IOP Publishing Physica Scripta
Phys. Scr. **87** (2013) 058504 (6pp) doi:10.1088/0031-8949/87/05/058504

Multicanonical simulation of biomolecules and microcanonical statistical analysis of conformational transitions

Michael Bachmann

Center for Simulational Physics, The University of Georgia, Athens, GA 30602, USA

E-mail: bachmann@smsyslab.org

Received 10 October 2012
Accepted for publication 28 November 2012
Published 11 April 2013
Online at stacks.iop.org/PhysScr/87/058504

Abstract
The simulation of biomolecular structural transitions such as folding and aggregation does not only require adequate models that reflect the key aspects of the cooperative transition behaviour. It is likewise important to employ thermodynamically correct simulation methods and to perform an accurate subsequent statistical analysis of the data obtained in the simulation. The efficient combination of methodology and analysis can be quite sophisticated, but also very instructive in their feedback to a better understanding of the physics of the underlying cooperative processes that drive the conformational transition. We here show that the density of states, which is the central result of multicanonical sampling and any other generalized-ensemble simulation, serves as the optimal basis for the microcanonical statistical analysis of transitions. The microcanonical inflection-point analysis method, which has been introduced for this purpose recently, is a perfect tool for a precise, unique identification and classification of all structural transitions.

PACS numbers: 05.10.−a, 05.20.Gg, 82.35.Lr, 83.10.Tv, 87.10.−e

(Some figures may appear in colour only in the online journal)

1. Introduction

The investigation of complex systems on mesoscopic scales is so difficult that only experimental and computational studies can help gain deeper insights into the physical mechanisms that guide cooperative, qualitative changes of the system's macrostate. For biomolecular systems, folding, aggregation and substrate-adhesion transitions are certainly among the most relevant processes. Successful medical treatment of epidemic diseases that have a microbiological origin, such as Alzheimer's disease, are unthinkable without a thorough understanding of the molecular behaviour in a diverse environment. Whereas experimental techniques cannot yet track structural transitions in sufficiently high resolution, computer simulations suffer from a variety of problems, among which the lack of precise models and the failure of molecular dynamics methods to simulate nonlocal structural changes in a thermodynamically correct way [1] are most prominent. However, also the statistical analysis of the data obtained in simulation is intricate, because finite-size effects affect thermodynamic response quantities such as fluctuations of energy (specific heat) and structural order parameters in a different way, yielding somewhat diffuse information about transition points or 'transition ranges'.

The currently most efficient simulation methods are based on Monte Carlo sampling in generalized ensembles. This includes simulated [2, 3] and parallel tempering [4–6], as well as multicanonical [7–10] and Wang–Landau sampling [11].

The most popular among these methods, and most frequently used in biomolecular simulations, is probably parallel tempering, also known as replica-exchange Monte Carlo. In this method, conventional simulations like Metropolis Monte Carlo or Langevin molecular dynamics are performed independently of each other at different temperatures in parallel threads until an exchange of replicas (conformations) in neighbouring threads is attempted. This method can easily be parallelized on conventional computers or even on more specialized architectures designed for parallel operation, such as graphics processing

units [12]. Nonetheless, the method has a major drawback and this is the problem of stochastic tunnelling through entropic transition barriers, which are characteristic for first-order-like transitions. In these cases, two (or more) different phases coexist, but the pathways between these phases are suppressed entropically. In the free-energy landscape picture, this entropic gap corresponds to a barrier. Since the replica-exchange step is based on a simple Metropolis criterion in an extended (actually two-fold) ensemble represented by the threads affected by the exchange, 'climbing' the free-energy barrier is a local process (in energy space), i.e., the barrier has to be taken step by step. This results in a slowdown of the simulation at the transition point, or, in the worst case, in a reflection at the barrier. If the latter happens, the numerical results of the entire simulation are questionable because of inadequate sampling at the transition point. Although the principal problem is known, the actual origin of the barrier is typically not, because it is highly system-specific. Therefore, although many attempts have been undertaken to improve the sampling efficiency near the transition point or 'to work around it', this problem is an intrinsic feature of the method and generally cannot be resolved, because, in effect, one still simulates individual canonical ensembles. Thanks to the multiple-histogram reweighting method [13, 14], the fragments of the density of states, obtained in the individual threads, can be assembled. This yields an estimate for this fundamental quantity which is extremely useful for the analysis of structural transitions [15–18] and scaling properties of phase transitions [19].

