Determination of low-energy structures of a small RNA hairpin using Monte Carlo–based techniques

SUDHANSHU SHANKER and PRADIPTA BANDYOPADHYAY*

School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi 110 067, India

**Corresponding author (Fax, 011-26741586; Email, praban07@gmail.com)*

The energy landscape of RNA is known to be extremely rugged, and hence finding low-energy structures starting from a random structure is a challenging task for any optimization algorithm. In the current work, we have investigated the ability of one Monte Carlo–based optimization algorithm, Temperature Basin Paving, to explore the energy landscape of a small RNA T-loop hairpin. In this method, the history of the simulation is used to increase the probability of states less visited in the simulation. It has been found that using both energy and end-to-end distance as the biasing parameters in the simulation, the partially folded structure of the hairpin starting from random structures could be obtained.

[Shanker S and Bandyopadhyay P 2012 Determination of low-energy structures of a small RNA hairpin using Monte Carlo–based techniques. *J. Biosci.* **37** 533–538] **DOI** 10.1007/s12038-012-9209-3

1. Introduction

RNAs are known to have a variety of conformations due to their single-stranded structure. One RNA molecule can have several different interconverting secondary structural elements, making its structure prediction an extremely difficult problem (Tinoco and Bustamante 1999). RNA structure prediction is also complicated by the presence of non-canonical base pairing (Chandrasekhar and Malathi 2003; Bhattacharyya *et al.* 2007). Prediction of the secondary structure of RNA has a long history due to the pioneering works of Zucker and Nussinov, among others (Nussinov and Jacobson 1980; Mathews *et al.* 1999; Hofacker 2003). However, prediction of the tertiary structure of RNA is still in its infancy, although there have been several breakthroughs in recent years (Das and Baker 2007; Ding *et al.* 2008; Seetin and Mathews 2011). RNA energy landscape is rugged, and standard computer simulation techniques such as molecular dynamics (MD) or Monte Carlo (MC) may fail to sample the landscape efficiently. The exploration of RNA energy landscape can be an ideal testing ground for new optimization algorithms. With this in mind, we have investigated the use of one MC-based optimization algorithm to determine the low-energy structures of RNA starting from completely random structures. Among various methods available for geometry optimization on a rugged energy landscape, the Basin Hopping (BH) method (Li and Scheraga 1987; Wales and Hodges 1998) has been widely used for diverse systems such as water clusters, Argon clusters and different biomolecules. In this method, MC moves are combined with local gradient-based optimization. Recently, improvement of the BH method has been proposed by Zhan *et al.* (2006) by combining the Energy Landscape Paving (ELP) method (Hansmann and Wille 2002) with local optimization, which they termed as the Basin Paving (BP) method. In the ELP method, the probability of visiting different structures is dynamically changed based on the history of the simulation. Bandyopadhyay showed that the ELP and BP methods can be used to determine the low-energy structure of a small RNA T-loop hairpin starting from random structures (Bandyopadhyay and Kharerin 2011). More recently, Shanker and coworkers have further improved the BP method and have successfully applied this method for large water clusters. The improved method is termed as Temperature Basin Paving (TBP) (Shanker and Bandyopadhyay 2011).

In the current work, the efficiency of the TBP method to determine the low-energy structures of a small RNA T-loop hairpin has been investigated. A 12-nucleotide-long T-loop hairpin motif of RNA with an all-atom model was utilized for the calculations in the present study. We have focused on

