Efficient Combination of Wang–Landau and Transition Matrix Monte Carlo Methods for Protein Simulations

RUBEN G. GHULGHAZARYAN,1 2 SHURA HAYRYAN,1 CHIN-KUN HU
1 Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan
2 Department of Theoretical Physics, Yerevan Physics Institute, Alikhanian Brothers 2, 375036 Yerevan, Armenia

Received 22 February 2006; Revised 8 June 2006; Accepted 12 June 2006
DOI 10.1002/jcc.20597
Published online 28 December 2006 in Wiley InterScience (www.interscience.wiley.com).

Abstract: An efficient combination of the Wang-Landau and transition matrix Monte Carlo methods for protein and peptide simulations is described. At the initial stage of simulation the algorithm behaves like the Wang-Landau algorithm, allowing to sample the entire interval of energies, and at the later stages, it behaves like transition matrix Monte Carlo method and has significantly lower statistical errors. This combination allows to achieve fast convergence to the correct values of density of states. We propose that the violation of TTT identities may serve as a qualitative criterion to check the convergence of density of states. The simulation process can be parallelized by cutting the entire interval of simulation into subintervals. The violation of ergodicity in this case is discussed. We test the algorithm on a set of peptides of different lengths and observe good statistical convergent properties for the density of states. We believe that the method is of general nature and can be used for simulations of other systems with either discrete or continuous energy spectrum.


Key words: Wang–Landau algorithm; transition matrix Monte Carlo; density of states; statistical errors; TTT identities; proteins

Introduction

After the seminal work by Metropolis et al.1 the Monte Carlo method has become one of the most powerful and fruitful methods for simulation of a large variety of different physical systems.2–7 However, in the course of time, it becomes evident that the traditional Metropolis dynamics for Monte Carlo (MC) moves cannot work efficiently in many systems. The typical example is exponential increase of the correlation times near the phase transition temperatures in spin systems.6 Other examples are extremely insufficient sampling of phase space in such a complex systems as spin glasses2, 5 or proteins.

That is why the search for efficient algorithms enhancing the sampling of configuration space in MC simulations of complex systems is one of the central problems in modern computational physics, chemistry, and biology. Over the last decade, several advanced MC algorithms have been proposed to overcome the typical problems which arise in traditional MC dynamics. It would be a difficult task to list all methods, so here we mention just some of them. Among these algorithms are the cluster algorithm9 for the simulation of clusters for the lattice spin models,10, 11 parallel tempering,12–14 broad histogram (BH) method,15, 16 transition matrix Monte Carlo (TMMC),17 entropy sampling (ES),18 multicanonical (MUCA) algorithm,19–22 and the more recently proposed Wang–Landau (WL) algorithm.23, 24 In general, the ES, MUCA methods, and WL algorithm are based on the idea of sampling the microstates of the system uniformly in the space of some macroscopic observable, i.e. to perform a random walk in the space of given observable, usually energy. The probability P(Q) of sampling a macroscopic state Q is proportional to its density (or degeneracy) n(Q), i.e. the number of corresponding microstates q, times the probability w(Q) of visiting the macrostate Q during a MC simulation, i.e. P(Q) ∝ n(Q)w(Q). Thus, if one chooses w(Q) ∝ n−1(Q), then one is supposed to achieve a uniform sampling for the observable Q, i.e. P(Q) ∝ const. This means that the accumulated histogram for observable Q during such simulation will be almost flat (up to some statistical errors due to finite number of Monte Carlo steps). These methods are sometimes called

Correspondence to: C.-K. Hu; e-mail: huck@phys.sinica.edu.tw
Contract/grant sponsor: National Science Council (Taiwan); contract/grant numbers: NSC 94-2112-M001-014, 94-2811-M001-014, and 94-2119-M-002-001
Contract/grant sponsor: Academia Sinica (Taiwan); contract/grant numbers: AS-92-TP-A09 and AS-95-TP-A07

© 2006 Wiley Periodicals, Inc.
“flat-histogram” methods. In all these methods, the density of states is not known a priori and one needs a procedure to find proper estimates in order to perform a “flat-histogram” simulation. Usually every “flat-histogram” methods suggests such a procedure. For example, the ES and MUCA methods use an iterative procedure for improving density of states from the accumulated histograms during each iteration. The advantage of the MUCA method over the ES is that it uses information of all previous iterations in a specially designed iterative procedure for microcanonical temperature, thus reducing the statistical errors. One general drawback of these methods is their bad scaling for large systems. The efficiency of some of these algorithms, like parallel tempering or replica exchange, MUCA (and references therein), ES in protein simulations is more or less well studied while others need to be investigated in more details. The conceptual differences between the broad histogram and MUCA methods for cubic Ising model have been studied in detail in ref. 30.