Much more powerful in this regard are, however, flat-histogram methods, where sampling is nonlocal and effectively enables the system to *tunnel* through barriers. The most frequently used methodologies are multicanonical and Wang–Landau sampling, or the combination of both. The reason why these methods perform so much better is that the algorithms directly aim at the estimation of the density of states.

In this paper, we will review the multicanonical recursion in detail and then discuss the central role of its major results, the accurately estimated density of states and the inverse microcanonical temperature, for the unique identification and classification of structural transitions.

2. The multicanonical recursion

The multicanonical simulation method [7–9] is entirely based on microcanonical statistical mechanics. This method does not only aim at the estimation of the density of states as most generalized-ensemble methods do. Multicanonical sampling dynamics is actually based on entropic sampling governed by the density of states. The (canonical) temperature is not considered a relevant parameter of the simulation anymore. In multicanonical simulations the state space is continuously integrated over the (microcanonical) temperature. This close relationship between the fundamental quantities of statistical mechanics and the decoupling of multicanonical sampling from individual canonical ensembles makes it possible to scan the whole phase space within a single simulation with very high accuracy, even if first-order transitions occur.

The principle Boltzmann energy distribution $p_{\mathrm{can}}(E;T) \propto g(E)\mathrm{e}^{-\beta E}$ is deformed in such a way that the sampling rates of the entropically strongly suppressed energetic coexistence regimes in first-order-like transitions and of lowest-energy conformations are artificially enhanced and the modified, multicanonical ensemble possesses a flat energy distribution. Hence, the multicanonical weight $W_{\mathrm{muca}}(E;T)$ is introduced in the following way:

$$p_{\mathrm{can}}(E;T)W_{\mathrm{muca}}(E;T) \sim h_{\mathrm{muca}}(E) = \mathrm{const}_{E;T}. \quad (1)$$

Ideally, the multicanonical histogram $h_{\mathrm{muca}}(E)$ is flat and thus a constant in energy and temperature space. By this construction, the multicanonical simulation performs a random walk in energy space which leads to a rapid decrease of the autocorrelation time in entropically suppressed regions. Recalling that the simulation temperature T does not possess any meaning in the multicanonical ensemble as, according to equation (1), the energy distribution is always constant, independently of temperature. Actually, it is convenient to set it to infinity. As we will see in the following, the multicanonical recursion starts with an ordinary Metropolis run. Therefore, the performance of the multicanonical simulation in the initial recursions can be improved substantially by setting the simulation temperature to a finite value, such that the most interesting energy regions are sampled right from the beginning of the simulation. Otherwise, it can take several recursions before the random walker has dug through and reached the most interesting regions of energy space. At infinite temperature, it starts in the purely random phase, where the density of states is largest, which is energetically typically far away from transition points.

In the infinite-temperature limit, $\lim_{T\to\infty} p_{\mathrm{can}}(E;T) \sim g(E)$ and thus $\lim_{T\to\infty} W_{\mathrm{muca}}(E;T) \sim g^{-1}(E)$. Then, the acceptance probability for a conformational change $\mathbf{X} \to \mathbf{X}'$ is governed by

$$\begin{aligned} w(\mathbf{X}\to\mathbf{X}') &= \min(1, W_{\mathrm{muca}}(E(\mathbf{X}'))/W_{\mathrm{muca}}(E(\mathbf{X}))) \\ &= \min(1, g(E(\mathbf{X}))/g(E(\mathbf{X}'))). \end{aligned} \quad (2)$$

The weight function can suitably be parameterized as (in the following we set $k_{\mathrm{B}} \equiv 1$)

$$W_{\mathrm{muca}}(E) \sim \exp[-S(E)] = \exp\{-\beta(E)[E - F(E)]\}, \quad (3)$$

where $S(E)$ is the microcanonical entropy

$$S(E) = \ln g(E). \quad (4)$$

Since

$$\beta(E) = \mathrm{d}S(E)/\mathrm{d}E \quad (5)$$

is the inverse microcanonical temperature $\beta(E) = 1/T(E)$, the microcanonical free-energy scale $f(E) = \beta(E)F(E)$ and $\beta(E)$ are related to each other by the differential equation

$$\frac{\mathrm{d}f(E)}{\mathrm{d}E} = \frac{\mathrm{d}\beta(E)}{\mathrm{d}E}E. \quad (6)$$

