Hybrid Monte Carlo with adaptive temperature choice: efficient conformational analysis of RNA

A. Fischer, F. Cordes, C. Schütte [1]

Konrad Zuse Zentrum für Informationstechnik, Takustrasse 7, 14195 Berlin, Germany

Abstract

A hybrid Monte Carlo method with adaptive temperature choice is presented, which exactly generates the distribution of a mixed-canonical ensemble composed of two canonical ensembles at low and high temperature. The analysis of resulting Markov chains with the reweighting technique shows an efficient sampling of the canonical distribution at low temperature, whereas the high temperature component facilitates conformational transitions, which allows shorter simulation times. © 1999 Published by Elsevier Science B.V. All rights reserved.

1. Introduction

Efficient sampling of phase space for complex biological systems requires crossing of energy barriers, which are large compared to the thermal energy. Different theoretical approaches exists in this context [1, 2]. The energy barriers separate clusters in configuration space [3], biochemically known as conformations with respect to similarities in structure and function. RNA molecules are highly flexible and appear in different conformations depending on environmental conditions. Some conformations exhibit biochemical functions as catalytic activity, specific recognition and inhibition of other molecules. The triribonucleotide adenylyl$(3'–5')$cytidylyl$(3'–5')$cytidin $[r(ACC)]$, a small RNA segment, serves as the model system of this study. Its global structure can be roughly described by eight parameters per nucleotide (Fig. 1).
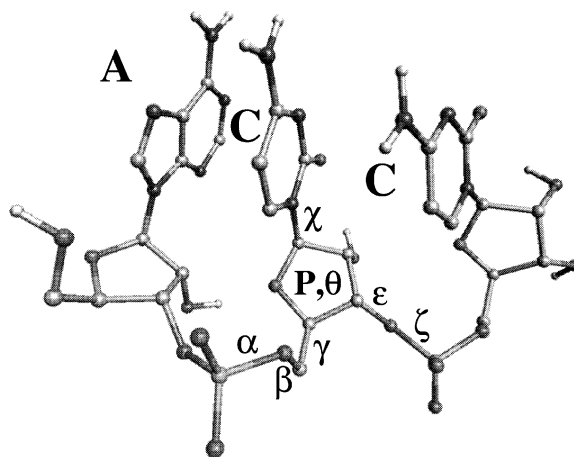

Fig. 1. The triribonucleotide adenylyl$(3'–5')$cytidylyl$(3'–5')$cytidin $[r(ACC)]$. Small greek letters refer to the set of torsion angles, which is necessary for a rough reconstruction of the nucleotide's configuration.

2. Method

Given a separable Hamiltonian H and a potential V, which is parameterized in this study according to the
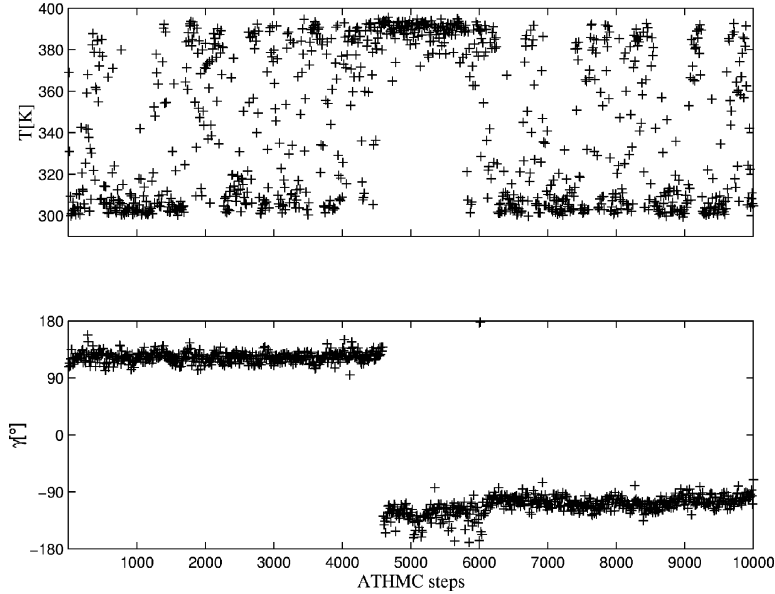
---

[1] E-mail: cordes@zib.de.

Fig. 2. ATHMC for $r(ACC)$ in the mixed-canonical ensemble. The simulation was performed for $\mathcal{T}^- = 295$ K, $\mathcal{T}^+ = 400$ K and $c = -1121$ kJ/mol. The temperature $\mathcal{T}$, and for the cytidylyl group the torsion angle γ are displayed at every tenth step.

semi-empirical force field GROMOS96 [4], adaptive temperature hybrid Monte Carlo (ATHMC) samples an ensemble of configurations described by a mixed-canonical distribution $\mu(x) = \mu^*(x)/Z_\mu$ with [2]

$$\mu^*(x) = \tfrac{1}{2}\big(\exp\!\big[-\beta^-\big(V(x) - c\big)\big] \\ + \exp\!\big[-\beta^+\big(V(x) - c\big)\big]\big), \qquad (1)$$

calibration parameter c and normalizing constant Z_μ, according to the following scheme:

(1) Draw momenta p from a normal distribution at inverse temperature $\beta(x)$, which is known from the previous step.

