Multicanonical jump walk annealing: An efficient method for geometric optimization

Huafeng Xu and B. J. Berne[a)]
Department of Chemistry and Center for Biomolecular Simulation, Columbia University, 3000 Broadway, New York, New York 10027

(Received 6 October 1999; accepted 16 November 1999)

A new global optimization method, multicanonical jump walk annealing (MJWA), is proposed and applied to the geometric optimization of Lennard-Jones and Morse clusters and the hydrophobic (B), hydrophilic (L), and neutral (N) (BLN) protein model. The method efficiently finds the global minima of these systems. In four comparative studies, MJWA greatly outperforms the conventional simulated annealing in locating the global minima. Theoretical comparison with other global optimization methods is discussed. Through this paper, we demonstrate a criterion for devising stochastic global optimization schemes. Namely, a stochastic global optimization method must favor the global minimum thermodynamically and at the same time be able to cross the high energy barriers. © *2000 American Institute of Physics.* [S0021-9606(00)00706-6]

I. INTRODUCTION

Efficient global optimization methods are crucial to the study of many scientific and engineering problems. For a specific class of problems, known as NP-complete problems,[1] however, a deterministic polynomial time solution doesn't exist. In other words, if the size of the system is denoted as N, then the computational cost required to deterministically find the global optimum of the system is not bounded by a polynomial of N. Such problems are encountered in the prediction of the native structures of proteins,[2] the design of very large scale integrated (VLSI) layout,[3] the traveling salesman problem,[4] the scheduling of machine time,[5] to name only a few. It is shown that all the NP-complete problems are equivalent in that if a deterministic polynomial time solution is given to one NP-complete problem, then all the other problems in this class are simultaneously solved by a certain mapping of the solution. On the other hand, if a deterministic polynomial solution does not exist for a particular NP-complete problem, none of the problems in the class can be solved within polynomial time.

Of the many global optimization problems that are NP-complete, minimization of the potential energies of molecules with respect to their structures is of particular interest to chemists. Deterministic approaches such as the branch and bound method[6–8] have been adopted in the minimization of many systems. But, because of the prohibitive computational cost required in the deterministic methods, significant effort has been invested in seeking heuristic methods, which, although not guaranteed, find the global minimum at smaller costs. A variety of stochastic global minimization methods based on Monte Carlo simulations has been developed and successfully applied to a wide spectrum of problems. Among the noticeable achievements are simulated annealing (SA),[9] quantum annealing,[10–12] basin hopping (Monte Carlo minimization),[13,14] J-walking,[15] parallel tempering,[16,17] the multicanonical methods,[18–20] and very recently, the stochastic tunneling method (STUN).[21]

A major difficulty for global minimization is to single out the true global minimum from an enormous number of local minima. In order to efficiently search for the global minimum on a rough potential energy surface, a stochastic global optimization method must be able to overcome the high energy barriers adequately so as not to be trapped in some local energy minimum, and explore the low energy regions sufficiently at the same time. These two requirements seem to contradict each other at first glance, and the unsatisfactory performance of most global optimization methods can be ultimately attributed to the failure to strike an optimal balance between the two. This problem is manifest in the classical simulated annealing method. At the early stages of the annealing process, the high temperature facilitates the barrier crossing, but the low energy states are inadequately sampled, while at later stages the low temperature quenches the system to a local energy minimum, and the method fails to explore other low energy regions. This is the reason that the classical simulated annealing performs unsatisfactorily when applied to sophisticated systems.[12] A successful global optimization method should be crossing the energy barriers and exploring the low energy regions simultaneously, not separately.

In this work we present the multicanonical jump walk annealing (MJWA) method, which couples classical simulated annealing with multicanonical sampling. The canonical Monte Carlo (MC) sampling digs deep into energy landscape for low energy minima while the multicanonical sampling surmounts the energy barriers and prevents the MC process being locally trapped. The MJWA is applied to the geometric optimization of Morse and Lennard-Jones clusters to demonstrate its efficiency.

a)Electronic mail: berne@chem.columbia.edu

II. METHOD

In the multicanonical ensemble,[18,19] the configurations are sampled with the probability inversely proportional to the density of states of the corresponding energy, namely,

$$\rho_{mu}(\mathbf{r}) \propto \frac{1}{\Omega(E(\mathbf{r}))}. \tag{1}$$

With this weight factor, the energies are sampled with a flat distribution

$$P(E) \propto \Omega(E)\rho_{mu}(E) = \Omega(E)\frac{1}{\Omega(E)} = 1. \tag{2}$$

Consequently, the high energy barriers and the low energy minima are each explored with sufficient frequency.