Keywords. Energy landscape; Monte Carlo; optimization; RNA; tertiary structure

Published online: 25 June 2012

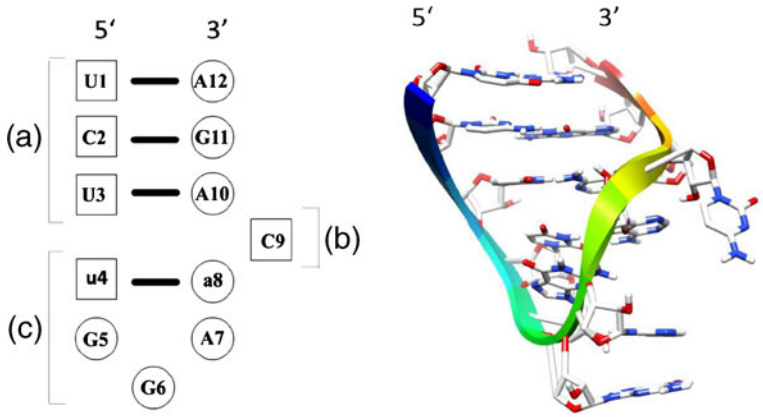

Figure 1. The structure of the RNA T-loop hairpin used in this work. Both schematic and three-dimensional structures are shown. (a), (b) and (c) on the left-hand-side figure represents stem, bulge and loop, respectively.

optimizing the parameters present in the TBP method. Both energy and end-to-end distance of the RNA were used to extract the history of the simulation in the TBP method. It was found that incorporation of the end-to-end distance as one of the biasing terms has improved the results significantly. Our preliminary investigation could find several folded structures of the RNA that are close to the crystal structure. However, only backbone folding was considered in this work. The article is arranged in the following manner. The next section describes the details of the computational algorithm. The details of simulation are given in section 3. This is followed by results and discussion and ends with our conclusions.

2. Method

A brief introduction of the Basin Hopping, Basin Paving and Temperature Basin Paving methods are given in this section. In this work, the MC moves are given by running a short constant-energy MD trajectory, as is done in the hybrid Monte Carlo (HMC) algorithm (Duane *et al.* 1987).

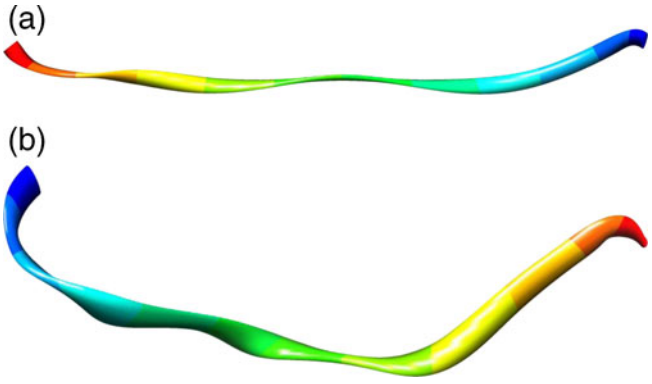

Figure 2. (**a**) One of the RNA structures used as the starting structure in the simulation. (**b**) One of the representative structures obtained from Temperature Basin Paving simulation with only energy as the biasing term.

2.1 *Basin Hopping*

In this method, the MC moves consist of two steps. At first, a regular random MC move is given to the molecule. Then, the molecule is minimized to its nearest minimum. The obtained minimum is considered as a state of the MC scheme. The acceptance/rejection of the states is done between the minima only. In favourable cases, this scheme can eliminate the barriers in the landscape.

2.2 *Basin Paving*

In the Basin Paving method, the acceptance/rejection of states is done with a non-Boltzmann weight factor, which is determined from the history of the simulation. The weight factor changes during the course of simulation and changes in such a way that frequently visited states have a lower probability and vice versa. One specific choice of the weight factor $w(E,t)$ can be the following:

$$w(E,t) = \exp(-\beta(E + CH(E,t)) \quad (1)$$

Where β is $1/kT$, k and T are the Bolztmann constant and temperature respectively. $H(E,t)$ is the histogram of energy E at MC step t. C is a normalization constant. Here the biasing term $H(E,t)$ is taken as a function of energy and, of course, it is also a function of MC step t.