Since $\beta(E)$ and $f(E)$ are unknown in the beginning of the simulation, this relation must be solved recursively [8, 9]. If not already being discrete by the model definition, the energy

spectrum must be discretized, i.e. neighbouring energy bins have an energetic width ΔE. Thus, for the estimation of $\beta(E)$ and $f(E)$, the following system of difference equations needs to be solved recursively. The starting point is equation (1) with $p_{\mathrm{can}}(E;T) \sim g(E)$ for an infinite simulation temperature. Since the simulated histogram will not be perfectly flat, the estimate for the density of states after the nth recursion will read $\hat{g}^{(n)}(E) \sim h^{(n)}_{\mathrm{muca}}(E)/W^{(n)}_{\mathrm{muca}}(E)$ such that the entropy can be written as

$$S^{(n)}(E) = \ln \hat{g}^{(n)}(E) = \ln h^{(n)}_{\mathrm{muca}}(E) - \ln W^{(n)}_{\mathrm{muca}}(E) + c, \quad (7)$$

where c is an unimportant constant. Then, the discrete recursive scheme to solve the above set of continuous equations for β, f, S and finally W_{muca} iteratively looks like this

$$\begin{aligned}\beta^{(n+1)}(E) &= [S^{(n)}(E) - S^{(n)}(E-\Delta E)]/\Delta E \\ &= \beta^{(n)}(E) + [\ln h^{(n)}_{\mathrm{muca}}(E) - \ln h^{(n)}_{\mathrm{muca}}(E-\Delta E)]/\Delta E,\end{aligned} \quad (8)$$

$$\begin{aligned}f^{(n+1)}(E) &= f^{(n+1)}(E-\Delta E) \\ &\quad + [\beta^{(n+1)}(E) - \beta^{(n+1)}(E-\Delta E)](E-\Delta E),\end{aligned} \quad (9)$$

$$S^{(n+1)}(E) = \beta^{(n+1)}(E)E - f^{(n+1)}(E), \quad (10)$$

$$W^{(n+1)}_{\mathrm{muca}}(E) = \exp[-S^{(n+1)}(E)]. \quad (11)$$

If no better initial guess is available, one typically sets $W^{(0)}_{\mathrm{muca}}(E) = 1$ in the beginning. The zeroth iteration thus corresponds to a Metropolis run at infinite temperature, which generates the histogram $h^{(0)}_{\mathrm{muca}}(E) = h_{\mathrm{can}}(E) =: \hat{g}^{(0)}(E)$. Thus, the histogram is already an estimate for the density of states $\hat{g}^{(0)}(E)$ such that $S^{(0)}(E) = \ln \hat{g}^{(0)}(E)$. The recursion (8)–(11) then yields the first estimate for the multicanonical weight function $W^{(1)}_{\mathrm{muca}}(E)$, which is used to initiate the second recursion, etc. The recursion procedure can be stopped after I recursions, if the weight function has sufficiently converged to a stationary distribution. The number of necessary recursions and also the number of sweeps to be performed within each recursion is model dependent. Since the sampled energy space increases from recursion to recursion and the effective statistics of the histogram in each energy bin depends on the number of sweeps, it is a good idea to increase the number of sweeps successively from recursion to recursion. Since the energy histogram should be 'flat' after the simulation run at a certain recursion level, an alternative way to control the length of the run is based on a flatness criterion. If, for example, minimum and maximum value of the histogram deviate from the mean histogram value by less than 20%, the run is stopped.