In this article, we concentrate on the TMMC and the most recently proposed Wang–Landau methods. The Wang–Landau algorithm was proposed as an efficient method to calculate the density of states. It implements an uniform sampling or random walk in the energy space. Since the densities of states are usually very large numbers, in order to avoid overflows in computer registers during the simulation, it is convenient to formulate the WL algorithm in terms of microcanonical entropy \( S(E) = \ln n(E) \), where \( E \) is the energy of the system and the Boltzmann constant \( k_B \) before \( \ln(n(E)) \) is set to 1. The algorithm is as follows: The proposed new state with energy \( E' \) is accepted with probability \( P(E, E') = \min(1, \exp[S(E) - S(E')] ) \) and the corresponding entropy is increased by some small number \( \gamma \): \( S(E') \rightarrow S(E') + \gamma \). Otherwise, the entropy of the previous state \( E \) is increased by the same \( \gamma \). Initially one can take \( \gamma = 1 \), which allows to reach all energy levels quite fast. Larger values for \( \gamma \) lead to large statistical errors and too small values for \( \gamma \) lead to slow sampling of the energy levels. Since at every step the microcanonical entropy or density of states are changed, the WL algorithm violates the detailed balance condition, and only for \( \gamma = 0 \), the detailed balance condition is satisfied. Wang and Landau proposed to accumulate the histogram of visits of energy levels and reduce the value of \( \gamma \) every time when each histogram entry is greater than 80% of the average histogram (the histogram flatness condition). In doing so, the value of \( \gamma \) is eventually reduced to machine precision zero and the detailed balance condition will be satisfied with this precision. It is supposed that the density of states should converge to their true values. Unfortunately, in WL method, the statistical errors in the density of states reach some saturation values and they are not reduced in the course of further simulations. This phenomenon is known as freezing statistical errors and it creates significant difficulties in calculating the density of states for large systems. Currently, the dynamics of WL algorithm and search for optimal schedule for updating \( \gamma \) in order to achieve fast convergence to the true density of states with small statistical errors are under intensive studies.

Alternative methods for calculating the density of states are the broad histogram and the transition matrix Monte Carlo (TMMC) methods. It can be shown that the broad histogram method may be considered as a particular case of infinite temperature TMMC method (see further in the text). The advantage of the TMMC method is that the quality of the obtained density of states does not depend on the underlying MC dynamics, i.e. on the probability of accepting or rejecting the proposed configurations, and the statistical errors are small. Nevertheless, the method is sensitive to violation of the detailed balance condition. Unlike the WL algorithm where the violation of the detailed balance condition leads to large statistical errors, in the TMMC method the statistical errors remain small, but a bias in the density of states arises. One of the interesting properties of the TMMC is that the elements of the transition matrix accumulated during the simulation should satisfy total transition matrix (TTT) identities if the detailed balance condition is satisfied during the simulation. Thus one can expect that the violation of the TTT identities may serve as a measure of the violation of the detailed balance condition which in turn is related to the quality of the density of states. We will discuss the theoretical foundations of the TMMC method and TTT identities in the next section.

Recently, Shell et al. proposed to combine the WL and TMMC methods to improve the statistical properties of the calculated density of states. The main idea here is to run a pure WL simulation first, and when \( \gamma \) becomes small enough, to replace the density of states by the ones calculated from the TMMC method or use the TM elements directly in the acceptance probabilities. Also a faster schedule was proposed for reducing \( \gamma \) as \( \gamma \rightarrow 0.1 \gamma \) every time when all energy bins are visited at least once. The algorithm was tested by simulating the Lennard-Jones liquid and the obtained data showed good statistical convergent behavior.

However, when we tried to apply the method of ref. 36 to simulate peptides or proteins of more than 10 amino acids, we found that simulations got stuck at low energy levels (producing a Gaussian like histograms) for too long time and sample poorly the transition region. The purpose of the present paper is to investigate in detail the performance of combined WL and TMMC algorithms in protein simulations. Our basic goal is to find an efficient combination of two methods, which would allow fast sampling and produce small statistical errors.

We use the standard geometry model of protein chains with ECEPP/3 forcefield as implemented in the protein simulation package SMMP. We use the improved version of our algorithm for parallel calculation of the protein energy with all atom representation and carry-out simulations in Linux clusters.

We compare our method with the one originally proposed in ref. 36 and find that our method can achieve almost the same quality for the density of states, and is, at the same time, about 50% faster. We also propose and check a new idea to use the TTT identity violation as a qualitative criterion for checking the convergence of the density of states. The algorithm is tested by simulating the helix–coil transition in polyalanine of different lengths. The optimal ways of parallelization of simulations are discussed as well.

This paper is organized as follows. In the next (second) section, we first briefly describe the system which we simulate and the strategy of Monte Carlo moves. We then briefly review the algorithms which we use in constructing our method. Finally, we present our improved combination of WL and TMMC methods for protein simulations. In the third section, we present simulated results and compare the performance of different algorithms. In the last section, we discuss our results.
Methods

The Protein Model and the Monte Carlo Moves

In SMMP, valence bond lengths and valence angles are fixed and the only free variables are the rotation angles around single chemical bonds (torsions). The rotations around N–C\(^\#\) (\(\psi\)-angles) and C\(^\#\)–C (\(\zeta\)-angles) are free, while the rotations around the bonds in side chains (\(\chi\)-angles) are subject to some, relatively soft, energetic barriers. Rotations around the peptide bonds C–N (\(\omega\)-angles) are also permitted but with very high barriers. In some simulations, we allow the \(\omega\) angle to vary only in the range \(\pm 20^\circ\) around the trans-conformation. The simulations are started usually from some random configuration obtained by assigning random values to all rotation angles. The elementary MC move consists of assigning a new random value from the interval \([-\pi; +\pi]\) to the given variable. This configuration-proposal procedure is reversible since by rotating the same angle in the opposite direction one can restore the initial configuration. The variable may be chosen randomly or, alternatively, each variable may be changed in one complete sweep over all variables. To reduce the correlations between two successive steps we measure the quantities after every \(m_i > 1\) sweeps. This move proposal mechanism enables to construct the infinite temperature transition matrix correctly (see further sections).