In practice, the density of states $\Omega(E)$ is not known *a priori*, and has to be numerically estimated via iterative simulations.[18,19,22] Since the probability that a certain energy E is sampled in a simulation is proportional to the weight factor $\rho(E)$ used in the simulation multiplied by the density of states $\Omega(E)$ of that energy, the energy histogram $H(E)$ of the sampling should satisfy

$$H(E) \propto \Omega(E)\rho(E). \tag{3}$$

If the lowest and highest energies sampled in the simulation are $E_{\min}$ and $E_{\max}$, respectively, we can estimate the density of states $\Omega(E)$ for $E_{\min} \leqslant E \leqslant E_{\max}$ from the above relation. Defining the entropy function

$$S(E) = \ln(\Omega(E)), \tag{4}$$

we have

$$S(E) = \ln(\Omega(E)) = \ln(H(E)) - \ln(\rho(E)) \quad (\text{for } E_{\min} \leqslant E \leqslant E_{\max}). \tag{5}$$

The above equations are used to update the estimate of the density of states in each multicanonical iteration, which in turn is used in the weight factor for the next iteration. In each iteration the range of the sampled energies increases in width, and extends to the low energy region. The multicanonical annealing method[20] was also proposed as a specialized scheme for global optimization. In the multicanonical annealing method, the width of the sampled energy range is fixed at a prescribed value, and the energy band moves toward the low energy region in each iteration.

Multicanonical methods have difficulties in exploring the low energy regions.[23] Because the weight factor used is derived from the previous iterations, it has no information concerning the unexplored low energy regions of the landscape. Therefore, pure multicanonical methods do not efficiently sample low energy configurations.

Our proposed multicanonical jump walk annealing (MJWA) method solves this problem by coupling a simulated annealing procedure to the lower energy end of the multicanonical sampling. Below the lowest energy $E_{\min}$ sampled in the previous multicanonical iterations, canonical samplings at decreasing temperatures are used, while above $E_{\min}$, the configurations are sampled with the multicanonical method. The canonical annealing efficiently searches the low energy regions and the multicanonical sampling helps barrier crossing. The procedure of MJWA is outlined as follows:

(1) Starting from a sufficiently high temperature T_0, where the simulation can cross the high energy barriers frequently, run a canonical Monte Carlo simulation and construct the energy histogram $H^0(E)$ during the simulation. In the canonical simulation the weight factor used is just the Boltzmann factor $\rho^0(E) = \exp(-\beta_0 E)$, where $\beta_0 = 1/T_0$, therefore from Eq. (5), our first estimate of the entropy function is

$$S^0(E) = \ln(H^0(E)) - \ln(\rho^0(E)) = \ln(H^0(E)) + \beta_0 E \quad (E^0_{\min} \leqslant E \leqslant E^0_{\max}) \tag{6}$$

where $E^0_{\min}$ and $E^0_{\max}$ are the lowest and highest energies sampled in the canonical simulation, respectively.

(2) Anneal the temperature by a preselected cooling factor $\xi < 1$,

$$T_{k+1} = T_k \xi, \tag{7}$$

where $k = 0,1,\ldots,$ and construct the following weight factor from the estimated entropy function $S^k(E)$ obtained in the previous iteration, and temperature T_{k+1}:

$$\rho^{k+1}(E) = \begin{cases} \exp(-\beta_{k+1}E) & \text{if } E < E^k_{\min} \\ \exp(-S^k(E)) & \text{if } E^k_{\min} \leqslant E \leqslant E^k_{\min} + E_{\text{window}}, \\ 0 & \text{if } E > E^k_{\min} + E_{\text{window}} \end{cases} \tag{8}$$

where E_{window} is the prescribed energy band width, and $\beta_{k+1} = 1/T_{k+1}$.

(3) Carry out a Monte Carlo simulation using the above weight factor. The random trial move $\mathbf{r} \to \mathbf{r}'$ in the MC simulation is accepted with the following probability:

$$\text{acc}(\mathbf{r}'|\mathbf{r}) = \min\left\{1, \frac{\rho^{k+1}(E(\mathbf{r}'))}{\rho^{k+1}(E(\mathbf{r}))}\right\}. \tag{9}$$

The energy histogram $H^{k+1}(E)$ is constructed during the simulation, and the lowest energy configuration encountered so far is saved.

(4) Update the entropy function estimate $S^{k+1}(E)$ using

$$S^{k+1}(E) = ln(H^{k+1}(E)) - ln(\rho^{k+1}(E)) (\text{for } E_{\min} \leqslant E \leqslant E_{\max}). \tag{10}$$

(5) Repeat steps (2)–(4) for a certain number of iterations till the temperature T_k anneals to near zero.

We call the method ''jump'' walk because it switches from a canonical annealing to a multicanonical sampling as the energy crosses $E^k_{\min}$, resembling a jump between the canonical ensemble and the multicanonical ensemble. Our

method differs from the pure multicanonical annealing method in that we systematically anneal the sampling temperature in the low energy region.