(2) Propose new coordinates x' and momenta p' by integrating the system with a reversible and volume preserving discretized flow, e.g., choosing the Verlet integrator.

(3) Compute new inverse temperature due to

$$\beta(x) = -\frac{\ln \mu^*(x)}{V(x) - c}.$$

(4) Accept new coordinates x' with a probability

---

[2] $\beta = 1/(k_{\mathrm{B}}T)$, $k_{\mathrm{B}} =$ Boltzmann's constant.

$$P_{\mathrm{acc}} = \min\!\left(1, \frac{\mu^*(x')\exp[-\beta(x')\,T(p')]}{\mu^*(x)\exp[-\beta(x)\,T(p)]} \right. \\ \left. \times \left(\frac{\beta(x')}{\beta(x)}\right)^{s/2}\right), \qquad (2)$$

otherwise stay in old coordinates x; s denotes the number of degrees of freedom.

Temperature choice and acceptance probability exploits the special structure of the mixed-canonical ensemble. Because detailed balance is satisfied for any arbitrary temperature, we search for a temperature function $\beta = \beta(x)$, which depends on the actual potential energy in such a way, that the Boltzmann factor at $\beta(x)$ is equal to $\mu^*(x)$. Whenever the potential energy increases in the vicinity of c, kinetic energy is pumped into the system according to the choice of higher temperature for the generation of momenta. Additionally, proposals with higher energy are accepted more easily in μ and the system can move towards higher energy regions, where conformational changes happen more often. Because thermodynamical averages of the canonical distribution have to be calculated by the reweighting method, the low temperature part is important for the statistical reliability. For optimal temperature fluctuation in terms of transitions and sam-
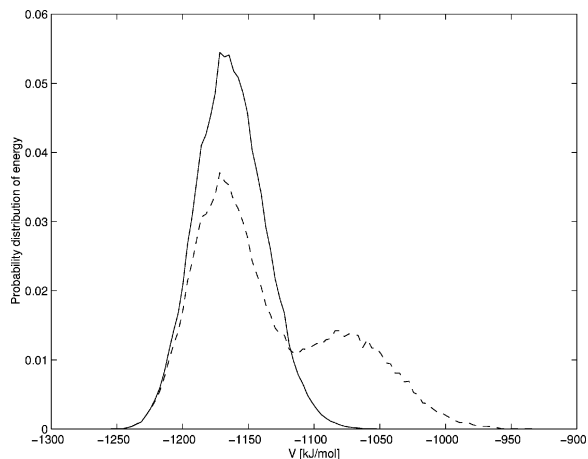
Fig. 3. ATHMC for $r(ACC)$ in the mixed-canonical ensemble. Probability distribution of energy (- - -) before and (—) after reweighting.

pling, c should be chosen slightly above the averaged potential energy.

3. Results

The dependence between temperature choice and conformational transitions in ATHMC is shown in Fig. 2, where we have focused on a section of a longer run (for details see [5]). The γ torsion angle at the beginning is around $120°$. The heating of the system due to the choice of momenta according to higher temperature induces the necessary transition of γ around step 4500. A conventional HMC run was unable to induce a transition to another state [5].

The fact, that ATHMC samples at different temperatures with sufficient rates, is furthermore illustrated by the probability distribution of energy before and after reweighting (Fig. 3). Without reweighting we ob-

serve a maximum around the averaged potential energy, but another distribution peak for higher energies, exactly enforced by the choice of higher temperatures and the non-negligible acceptance at higher energies. Fig. 3 makes the strategy of generalized ensembles very clear, that is to overcome energy barriers by sampling over an extended energy range, which finds its extreme realization in the multi-canonical approach [2]. ATHMC also stretches the energy distribution, but it is conceptually still connected to the canonical distribution of interest.

4. Conclusion

The adaptive temperature choice coupled with a generalized, mixed-canonical ensemble was discussed to be responsible for efficient sampling properties. Like all other strategies based on generalized ensembles the proposed algorithm cannot relinquish pre- and postprocessing procedures. But preprocessing in ATHMC needs only one parameter, c, which corresponds to the averaged potential energy of the system.

References

[1] B.J. Berne, J.E. Straub, Curr. Opinion Struct. Biol. 7 (1997) 181.
[2] U.H.E. Hansmann, Y. Okamoto, J. Comput. Chem. 18 (1997) 920.
[3] C. Schütte, A. Fischer, W. Huisinga, P. Deuflhard, Preprint SC 98-04, Konrad-Zuse-Zentrum Berlin (1998).
[4] W.F. van Gunsteren et al., Biomolecular Simulation: The GROMOS96 Manual and User Guide, vdf Hochschulverlag AG (1996).
[5] A. Fischer, F. Cordes, C. Schütte, J. Comput. Chem. 19 (1998), to appear.