The recursive scheme (8)–(11) clearly shows how the multicanonical weights are naturally connected to microcanonical thermodynamic quantities such as temperature, entropy and free energy as functions of energy. For the implementation of the multicanonical method, however, one typically makes direct use of equation (3), which reduces the scheme by establishing a direct connection between $W^{(n+1)}_{\mathrm{muca}}(E)$ and β [8]. We simply consider the ratio

$$\frac{W^{(n+1)}_{\mathrm{muca}}(E)}{W^{(n+1)}_{\mathrm{muca}}(E-\Delta E)} = \mathrm{e}^{-[S^{(n+1)}(E) - S^{(n+1)}(E-\Delta E)]} = \mathrm{e}^{-\beta^{(n+1)}(E)\Delta E} \quad (12)$$

and immediately obtain by using equation (8)

$$\frac{W^{(n+1)}_{\mathrm{muca}}(E)}{W^{(n+1)}_{\mathrm{muca}}(E-\Delta E)} = \frac{W^{(n)}_{\mathrm{muca}}(E)}{W^{(n)}_{\mathrm{muca}}(E-\Delta E)} \frac{h^{(n)}_{\mathrm{muca}}(E-\Delta E)}{h^{(n)}_{\mathrm{muca}}(E)}. \quad (13)$$

This recursion formula for the multicanonical weights avoids the intermediate calculations in the scheme above.

The quality of the $(n+1)$th estimator for β in equation (8) can be substantially improved, if the statistics gained in the n previous recursions is also considered and an optimized estimator β_{opt} is introduced. Rewriting equation (8) as

$$\beta^{(n+1)}(E) = \beta^{(n)}_{\mathrm{opt}}(E) + [\ln h^{(n)}_{\mathrm{muca}}(E) - \ln h^{(n)}_{\mathrm{muca}}(E-\Delta E)]/\Delta E, \quad (14)$$

the optimized estimator is obtained by error-weighted superposition of $\beta^{(n)}_{\mathrm{opt}}$ and $\beta^{(n+1)}$:

$$\beta^{(n+1)}_{\mathrm{opt}}(E) = \alpha^{(n)}(E)\beta^{(n+1)}(E) + (1-\alpha^{(n)}(E))\beta^{(n)}_{\mathrm{opt}}(E) \quad (15)$$

$$\begin{aligned}&= \beta^{(n)}_{\mathrm{opt}}(E) + \alpha^{(n)}[\ln h^{(n)}_{\mathrm{muca}}(E) \\ &\quad - \ln h^{(n)}_{\mathrm{muca}}(E-\Delta E)]/\Delta E.\end{aligned} \quad (16)$$

A recursive histogram error analysis (for details see [8, 9]) yields

$$\alpha^{(n)}(E) = \frac{w_n(E)}{\sum_{i=0}^{n} w_i(E)} \quad (17)$$

with error weights

$$w_i(E) = \frac{h^{(i)}_{\mathrm{muca}}(E) h^{(i)}_{\mathrm{muca}}(E-\Delta E)}{h^{(i)}_{\mathrm{muca}}(E) + h^{(i)}_{\mathrm{muca}}(E-\Delta E)}. \quad (18)$$

Then, the optimized weights are given by

$$\frac{W^{(n+1)}_{\mathrm{muca}}(E)}{W^{(n+1)}_{\mathrm{muca}}(E-\Delta E)} = \frac{W^{(n)}_{\mathrm{muca}}(E)}{W^{(n)}_{\mathrm{muca}}(E-\Delta E)} \left(\frac{h^{(n)}_{\mathrm{muca}}(E-\Delta E)}{h^{(n)}_{\mathrm{muca}}(E)}\right)^{\alpha^{(n)}(E)}. \quad (19)$$

Finally, after the best possible estimate for the multicanonical weight function is obtained, a long multicanonical production run is performed, including all measurements of quantities of interest. From the multicanonical trajectory, the estimate of the canonical expectation value of a quantity O is then obtained at any (canonical) temperature T by

$$\overline{O}_T = \frac{\sum_t O(\mathbf{X}_t) W^{-1}_{\mathrm{muca}}(E(\mathbf{X}_t)) \mathrm{e}^{-E(\mathbf{X}_t)/k_{\mathrm{B}}T}}{\sum_t W^{-1}_{\mathrm{muca}}(E(\mathbf{X}_t)) \mathrm{e}^{-E(\mathbf{X}_t)/k_{\mathrm{B}}T}}. \quad (20)$$

Since the accuracy of multicanonical sampling is independent of the canonical temperature and represents a random walk in the entire energy space, the application of reweighting procedures is lossless. This is a great advantage of the multicanonical method, compared with Metropolis Monte Carlo simulations. A multicanonical simulation virtually samples the system behaviour at *all* temperatures simultaneously. In other words, the direct estimation of the density of states is another advantage, because multiple-histogram reweighting is not needed for this (in contrast to replica-exchange methods).