The total conformational energy of the intramolecular interactions \(E_{\text{pot}}\) is defined as the sum of the electrostatic \(E_{\text{el}}\), the Lennard–Jones \(E_{\text{LJ}}\), the hydrogen bond \(E_{\text{hb}}\), and the torsional rotation potential \(E_{\text{tor}}\)

\[
E_{\text{pot}} = E_{\text{el}} + E_{\text{LJ}} + E_{\text{hb}} + E_{\text{tor}},
\]

where

\[
E_{\text{el}} = 332 \sum_{j \neq i} \frac{q_i q_j}{r_{ij}}, \\
E_{\text{LJ}} = \sum_{j > i} \left( A_{ij} \left( \frac{r_{ij}}{r_{ij}^1} \right)^{12} - B_{ij} \left( \frac{r_{ij}}{r_{ij}^1} \right)^{6} \right), \\
E_{\text{hb}} = \sum_{j > i} \left( C_{ij} \left( \frac{r_{ij}}{r_{ij}^1} \right)^{10} - D_{ij} \left( \frac{r_{ij}}{r_{ij}^1} \right)^{10} \right), \\
E_{\text{tor}} = \sum_n U_n \left( 1 \pm \cos(k_n \phi_n) \right).
\]

Here \(r_{ij}\) is the distance between atoms \(i\) and \(j\) measured in \(\text{\AA}\); \(A_{ij}, B_{ij}, C_{ij}, D_{ij}\) are parameters of the empirical potentials; \(q_i\) and \(q_j\) are the partial charges on the atoms \(i\) and \(j\), respectively, in the units of electronic charges; \(\varepsilon\) is the dielectric constant of environment and is usually set to \(\varepsilon = 2\), its value in the protein interior space. The factor 332 in eq. (2) is used to express the energy in kcal/mol. \(U_n\) is the energetic torsion barrier of rotation about the bond \(n\) and \(k_n\) is the multiplicity of the torsion angle \(\phi_n\).

The energy of protein–solvent interaction is calculated by the solvation-parameters method, which is based on the assumption that the free energy difference between solvated and unsolvated groups is proportional to the surface area of the molecule, exposed to the water.

\[
E_{\text{solv}} = \sum_i \sigma_i A_i, 
\]

where \(A_i\) is the solvent accessible surface area and \(\sigma_i\) is the atomic solvation parameter of the atom \(i\). In the present study, we use the OONS parameter set. Since the energy spectrum of our system is continuous, for computational purposes, we divide the energy with bin size 1 kcal/mol (the bin size is, in general, system-dependent). If the bin size is too big, then large systematic errors appear. If, in contrary, it is too small, then too many MC steps are required to accumulate enough statistics in every bin.

The Transition Matrix

We give the definition of transition matrix and describe the method for calculating the density of states from the transition matrix elements. Here we follow refs. 17 and 36 to describe the method and focus mainly on simulation in energy space, but the same arguments may be applied for other physical quantities.

The microstate \(i\) is defined as a configuration of the system with the given energy \(E_i\). The macrostate \(I\) is the set of all microstates \(i\) in the energy bin \(J\). The transition matrix is composed of the elements \(T(i, J)\) which represent the probabilities of moving from macrostate \(I\) to \(J\) provided that the system was initially in \(I\). Let us denote by \(t(i, j)\) the transition probability of moving from the given microstate \(i\) to the target microstate \(j\). \(t(i, j)\) is a stochastic matrix, and the normalization conditions must hold

\[
\sum_j t(i, j) = 1, \quad t(i, j) \geq 0,
\]

where the sum goes over all microstates. In terms of microstate transition probabilities, the transition matrix elements are written as

\[
T(I, J) = \frac{1}{n(I)} \sum_{i \in I} \sum_{j \in J} t(i, j),
\]

where the sums extend over all microstates, corresponding to macrostates \(I\) and \(J\), respectively, and \(n(I)\) is the total number of microstates corresponding to the macrostate \(I\), which will be called the density of states or degeneracy of the macrostate \(I\) below. This definition of the transition matrix implies that the configurational probability distribution is a function of energy only, i.e. the probability for realization of any microstate depends on its energy only. By the definition (8) the macroscopic transition probability is simply the microcanononical average of all corresponding microstate transition probabilities. The matrix \(T(I, J)\) satisfies a similar normalization conditions

\[
\sum_j T(I, J) = 1, \quad T(I, J) \geq 0,
\]
where summation extends over all allowed energy bins (macrostates). By dividing the transition matrix element \( T(I, J) \) by \( T(J, I) \), i.e. the element corresponding to the move in the reverse direction, \( J \rightarrow I \), an important relation between transition matrix elements and the density of states is derived

\[
\frac{T(I, J)}{T(J, I)} = \frac{n(J) \sum_{i \in I} \sum_{j \in J} t(i, j)}{n(I) \sum_{i \in I} \sum_{j \in J} t(j, i)}.
\] (10)

The elements of the stochastic matrix can be represented as a product of two terms \( t(i, j) = a(i, j)P_{acc}(i, j) \), where \( a(i, j) \) is the probability of proposing a transition from microstate \( i \) to \( j \), and \( P_{acc}(i, j) \) is the probability of accepting the proposed move.\(^{36} \) The probabilities \( a(i, j) \) depend only on the type of MC moves used in the simulation, and \( P_{acc}(i, j) \) is defined by the probability distribution with which the states are sampled in the given ensemble.

Our aim is to calculate the density of states which are ensemble-independent and unique properties of the system. Let us consider the case when all \( P_{acc}(i, j) \) are constant and equal to each other, for example \( P_{acc}(i, j) = 1 \) for all \( i, j \). This means that all proposed moves will be accepted which corresponds to the classical Metropolis dynamics at infinite temperature.\(^1 \) We add an index \( \infty \) to the transition matrix elements to indicate this strict move acceptance and call \( T_\infty(I, J) \) the infinite temperature transition matrix (ITTM).