III. RESULTS AND DISCUSSION

The MJWA method is first applied to the geometric optimization of Lennard-Jones and Morse clusters. These model clusters serve as prototypes in the study of real atomic clusters. For example, argon clusters in gas phase are very well modeled by the Lennard-Jones clusters,[24] while sodium, potassium,[25] and C_{60}[26] clusters can be modeled by Morse clusters. These two model systems also serve as a standard test field for global optimization methods.[13,27]

A. Lennard-Jones clusters

The Lennard-Jones clusters are bound by the potential

$$V(\mathbf{r}^N)=4\varepsilon\sum_{i<j}\left(\left(\frac{\sigma}{r_{ij}}\right)^{12}-\left(\frac{\sigma}{r_{ij}}\right)^{6}\right), \tag{11}$$

where r_{ij} is the interatomic distance between the ith and the jth atoms. In this work, the reduced units are used, i.e., $\varepsilon=1.0$, $\sigma=1.0$.

Lennard-Jones clusters with up to 30 atoms were minimized using MJWA. Five MJWA simulations were carried out for each cluster size. For clusters with size up to 20, as few as 200 000 MC sweeps were used in each MJWA simulation, while for clusters with size between 21 and 30, 1 000 000 total MC sweeps were used. The starting temperature of the annealing process was $T_0=1.0$ (in reduced units). The performance of the simulations proved not sensitive to the annealing schedule. Here, we present the results obtained through 100 iterations with the cooling factor $\xi=0.933\,25$, which are summarized in Table I. For most of the clusters, the MJWA method finds the global minima with satisfactory success ratio. Through the trials, however, we found that the Lennard-Jones clusters with 17 and 27 atoms pose an interesting challenge to global optimization methods. As shown in the table, only one out of five MJWA trials in these simulations located the global minima of 17- and 27-atom Lennard-Jones clusters successfully. We had to increase the computational cost to find the global minima of these two clusters with satisfactory success ratio. We also made a comparative study between the conventional simulated annealing and MJWA on these two clusters.

TABLE I. Global minima of Lennard-Jones clusters found by multicanonical jump walk annealing. For each cluster a total of five trials was carried out, and the number of trials that successfully locate the global minimum is presented in the table. The annealing was implemented with 100 iterations. For clusters with size no larger than 20, a total of 200 000 MC sweeps was used in each simulation, otherwise 1 000 000 MC sweeps were used.

Cluster size n	Global minimum found	Number of successful trials
5	−9.103 852	5
6	−12.712 062	5
7	−16.505 384	5
8	−19.821 489	5
9	−24.113 360	5
10	−28.422 532	5
11	−32.765 970	5
12	−37.967 600	5
13	−44.326 801	5
14	−47.845 157	5
15	−52.322 627	5
16	−56.815 742	5
17	−61.317 995	1
18	−66.530 949	3
19	−72.659 782	5
20	−77.177 043	4
21	−81.684 571	2
22	−86.809 782	5
23	−92.844 472	5
24	−97.348 815	4
25	−102.372 663	5
26	−108.315 616	5
27	−112.873 584	1
28	−117.822 402	4
29	−123.587 371	5
30	−128.286 571	2

B. Comparison with SA

1. 17-atom Lennard-Jones cluster

The 17-atom Lennard-Jones cluster has three lowest energy structures that have very similar geometries and are very close in potential energy. All three structures have 13 atoms forming an icosahedron, with the other 4 atoms forming a ''cap'' on its top. The lowest energy structure has the energy $E_1=-61.317\,995$. Its cap atoms are positioned in a zigzag fashion, and it belongs to the point group C_2. The second lowest energy structure has the energy $E_2=-61.307\,146$, with the cap atoms forming a trapezoid, and belongs to the point group C_v. The third lowest energy structure has the energy $E_3=-61.296\,768$ and belongs to the point group C_{3v}. Three of its cap atoms forming an equilateral triangle, with the fourth atom sitting at the center. The energy difference between the three structures mainly comes from the subtle difference in the interaction of the four cap atoms. In other words, the energy difference results from different surface tension.

The annealing schedules for both the SA and MJWA were first optimized via numerous trials. Ten independent trials for each method were then carried out using the optimized schedule. In SA the temperature was annealed exponentially in 8000 stages from 1.0 to 0.001, and the system was equilibrated for 500 MC sweeps at each stage. Therefore, a total of 4 000 000 MC sweeps was used in each SA simulation. Out of the ten SA trials only one trial located the correct global minimum, while eight trials annealed the cluster to the second lowest energy structure. A partial reason for favoring the second lowest energy structure is that we eliminated the rotational degrees of freedom in our simulations, making the lowest energy nondegenerate, while leaving the second lowest energy level twofold degenerate. Since these two structures are very close in energy, a conventional annealing process will lead to a structure with the probability proportional to its degeneracy, thus favoring the second lowest energy structure.

