

## Metropolis Importance Sampling for Rugged Dynamical Variables

Bernd A. Berg\*

*Department of Physics, Florida State University, Tallahassee, Florida 32306, USA*

*School of Computational Science and Information Technology, Florida State University, Tallahassee, Florida 32306, USA*

(Received 19 September 2002; published 8 May 2003)

A funnel transformation is introduced, which acts recursively from higher towards lower temperatures. It biases the *a priori* probabilities of a canonical or generalized ensemble Metropolis simulation, so that they zoom in on the global energy minimum, if a funnel exists indeed. A first, crude approximation to the full transformation, called rugged Metropolis one (RM<sub>1</sub>), is tested for Met-Enkephalin. At 300 K the computational gain is a factor of 2 and, due to its simplicity, RM<sub>1</sub> is well suited to replace the conventional Metropolis updating for these kinds of systems.

DOI: 10.1103/PhysRevLett.90.180601

PACS numbers: 05.10.Ln, 87.15.-v, 87.14.Ee

To explain important aspects of protein folding, Bryngelson and Wolynes introduced a funnel picture [1], which is supported by various numerical results [2–4]. Nevertheless, the understanding as well as the practical relevance of the funnel concept has remained somewhat limited. The reason is that the funnel lives in the high-dimensional configuration space, while numerical studies have been confined to projections onto so-called reaction coordinates, and there is no generic definition of a good reaction coordinate. Here I follow a different path and introduce a general funnel description from higher towards lower temperatures. It yields a powerful new method for designing Metropolis [5] weights.

In protein models the energy  $E$  is a function of a number of dynamical variables  $v_i$ ,  $i = 1, \dots, n$  whose fluctuations in the Gibbs canonical ensemble are described by a probability density (pd)  $\rho(v_1, \dots, v_n; T)$ , where  $T$  is the temperature. To be definite, we use in the following the all-atom energy function [6] ECEPP/2 (empirical conformational energy program for peptides). Our dynamical variables  $v_i$  are the dihedral angles, each chosen to be in the range  $-\pi \leq v_i < \pi$ , and the volume of the configuration space is  $K = (2\pi)^n$ .

Let us define the *support* of a pd of the dihedral angles. Loosely speaking, the support of a pd is the region of configuration space where the protein wants to be. Mathematically, we define  $K^p$  to be the smallest sub-volume of the configuration space for which

$$p = \int_{K^p} \prod_{i=1}^n dv_i \rho(v_1, \dots, v_n; T) \quad (1)$$

holds. Here  $0 < p < 1$  is a probability, which ought to be chosen close to 1, e.g.,  $p = 0.95$ . The free energy landscape at temperature  $T$  is called *rugged*, if the support of the pd consists of many disconnected parts (this depends, of course, a bit on the adapted values for  $p$  and “many”). That a protein folds at room temperature, say, 300 K, into a unique native structure  $v_1^0, \dots, v_n^0$  means that its pd  $\rho(v_1, \dots, v_n; 300 \text{ K})$  describes small fluctuations around

this structure. We are now ready to formulate the funnel picture in terms of pds. Let us choose a protein and consider for it a sequence of pds

$$\rho_r(v_1, \dots, v_n) = \rho(v_1, \dots, v_n; T_r), \quad r = 1, \dots, s, \quad (2)$$

which is ordered by the temperatures  $T_r$ , namely,

$$T_1 > T_2 > \dots > T_f. \quad (3)$$

The sequence (2) constitutes a protein *funnel* when, for a reasonable choice of the probability  $p$  and the temperatures (3), the following holds:

- (i) The pds are rugged.
- (ii) The support of a pd at lower temperature is contained in the support of a pd at higher temperature

$$K_1^p \supset K_2^p \supset \dots \supset K_f^p, \quad (4)$$

e.g., for  $p = 0.95$ ,  $T_1 = 400 \text{ K}$ , and  $T_f = 300 \text{ K}$ .

(iii) With decreasing temperatures  $T_r$  the support  $K_r^p$  shrinks towards small fluctuations around the native structure.

Properties (ii) and (iii) are fulfilled for many systems of statistical physics, when some ground state stands in for the native structure. The remarkable point is that they may still hold for certain complex systems with a rugged free energy landscape, i.e., with property (i) added. In such systems one finds typically local free energy minima, which are of negligible statistical importance at low temperatures, while populated at higher temperatures. In simulations at low temperature the problem of the molecular dynamics (for a review, see [7]) as well as of the Metropolis [5] canonical ensemble approach is that the updating tends to get stuck in those local minima. On realistic simulation time scales this prevents convergence towards the native structure. On the other hand, the simulations move quite freely at higher temperatures, where the native structure is of negligible statistical weight. Nevertheless, the support of a protein pd may already be severely restricted, as we illustrate. The idea is to use a

relatively easily calculable pd at a higher temperature to improve the performance of the simulation at a lower temperature. In the following we investigate this idea for the Metropolis algorithm.