For \( T_\infty(I, J) \) the equation (10) has the form

\[
\frac{T_\infty(I, J)}{T_\infty(J, I)} = \frac{n(J) \sum_{i \in I} \sum_{j \in J} a(i, j)}{n(I) \sum_{i \in I} \sum_{j \in J} a(j, i)}.
\] (11)

The elements of \( T_\infty(I, J) \) depend only on the symmetry of the particular move type. If the moves are reversible and symmetric, \( a(i, j) = a(j, i) \), then

\[
n(I)T_\infty(I, J) = n(J)T_\infty(J, I).
\] (12)

Examples of symmetric reversible moves are single spin flips, single particle displacements and single torsion angle rotation, in the case of protein simulations. \( T_\infty(I, J) \) gives the probability of proposing move from macrostate \( I \) to \( J \) and this probability does not depend on the particular Monte Carlo dynamics. The set of all \( T_\infty(I, J) \) elements gives a measure of connectivity of macrostates for a given move type. Also, it can be shown that eqs. (10) and (12) are directly related to the detailed balance condition,\(^{17} \) which must be satisfied for correct estimation of \( T_\infty(I, J) \) elements and the density of states. Equation (12) coincides exactly with the broad histogram equations, initially proposed by Oliveira et al.\(^{15} \) for calculating the density of states.

The density plot of a typical transition matrix found in our simulations of the peptide Ala\(_{16}\) with 16 alanine amino acids is shown in Figure 1. One can see that by a single dihedral angle rotation mechanism of proposing new conformations the transition probabilities from any low energy state to the higher energy states are nonzero.

Meanwhile, one can see that when the system is in a low energy state, the number of proposed moves to lower energy levels is negligibly small. This might mean that by a single dihedral angle rotation it is prohibitively hard to change the low energy configurations significantly.

### TTT Identities for the Transition Matrix Elements

Consider three distinct macrostates \( I, J, \) and \( K \) for which the transition matrix elements of all possible mutual moves are nonzero, and each of them contains at least one microstate, i.e. their degeneracy of states are nonzero. For each pair of the mutual moves, the eq. (12) holds:

\[
n(I)T_\infty(I, J) = n(J)T_\infty(J, I),
\] (13)

\[
n(J)T_\infty(J, K) = n(K)T_\infty(K, J),
\]

\[
n(K)T_\infty(K, I) = n(I)T_\infty(I, K).
\] (14)

By multiplying the left and right sides of eqs. (13) and cancelling the nonzero factor \( n(I)n(J)n(K) \) from both sides of the obtained equation, we have

\[
T_\infty(I, J)T_\infty(J, K)T_\infty(K, I) = T_\infty(I, K)T_\infty(K, J)T_\infty(J, I).
\] (14)

The relation (14) is called TTT identity.\(^{17} \) Figure 2 shows relationships among macrostates \( I, J, \) and \( K \), which are used to construct the TTT identity. The identity allows to verify the agreement of the transition matrix data with the exact density of states because eqs. (12) imply that the density of states are exact. It also allows to check the agreement of the transition matrix data with the detail balance condition even in the lack of \textit{a priori} knowledge about the stationary distribution.
move from $I$ to $J$, $T_\infty(I, J)$, one needs to divide $C(I, J)$ by the total number of proposed moves from $I$ to any macrostate, accessed by given move proposal procedure, i.e.

$$T_\infty(I, J) = \frac{C(I, J)}{H(I)}.$$  \hfill (17)

where $H(I) = \sum_k C(I, K)$, and summation extends over all macrostates. After the estimate for the ITTM is found, one can proceed to calculate the density of states, which can be done in a number of ways. By substituting estimates of ITTM into (12), a set of equations for the ratios of density of states is obtained.

$$\frac{n(I)}{n(J)} = \frac{T_\infty(J, I)}{T_\infty(I, J)}.$$  \hfill (18)

The simplest way to find the density of states is to solve the set of eq. (18) iteratively for only nearest energy levels by rewriting (18) in terms of microcanonical entropies $S(I)$

$$S(I + 1) = S(I) + \ln \frac{T_\infty(I, I + 1)}{T_\infty(I + 1, I)}.$$  \hfill (19)

Equation (19) corresponds to the broad histogram equation by Oliveira et al.\textsuperscript{15} and gives only relative values of the density of states. The density of states may be found if the degeneracy of the ground state is known \textit{a priori} (as for the Ising model) or one of the quantities $n(I)$, $S(I)$ is fixed to some value.
Another approach is based on considering the problem of finding the density of states as an optimization problem. Indeed the set of eq. (18) is over-specified, i.e. the number of equations is larger than the number of unknown variables. If the number of macro-states is \( N \) then there are \( N(N-1)/2 \) equations and only \( N-1 \) unknown density of states, because equations (18) define only the ratios of density of states. The least-square method may be used for solving the set of all \( N(N-1)/2 \) equations as an optimization problem for finding the set of entropies \( S(E) \) with minimum total variance \( \sigma^2_{tot} \) of the predicted values for \( S(I) \) and corresponding \( T_\infty(I,J) \). The total variance \( \sigma^2_{tot} \) is defined as

\[
\sigma^2_{tot} = \sum_{I,J} \frac{1}{\sigma^2_{IJ}} \left( S(I) - S(J) + \ln \frac{T_\infty(I,J)}{T_\infty(J,I)} \right), \tag{20}
\]

where

\[
\sigma^2_{IJ} = C(I,J)^{-1} + C(J,I)^{-1} + H(I)^{-1} + H(J)^{-1}. \tag{21}
\]