The MJWA method was implemented in 100 iterations, with 40 000 MC sweeps in each iteration, bringing the total MC sweeps up to 4 000 000, the same as that in SA. Out of

the ten trials, eight found the global minimum successfully, which is eight times better than SA. The advantage of MJWA over SA comes from its ability to cross the high energy barriers between the local minima. The lowest two structures can interconvert between each other, enabling the method to find the true global minimum.

2. 27-atom Lennard-Jones cluster

The 27-atom Lennard-Jones (LJ) cluster has two low lying energy structures that are close in potential energy. The global minimum lies at $E_1 = -112.873\,584$ and has the symmetry C_{2v}. The second lowest structure has the energy $E_2 = -112.825\,518$ and belongs to the point group C_s. The simulated annealing was implemented in 8000 annealing stages, with 1000 MC sweeps at each stage. Thus, each SA simulation consisted of 8 000 000 MC sweeps in total. MJWA was implemented with 100 iterations, and in each iteration 80 000 MC sweeps were used, so a total 8 000 000 MC sweeps was used in each simulation, the same as used in SA. Ten independent MJWA and SA simulations were carried out. Out of the ten SA trials, only one found the true global minimum, while all the other nine trials annealed to the second lowest structure. On the other hand, five out of the ten MJWA trials found the global minimum, which is five times better than SA.

C. Morse clusters

The Morse potential can be expressed as

$$V(\mathbf{r}^N) = \sum_{i<j} e^{\lambda(1-r_{ij}/r_e)}[e^{\lambda(1-r_{ij}/r_e)} - 2]\varepsilon. \tag{12}$$

After adopting the reduced units and setting r_e and ε to be 1, we are left with a free parameter λ which decides the range of the pairwise interaction. Large λ corresponds to short-ranged interactions and small λ to long-ranged.

The MJWA method was applied to Morse clusters with long- and short-ranged interactions, i.e., $\lambda = 3,6,10$. The global minima were found for clusters of up to 20 atoms. Global minimization for long ranged interaction Morse clusters is a relatively easy task. For clusters with $\lambda = 3,6$ short MJWA simulations with 200 000 total MC sweeps sufficed to locate the global minima with good performance. The global minima for clusters with short-ranged interactions are considerably more difficult to find because of the increased energy barriers and the rough energy landscapes associated with short-range interactions.[28] Therefore, for Morse clusters with $\lambda = 10$, a total of 1 000 000 MC sweeps was used in each simulation. As in the study of LJ clusters, the system was annealed in 100 iterations. The starting temperature was $T = 1.0$, and the cooling factor was $\xi = 0.933\,25$. The results are summarized in Table II. MJWA located the global minima for most clusters with good success ratio, except for the short ranged ($\lambda = 10$) 18-atom cluster. For this singular case, a total of 5 000 000 MC sweeps was used in each MJWA simulation to obtain an acceptable success ratio of 60%.

TABLE II. The global minima of Morse clusters found by multicanonical jump walk annealing. λ is the range parameter in the Morse potential [cf. Eq. (12)]. The system was annealed via 100 iterations. For $\lambda = 3,6$, a total of 200 000 MC sweeps was used in each simulation, while for $\lambda = 10$, a total of 1 000 000 MC sweeps was used. Five trials were carried out for clusters of each size. As shown in the table, for most clusters, MJWA finds the correct global minima with high success ratio. For $n = 18$, $\lambda = 10$, however, an MJWA with 1 000 000 total MC sweeps failed to find the global minimum, and we had to increase the total number of MC sweeps to 5 000 000 to have a satisfactory success ratio.

Cluster size n	λ	Global minimum found	Number of successful trials
5	3	−9.299 500	5
5	6	−9.044 930	5
5	10	−9.003 565	5
6	3	−13.544 229	5
6	6	−12.487 810	5
6	10	−12.094 943	5
7	3	−17.552 961	5
7	6	−16.207 580	4
7	10	−15.956 512	5
8	3	−22.042 901	5
8	6	−19.327 420	5
8	10	−18.964 638	5
9	3	−26.778 449	5
9	6	−23.417 190	5
9	10	−22.850 758	5
10	3	−31.888 630	5
10	6	−27.473 283	5
10	10	−26.583 857	5
11	3	−37.930 817	5
11	6	−31.521 880	5
11	10	−30.265 230	4
12	3	−44.097 880	5
12	6	−36.400 278	5
12	10	−34.366 755	5
13	3	−51.737 046	5
13	6	−42.439 863	5
13	10	−39.662 975	3
14	3	−56.754 744	5
14	6	−45.619 277	5
14	10	−42.675 222	5
15	3	−63.162 119	5
15	6	−49.748 409	5
15	10	−46.541 404	5
16	3	−69.140 648	5
16	6	−53.845 835	5
16	10	−50.261 947	5
17	3	−75.662 417	5
17	6	−57.941 386	3
17	10	−53.983 559	2
18	3	−82.579 266	4
18	6	−62.689 245	3
18	10	−57.657 135	3[a]
19	3	−90.647 461	5
19	6	−68.492 285	3
19	10	−62.166 843	3
20	3	−97.417 393	5
20	6	−72.507 782	5
20	10	−65.679 115	3

[a]For this system, 5×10^6 MC sweeps were used in each simulation.