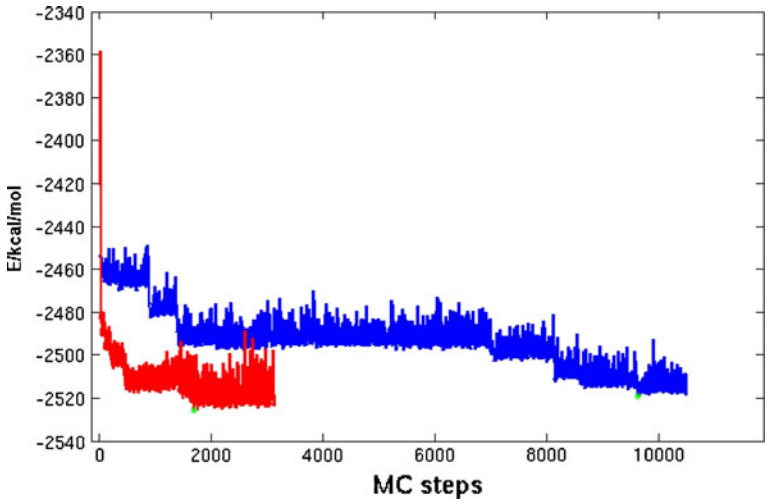

Figure 3. A comparison between the trajectories obtained from two different biasing terms. The blue trajectory is with only energy as the biasing term and the red trajectory is with both energy and end-to-end distance as the biasing terms.

2.3 *Temperature Basin Paving*

In this method, the weight factor is changed in the following manner: The energy range to be used in the simulation is discretized into bins, as is done for the Basin Paving simulation. Each bin of energy is assigned a temperature, which is same for all bins at the start of the simulation. Afterwards, with each visit to the bin, the temperature of that bin increases. In our implementation, the temperature of each bin is expressed as

$$T(E,t) = T_{initial} + C' T_{initial} H(E,t) \quad (2)$$

Here $T_{initial}$ is the initial temperature, which is the same for all bins. $H(E,t)$ is histogram of energy at MC step t. C' is a normalization factor. If the energy decreases in the MC move, then the new structure is always accepted. Otherwise, the following acceptance/rejection condition is used in the TBP method:

$$\min(1,\ \exp(-\beta_{old}(E_{new} - E_{old}))) \quad (3)$$

The subscripts (*old* and *new*) represent the initial and final state of the MC move respectively. It is to be noted that in the HMC method, total energy (potential energy plus kinetic energy) must be used in the acceptance/rejection procedure of the MC scheme to generate a proper statistical mechanical ensemble. However, here we have used only the potential energy in the acceptance/rejection part of the MC scheme, since no thermodynamic quantity was calculated from the simulation. The simulations were used to get low-energy structures only. To ensure that temperature of a bin does not increase to a very high value, a cooling scheme is used, which decreases the bin temperature slowly. In one implementation of the TBP method used in this work, the end-to-end distance of the RNA (P-P distance) is used as a biasing parameter along with energy. The acceptance probability used in this case is

$$\min(1, \exp(-\beta_{old}((E_{new} - E_{old}) + C''(R_{new} - R_{old})))) \quad (4)$$

where R represents the end-to-end distance of the RNA. If energy of the new structure is lower than that of the old one, then the structure is always accepted; otherwise, acceptance of the new structure is done by equation (4). Here C'' is a constant and is given a value of 10 when $R_{new} < R_{old}$ as compared to the other case, where C'' is given the value of 5. Due to this change in the acceptance probability, the structures with smaller end-to-end distance increases even if they have higher energy and the simulation tries to decrease the acceptance probability of the structure having higher end-to-end distance.

3. Details of the simulation

The RNA was represented with all atom amber99 force field with GB continuum solvent model. All the MD updates are done by the AMBER program package (Case *et al.* 2006). The coordinates of the T-loop hairpin structure were obtained from the crystal structure of *Haloarcula marismortui* large ribosomal subunit (Residues 310–321, PDB ID: 1JJ2). End-to-end distance of RNA T-loop was calculated between the two terminal phosphoruses. The MC move was given using a hybrid MC and MD approach. A short constant-energy MD run was used to give the MC move.