The weights \( \sigma^2_{IJ} \) are introduced to take into account the statistics of visits to macro-states and the number of proposed moves during the simulation. It is assumed that \( \text{var}[C(I,J)] \propto C(I,J) \) and hence \( \text{var}[H(I)] \propto H(I) \). In solving this optimization problem, the value of one of the \( S(I) \) must be fixed, for example \( S(I_0) = 0 \). The necessary condition for the minimum of \( \sigma^2_{tot} \) is

\[
\frac{\partial \sigma^2_{tot}}{\partial S(I)} = 0 \quad \text{for all} \quad I \neq I_0. \tag{22}
\]

After differentiation and some algebra, the following set of \((N-1)\) linear equations are obtained

\[
S(K) \sum_l \frac{1}{\sigma^2_{lk}} - \sum_l \frac{1}{\sigma^2_{lk}} S(l) = \sum_l \frac{1}{\sigma^2_{lk}} \ln \frac{T_\infty(I,K)}{T_\infty(K,I)}, \tag{23}
\]

where summations go over all macrostates and \( K \neq I_0 \). The system of eq. (23) may be solved iteratively or by matrix inversion method.\(^{45}\) Note that the optimization method provides better quality for the density of states when the statistics of sampling of energy levels is not good enough. But with accumulation of a larger statistics, both (20) and (19) give almost the same results.

**Efficient Combination of WL and TMMC Methods**

The advantage of the WL algorithm is that at initial stages (large \( \gamma \)) the simulation easily jumps over energy barriers and samples all energy levels almost uniformly. However, as already mentioned in **Introduction**, at the later stages (small \( \gamma \)), the statistical errors freeze down, and this is the main drawback of the method. In contrast to WL algorithm, TMMC methods have better statistical properties and give estimates of the density of states with small statistical errors. But the TMMC methods, in their turn, have a disadvantage that they require a good initial guess for the density of states in order to start the simulations. Otherwise, a systematic bias in the density of states appears due to improper sampling of microstates leading to errors in microcanonical averages of the number of proposed moves corresponding to ITTM elements. Thus, TMMC method provides a good estimate of the density of states when all energy levels in the given interval are visited repeatedly. From these considerations, one might expect that the two methods can be combined in a proper way to sample efficiently all energy levels, and, at the same time, to calculate the density of states with high accuracy.

We propose the following strategy for combining WL and TMMC methods: Start a WL simulation with \( \gamma = \gamma_0 \), where \( 0 < \gamma_0 \leq 1 \), and accumulate move proposal array \( C(I,J) \). Update periodically (after each \( N_{up} \) MC sweeps) the microcanonical entropy of WL simulation, \( S_{WL}(I) \), with the estimate of microcanonical entropy \( S_{TM}(I) \) [calculated from ITTM by eq. (19)] as follows

\[
S_{WL}(I) + \eta[S_{TM}(I) - S_{WL}(I)] \rightarrow S_{WL}(I), \tag{24}
\]

where \( 0 < \eta < 1 \), until all energy levels are visited at least once. Then, reduce the value of \( \gamma \), by some factor \( \varepsilon \): \( \gamma \leftarrow \varepsilon \gamma \). Continue this procedure until \( \gamma \) becomes small enough, say \( \gamma \leq 10^{-8} \). In our simulation, we used the values \( \gamma_0 = 2^{-1}, N_{up} = 5000, \eta = 0.1 \) and \( \varepsilon = 0.5 \).

One can see that the outlined algorithm is essentially the WL algorithm but the improvement of the density of states in (24) eventually reduces the statistical errors to the level of errors in TMMC method which is of the order of magnitude less than that for the pure WL algorithm.\(^{17}\) At the initial stage, the method behaves like a pure WL algorithm and samples effectively the configurations in the entire range of energies. At the later stages, when the detailed balance violation is small, both \( S_{WL}(I) \) and \( S_{TM}(I) \) data eventually converge to each other. Note that in our scheme \( S_{WL}(I) \) is updated even for macrostates that have not been sampled during given \( N_{up} \) MC sweeps. This of course may be changed by adding an extra multiplicative term in front of the brackets on the left-hand side of the eq. (24) which equals to 1 if the state is visited, and 0, otherwise. The advantage of this algorithm is that it uses the advanced properties of WL (fast sampling of all energy levels) and TM (small statistical errors) methods to achieve fast convergence for the density of states.

By changing the values for parameters \( \eta \) and \( \varepsilon \) the dynamics of the simulation may be changed. Recall that at the initial stages of WL simulation the errors in \( S_{WL}(I) \) data are large. These large errors allow the simulation process to stay at low energy states a proper number of MC steps and to generate configurations with lower energies when \( \gamma \) is large. In contrast to \( S_{WL}(I) \), the statistical errors in the \( S_{TM}(I) \) data are relatively small, and the values of microcanonical entropies are closer to the true values. But now these values carry a bias due to detailed balance condition violation. If \( \eta \) is close to unity, i.e. \( S_{WL}(I) \) are replaced by \( S_{TM}(I) \), then large values of \( \gamma \) will not allow to sample low energy states well because the simulation cannot stay long enough at the low energy states. This is one of the reasons why the errors in density of states is large for large \( \gamma \) in WL simulation. Hence, for values of \( \eta \) close to unity the values of \( \gamma \) have to be reduced accordingly. But it is a difficult problem to find such a fine tuned schedule for adjusting \( \gamma \) for such a dramatic changes in entropy. So, by keeping \( \eta \) small one
will eventually reduce statistical errors in the WL simulation while not worsening the sampling procedure of WL algorithm. Note, that \( \eta = 0 \) corresponds to the pure WL simulation, and \( \eta = 1 \) with \( \gamma = 0 \) corresponds to TMMC method.