D. Comparison between MJWA and SA

Morse clusters bound by extremely short-ranged interactions pose a severe challenge to any global optimization methods. The energy landscape of such clusters resembles

that of a golf course, consisting of a multitude of local minima that are very close in energy and separated by high energy barriers.[28] The conventional simulated annealing method can be trapped in any one of these local minima and fails to anneal to the true global minimum. The MJWA method, on the other hand, can always escape any local energy minimum trap, due to the coupled multicanonical sampling, and will thus eventually anneal to the global minimum.

A comparative study between SA and MJWA is done on a cluster of 13 atoms, interacting through the very short ranged Morse potential with $\lambda = 14$. [cf. Eq. (12)] This system is chosen because its energy landscape has been thoroughly studied[28] and its icosahedral global minimum structure ($E = -37.258\,877$ in reduced unit) has been determined with fair confidence.

Optimized annealing schedules for both the conventional simulated annealing and MJWA were first determined via a large number of trials. Ten independent SA and MJWA simulations were then carried out using the optimized annealing schedule. In SA the temperature was annealed exponentially in 8000 stages from 1.0 to 0.001, and at each stage the system was equilibrated for 250 MC sweeps. Therefore, the total number of sweeps was 2 000 000 in each SA simulation. Out of the ten trials, only two successfully located the global minimum. In MJWA 50 iterations were used, and in each iteration the system was advanced for 40 000 MC sweeps, therefore, 2 000 000 MC sweeps in total were used in each simulation, the same as that in SA. Out of the ten MJWA trials, six of them successfully located the global minimum. The MJWA method is thus found to be three times more efficient than the conventional SA method in this case.

E. BLN-protein model

Simplified protein models are often used to test computational methods for native structure prediction. In this work, we applied the MJWA method to the global minimization of the BLN-protein model.[29–31] The model protein sequence consists of three kinds of residues: hydrophobic (B), hydrophilic (L), and neutral (N). The residues contribute to the potential energy of the protein in the following terms:

$$E_p(\{\mathbf{r}_i\}) = V_{\mathrm{bl}}(\{\mathbf{r}_i\}) + V_{\mathrm{ba}}(\{\theta_i\}) + V_{\mathrm{dih}}(\{\phi_i\}) + V_{\mathrm{nb}}(\{r_{ij}\}), \tag{13}$$

where

(i) Bond length potential

$$V_{bl}(\{\mathbf{r}_i\}) = \sum_{i=1}^{N-1} \frac{k_r}{2}(|\mathbf{r}_{i+1} - \mathbf{r}_i| - a)^2, \tag{14}$$

where $k_r = 400\varepsilon_h/a^2$. a is the average bond length between two residues, taken to be 1 in our study. ε_h is the average strength of the hydrophobic interaction, also taken to be 1 to simplify the computation. The bond length potential restricts the bonds to stretch and compress around the equilibrium length.

(ii) Bond angle potential

$$V_{\mathrm{ba}}(\{\theta_i\}) = \sum_{i=1}^{N-2} \frac{k_\theta}{2}(\theta_i - \theta_0)^2, \tag{15}$$

where θ_i is the bond angle between $(\mathbf{r}_i - \mathbf{r}_{i+1})$ and $(\mathbf{r}_{i+2} - \mathbf{r}_{i+1})$, $k_\theta = 20\varepsilon_h$ and $\theta_0 = 1.8326$. Bond angle potential controls the bending vibration of the bonds.

(iii) Dihedral angle potential

$$V_{\mathrm{dih}}(\{\phi_i\}) = \sum_{i=1}^{N-3} [A_i(1 + \cos\phi_i) + B_i(1 + \cos 3\phi_i)], \tag{16}$$

where ϕ_i is the dihedral angle formed by the four successive residues $i, i+1, i+2$, and $i+3$. If two or more of the four residues are neutral (N), $A_i = 0$ and $B_i = 0.2\varepsilon_h$, otherwise $A_i = B_i = 1.2\varepsilon_h$. Dihedral angle potential arises from the steric hindrance of the bonds' rotation.