The Metropolis importance sampling would be perfected, if we could propose new configurations  $\{v'_i\}$  with their canonical pd  $\rho(v'_1, \dots, v'_n; T)$ . Because of the funnel property (ii) we expect that an *estimate*  $\bar{\rho}(v_1, \dots, v_n; T')$  from some sufficiently close-by higher temperature  $T' > T$  will feed useful information into the simulation at temperature  $T$ . The potential for computational gains is large because of funnel property (iii). The suggested scheme for the Metropolis updating at temperature  $T_r$  is to propose new configurations  $\{v'_i\}$  with the pd (2)  $\bar{\rho}_{r-1}(v'_1, \dots, v'_n)$  and to accept them with the probability

$$P_a = \min \left[ 1, \exp \left( - \frac{E' - E}{k T_r} \right) \frac{\bar{\rho}_{r-1}(v_1, \dots, v_n)}{\bar{\rho}_{r-1}(v'_1, \dots, v'_n)} \right]. \quad (5)$$

This equation biases the *a priori* probability of each dihedral angle with an estimate of its pd from a higher temperature. In previous literature [8,9] such a biased updating has been used for the  $\phi^4$  theory, where it is efficient to propose  $\phi(i)$  at each lattice site  $i$  with its single-site probability.

For our temperatures  $T_r$  the ordering (3) is assumed. With the definition  $\bar{\rho}_0(v_1, \dots, v_n) = (2\pi)^{-n}$  the simulation at the highest temperature,  $T_1$ , is performed with the usual Metropolis algorithm. We have thus a recursive scheme, called rugged Metropolis (RM) in the following. When  $\bar{\rho}_{r-1}(v_1, \dots, v_n)$  is always a useful approximation of  $\rho_r(v_1, \dots, v_n)$ , the scheme zooms in on the native structure, because the pd at  $T_f$  governs its fluctuations.

To get things working, we need to construct an estimator  $\bar{\rho}(v_1, \dots, v_n; T_r)$  from the numerical data of the RM simulation at temperature  $T_r$ . Although this is neither simple nor straightforward, a variety of approaches offer themselves to define and refine the desired estimators. In the following we work with the approximation

$$\bar{\rho}(v_1, \dots, v_n; T_r) = \prod_{i=1}^n \bar{\rho}_i^1(v_1, \dots, v_n; T_r), \quad (6)$$

where the  $\bar{\rho}_i^1(v_1, \dots, v_n; T_r)$  are estimators of reduced one-variable pds defined by

$$\rho_i^1(v_i; T) = \int_{-\pi}^{+\pi} \prod_{j \neq i} dv_j \rho(v_1, \dots, v_n; T). \quad (7)$$

The resulting algorithm, called RM<sub>1</sub>, appears to constitute the simplest RM scheme possible. Its implementation is straightforward, as estimators of the one-variable reduced pds are easily obtained from the time series of a simulation. The computer time consumption of RM<sub>1</sub> is practically identical with the one of the conventional Metropolis algorithm. In the following RM<sub>1</sub> is used to demonstrate the correctness of our basic assumptions.

The scope of this Letter limits us to one illustration. We rely on a simulation of the brain peptide Met-Enkephalin (Tyr-Gly-Gly-Phe-Met), because it is a numerically well studied system, which allows for comparison with results of the literature [10–13]. Our Metropolis simulations are performed with a variant of SMMP [14] (simple molecular mechanics for proteins). We keep the  $\omega$  torsion angles unconstrained and thus have 24 fully variable dihedral angles. In the previous literature the omega angles were either fixed to  $\pi$  or restricted to  $[\pi - \pi/9, \pi + \pi/9]$ . This leads to statistically significant differences in the energy, while the structural differences remain negligible. The performance of the RM<sub>1</sub> updating was tested at 300 K using input from a simulation at 400 K. The value of 300 K is chosen, because it is in the temperature range on which simulations of biological molecules eventually have to focus. The temperature of 400 K is high enough so that the Metropolis algorithm is efficient as the autocorrelation times are small, while it is low enough to provide useful input for the 300 K simulation.

