical simulation provides sampling of all available energies and enables one to study the thermodynamical aspects of the system at a wide range of temperature from a single production run. In Fig. 1, time series of the multicanonical algorithm was shown for a

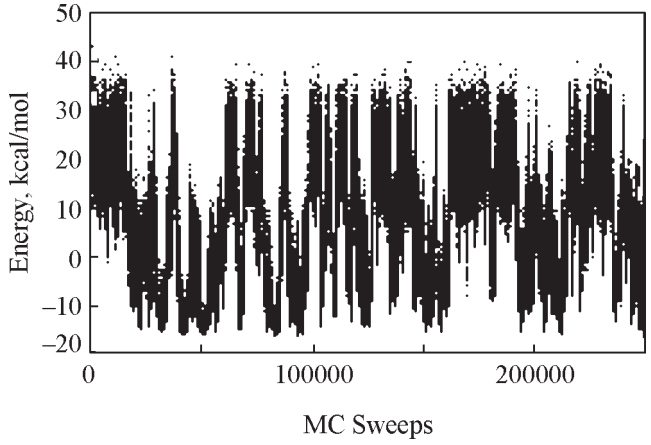

Fig. 1. Monte Carlo time series of the multicanonical simulation for a 12-residue peptide chain.

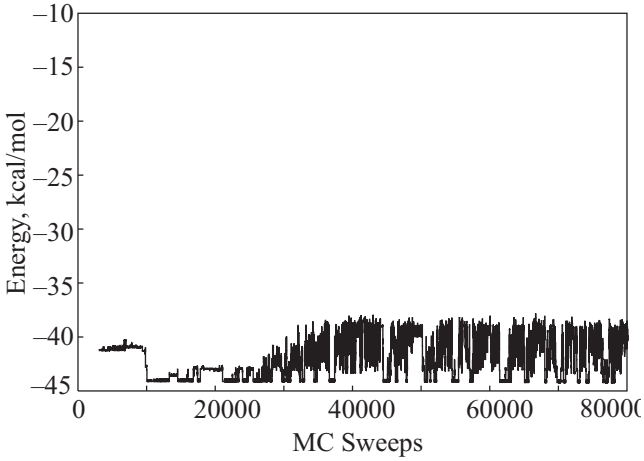

Fig. 3. Monte Carlo time series of the hybrid algorithm.

small peptide. Whole energy range of the system is freely sampled as a characteristic property of this technique.

Energy Landscape Paving (ELP) algorithm combines ideas from tabu search and energy landscape deformation approaches and configurations are searched with the time-dependent weights [7]

$$w(E,q,t) = \exp\left[-(E + f(H(q,T)))/k_B T\right].$$

The low temperature T leads to drive towards low energies. The function f enables to drive simulation out of local minima. Function $f(H(q,T))$ is chosen generally as:

$$f(H(q,t)) = H(q,t)$$

and simulation produces a histogram distribution shown in Fig. 2.

We proposed a hybrid algorithm, which combines the features of energy landscape paving and the Monte Carlo Minimization methods [8]. Namely, we have implanted a Monte Carlo minimization (MCM) step in between the two updates of dihedral angles in the ELP algorithm. This algorithm has a fast convergence property to global minimum state of the system and samples freely the low energy states. Time series of this algorithm was shown in Fig. 3.

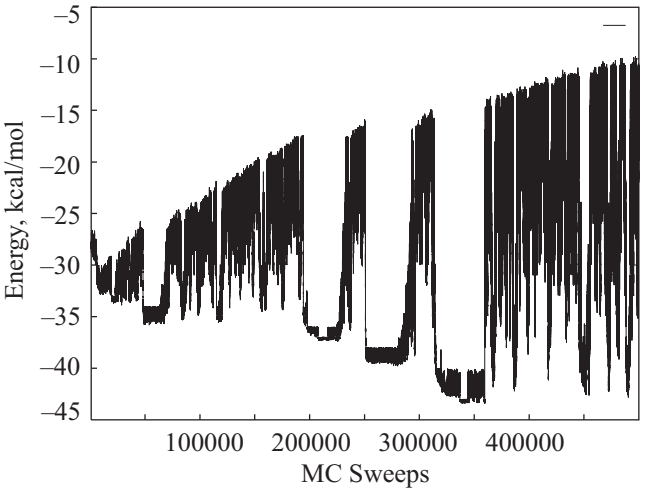

Fig. 2. Monte Carlo time series of a typical energy landscape paving simulation.

Proteins are polymers of amino acid molecules. The complexity of the conformation space of proteins and peptides makes the protein folding problem so difficult. The main goal of the protein folding studies is to determine the 3D global minimum energy stable conformation.

In calculation procedure of the energy of a peptide chain, a chosen force field is used. One of the most commonly used all atom force field is called Empirical Conformational Energies of Proteins and Polypeptides (ECEPP) which is composed of following energy terms:

$$E_C = \sum_{\langle ij \rangle} \frac{332 q_i q_j}{\varepsilon r_{ij}},$$

$$E_{LJ} = \sum_{\langle ij \rangle} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right),$$

$$E_{HB} = \sum_{\langle ij \rangle} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right),$$

$$E_{\text{tor}} = \sum_{\langle ij \rangle} U_l (1 + \cos(n_l \chi_l)).$$

Here r_{ij} is the distance between the atoms i and j, χ is the torsion angle for the chemical bond l. The parameters are adopted from ECEPP/2 provided by Brookhaven PDB databank. The backbone torsion angles φ, Ψ and ω and the side chain angles χ are the degrees of freedoms of our simulations.