(iv) Non-bonded potential

$$V_{\mathrm{nb}}(\{r_{ij}\}) = \sum_{i=1}^{N-3} \sum_{j=i+3}^{N} V_{\alpha\beta}(r_{ij}), \tag{17}$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ and $\alpha, \beta = B, L$, or N. $V_{\alpha\beta}$ represents the interaction between the residues that are not covalently bonded and is the main factor for the folding process.

$$V_{L\beta}(r_{ij}) = 4\varepsilon_l\left[\left(\frac{a}{r_{ij}}\right)^{12} + \left(\frac{a}{r_{ij}}\right)^{6}\right] \quad (\beta = B \text{or } L, \varepsilon_l = \tfrac{2}{3}\varepsilon_h), \tag{18}$$

$$V_{N\beta}(r_{ij}) = 4\varepsilon_h\left[\left(\frac{a}{r_{ij}}\right)^{12}\right] \quad (\beta = B, N, \text{or } L), \tag{19}$$

$$V_{BB}(r_{ij}) = 4\varepsilon_h\left[\left(\frac{a}{r_{ij}}\right)^{12} - \left(\frac{a}{r_{ij}}\right)^{6}\right]. \tag{20}$$

In this paper we used the MJWA method to search for the global minimum of a BLN-protein consisting of 46 residues with the sequence $B_9N_3(LB)_4N_3B_9N_3(LB)_5L$. This system has been shown as a highly frustrated system.[32] The lowest four minimum structures found have similar barrel-like geometry and their energies differ only in the first decimal place.[12] The true global minimum lies at $E = -49.263\,512\varepsilon_h$.

Simulations of considerable length have to be used to locate the global minimum of this system. We carried out 20 independent MJWA simulations, each consisting of 200 annealing iterations and 150 000 MC sweeps per iteration. Thus a total of 30×10^6 MC sweeps was used in each of these simulations. The starting temperature T_0 was fixed at $T_0 = 3.0\varepsilon_h/k_B$, where k_B is the Boltzmann constant, and the cooling factor used was $\xi = 0.960\,76$. In the first ten simulations, a zigzag structure was used as the initial configuration. The residues were placed so that all the bond lengths were a, all the bond angles were θ_0, and all the dihedral angles were

TABLE III. Comparison between the MJWA method and simulated annealing. Same cost (total MC sweeps) was employed in both methods for each system studied. The successful trials are those that located the true global minima. The ratios of the number of successful trials versus the total number of trials are presented for both methods. The advantage of the MJWA method over SA is manifest.

			Successful trials/Total trials	
Molecular system	Global minimum	Total MC sweeps	MJWA	SA
LJ$\{n=17\}$	−61.317 995	4×10^6	8/10	1/10
LJ$\{n=27\}$	−112.873 584	8×10^6	5/10	1/10
Morse$\{n=13,\lambda=14\}$	−37.258 877	2×10^6	6/10	2/10
LNB-46-mer	−49.263 512	30×10^6	14/20	4/20 (Ref.12)

π. Because different random number sequences were used in each of these simulations, the results were uncorrelated. Nine out of these ten simulations successfully located the global minimum. In the other ten simulations, the residues were randomly positioned in a box at the beginning of the simulation. In this case five out of the ten simulations located the true global minimum. The combined success rate is thus 14 out of 20, or 70%, which is a significant improvement compared to the optimal success rate of 20% for simulated annealing at the same cost.[12] Comparison between the MJWA and SA methods is summarized in Table III.

IV. DISCUSSION

From the point of view of thermodynamics, Monte Carlo methods more frequently sample the states corresponding to low free energies. For many molecular systems, the global potential energy minimum becomes the free energy minimum only at a relatively low transition temperature. At a higher temperature other local minima dominate thermodynamics because of the associated larger entropy.[33] The global energy minimum becomes thermodynamically favored only at temperatures below the transition point. Suppose a local energy minimum $\mathbf{r}_2^N$ has the energy E_2 and the entropy S_2, while the global energy minimum $\mathbf{r}_1^N$ has the energy $E_1 < E_2$ and the entropy S_1. The two minima are separated by a high energy barrier. In the case where $S_1 < S_2$, the global energy minimum $\mathbf{r}_1^N$ has lower free energy $F_1 = E_1 - TS_1$ than the local energy minimum $\mathbf{r}_2^N$ only when the temperature is below the transition point $T_c = \Delta E/\Delta S$, where $\Delta E = E_2 - E_1 > 0$ and $\Delta S = S_2 - S_1 > 0$. Consequently, at temperatures above the transition temperature, Monte Carlo samples the local minimum $\mathbf{r}_2^N$ more frequently than the global minimum. At low temperatures, however, the system cannot cross the energy barrier and commute between the minima, so a simulation starting from a local minimum configuration may fail to visit the global minimum in a simulation of finite length. The conventional simulated annealing method fails for such cases because it traps the system in the regions of large entropy as the temperature cools down, resulting in a ''supercool'' system. It has been suggested that slowing down the cooling procedure around the transition temperature will help SA locate the global minimum. This practice, however, entails pinpointing the transition point for every system to be studied. The MJWA method eliminates these supercool traps by forcing the system to always sample the high energy regions. Through the annealing of the temperature at the low energy end, the MJWA method enforces the transition in which the global potential energy minimum becomes the free energy minimum as the temperature drops below the transition point T_c. In MJWA there is no kinetic barrier as in SA, consequently, the method should always locate the global minimum in a simulation of reasonable length.