Instead of eq. (19), one may solve the optimization problem (20) for the entropies \( S_{\text{TM}}(I) \). Since in our algorithm we only slightly change the \( S_{\text{WL}}(I) \) data, we do not need high accuracy at initial stages when statistics is poor. At the later stages when large statistics is accumulated in \( C(I,J) \) array, both methods (20) and (19) give compatible results. Moreover, at the initial stages many \( C(I,J) \) elements may be zero and usage of (20) method becomes problematic because zeros in denominators of the corresponding linear system of equations may appear in eq. (23). Also, computational optimization method (20) is more time consuming than the iterative scheme (19), especially for the large systems.

In the next section, we present the results of testing our algorithm on polyalanine of different lengths.

### Results

For testing our algorithm, we calculate the density of states for peptide polyalanine of different lengths both in gas phase and in the presence of water. In all-atom model, the protein can have energies from some minimum value \( E_G \), usually called ground state energy, and up to \( \infty \). Hence, first of all, we need to find the relevant energy interval \([E_{\text{min}}, E_{\text{max}}]\) at which the density of states are to be calculated. Because of the rugged energy landscape of the energy function, it is very hard to find the exact ground state energy \( E_G \). In this paper, we do not set a goal to search for the true ground state. Instead we concentrate on some (though broad enough) physically and biologically relevant energy range \([E_{\text{min}}, E_{\text{max}}]\) around the cooperative helix/coil transition point which is of order of the room temperature. The ground state is far below this point, and its contribution to this transition is negligible. To find \( E_{\text{min}} \), we have performed several simulated annealing simulations from different random initial conformations with starting temperature 1000 K and final temperature 100 K. Then, we take the lowest value of the energy ever observed in these simulations as \( E_{\text{min}} \). The upper limit of the energy interval \( E_{\text{max}} \) was found by performing a canonical simulation at very high temperature, say 1000 K. The highest value of the energy in this simulation was taken as \( E_{\text{max}} \). All energy levels higher than \( E_{\text{max}} \) can be safely neglected since their contributions to the properties of the system at temperatures near the room temperatures (much lower than 1000 K) is negligible. The temperature 1000 K is very high for proteins but it was chosen in order to avoid trapping of the canonical simulation at any local minimum. After the range of energy is defined we divide the energy into bins with width \( \Delta E = 1 \) kcal/mol. This value seems to be optimal because the energy of the thermal fluctuations in the proteins are of order of kilocalories. All moves that propose configurations with energies out of the range \([E_{\text{min}}, E_{\text{max}}]\) are rejected and the corresponding diagonal element is updated: \( C(I, I) = C(I, I) + 1 \), where \( I \) is the initial state.

To test our algorithm, we performed 10 independent simulations of polyalanine \( \text{AL} \) (the index shows the number of amino acids in the chain) with pure WL, this algorithm (WL-TM1) and the one proposed in ref. 36 (WL-TM2). For all three algorithms, we start simulations with \( \gamma_0 = 2^{-7} \). The difference between WL-TM1 and WL-TM2 is that in WL-TM2 one replaces the density of states \( S_{\text{WL}}(I) \) by \( S_{\text{TM}}(I) \) after each number of sweeps and after all energy bins are visited, \( \gamma \) is reduced by factor 10, i.e. \( \eta = 1 \) and \( \varepsilon = 0.1 \), while \( \eta = 0.1 \) and \( \varepsilon = 0.5 \) for WL-TM1 algorithm. It will be shown below that although our algorithm only slightly improves \( S_{\text{WL}} \) after each number of sweeps, its performance is superior to both pure WL and WL-TM2 algorithms.

The microcanonical temperature \( \beta(E) \) plays a central role in flat histogram simulations. In a discrete form

\[
\beta(E) = \frac{\Delta S(E)}{\Delta E} \geq \frac{S(E + \Delta E) - S(E)}{\Delta E},
\]

it defines the acceptance probability of moving from energy state \( E \) to the nearest higher energy state \( E + \Delta E \). The probability of moving to the higher energy state. In general, the acceptance probability of a move from \( E \) to \( E' > E \) is \( P_{\text{acc}}(E, E') = \exp[-\Delta E \sum_{I=1}^{E'-E} \beta(I)] \). Knowing \( \beta(E) \), we can construct \( S(E) \)

\[
S(E) = S(E_{\text{min}}) + \Delta E \sum_{I=E_{\text{min}}}^{E} \beta(I).
\]

In Figure 4 we show the dependence of the total average (statistical) error on the number of MC sweeps for microcanonical temperature \( \beta(E) \). The total average error is calculated as the variance of \( \beta(E) \) averaged over all energy bins. A similar procedure for estimating errors was used by Shell et al.36 See also57 for discussion of the statistical errors in the broad histogram method. In WL-TM1 and WL-TM2 methods, the errors of \( \beta(E) \) were found from \( S_{\text{ITTM}}(E) \) data calculated from the ITTM through (19). To avoid division by zero in (19) we start to calculate errors from the second iteration when all energy bins are already visited at least once. Since different simulations complete the first iteration in different numbers of MC sweeps, we do not count the data from the first iteration in Figure 4.