At each simulation temperature a time series of  $2^{17} = 131\,072$  configurations is kept, in which subsequent configurations are separated by 32 sweeps. A sweep is defined by updating each dihedral angle once. Before starting with the measurements  $2^{18} = 262\,144$  sweeps are performed for reaching equilibrium. Thus, the entire simulation at one temperature relies on  $2^{18} + 2^{22} = 4\,456\,448$  sweeps. On a modern PC (1.9 GHz Athlon) this takes under 12 h for the vacuum system and less than 2 d with the inclusion [15] of solvent effects. For each dihedral angle the acceptance rate of the Metropolis algorithm was monitored at run time and the integrated autocorrelation time  $\tau_{\text{int}}$  (see [16] for its definition) was calculated from the recorded time series. Values around 0.5 are desirable for the acceptance rate, but the decisive quantity for the performance of an algorithm is the integrated autocorrelation time. To achieve a predefined accuracy, the computer time needed is directly proportional to  $\tau_{\text{int}}$ . In the following results from vacuum simulations are summarized. Computations which include solvent effects will be reported elsewhere [17].

In Table I results for the energy and two dihedral angles are presented. Error bars are given in parentheses. The acceptance rates are accurate to  $\pm 1$  in their last digit. Using the SMMP [14] conventions, which differ from previous literature [10–12], the angles are Gly-2  $\omega$  ( $v_6$ ) and Gly-3  $\phi$  ( $v_{10}$ ). They are well suited to illustrate important features of our approach.

The Gly-3  $\phi$  angle and four more angles (Gly-2  $\phi$ , Gly-2  $\psi$ , Phe-4  $\phi$ , and Phe-4  $\psi$ ) exhibit very large autocorrelation times. In Fig. 1 estimates of the one-variable pds (6) for the Gly-3  $\phi$  angle at 400 and 300 K are depicted. The rugged nature of the distribution is already obvious at 400 K. While the shapes of the other 23 dihedral angle pds vary greatly, featuring from one to three local maxima, a ruggedness and our funnel property (ii) are found

TABLE I. Acceptance rates and integrated autocorrelations times for the energy  $E$  and, in the SMMP notation, the dihedral angles Gly-2  $\omega$  ( $v_6$ ) and Gly-3  $\phi$  ( $v_{10}$ ).

Variable	Method	400 K	400 K	300 K	300 K
		acpt	$\tau_{\text{int}}$	acpt	$\tau_{\text{int}}$
$E$	Metro	0.168	4.98 (20)	0.120	49.6 (5.0)
$E$	RM <sub>1</sub>	...	...	0.375	26.2 (1.6)
$E$	PT	0.167	3.67 (20)	0.119	19.9 (1.6)
$E$	PT + RM <sub>1</sub>	0.460	2.56 (34)	0.375	9.94 (60)
$v_6$	Metropolis	0.049	3.09 (10)	0.034	21.1 (1.8)
$v_6$	RM <sub>1</sub>	...	...	0.416	9.68 (66)
$v_6$	PT	0.049	2.24 (07)	0.034	7.85 (36)
$v_6$	PT + RM <sub>1</sub>	0.553	1.34 (04)	0.413	4.62 (55)
$v_{10}$	Metro	0.088	7.49 (47)	0.034	167 (27)
$v_{10}$	RM <sub>1</sub>	...	...	0.070	80.6 (7.0)
$v_{10}$	PT	0.087	6.13 (30)	0.034	32.7 (3.1)
$v_{10}$	PT + RM <sub>1</sub>	0.141	4.43 (26)	0.070	22.6 (2.7)

for each case. Towards low temperatures a shrinking of the one-variable pds to their global energy minimum (GEM) implies property (iii). The arrow in Fig. 1 indicates the GEM value of this particular angle. Note that for the Metropolis algorithm the  $\rho_i^! \approx 0$  regions do *not* constitute free energy barriers, because they can be jumped by a single Metropolis updating step. This is different for molecular dynamics simulations.

Despite the similarities of the Gly-3  $\phi$  pds at 400 and 300 K, there is a big increase of the Gly-3  $\phi$  integrated autocorrelation time. In the conventional, canonical Metropolis simulation it is by more than a factor of 20 from 7.5 to 167. The RM<sub>1</sub> updating reduces this by a factor of 2.1, from 167 to 81. The integrated autocorrelations times of Table I are given in units of 32 sweeps, as this is the step size of our time series data recorded.

