PHYSICAL REVIEW E VOLUME 53, NUMBER 4 APRIL 1996

Generalized simulated annealing algorithms using Tsallis statistics: Application to conformational optimization of a tetrapeptide

Ioan Andricioaei and John E. Straub
Department of Chemistry, Boston University, Boston, Massachusetts 02215
(Received 18 December 1995)

A Monte Carlo simulated annealing algorithm based on the generalized entropy of Tsallis is presented. The algorithm obeys detailed balance and reduces to a steepest descent algorithm at low temperatures. Application to the conformational optimization of a tetrapeptide demonstrates that the algorithm is more effective in locating low energy minima than standard simulated annealing based on molecular dynamics or Monte Carlo methods.

PACS number(s): 02.70.−c, 02.60.Pn, 02.50.Ey

Finding the ground state conformation of biologically important molecules has an obvious importance, both from the academic and pragmatic points of view [1]. The problem is hard for biomolecules, such as proteins, because of the ruggedness of the energy landscape which is characterized by an immense number of local minima separated by a broad distribution of barrier heights [2,3]. Algorithms to find the global minimum of an empirical potential energy function for molecules have been devised, among which a central role is played by the simulated annealing methods [4]. Once a cooling schedule is chosen, representative configurations of the allowed microstates are generated by methods either of the molecular dynamics (MD) or Monte Carlo (MC) types. For biomolecular simulations, simulated annealing is traditionally built on an MD approach [5] where the dynamics of the system is simulated by integrating the Newtonian equations of motion and the temperature is controlled through coupling to a heat bath. If the MC approach is used, after having drawn a new configuration, it is accepted or rejected according to a probability of, for example, the Metropolis type [6]

$$p=\min[1,\exp(-\beta\Delta E)], \tag{1}$$

where $\beta=1/kT$ and ΔE is the change in potential energy. This acceptance probability has the desirable features that (i) it obeys detailed balance and (ii) it reduces to a steepest descent minimizer at low temperature (where only moves which decrease the potential energy are accepted). In addition to the standard Metropolis Monte Carlo protocol, several other smarter MC algorithms have been designed using atomic moves biased by the forces acting upon the atoms in the molecule [7,8] or by relaxing the restriction to Markov processes [9].

Recently, a new type of updating criterion has been proposed based upon the generalized statistical mechanics of Tsallis [10]. In the Tsallis formalism, a generalized statistics is built from the generalized entropy

$$S_q=k\,\frac{1-\Sigma p_i^q}{q-1} \tag{2}$$

where q is a real number and S_q tends to the information entropy

$$S=-k\sum p_i\ln p_i \tag{3}$$

when $q\to 1$. Maximizing the Tsallis entropy with the constraints

$$\sum p_i=1 \quad \text{and} \quad \sum p_i^q\epsilon_i=\text{const}, \tag{4}$$

where ϵ_i is the energy spectrum, the generalized probability distribution is found to be

$$p_i=[1-(1-q)\beta\epsilon_i]^{1/(1-q)}/Z_q, \tag{5}$$

where Z_q is the generalized partition function. This distribution goes to the Gibbs-Boltzmann distribution when q tends to 1. It has been demonstrated that this generalized statistics preserves the Legendre transformations between thermodynamic state functions [11], leaves form-invariant, for any q, the von Neumann equation [12] and the Ehrenfest theorem [13], can be used to write, among others, a generalized Boltzmann H-theorem [14], Langevin and Fokker-Planck equations [15], and the fluctuation-dissipation theorem [16].

Inspired by this generalized statistics, a new generalized simulated annealing algorithm has been suggested based on the acceptance probability

$$p=\min[1,[1-(1-q)\beta\Delta E]^{1/(1-q)}] \tag{6}$$

where ΔE is the change in the potential energy. This probability has the desirable property that it reduces to a steepest descent in energy for $T\to 0$. This method was shown [10] to be faster than both the classical simulated annealing and the fast simulated annealing methods [17] and was employed to find close to optimal solutions to the traveling salesman problem [18].

However, it is easy to demonstrate that the acceptance criterion (6) does not obey detailed balance. It is true that detailed balance is a sufficient but not necessary condition for the convergence to the equilibrium distribution. Even so, the acceptance in (6) does not, in general, converge towards the generalized distribution of (5).