We see that the total average error in our WL-TM1 method decreases faster than those for the corresponding WL-TM2 and WL simulations. In the case of WL simulation, the errors are simply saturated at some value and do not decrease in further simulations. In all simulations, we did 27 iterations (this means that the value of \( \gamma \) was changed 27 times). On average, our method performs 20% and 10% less MC sweeps when compared with WL and WL-TM2 methods, and produces smaller statistical errors for density of states. Errors calculated from \( S_{\text{ITTM}}(E) \) and \( S_{\text{ITTM}}(E) \) coincide after one million sweeps as shown by open and filled circles in Figure 4. Dependence of the total average error on \( \gamma \) is given in Figure 5. We see that at initial iterations, when \( \gamma \) is large, the errors in \( S_{\text{ITTM}}(E) \) data of WL-TM2 simulation (open circles) and in pure WL simulation (filled triangles) are almost of the same order, while for the smaller \( \gamma \), \( S_{\text{ITTM}}(E) \) and \( S_{\text{ITTM}}(E) \) data coincide for WL-TM1 simulation. Note that in order to reduce \( \gamma \) to the order of \( 10^{-10} \) the WL-TM2 method needs less number of iterations than WL-TM1 algorithms. But Figure 4 shows that the numbers of MC sweeps are almost the same. This means that at initial stages the WL-TM2 simulation performs more MC sweeps per iteration than WL-TM1.
Figure 4. Dependence of the total average error in $\beta(E)$ on the number of MC sweeps. Filled circles, squares, and triangles correspond to WL-TM1, WL-TM2 and pure Wang–Landau simulations, respectively. In WL-TM1 and WL-TM2 methods, the errors were calculated from the infinite temperature transition matrix, i.e. from $S_{TM}(E)$ data. Open circles correspond to the errors calculated from $S_{WL}(E)$ data in WL-TM1 simulation.

Figure 5. Dependence of the total average error in $\beta(E)$ on $\gamma$ (see the text).
method. Indeed the WL-TM2 algorithm has tendency to get stuck at low energy levels for long time since its \( \gamma \) is small, while the WL-TM1 method samples relatively easy all energy bins and updates all elements of the transition matrix much more frequently. Our simulations show that the statistical errors of transition matrix data are not much sensitive to detailed balance violation if \( \gamma \) is small enough (\( \gamma < 10^{-3} \) or so), and sampling many times all energy bins with relatively large \( \gamma \) reduces statistical errors faster than sampling all energy bins few times with smaller \( \gamma \).

Now we discuss the dependence of data on energy. In Figure 6, we plot the dependence of the average microcanonical temperature \( \beta(E) \) on energy for WL-TM1 and WL-TM2 and pure WL methods. In the same figure, we show the energy dependence of variance of \( \beta(E) \) rescaled by factor 30 for visualization convenience. Starting from the highest energy level and downward, the errors for all three methods initially remain almost unchanged (but WL data have some fluctuations) in the interval \( 50–120 \) kcal/mol. In the intermediate interval \( 30–50 \) kcal/mol the errors slightly increase, and then decrease again in the interval \( 0–30 \) kcal/mol for WL-TM1 and WL-TM2 methods. After that the errors start to increase again toward the lowest energy. The phenomenon of increasing errors near the lowest energy is well known and it is explained by the fact that the density of states of low energy levels is strongly suppressed, i.e. the configurations with these energies are visited poorly. The increase of the errors in the intermediate interval \( 30–50 \) is quite interesting. We believe that this can be explained by the well-known phenomenon of critical fluctuations near the critical (or transition) point.48 In Figure 6 we see that due to relatively large errors in \( \beta(E) \)'s in the interval \( 0–60 \), plots of \( \beta(E) \) for different simulations are slightly different within the range of statistical errors. For the sake of clearness of the picture, we do not plot the data points of \( \beta(E) \) for low energies.

Now we address an interesting question: is there any correlation between the statistical errors in \( \beta(E) \) and TTT identity violation \( \nu(E) \)? By comparing energy dependence of the average variance of \( \beta(E) \) and average TTT identity violation \( \nu(E) \) at the end of the last iteration we observe that these quantities are indeed correlated (Fig. 7). To study the correlations, we calculated the ratio \( r(E) = \text{var}[\beta(E)]/\nu(E) \) for all energy levels. We found that within 40–60%

![Figure 6. Dependence of microcanonical temperature \( \beta(E) \) and rescaled average errors in \( \beta(E) \) on energy.](image)

![Figure 7. Dependence of the variance of \( \beta(E) \) and TTT identity violation, averaged over 10 independent simulations by WL-TM1 and WL-TM2 methods for ALA16.](image)
error on average this ratio may be considered as constant greater than two for all energy levels. Linear fitting of the ratio gave results $r(E) = 2.2954 + 0.0083E$ and $r(E) = 2.0119 + 0.0065E$ for WL-TM1 and WL TM2 methods, respectively. This is a good result if we take into account how big are fluctuations in $\nu(E)$ and $\text{var}[\beta(E)]$ data (Fig. 7). This observation allows us to assume that TTT identity violation may serve as a measure of dependence of the error on energy in the simulation, at least qualitatively.

To speed up the simulations, one can divide the total energy range $E_{\text{min}}; E_{\text{max}}$ into subintervals as proposed in.24 On the basis of the pattern of the error distribution we divided the entire energy region into three parts: region I—the high energy region, region II—the transition region and region III, the low energy region, as shown in Figure 8.

We performed four WL-TM1 simulations for Ala24 in gas phase on each of the three subintervals and on the total interval. To study convergence properties, we investigate the average errors and TTT identity violation for the total interval simulation as compared to subintervals simulations. For all simulations, we performed 27 iterations except for the simulation in region III, where 54 iterations where performed for improved statistics. On average 27 iterations over the total interval required about 4 million sweeps, while total average number of sweeps for partial interval simulations (i.e. the sum of the average numbers of sweeps for each subinterval), took only 1.5 millions sweeps. The energy dependence of the microcanonical temperature $\beta(E)$, $\text{var}[\beta(E)]$, and $\nu(E)$ are shown in Figure 8. From figures (a) and (b) we see that although the number of sweeps performed for parallel simulation is more than two times less than for the total interval simulation, the distributions of $\text{var}[\beta(E)]$ and $\nu(E)$ almost coincide. Only in the region (III), the errors are a little smaller, which means that during the simulation over the whole interval the system spends most of the time in the low energy region. The advantage of parallelization is that it allows to perform many round trip simulations in all three intervals without wasting CPU time on sampling mainly the low energy levels.