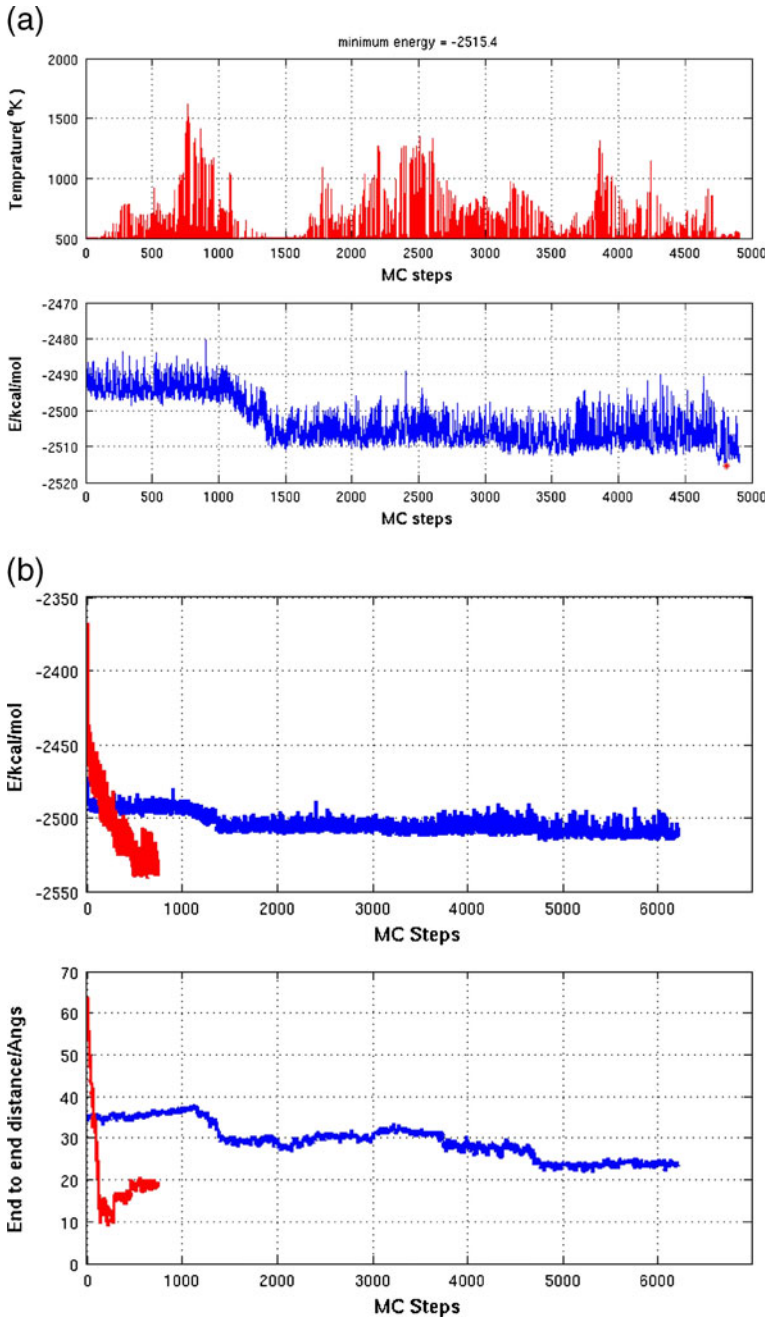

Figure 4. (**a**) Variation of temperature and energy in a representative TBP trajectory. (**b**) Variation of energy and end-to-end distance for two representative TBP trajectories.