3. Microcanonical inflection-point analysis

System energy E and microcanonical entropy S are the driving forces of any transition. Therefore, it is simple to build up a systematic analysis tool for phase transitions of any system on this basis. The density of states $g(E)$ and the microcanonical entropy $S(E)$ are related to each other, according to equation (4). The basic set of fundamental statistical quantities is complete with the inverse microcanonical temperature $\beta(E) = \mathrm{d}S(E)/\mathrm{d}E$. The production run $n = I$ of the multicanonical method yields, as we have discussed in the previous section, a direct estimate $\hat{\beta}$ for the inverse temperature, which is, in accordance with equation (8), given by

$$\hat{\beta}(E) = \beta^{(I)}_{\mathrm{opt}}(E) + \ln(h^{(I)}_{\mathrm{muca}}(E)/h^{(I)}_{\mathrm{muca}}(E - \Delta E)])/\Delta E. \tag{21}$$

This single expression contains everything needed to identify and classify a thermodynamic phase transition. The reason is that any such transition is accompanied by strong energetic fluctuations while the temperature hardly changes. This empirical observation is made manifest by the fact that the specific heat possesses a maximum or a 'shoulder' at the transition temperature. Thus, the (system) energy E can generally be considered as a kind of generic 'order parameter'. The inverse microcanonical temperature β, as equation (5) tells us, describes the change of entropy S with respect to energy. If we define phase transitions, or more carefully qualitative changes of macrostates (pseudo-phase transitions), as points in β–E space, where β variations are minimal, we hold in our hands the key for a systematic approach to classify all kinds of transitions with this characteristics [18]. This is a variant of the *variational principle of least sensitivity*. What we have to look for are those inflection points of the $\beta(E)$ curve, whose derivatives are maximal. Depending on whether the value of this maximum is positive or negative, we discriminate first- and second-order transitions. Thus, if we introduce

$$\gamma(E) = \frac{\mathrm{d}\beta(E)}{\mathrm{d}E} \tag{22}$$

the order of the transition is determined by

$$\gamma_{\max} \begin{cases} > 0: & \text{first-order transition,} \\ < 0: & \text{second-order transition.} \end{cases} \tag{23}$$

The function $\gamma(E)$ describes the variation of the inverse temperature with respect to energy at a given energy value. As such it is related with the microcanonical heat capacity via $C_V(E) = [\mathrm{d}T(E)/\mathrm{d}E]^{-1} = -\beta^2(E)/\gamma(E)$.

In very small systems, it can occur that first- and second-order derivatives are not sufficiently sensitive to help identify a conformational transition. In such cases it is necessary to analyse the monotony of higher-order derivatives to locate transition points. The identification scheme represented by (23) is unique and it is independent of the thermodynamic limit. Therefore, it applies to any system of any size. In the thermodynamic limit, the thus identified transition points coincide with those determined by scaling analyses. For a second-order transition, $\gamma_{\max}$ approaches zero in this limit. First-order transitions separate into an infinite number of hierarchical subphase transitions, which extend over a finite transition range in energy space. This range corresponds to the latent heat [15].

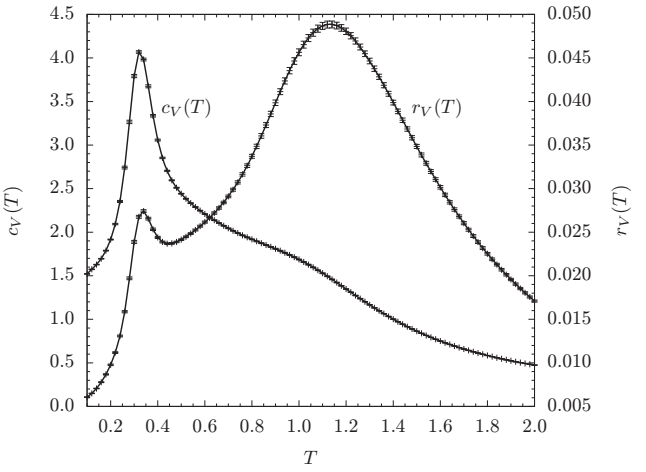

Figure 1. Specific heat $c_V(T)$ and fluctuations of radius of gyration $r_V(T)$ for a flexible elastic polymer with 20 monomers. Statistical errors were calculated by using the jackknife method.