The peaks in the distribution of the errors and TTT identity violation in the region (II) (Figs. 7 and 8) are explained by the cooperative nature of the helix–coil transition that takes place in the system and energy levels corresponding to the transition states lie in this region. Reducing statistical errors in this region requires many millions sweeps and is time consuming.

From the simulations in the regions I and II, we conclude that large statistical errors at the low energy levels of the interval III are not a boundary effect and arise mainly from low statistics at low energy levels, as mentioned earlier. After the simulations on each subinterval are performed one needs to glue the density of states or $\beta(E)$ from all subintervals. An efficient gluing of the density of states at the boundaries of overlapping subintervals is an open problem, and some methods are described in ref. 49.

We propose an alternative method based on reconstruction of the $C_{\text{tot}}(I,J)$ array for the entire interval from corresponding $C(I,J)$ arrays from the subinterval simulations by summing the
corresponding overlapping rows. Owing to rejection strategy used at each subinterval, it is important to omit from the partial $C(I, J)$ arrays several rows at the top or/and the bottom (see Fig. 1 and eq. (19)). Then one can solve the optimization problem (20) or apply recursive methods (19) for the $T_{tot}(I, J)$ and find the density of states over the entire interval. Another interesting method for improving the statistical properties of the density of states is to sum all corresponding density of states and then apply the gluing procedure described above.

**Discussion**

In this paper, we studied statistical properties and performance of different combinations of the Wang–Landau and transition matrix Monte Carlo methods, applied to protein simulation problem. We found that reducing continuously the statistical errors during Wang–Landau simulation as given in WL-TM1 algorithm can speed up the convergence of density of states more than drastic changes in the density of states as given in algorithm WL-TM2. We tested the algorithm (WL-TM1) by simulating the peptide Ala$_1$ and Ala$_4$ and compared its performance with the pure WL and with the algorithm proposed by Shell et al. (WL-TM2). The results of simulations show a superior performance of our algorithm with respect to both WL and WL-TM2 methods.

By analyzing the convergence of microcanonical temperature, we found correlations between the TTT identity violation for the infinite temperature transition matrix and the statistical errors in microcanonical temperature $\beta(E)$ (or microcanonical entropy $S(E)$) data calculated from the infinite temperature transition matrix. This means that TTT identity violations $\nu(E)$ may serve, at least qualitatively, as a measure of errors in the $\beta(E)$ data or, equivalently, in density of states. To measure the statistical errors in $\beta(E)$, one has to perform several simulations, which is time consuming especially for large systems. The advantage of calculating $\nu(E)$ is that it can be done at any stage of a single simulation and relatively large values of $\nu(E)$ serve as indicators of large statistical errors in density of states at those energies. During a simulation of proteins or other systems for which exact density of states is not known a priori one needs information about convergence of the density of states and $\nu(E)$ can provide such information. The investigation of TTT identity violation for WL simulation seems an interesting problem and needs more detailed study.

For the correct calculation of density of states from the transition matrix data the method of proposing moves should be symmetric and reversible, thus we choose a random angles in the interval $[-\pi, \pi]$ during simulations. Actually the acceptance rate of the simulation is quite low. It is about 18%. If we introduce some discretization of the interval $[-\pi, \pi]$ then we will break the ergodicity of the simulation which will lead to bias in the transition matrix data and, hence, large errors in the calculated density of states. The problem to overcome the low acceptance ratios in polymer simulations with all atom potentials is notoriously hard. The move proposal strategy we used in the present paper is common for standard geometry model and is similar in spirit with the well known pivot algorithm and (references therein) which was proposed for lattice simulations of polymers. This algorithm is known to have a very low acceptance rate. For a polymer with $N$ monomers the acceptance ratio goes to zero as $N \rightarrow \infty$ roughly like $N^{-p}$, where $p = 2$ or $3$ depending on the lattice dimensionality. However, the use of the pivot algorithm and the move proposal strategy in our algorithm are justified due to two facts: (1) more careful analysis shows that one accepted move may be generated in $\sim N$ moves; (2) each accepted move generates a radically new configuration. As a result the autocorrelation time is quite “acceptable.” Nevertheless, many researches have been done to modify the move proposal strategy in the standard geometry model to enhance the acceptance rate. One of the more successful strategies is the method of concerted moves based on the loop closure geometry. The corresponding computational tool has been included in our simulation package SMMP only recently and we do not use it in current work. We hope to include this technique in the described algorithm in future.

The algorithm may be parallelized by dividing the whole energy interval into three subintervals and final density of states may be found by reconstruction of the total transition matrix from transition matrices of subintervals. We propose a new gluing technique for $\beta(E)$ obtained in subintervals. In principle, one can divide the interval of simulations into more than three parts but the intervals must be large enough not to break the ergodicity of the algorithm, which may lead to large bias in the final density of states. In algorithm WL-TM1, we took $\eta = 0.1$, but this is of course not the only choice. We make some tests for larger values of $\eta = 0.5$ and found that the method behaves like WL-TM2 algorithm ($\eta = 1$). The algorithm described in this paper is quite general and can be generalized for physical quantities other than energy as well as for other systems with discrete or continuous energy spectrum. A method for gluing data from neighboring intervals proposed by Swendsen et al. can also be combined with WL-TM1 algorithm of the present paper to simulate proteins.

In conclusion, our study showed that a slight modification of density of states calculated with WL algorithm by transition matrix estimate of the density of states may significantly improve the performance of WL algorithm.

**References**