We propose a generalized acceptance probability,

$$p=\min\left[1,\left(\frac{1-[1-q(T)]\beta E_{\text{new}}}{1-[1-q(T)]\beta E_{\text{old}}}\right)^{q(T)/[1-q(T)]}\right],$$

$$\lim_{T\to 0} q(T)=1 \tag{7}$$

that obeys detailed balance. Moreover, this acceptance probability does tend towards the generalized equilibrium distribution in (5). The parameter q is varied as a monotonically decreasing function of temperature. Starting with a convenient value of q at the initial temperature, q tends towards 1 as the temperature decreases during annealing. Since $q\to 1$ as $T\to 0$, the desirable reduction to a steepest descent at low temperature is preserved.

Note that the probability is raised to the power of q as required by the generalized statistical mechanics (5). In this generalization, the average of an observable O is defined as $O=\Sigma p_i^q O_i$ and thus the detailed balance must be written as $W_{ij}p_i^q=W_{ji}p_j^q$, where W_{ij} are the elements of the transition matrix. The fact that the probability distribution is raised to the power of q allows one to use simulation binning to compute average properties of the system according to the standard definition of the statistical average [19].

This acceptance probability is, in the spirit of the generalized statistical mechanics of Tsallis, invariant under the transformation

$$\bar{E}_i\to\bar{E}_i+\bar{E}_0, \tag{8}$$

where $\bar{E}_0$ is a constant shift in energy and, by definition,

$$\bar{E}=\frac{q}{\beta(q-1)}\ln[1-(1-q)\beta E]. \tag{9}$$

For $q\to 1$ the property that the probability distribution does not depend on the choice of the zero of energy is thus recovered. Also, using the definition (9) the acceptance probability (7) can be written in the more familiar form

$$p=\min[1,\exp(-\beta\Delta\bar{E})]. \tag{10}$$

It was shown [20] that when the maximum entropy formalism is applied to the entropy postulated by Tsallis (2) one is able to recover the general Lévy probability distribution (corresponding to a fractal random walk the dimension of which is determined by q). A variational entropic formalism based on the Boltzmann entropy is unable to do this. Using the acceptance probability proposed here and $q(T)>\frac{5}{3}$, a Markov chain generated at constant temperature will converge to a Lévy distribution of the length of flights. For the particular case of $q=2$, the Lévy distribution is a Cauchy-Lorentz distribution which is the same distribution used for the fast simulated annealing method of Szu and Hartley [17].

The Metropolis acceptance criterion in (7) is not the only solution that guarantees convergence towards the probability distribution in (5). It is known that a solution of the type proposed by Barker

$$W_{ij}=\alpha_{ij}\frac{\rho_j}{\rho_j+\rho_i}, \tag{11}$$

where α_{ij} is the *a priori* symmetric transition matrix and ρ_i the probability distribution function for the ith state, also satisfies detailed balance [21]. Thus, by constructing the acceptance probability using (5) and (11) we recover the familiar form of acceptance widely used in simulations of spin systems

$$p=\tfrac{1}{2}[1-\tanh(\beta\Delta\bar{E}/2)]. \tag{12}$$

A simulation using this acceptance probability will tend towards the generalized probability distribution of Tsallis, as does (7). It will obey detailed balance and, by making $q\to 1$ as the temperature decreases, it will behave like a steepest descent at low temperatures.

In this work we have applied simulated annealing protocols to the conformational optimization of the 48 atom tetraalanine peptide [isobutyl-(ala)$_3$-methlyamide]. We have generated 50 random (uncorrelated) initial configurations of the tetraalanine by "amnesiating" the system using constant temperature MD runs at 3000 K. The annealing was performed using both molecular dynamics simulations and stochastic simulations of the Monte Carlo type.

A Monte Carlo simulation using individual atomic moves drawn from a certain distribution function has been shown to be inefficient for polymeric systems [5]. Most of the substantial moves are rejected since they stretch the molecule to positions with high potential energies. We have utilized a method that uses the features of both stochastic methods, such as MC, and deterministic methods, such as MD. This hybrid MD-MC method was first devised in quantum chromodynamics for simulations of lattice field theory. It combines the convenience of integrating Newton's equations of motion with the absence of discretization errors usually involved in such calculations to give a Monte Carlo–like program [22]. We perform the simulation in three steps. (i) We select velocities from a Maxwell distribution compatible with the system being in contact with a heat bath at a well defined temperature. (ii) We evolve the positions and the velocities by integrating the Newtonian equations of motion over a small time interval (50 fs) to generate a new point in the phase space. (iii) We accept or reject this new phase point according to a Metropolis procedure to return to (i). It is well known that, as long as the discretization of the equations of motion is reversible and preserves the phase space volume, an acceptance probability such as (1) will exactly satisfy detailed balance and the system will tend towards the Maxwell-Boltzmann distribution [23].