As an example, we will discuss the transitions of an elastic, flexible polymer with $N = 20$ monomers. The polymer is modelled by a coarse-grained approach, in which monomers interact via a truncated–shifted Lennard-Jones potential, $E^{\mathrm{mod}}_{\mathrm{LJ}}(r_{ij}) = E_{\mathrm{LJ}}(\min(r_{ij}, r_{\mathrm{c}})) - E_{\mathrm{LJ}}(r_{\mathrm{c}})$ with $E_{\mathrm{LJ}}(r_{ij}) = 4\epsilon[(\sigma/r_{ij})^{12} - (\sigma/r_{ij})^{6}]$, where r_{ij} is the distance between two monomers located at $\mathbf{r}_i$ and $\mathbf{r}_j$ $(i, j = 1, \ldots, N)$, and $\epsilon = 1$ and $\sigma = 2^{-1/6}r_0$, with the potential minimum at $r_0 = 1.0$ and the cutoff at $r_{\mathrm{c}} = 2.5\sigma$. Adjacent monomers are connected by finitely extensible nonlinear elastic (FENE) anharmonic bonds [20–22], $E_{\mathrm{FENE}}(r_{i\,i+1}) = -K R^2 \ln\{1 - [(r_{i\,i+1} - r_0)/R]^2\}^{1/2}$. The FENE potential minimum is located at r_0 and diverges for $r \to r_0 \pm R$ (in our simulations $R = 3/7$). The spring constant K is set to 98/5. The total energy of a polymer conformation $\mathbf{X} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$ is given by $E(\mathbf{X}) = \sum_{i=1}^{N}\sum_{j=i+1}^{N} E^{\mathrm{mod}}_{\mathrm{LJ}}(r_{ij}) + \sum_{i=1}^{N-1} E_{\mathrm{FENE}}(r_{ii+1})$. Simulations of this model were performed using the multicanonical sampling method as described above.

As a first result, figure 1 shows the specific heat $c_V(T) = N^{-1}\mathrm{d}\langle E\rangle/\mathrm{d}T$ and the fluctuations of the radius of gyration, $r_V(T) = N^{-1}\mathrm{d}\langle R_{\mathrm{gyr}}\rangle/\mathrm{d}T$, as functions of the canonical temperature T. Both curves exhibit a peak at $T^{\mathrm{can}}_{\mathrm{ls}} \approx 0.33$, which signals the liquid–solid nucleation transition from globular to crystalline structures. The other transition is clearly indicated by the peak of r_V at $T^{\mathrm{can}}_{\Theta} \approx 1.13$. However, the specific heat has only a shoulder in this region that renders a precise location of the transition point complicated. This diffuse transition is the coil–globule transition that separates random structures ('vapour') from compact, but unstructured globules ('liquid'). In fact, from a canonical statistical perspective, finite systems do not possess unique transition points but rather span transition regions in temperature space. The collapse of fluctuations occurs only in the thermodynamic limit; the transition region shrinks to a transition point only in that limit. These transitions are generic for any kind of single polymer, including proteins [23]. Depending on size and interaction range, both transitions can be close to each other or even fall together [24–26]. Additional transitions are possible

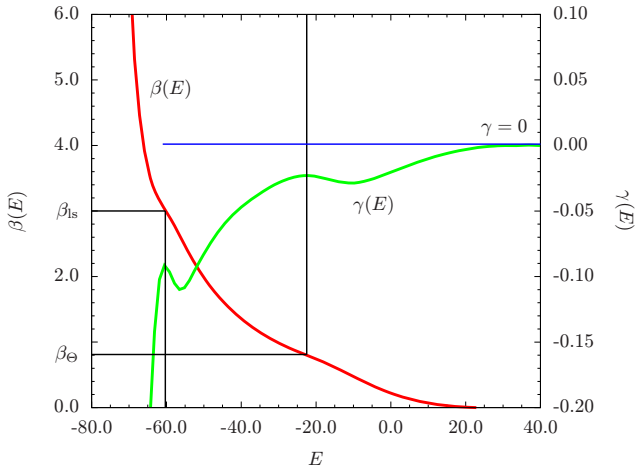

Figure 2. Microcanonical inflection-point analysis. Inverse microcanonical temperature $\beta(E)$ and its derivative $\gamma(E) = \mathrm{d}\beta(E)/\mathrm{d}E$ as functions of system energy E.

and can be classified as solid–solid transitions [18, 27]. No solid–solid transition is experienced by the system we discuss here.