4. Results and discussion

Figure 1 shows the structure of the T-loop hairpin considered in this work. The presence of stem, bulge and loop can be seen from the figure. The first issue in our MC scheme was how to give an optimum MC move by the short MD trajectory. Ideally we would like to have diverse structures generated by the MD-based MC move. We used the following protocol: If the structure during optimization was found to be confined in an energy bin for some time, then the temperature of that bin increased due to the increase of $H(E,t)$ for that bin. In that case, we used a higher temperature for MD, since this may help the simulation to come out of the bin. Otherwise, MD temperature was kept low. The second issue was which terms to use in the biasing term to get the history of the simulation. Initially we used only the energy as the biasing term and used

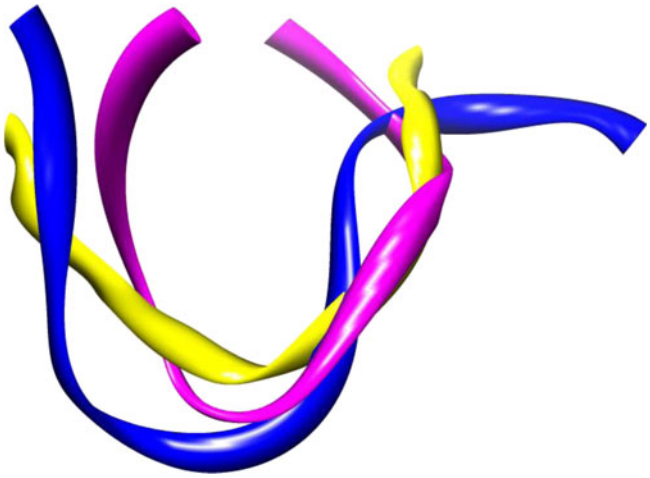

Figure 5. Two good structures (yellow and blue) obtained from TBP simulation superimposed with the crystal structure shown in magenta.

equation (3) to accept/reject structures. However, simulations with energy as the only biasing term gave structures only partially folded and the system could not move to other types of structure. One starting structure and a typical structure obtained from only the energy as the biasing term are shown in figure 2a and figure 2b respectively. From figure 2a it can be seen that our starting structure is almost linear without any base paring. The structure shown in figure 2b has folded only partially. Most of the structures obtained from this set of simulations were like the one shown in figure 2b.

In the next set of investigations, we used both energy and end-to-end distance of RNA as the biasing terms and used equation (4) as our MC acceptance/rejection condition. It dramatically improved the performance. A comparison between the two trajectories with and without the end-to-end biasing term is shown in figure 3. The lowest-energy structures obtained from these two runs are shown by green dots. Both the trajectories start from −2360 kcal/mol energy. However, there is a sharp decrease of the energy for the case where end-to-end distance was used. It reached the lowest-energy structure (which is less than −2520 kcal/mol) in less than 2000 steps. However, the trajectory with only energy as the biasing term reached the low-energy structure more incrementally. It took almost 10000 steps to reach its lowest energy structure, which is around −2510 kcal/mol. Figure 4a shows a typical trajectory, where both energy and temperature are plotted. It illustrates how the temperature changes in the TBP method. Between 500 and 1000 steps, where energy was confined in a narrow region, the temperature increased. When it came out of that region after 1000 steps, the temperature again decreased. Slowly the system, due to the increase of temperature, came out of the minima and reached lower-energy states. The lowest-energy structure in this run is shown by a red dot in the figure. Figure 4b shows two trajectories, both with end-to-end distance and energy as biasing terms. Both energy and end-to-end distance are plotted in this figure. It can be seen that in one trajectory the end-to-end distance decreased very fast. It reached ≤10 Å in first few hundred steps. The other trajectory took much longer time to reach structures with end-to-end distance close to 20 Å. This shows that thorough statistics of the optimization runs is necessary to comment on the efficiency of the new algorithm in a quantitative manner. Figure 5 shows some of the best structures obtained from the simulation superimposed on the crystal structure (shown in magenta). Since the crystal structure was cut from a larger RNA, the end terminals are closer than what was obtained from optimization of our truncated system. The yellow structure has a smaller backbone RMSD (5.0 Å) compared to the blue structure (RMSD 7.1 Å). However, the yellow structure is about 10 kcal/mol higher in energy than the blue structure.