The acceptance rates for moves of the Gly-3  $\phi$  angle are very small. As the support of its pds in Fig. 1 covers

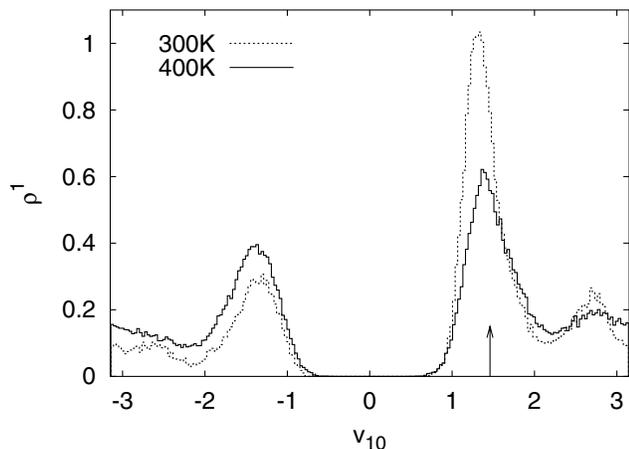


FIG. 1. Vacuum probability densities at 400 and 300 K for the Met-Enkephalin dihedral angle Gly-3  $\phi$ .

more than 50% of the full range, this has to be due to correlations with other dihedral angles. To exhibit a rather distinct case of an angle with a low acceptance rate, results for one of the six  $\omega$  angles are also included in Table I. For the conventional Metropolis simulation this angle has the same acceptance rate as the Gly-3  $\phi$  angle (incidentally to all digits given). A look at the Gly-2  $\omega$  pds reveals an obvious reason: They are narrowly peaked around the (identified) values  $\pm\pi$ , which is explained by the specific electronic hybridization of the CO-N peptide bond. Autocorrelation times are about 8 times smaller than for the Gly-3  $\phi$  angle. Applying RM<sub>1</sub> updating to the Gly-2  $\omega$  pds cures entirely the problem of its low acceptance rate. At 300 K the increase is from 0.034 to 0.416, and similar numbers are found for the other  $\omega$  angles. In contrast to that, the Gly-3  $\phi$  acceptance rate increases only to the modest value of 0.07, while the improvements of  $\tau_{\text{int}}$  are kind of similar. For Gly-2  $\omega$  it is by a factor of 2.7 from 21 to 7.9.

It is straightforward to combine the RM<sub>1</sub> updating with generalized ensemble methods (see [18] for a review). They are enabling techniques for studying equilibrium physics of complex systems at very low temperatures [19]. In the parallel tempering (PT) approach [20,21] two or more replica are simulated in parallel at distinct temperatures and Metropolis exchanges of the temperatures are offered occasionally. Autocorrelations are reduced, when suitable excursions to higher temperatures become feasible. In Table I results for a PT simulation with replica at 400 and 300 K are included. To compare with RM<sub>1</sub>, we use the integrated autocorrelation time of the energy, which characterizes the overall performance. The PT method decreases  $\tau_{\text{int}}$  by a divisor of 2.5, from 50 to 20. When parallel nodes with a reasonably fast communication are available, this is the gain in real time. For RM<sub>1</sub> the improvement factor is 1.9, from 50 to 26. Thus, PT outperforms RM<sub>1</sub> in real time, while RM<sub>1</sub> wins in the total CPU time. Most remarkable is that the improvement factors multiply. Running PT with the RM<sub>1</sub> updating yields another factor of 2,  $\tau_{\text{int}}$  is reduced from 20 to 10 and, altogether, from 50 to 10. Note that the acceptance rates are not changed by PT.

In previous simulations [3,12] it has been shown that considerably lower temperatures than 300 K are needed to reduce the fluctuations of Met-Enkephalin in vacuum to fluctuations around its GEM. Here, I point out that each of our simulations at 300 K reaches the valley of attraction of the GEM sufficiently often, so that the GEM can be found by local minimization. To be specific, the energy spectrum of Fig. 2 is obtained from our RM<sub>1</sub> simulation at 300 K in the following way: First we isolate all configurations of the time series which are minima in the lower 10% quantile (for the definition, see [22]) of the energy distribution and separated by an excursion of the time series into the upper 10% quantile. In this way we obtain 1357 configurations and the SMMP minimizer

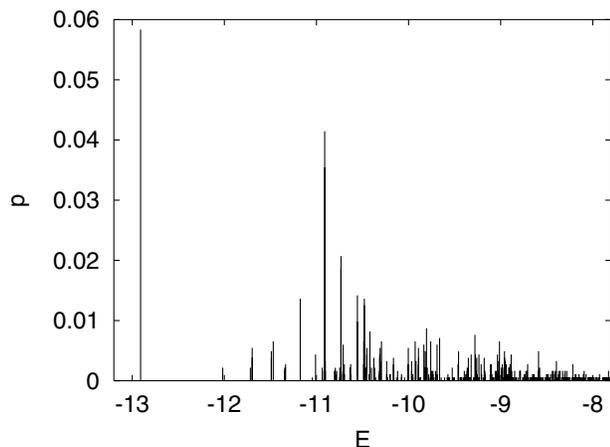


FIG. 2. Spectrum of energy-minimized Met-Enkephalin configurations from the 300 K RM<sub>1</sub> simulation in vacuum.

is run on each of them. The probabilities of the resulting spectrum levels are plotted in Fig. 2. The GEM occurs at  $E = -12.91$  Kcal/mol with about 6% probability (107 times) followed by a conformation at  $E = -10.92$  Kcal/mol with about 4% probability. The frequency of finding the GEM for the other simulations of Table I is approximately proportional to  $\tau_{\text{int}}^{-1}$  of the energy.