The major results of the microcanonical inflection-point analysis are shown in figure 2. The inverse microcanonical temperature $\beta(E)$, obtained from equation (21) as a direct result of the multicanonical simulation, possesses two inflection points. Comparing its derivative, $\gamma(E)$, with the $\gamma = 0$ line, we find that the maximum values of γ at these inflection points are negative. Hence, both transitions have to be classified as second-order transitions. This is not particularly surprising; apart from a few exceptional cases, structural transitions of small flexible, elastic polymers tend to be second-order-like [18]. This is due to the small energetic regions, in which transitions can occur only. A first-order transition requires an extended energetic region (typically bridged by the Maxwell line that allows for the definition of the latent heat), which is only available if the energetic space is large enough. This is not the case in the example considered here. Although the inflection point at low energies signals the nucleation transition, it is of second order. It shall be noted here that the character of the same transition can be different in larger systems, where more energetic states are available. Phases might energetically be more separate, thereby creating an entropic depletion zone, which turns a second- into a first-order-like transition. The effect can also be converse, if entropically suppressed zones are filled with additionally available states, in which case a first-order transition changes to second order. In the thermodynamic limit, order–disorder transitions such as liquid–solid transitions, are typically first-order phase transitions.

From figure 2, we read off that the low-energy transition occurs at $E \approx -60.3$, which, following the line towards the β curve, is associated with a microcanonical temperature $\beta_{ls} \approx 3.0$. Thus, the microcanonically identified liquid–solid transition temperature is $T_{ls} \approx 0.33$. This perfectly agrees with the earlier canonical estimate T_{ls}^{can}, obtained from the peak positions of specific heat and from the fluctuations of the radius of gyration. The second γ peak in figure 2 at $E \approx -22.5$ signals the Θ transition and corresponds to the inverse temperature $\beta_\Theta \approx 0.81$, or $T_\Theta \approx 1.23$, which is close

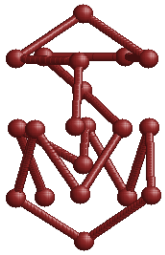

Figure 3. Lowest-energy conformation of the 20 mer ($E_{min} \approx -73.9$) found in the simulation. The density of states at this energy is by a factor 10^{80} smaller than in the random-coil phase. Note that it does not possesses a rotational symmetry (not even if the bonds are ignored).

to, but slightly larger than, the canonical estimate for the Θ temperature obtained from r_V. Whereas the microcanonical estimate is unique and does not leave any space for ambiguity, the canonical estimates depend on the fluctuating quantity chosen and are multivalued.

The precise estimation of the density of states is the main challenge in this combination of multicanonical simulation and subsequent microcanonical analysis. Even in the simple example discussed here, the density of states spans 80 orders of magnitude between the random states (at high energy) and the energetic states resembling the ground state or global energy minimum. The conformation with the lowest energy found in the simulation is shown in figure 3.

4. Summary

We have shown that the multicanonical Monte Carlo method and the microcanonical statistical analysis are closely intertwined in a natural way. Multicanonical sampling is based on the density of states and aims at a direct estimate of this quantity and its advanced version, the multicanonical recursion, intrinsically calculates the first derivative of the microcanonical entropy with respect to energy. The enormous importance of this quantity—the inverse microcanonical temperature—lies in the fact that all qualitative changes of the system are encoded in it. By systematic analysis of its inflection points, all phase transitions can be uniquely identified and even classified. We have demonstrated the power of this method for a simple homopolymer example which exhibited the usual coil–globule transition and a condensation transition towards a crystalline state. However, microcanonical inflection-point analysis can be applied to any system, independently of whether it is small or large. Therefore, this method is particularly useful for the unique identification of structural transitions in biomolecular systems.