For both the MD and hybrid MD-MC methods, we use the version 19 CHARMM (Chemistry at Harvard Molecular Mechanics) force field [24], containing harmonic bond and angle terms, dihedral, electrostatic and Lennard-Jones terms. There was no nonbonded potential energy truncation and the integration was performed using the velocity Verlet integrator. For the cooling schedule, the initial temperature is set to 3000 K and decreased exponentially according to the rule $T_{i+1}=(1-\alpha)T_i$, where $\alpha=10^{-3}$. This is done until the temperature reaches 30 K, after which a short refinement by conjugate gradient minimization is performed. The value of q is also decreased exponentially to 1, starting from values higher than 1 (typically $q=2$). The protocol of exponential reduction of temperature and variation of q employed here

are useful for comparing optimization methods but not optimal. In practice, it may be further optimized for a given problem. Note that since $q>1$, we never have to face the awkward problem of dealing with complex probabilities. Using the acceptance probability (6) of Tsallis with $q<1$, one needs to set the acceptance probability equal to zero whenever the argument of the power law acceptance function (6) is negative.

The generalized simulated annealing [10] extends classical simulated annealing and fast simulated annealing and introduces the parameters (q_a,q_v), where q_a corresponds to our q and q_v is related to the visiting distribution of the random phase points. It was suggested that an optimal q_a would be negative and large in absolute value, because in this case the acceptance would be biased towards acceptance of lower energies. However, in our application to the particular case of biomolecules, the choice is towards positive q where the objective is to explore phase space more effectively. Values of $q<1$ lead to a more effective *local optimization* and refinement, while values of $q>1$ are expected to be more effective in *global optimization.* This is consistent with the findings for a harmonic oscillator, where it was shown that the equilibrium configurational distribution broadens with increasing q [25,26].

Should we wish to relax the $q>1$ requirement for the purpose of testing negative values of q, we might have a complex probability when the new energy is higher and, respectively, the old one is lower than $kT/(1-q)$. However, because of the way in which the acceptance is written, we can choose to shift the potential in such a way that all the potential energies are higher than $kT/(1-q)$ even at the highest T. The potential shift can also be used to tune the acceptance ratio at high temperatures. In this work we take the zero of energy as the natural zero of the CHARMM potential.

We have compared annealing with (i) MD, (ii) Metropolis MC using the Gibbs-Boltzmann statistics (1), and Tsallis statistics using the protocol defined above (positive q, starting at 2 and converging to 1) using the (iii) generalized Metropolis MC with acceptance probability (7) or (iv) the generalized Barker MC with acceptance (12). The results are shown in Fig. 1 where the average dependence over 50 runs is presented. For the same number of CPU time steps, MD goes to a state that is a high local minimum (even after conjugate gradient refinement). For the same CPU time, the standard MC method barely has time to accept a few steps, while the method proposed here shoots down to a position that after refinement by a conjugate gradient minimizer is very close to the global minimum.

Simulated annealing on the tetrapeptide with the Barker acceptance probability (12) (see Fig. 1) requires a CPU time a factor of about 1.5 larger than the algorithm based on (7). This results from the fact that the off-diagonal elements of the transition probability are larger, for the same value of $\Delta\bar{E}$, in the case of the acceptance in (7) than in (12). This is in accord with the classical Monte Carlo schemes where it has been shown [27] that the Metropolis solution leads to a lower statistical inefficiency than the Barker solution. Even so, the acceptance in (12) is much faster than classical simulated annealing using the acceptance in (1). Even if in the particular case of the tetrapeptide the acceptance in (7) is employed, there might be cases (for example in two-state