Rugged distributions of the dynamical variables are typical for Metropolis simulations of proteins and RM<sub>1</sub> improves the importance sampling. It ought to become standard, as it yields a relevant gain in computer time for little extra efforts by the programmer. Met-Enkephalin is essentially solved by a simulation at 300 K. For some of its dihedral angles the RM<sub>1</sub> updating overcomes the problem of low acceptance rates entirely. For others the improvement remains more modest, because their low acceptance rates are due to correlations with other angles. Only multivariable moves can achieve importance sampling in such a situation. For the usual Metropolis algorithm the acceptance rates for multivariable moves are practically zero. Here, we gained novel physical insight into the funnel picture, which provides us with a recipe on how to design multivariable moves, so that the acceptance rate is expected to stay reasonably large. One will start with the angles with the worst autocorrelations and construct their two-angle moves according to their reduced two-variable pds, the RM<sub>2</sub> algorithm. Next, one may use three variables, and so on. In this way the outcome of our investigation of computational consequences of the funnel picture is that we do not expect a single, generically good Metropolis algorithm for protein simulations. Instead, we have developed a strategy to design an algorithm for each particular protein. This aspect of the RM approach promises advances for simulations of larger peptides.

I thank Yuko Okamoto for numerous useful discussions. This research was in part supported by the Okazaki National Institutes and by the U.S. Department of Energy under Contract No. DE-FG02-97ER41022.

\*Electronic address: berg@csit.fsu.edu

- [1] D. Bryngelson and P.G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987).
- [2] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, J. Chem. Phys. **104**, 5860 (1996).
- [3] U. H. Hansmann, Y. Okamoto, and J. N. Onuchic, Proteins: Struct., Funct., Genet. **34**, 472 (1999).
- [4] U. H. Hansmann and J. N. Onuchic, J. Chem. Phys. **115**, 1601 (2001).
- [5] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
- [6] M. J. Sippl, G. Nemethy, and H. A. Scheraga, J. Phys. Chem. **88**, 6231 (1984), and references therein.
- [7] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic Press, San Diego, 1996).
- [8] A. D. Bruce, J. Phys. A **18**, L873 (1985).
- [9] A. Milchev, D. W. Heermann, and K. Binder, J. Stat. Phys. **44**, 749 (1986).
- [10] Z. Li and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **85**, 6611 (1987).
- [11] Y. Okamoto, T. Kikuchi, and H. Kawai, Chem. Lett. **1992**, 1275 (1992).
- [12] U. H. Hansmann and Y. Okamoto, J. Comput. Chem. **14**, 1333 (1993).
- [13] H. Meirovitch, E. Meirovitch, A. G. Michel, and M. Vásquez, J. Phys. Chem. **98**, 6241 (1994).
- [14] F. Eisenmenger, U. H. Hansmann, S. Hayryan, and C.-K. Hu, Comput. Phys. Commun. **138**, 192 (2001).
- [15] T. Ooi, M. Oobatake, G. Nemethy, and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **84**, 3086 (1987).
- [16] A. Sokal, in *Functional Integration: Basics and Applications*, edited by C. DeWitt-Morette, P. Cartier, and A. Folacci, Cargese Summer School (Plenum Press, New York, 1997), pp. 131–192.
- [17] B. A. Berg, H. P. Hsu, and P. Grassberger (to be published).
- [18] A. Mitsutake, Y. Sugita, and Y. Okamoto, Biopolymers **60**, 96 (2001).
- [19] B. A. Berg and T. Celik, Phys. Rev. Lett. **69**, 2292 (1992).
- [20] G. J. Geyer, in *Computing Science and Statistics, Proceedings of the 23rd Symposium on the Interface*, edited by E. M. Keramidis (Interface Foundation, Fairfax, VA, 1991), pp. 156–163.
- [21] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
- [22] S. Brandt, *Statistical and Computational Methods in Data Analysis* (North-Holland Publishing Company, Amsterdam, 1983).