From the results of this work, we have found that the TBP method has the potential to obtain the folded structure of small RNA with an all atom model. However, at this point only backbone folding has been investigated. The efficiency of the algorithm depends on the choice of biasing term, which steers the system from unfolded to folded states. In the current work, two biasing terms, one with a function of energy and the other with a function of end-to-end distance, were used. It is shown that the biasing term with end-to-end distance improves the performance significantly. Based on the limited number of trajectories with and without end-to-end distance as a biasing term, it was found that when we added the end-to-end distance as a biasing term, it generally takes at least 10 times fewer number of steps to reach the lowest energy structure, which is also more than 10 kcal/mol lower in energy than the other case. In the present case of a hairpin, the choice of the biasing term other than energy was clear, the end-to-end distance. However, for RNA with more complicated secondary structural elements, it might be non-trivial to find appropriate biasing terms. Another issue is whether the MD-based MC move is the most optimum one. Thus, we plan to investigate this further by comparing MD move with other kinds of moves.

5. Conclusion

The newly developed Temperature Basin Paving (TBP) method is used to determine the low-energy structure of a RNA T-loop hairpin. The TBP method uses the history of the simulation to favour states that are less sampled. It has been found that the use of end-to-end distance as a biasing term improves the performance significantly. Several structures close to the crystal structure, as far as backbone folding is considered, have been found from our work.

Acknowledgements

This work is funded by a Department of Science and Technology (DST) grant awarded to PB (No. SR/S1/PC-45/2009). SS would like to acknowledge Department of Biotechnology (DBT) for the BINC fellowship awarded to him.

References

Bandyopadhyay P and Kharerin H 2011 Monte Carlo energy landscape paving and basin paving simulation of RNA T-loop hairpin. *Chem. Phys. Lett.* **502** 130–135

Bhattacharyya D, Koripella S C, Mitra A, Rajendran VB and Sinha B 2007 Theoretical analysis of noncanonical base pairing interactions in RNA molecules. *J. Biosci.* **32** 809–825

Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, *et al.* 2006 AMBER 9 (San Francisco: University of California)

Chandrasekhar K and Malathi R 2003 Non-Watson Crick base pairs might stabilize RNA structural motifs in ribozymes – A comparative study of group-I intron structures. *J. Biosci*. **28** 547–555

Das R and Baker D 2007 Automated de novo prediction of native-like RNA tertiary structures. *PNAS* **104** 14664–14669

Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE and Dokholyan NV 2008 Ab initio RNA folding by discrete molecular dynamics: From structure prediction to folding mechanisms. *RNA* **14** 1164–1173

Duane S, Kennedy AD, Pendleton BJ and Roweth D 1987 Hybrid Monte Carlo. *Phys. Lett. B* **195** 216–222

Hansmann UHE and Wille LT 2002 Global Optimization by Energy Landscape Paving *Phys. Rev. Lett.* **88** 068105–068108

Hofacker IL 2003 Vienna RNA secondary structure server. *Nucleic Acids Res.* **31** 3429–3431

Li Z and Scheraga HA 1987 Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *PNAS* **84** 6611–6615

Mathews DH, Sabina J, Zuker M and Turne DH 1999 Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288** 911–940

Nussinov R and Jacobson AB 1980 Fast algorithm for predicting the secondary structure of single-stranded RNA. *PNAS* **77** 6309–6313

Seetin MG and Mathews DH 2011 Automated RNA tertiary structure prediction from secondary structure and low-resolution restraints *J. Comp. Chem.* **32** 2232–2244

Shanker S and Bandyopadhyay P 2011 Monte Carlo temperature basin paving with effective fragment potential: An efficient and fast method for finding low-energy structures of water clusters $(H2O)_{20}$ and $(H2O)_{25}$. *J. Phys. Chem. A* **115** 11866–11875

Tinoco I Jr and Bustamante C 1999 How RNA folds. *J. Mol. Biol.* **293** 271–281

Wales DJ and Hodges MP 1998, Global minima of water clusters $(H2O)_n$, $n \leq 21$, described by an empirical potential. *Chem. Phys. Lett.* **286** 65–72

Zhan L, Chen JZY and Liu WK 2006 Monte Carlo basin paving: An improved global optimization method. *Phys. Rev. E* **73** 15701–15704