Acknowledgment

This work was partially supported by the NSF under grant no. DMR-1207437.

References

[1] Schluttig J, Bachmann M and Janke W 2008 *J. Comput. Chem.* **29** 2603
[2] Marinari E and Parisi G 1992 *Europhys. Lett.* **19** 451

[3] Lyubartsev A P, Martsinovski A A, Shevkunov S V and Vorontsov-Velyaminov P N 1992 *J. Chem. Phys.* **96** 1776
[4] Swendsen R H and Wang J S 1986 *Phys. Rev. Lett.* **57** 2607
[5] Hukushima K and Nemoto K 1996 *J. Phys. Soc. Japan* **65** 1604
Hukushima K, Takayama H and Nemoto K 1996 *Int. J. Mod. Phys.* C **7** 337
[6] Geyer C J 1991 Computing Science and Statistics *Proc. 23rd Symp. on the Interface* ed E M Keramidas (Fairfax Station: Interface Foundation) p 156
[7] Berg B A and Neuhaus T 1991 *Phys. Lett.* B **267** 249
Berg B A and Neuhaus T 1992 *Phys. Rev. Lett.* **68** 9
[8] Janke W 1998 *Physica* A **254** 164
Berg B A 2000 *Fields Inst. Commun.* **26** 1
[9] Berg B A 2004 *Markov Chain Monte Carlo Simulations* (Singapore: World Scientific)
[10] Bachmann M and Janke W 2003 *Phys. Rev. Lett.* **91** 208105
[11] Wang F and Landau D P 2001 *Phys. Rev. Lett.* **86** 2050
Wang F and Landau D P 2001 *Phys. Rev.* E **64** 056101
[12] Gross J, Janke W and Bachmann M 2011 *Comput. Phys. Commun.* **182** 1638
[13] Ferrenberg A M and Swendsen R H 1989 *Phys. Rev. Lett.* **63** 1195
[14] Kumar S, Bouzida D, Swendsen R H, Kollman P A and Rosenberg J M 1992 *J. Comput. Chem.* **13** 1011
[15] Junghans C, Bachmann M and Janke W 2006 *Phys. Rev. Lett.* **97** 218103
Junghans C, Bachmann M and Janke W 2009 *Europhys. Lett.* **87** 40002
Junghans C, Janke W and Bachmann M 2011 *Comput. Phys. Commun.* **182** 1937
[16] Möddel M, Janke W and Bachmann M 2010 *Phys. Chem. Chem. Phys.* **12** 11548
[17] Bereau T, Bachmann M and Deserno M 2010 *J. Am. Chem. Soc.* **132** 13129
Bereau T, Deserno M and Bachmann M 2011 *Biophys. J.* **100** 2764
[18] Schnabel S, Seaton D T, Landau D P and Bachmann M 2011 *Phys. Rev.* E **84** 011127
[19] Behringer H and Pleimling M 2006 *Phys. Rev.* E **74** 011108
[20] Bird R B, Curtiss C F, Armstrong R C and Hassager O 1987 *Dynamics of Polymeric Liquids* 2nd edn (New York: Wiley)
[21] Kremer K and Grest G S 1989 *J. Chem. Phys.* **92** 5057
[22] Milchev A, Bhattacharaya A and Binder K 2001 *Macromolecules* **34** 1881
[23] Shakhnovich E 2006 *Chem. Rev.* **106** 1559
[24] Rampf F, Paul W and Binder K 2005 *Europhys. Lett.* **70** 628
Paul W, Strauch T, Rampf F and Binder K 2007 *Phys. Rev.* E **75** 060801
[25] Taylor M P, Paul W and Binder K 2009 *J. Chem. Phys.* **131** 114907
Taylor M P, Paul W and Binder K 2009 *Phys. Rev.* E **79** 050801
[26] Gross J, Neuhaus T, Vogel T and Bachmann M 2013 *J. Chem. Phys.* **138** 074905
[27] Schnabel S, Vogel T, Bachmann M and Janke W 2009 *Chem. Phys. Lett.* **476** 201
Schnabel S, Bachmann M and Janke W 2009 *J. Chem. Phys.* **131** 124904