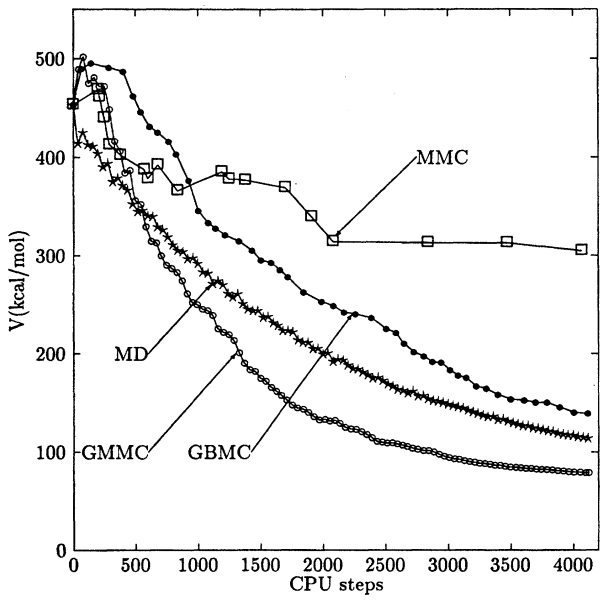

FIG. 1. The potential energy of tetraalanine is plotted versus computer time for the average of the 50 runs starting from random initial configurations. Comparison is made between annealings using the generalized Metropolis Monte Carlo (GMMC) of (7), the generalized Barker Monte Carlo (GBMC) of (12), molecular dynamics (MD), and the standard Metropolis Monte Carlo in (1) (MMC).

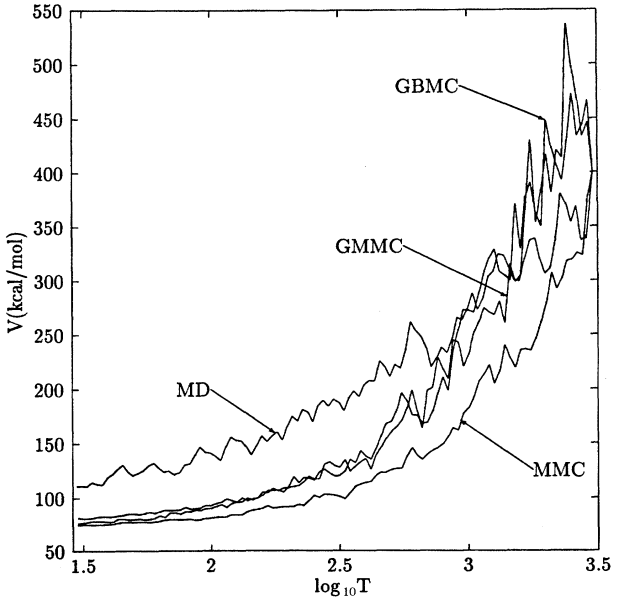

FIG. 2. The potential energy for tetraalanine is plotted versus the logarithm of the temperature. Note the expected smaller fluctuations in the generalized Barker Monte Carlo (GBMC) when compared to the generalized Metropolis Monte Carlo of (7) (GMMC). Metropolis Monte Carlo (MMC) took on the order of 30 times more computer time than GMMC and the molecular dynamics trajectory (MD) was trapped in high metastable states.

systems) where (12) would perform better. The low statistical inefficiency is just one of the criteria to be met in choosing between alternative transition matrices [21].

Nine of the runs done with the acceptance probability (7) led to the global minimum. The others led to lower lying minima, well below the minima the standard MD and MC simulated annealing methods found in the same CPU time. The comparison of our method with MD is shown in Fig. 2. Neither MD nor Metropolis MC found the global minimum in the same CPU time and both yielded final configurations higher in energy than the method proposed here.

One component of the success of annealing with $q>1$ is that by using the generalized acceptance criterion the effect is similar to using a less steep cooling schedule. Therefore, a larger fraction of the annealing run is spent at a high effective temperature and the probability of getting trapped in high local minima is reduced. As measured in annealing steps, lower lying energy minima are approached faster using the acceptance probability (1) than when using the method proposed here, as shown in Fig. 2. However, the amount of CPU time spent at each temperature step is *much higher* in the case of the standard Metropolis criterion than in the case of the acceptance criterion (7). For the data in Fig. 2 it took on the order of 30 times more computer time to anneal between the same temperatures using the same cooling schedule. This demonstrates the greater computational efficiency and improved sampling of this generalized simulated annealing algorithm.

J.E.S. received support from the Alfred P. Sloan Foundation and recognizes the generous support of the National Science Foundation (CHE-9306375). We thank Richard Friesner and Constantino Tsallis for helpful comments in the early stages of this work and Patricia Amara for comments on the manuscript.

[1] E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines* (Wiley, New York, 1990).