The criteria for an efficient stochastic global optimization method are therefore:

(1) In thermodynamics, the sampling should enforce the global potential energy minimum to be the free energy minimum.
(2) In kinetics, the method should be able to cross the high energy barriers.

Many stochastic global optimization methods have been developed in accordance with these criteria. Here, we briefly examine some of them.

In the *J*-walking method,[15] and its later modified version, parallel tempering method,[16,17] several samplings at temperatures ranging from very low to sufficiently high are carried out in parallel, and samplings at adjacent temperatures can occasionally exchange configurations with each other with certain criteria to maintain equilibrium. In this way, the sampling at low temperatures takes the responsibility of searching the low energy regions, while the sampling at the higher temperatures help the low temperature sampling to overcome high energy barriers through the exchange. The drawback of the *J*-walking method and the parallel tempering method, as has been pointed out in other papers,[15,23,34] is that for the exchanges to be adequately successful, the temperatures chosen have to be closely spaced. As a result a large number of samplings are needed, entailing huge computational cost. Moreover, the temperature spacing is a problem-dependent parameter that requires physical insight to be adjusted to optimize the method's performance, a troublesome factor for the method's general applicability.

The multicanonical methods force the sampling to perform a one-dimensional random walk in the scalar energy space, and sample all the energies within a certain range with a uniform distribution. In other words, the high energy barriers and the low energy minima are equally visited, therefore meeting both of the two criteria stated previously. The generic disadvantage of the purely multicanonical methods, as pointed out in Sec. II, is that in practice it is often inefficient in exploring low energy regions, which severely handi-

caps its candidacy for global optimization method.

Both the basin hopping method[13,14] and stochastic tunneling method[21] (STUN) transform the energy surface to eliminate or reduce the energy barriers while maintaining all the minima. In basin hopping, the energy surface is mapped to its locally minimized counterpart, namely, the sampling is done on the following transformed energy surface instead of on the original one

$$\tilde{E}(\mathbf{r}) = \min\{E(\mathbf{r})\}, \tag{21}$$

where $\mathbf{r}$ represents a configuration and min means that the energy is locally minimized starting from $\mathbf{r}$. With this transformation, the resistance to the intrawell motion is eliminated while the interwell barrier heights are reduced to be the energy differences between local minima. At the same time, all the minima are preserved. Basin hopping, however, requires local minimization at every step; therefore, it is very computationally expensive. This also forbids its use in problems where local minimization is impossible.

The stochastic tunneling method[21] dynamically transforms the energy surface, based on the lowest energy that has currently been sampled. The transformation is simple

$$\tilde{E}(\mathbf{r}) = 1 - \exp[-\gamma(E(\mathbf{r}) - E_0)], \tag{22}$$

where E_0 is the lowest energy sampled so far in the process, and γ is a problem-dependent parameter. This transformation squeezes all energies above E_0 into the small interval $[0,1]$, thus drastically decreasing the energy barriers above E_0, eliminating the possibility for the sampling to be trapped in any located local minima. On the other hand, it suppresses the energies below E_0 to be immensely lower, enhancing the search for lower energies. STUN, however, uses a uniform transformation that ignores the particulars of each individual system, and the method has very little control over distributing the sampling points evenly over high energy and low energy regions. (The method does this by adjusting the sampling temperature according to the energy distribution in the sampling.) In fact, the problem-dependent parameter γ determines the steepness of the transformed energy surface, and it effectively reflects a *guess* at the shape of the original energy landscape. In most cases it has to be determined by trial and error. We find that it is difficult to determine a proper γ in STUN for the test cases we studied in this work. The MJWA method, in contrast, uses the system-specific information, the density of states $\Omega(E)$, to insure that both the high and low energies are sampled evenly, and uses the decreasing temperature to help anneal the system to the global minimum.

The essential message we want to convey here is that stochastic global optimization methods should utilize as much information available from the system to strike an optimal balance between escaping local energy minimum traps and exploring the low energy configurations. We regard the multicanonical methods as a major step in this direction, in that they use the density of states $\Omega(E)$ as the extra information to aid the search. The MJWA method supplements this extra information with a little more control by introducing an annealing process at the lower energy end, which improves the sampling in the low energy regions.