In Fig. 4, 3D conformation of elastin pentapeptide VPGVG sequence obtained from x-ray experimental data and the global minimum obtained by our hybrid algorithm was shown [8].

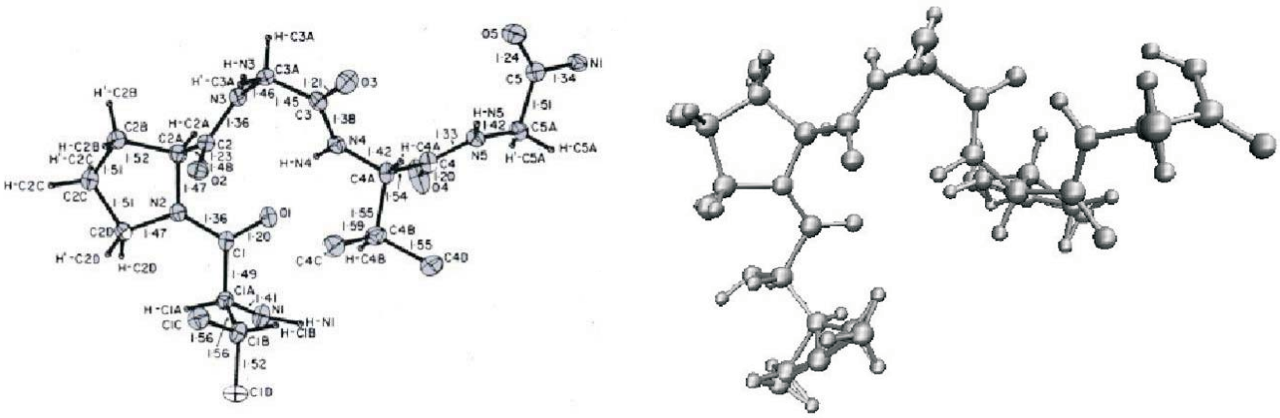

Fig. 4. 3D conformation of elastin sequnce VPGVG obtained from experimental data (left) and the global minimum (right) of our hybrid algorithm.

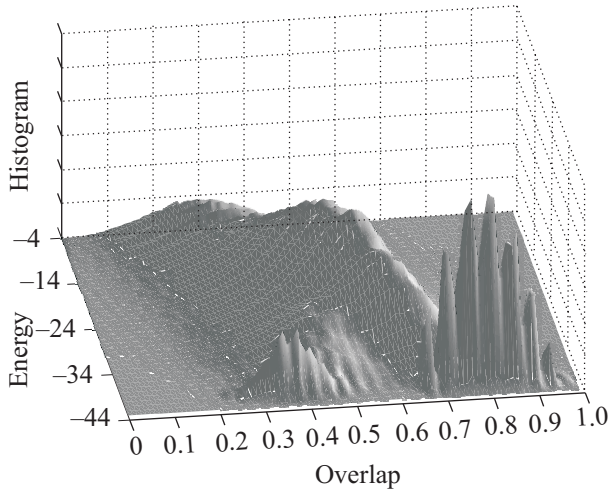

Fig. 5. 3D energy landscape of elastin sequnce vs energy and overlap parameter.

In polypeptide systems, an order parameter can be defined as [9];

$$OP = 1 - \frac{1}{90 n_F} \sum_{i=1}^{n_F} \left| \alpha_i^{(t)} - \alpha_i^{(RS)} \right|$$

where $\alpha_i^{(RS)}$ is the dihedral angles of the reference conformation (usually GEM) and $\alpha_i^{(RS)}$ is the dihedral angles of the conformation under consideration. This parameter measures how any sampled conformation is close to GEM state. In Fig. 5, energy landscape of elastin sequnce vs order parameter and energy was shown. This figure shows the number of sampled conformations in the entire energy range of the molecule as a function of the suitable defined order parameter. The representation of energy landscape entails one to study critical behavior of the complex system.

To summarize, the multicanonical ensemble enables one to display the distribution of all conformations in configuration space at all temperatures from a single simulation.

1. B.A. Berg and T. Çelik, *Phys. Rev. Lett.* **69**, 2292 (1992).
2. B.A. Berg, *Fields Institute Communications* **28**, 1 (1992).
3. U.H.E. Hansmann and Y. Okamoto, *J. Comp. Chem.* **14**, 1333 (1993).
4. M.H. Hao and H.A. Scheraga, *J. Phys. Chem*. **98**, 4940 (1994).
5. U.H.E. Hansmann and Y. Okamoto, *Ann. Rev. Comp. Phys.* **5**, 129 (1999).
6. A.M. Ferrenberg and R.H. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988); *ibid.* **63**, 1658 (1989).
7. U.H.E. Hansmann and L. Wille, *Phys. Rev. Lett.* **88**, 068105 (2002).
8. H. Arkin and T. Çelik, *Eur. Phys. J.* **B30**, 577 (2002).
9. U.H.E. Hansmann, Y. Okamoto, and J. Onuchic, *Proteins* **34**, 472 (1999).