[2] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, Science **254**, 1598 (1991); H. Frauenfelder, F. Parak, and R. D. Young, Annu. Rev. Biophys. Chem. **17**, 451 (1988).

[3] J. E. Straub and J.-K. Choi, J. Phys. Chem. **98**, 10 978 (1994); J. E. Straub and D. Thirumalai, Proc. Natl. Acad. Sci. USA **90**, 809 (1993); Proteins **15**, 360 (1993); J. E. Straub, A. Rashkin, and D. Thirumalai, J. Am. Chem. Soc. **116**, 2049 (1995).

[4] S. Kirkpatrick, C. D. Gellat, Jr., and M. P. Vecchi, Science **220**, 671 (1983).

[5] S. H. Northrup and J. A. McCammon, Biopolymers **19**, 1001 (1980).

[6] N. Metropolis, A. W. Rosenbluth, N. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[7] C. Pangali, M. Rao, and B. J. Berne, Chem. Phys. Lett. **55**, 413 (1978).

[8] P. J. Rossky, J. D. Doll, and H. L. Friedman, J. Chem. Phys. **69**, 4628 (1978).

[9] R. H. Swendsen, D. Bouzida, and S. Kumar, Carnegie Mellon University Technical Report No. 91-121-NAMS-27, 1991 (unpublished).

[10] D. A. Stariolo and C. Tsallis, *Annual Reviews of Computational Physics II*, edited by D. Stauffer (World Scientific, Singapore, 1995), p. 343.

[11] E. M. F. Curado and C. Tsallis, J. Phys. A: Math. Gen. **24**, L69 (1991)

[12] A. R. Plastino and A. Plastino, Physica A **202**, 438 (1994).

[13] A. R. Plastino and A. Plastino, Phys. Lett. A **177**, 177 (1993).

[14] A. M. Mariz, Phys. Lett. A **165**, 409 (1992).

[15] D. A. Stariolo, Phys. Lett. A **185**, 262 (1994).

[16] A. Chame and E. V. L. de Mello, J. Phys. Math A: Math. Gen. **27**, 3663 (1994).

[17] H. Szu and R. Hartley, Phys. Lett. A **122**, 157 (1987).

[18] T. J. P. Penna, Phys. Rev. E **51**, R1 (1995).

[19] If the acceptance probability is chosen to be $p^{1/q(T)}$, where p is the acceptance probability in (7), then the limiting configurational probability distribution will be that of (5). However, during an equilibrium simulation, average properties calculated as $O=\Sigma p_i O_i$ will diverge, whereas averages based on the acceptance (7), $O=\Sigma p_i^q O_i$, will converge [25]. As a simple example, consider the average potential energy of a harmonic oscillator and let $q=2$. Then $\int_0^\infty p(x)x^2dx \propto \int_0^\infty (1+\beta x^2)^{-1}x^2dx$ diverges, while $\int_0^\infty p^2(x)x^2dx \propto \int_0^\infty (1+\beta x^2)^{-2}x^2dx$ converges.

[20] P. A. Alemany and D. H. Zanette, Phys. Rev. E **49**, R956 (1994).

[21] J. P. Valleau and S. G. Whittington, in *Modern Theoretical Chemistry. Vol. 5. Statistical Mechanics A*, edited by B. J. Berne (Plenum, New York, 1977).

[22] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987).

[23] Our implementation differs from the hybrid MD-MC method [22], proposed in the context of lattice gauge theories, where the acceptance probability depends on the change of the full Hamiltonian. Instead, we are looking only at the change ΔV of the potential energy and use the velocities only as propagators of the Markov chain. We still sample the configurational space according to a Boltzmann distribution $e^{-\beta V}$ if we use the updating criterion in (1) or to a generalized Tsallis distribution, (5), if we use (7) for updating.

[24] B. R. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).

[25] C. Tsallis, A. M. C. de Souza, and R. Maynard, in *Lévy Flights and Related Phenomena in Physics*, edited by M. F. Shlesinger, U. Frisch, and G. M. Zaslavsky (Springer, Berlin, 1995), p. 269; C. Tsallis, S. V. F. Levy, A. M. C. Souza, and R. Maynard, Phys. Rev. Lett. **75**, 3589 (1995).

[26] Depending on the characteristics of both the energy landscape and the configurational phase point propagation one may tune the value of q to optimize the annealing success rate.

[27] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1992).