V. CONCLUSION

In this paper, we introduced the multicanonical jump walk annealing (MJWA) method as a new stochastic global optimization approach. We demonstrated its efficiency through the geometric optimization of Lennard-Jones and Morse clusters and the BLN-protein model. We also demonstrated the new method's superiority over conventional simulated annealing through a set of comparative studies. MJWA has an advantage over the conventional simulated annealing because it avoids the local energy minimum traps that trouble the SA method in cases of ''super cooling.'' It also has an advantage over the pure multicanonical methods because the decreasing temperature at the low energy end enforces the transition in which the global energy minimum becomes the free energy minimum and thus is the thermodynamically favored state.

Our aim was to devise an annealing method that could cross high energy barriers and explore low energy configurations at the same time. The multicanonical jump walk annealing method is a stochastic global optimization method that overcomes energy barriers and explores the low energy regions at the same time. In the language of thermodynamics this means that MJWA is able to cross energy barriers when the global potential energy minimum is also the free energy minimum.

In this work, we also want to show that a global optimization method should utilize as much information available about the system to guide the search. In the multicanonical ensemble, this information is the density of states $\Omega(E)$ at each energy level. Looking for the global minimum on a sophisticated energy landscape is analogous to finding a needle in a haystack. It helps to know more about the haystack and the needle.

ACKNOWLEDGMENTS

We would like to thank D. J. Wales of Cambridge University for helpful correspondence and Y. H. Lee of Columbia University and A. Roitberg of NIST for informative discussions. This work was supported by a grant from NIH.

[1] M. Gary and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness* (Freeman, San Francisco, 1979).
[2] J. Ngo and J. Marks, Protein Eng. **5**, 313 (1992).
[3] M. Breuer, *Design Automation of Digital Systems* (Prentice-Hall, Englewood Cliffs, NJ, 1972).
[4] R. Miller and J. Thatcher, *Complexity of Computer Computations* (Plenum, New York, 1972).
[5] S. Sahni, *Data Structures, Algorithms, and Applications in C++* (McGraw-Hill, New York, 1998).
[6] C. Maranas and C. Floudas, J. Chem. Phys. **97**, 7667 (1992).
[7] C. Maranas and C. Floudas, J. Chem. Phys. **100**, 1247 (1994).
[8] D. Standley, V. Eyrich, A. Felts, R. Friesner, and A. McDermott, J. Mol. Biol. **285**, 1691 (1999).
[9] S. Kirkpatrick, C. Gelatt, and M. Vecchi, Science **220**, 671 (1983).
[10] P. Amara, D. Hsu, and J. Straub, J. Phys. Chem. **97**, 6715 (1993).
[11] A. Finnila, M. Gomez, C. Sebenik, C. Stenson, and J. Doll, Chem. Phys. Lett. **219**, 343 (1994).
[12] Y. Lee and B. Berne, J. Phys. Chem. (in press).
[13] D. Wales and J. Doye, J. Phys. Chem. **101**, 5111 (1997).
[14] Z. Li and H. Scheraga, Proc. Natl. Acad. Sci. USA **84**, 6611 (1987).
[15] D. Frantz, D. Freeman, and J. Doll, J. Chem. Phys. **93**, 2769 (1990).
[16] U. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).
[17] M. Deem and M. Falcioni, J. Chem. Phys. **110**, 1754 (1999).

[18] B. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992).
[19] J. Lee, Phys. Rev. Lett. **71**, 211 (1993).
[20] M. Choi and J. Lee, Phys. Rev. E **50**, 651 (1994).
[21] W. Wenzel and K. Hamacher, Phys. Rev. Lett. **82**, 3003 (1999).
[22] U. Hansmann, Phys. Rev. E **56**, 6200 (1997).
[23] H. Xu and B. Berne, J. Chem. Phys. **110**, 10299 (1999).
[24] T. Beck and R. Berry, J. Chem. Phys. **88**, 3910 (1988).
[25] L. Girifalco and V. Weizer, Phys. Rev. **114**, 687 (1959).
[26] L. Girifalco, J. Phys. Chem. **96**, 858 (1992).
[27] D. Wales and J. Doye, J. Chem. Soc., Faraday Trans. **93**, 4233 (1997).
[28] M. Miller, J. Doye, and D. Wales, J. Chem. Phys. **110**, 328 (1999).
[29] J. Honeycutt and D. Thirumalai, Proc. Natl. Acad. Sci. USA **87**, 3526 (1990).
[30] J. Honeycutt and D. Thirumalai, Biopolymers **32**, 695 (1992).
[31] T. Veitshans, D. Klimov, and D. Thirumalai, Folding Des. **2**, 1 (1997).
[32] H. Nymeyer, A. García, and J. Onuchic, Proc. Natl. Acad. Sci. USA **95**, 5921 (1998).
[33] J. Doye and D. Wales, Phys. Rev. Lett. **80**, 1357 (1998).
[34] R. Zhou and B. Berne, J. Chem. Phys. **107**, 9185 (1997).