Kinetic Monte Carlo method applied to nucleic acid hairpin folding

Ben Sauerwine* and Michael Widom†
Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA
(Received 22 September 2011; revised manuscript received 8 November 2011; published 19 December 2011)

Kinetic Monte Carlo on coarse-grained systems, such as nucleic acid secondary structure, is advantageous for being able to access behavior at long time scales, even minutes or hours. Transition rates between coarse-grained states depend upon intermediate barriers, which are not directly simulated. We propose an Arrhenius rate model and an intermediate energy model that incorporates the effects of the barrier between simulated states without enlarging the state space itself. Applying our Arrhenius rate model to DNA hairpin folding, we demonstrate improved agreement with experiment compared to the usual kinetic Monte Carlo model. Further improvement results from including rigidity of single-stranded stacking.

DOI: 10.1103/PhysRevE.84.061912 PACS number(s): 87.15.Cc, 87.14.gk, 82.39.Pj, 87.15.R−

I. INTRODUCTION

Kinetic Monte Carlo [1] has been enormously successful in connecting thermodynamic processes to macroscopic events in systems. This method explores a state space through a sequence of stochastic transitions between states. Kinetics is introduced through the choice of transition rates between states. It is customary to assign rates to transitions that depend on the free-energy difference between initial state i and the final state j that obeys the principle of detailed balance [2], so as to ensure the process reaches equilibrium in the long-time limit. However, the free-energy barrier between i and j is not taken into account, other than through an overall time scale κ shared by all elementary steps, in the commonly used Kawasaki [3] and Metropolis [4] rate models. Thus the success of these models in simulating first passage processes relies in part on the rate-limiting barrier states being present in the state space that the simulated system explores, as the rate models are insensitive to variations among the barriers between states.

Secondary structure models of nucleic acids, which record the pairings of complementary nucleotides, are ideal for kinetic Monte Carlo simulation, being easily enumerable. However, for hairpin folding, since the bending free energy is generally unfavorable and the bonding energy generally favorable, there will be an "almost closed" state that is bent but not bonded that constitutes the true rate-limiting intermediate (see Fig. 1). Such an intermediate, which is not contained in the state space of secondary structures, will be present in general for processes in which loops are created or destroyed.

DNA and RNA structure plays crucial roles within the cell, including gene regulation by microRNAs [5] and riboswitches [6]. MicroRNA precursors and transcriptional terminators [7] both fold into hairpin structures, and indeed the hairpin is the fundamental element of nucleic acid secondary structure. Numerous tools exist that are able to study the thermodynamics, both with pseudoknots [8–10] and without [11–14], and even model kinetics of nucleic acids at the resolution of both secondary [15,16] and tertiary [17,18] structure. Kinetic Monte Carlo algorithms such as kinfold [16] are able to simulate significant changes in secondary structure of nucleic acids efficiently, since the number of states accessible to a nucleic acid of length L by an elementary move is easily enumerable. For instance, the number of elementary base pair changes available to a sequence of length L scales as L^2. By assigning a rate to each move, trajectories through the space of secondary structures can be simulated.

The goal of this research is to understand the simple process of loop closing [19] using secondary structure kinetic Monte Carlo. We will see in Sec. IV that the introduction of the almost closed state is essential to reproducing gross features of experimental results. The combination of entropy loss for bringing the ends of the loop into proximity and stacking rigidity of single-stranded DNA proves to be the dominant contributions to the free energy of this barrier state. Stacking rigidity is not customarily included in existing free-energy models, while the present study demonstrates its importance. Accurate free energies and transition rates are needed to properly model biological processes involving hairpin formation and more complex secondary structure.

The following section introduces the free-energy calculation for the almost closed state. We base this calculation on existing secondary structure free-energy models for nucleic acids (e.g., that of SantaLucia [20]), extended to include bending free energies for pairs brought into proximity but not bonded. Since nucleic acid secondary structure free-energy models separate the energetic contributions of pairing versus those from hairpin loops, bulges, and internal loops, it is a simple matter *in silico* to select which types of contribution to include and which to exclude once an appropriate model is developed.

For our purpose, kinetic Monte Carlo is implemented through the Gillespie algorithm. In the Gillespie algorithm, a transition from state i to state j occurs with probability

$$p_{i\to j} = \frac{R_{i\to j}}{\sum_k R_{i\to k}}, \tag{1}$$

and then the dwell time is increased by

$$\tau = \frac{1}{\sum_k R_{i\to k}} \ln\left(\frac{1}{A}\right), \tag{2}$$

where $A \in (0,1]$ is a uniform random variable. This is precisely the Gillespie algorithm "direct method" [1].

*bsauerwi@andrew.cmu.edu
†widom@andrew.cmu.edu

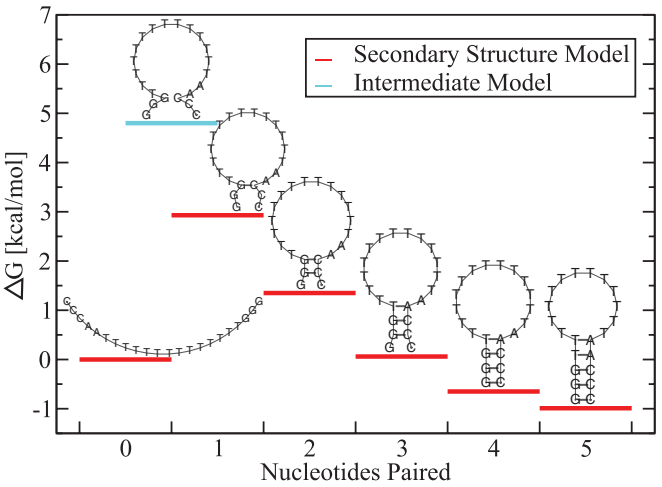

FIG. 1. (Color online) Free energies and secondary structures of the T_{12} sequence according to the SantaLucia free-energy model at $0.1M$ NaCl (see below). An almost closed unbound intermediate state forms a true barrier between open and closed states.

Frequent choices for the form of the reaction rates $R_{i\to j}$ in Eqs. (1) and (2) are the Metropolis and Kawasaki rules. For a transition from state i of free energy G_i to state j of free energy G_j, the Kawasaki [3]

$$R_{i\to j}^{\mathrm{Kaw}} = \kappa\, e^{-\Delta G_{ji}/2RT} \tag{3}$$

and Metropolis [4]

$$R_{i\to j}^{\mathrm{Met}} = \kappa\, \min(1, e^{-\Delta G_{ji}/RT}) \tag{4}$$

rate models assign rates $R_{i\to j}$ that are functions of only three free parameters: temperature T, the free-energy difference between states $\Delta G_{ji} = G_j - G_i$, and a rate constant κ with units of inverse time.

The overall time scale κ, which maps Monte Carlo time to real time, is most easily obtained through comparison to experiment. κ is certainly a function of the energy model used and could also be a function of temperature or ionic conditions. In principle, κ depends on the barrier separating states i and j, but this is not included in the Kawasaki and Metropolis rates given above. We shall soon factor that dependence out in an explicit form. Contemporary estimates for κ for nucleic acid base pairing events range from 10 ns [21] to 4.5 μs [22], with estimates as high as 100 μs for relaxation time of DNA bubbles [23]. The broad range of estimates does not imply some are necessarily wrong, but rather that κ can vary greatly depending on other simulation parameters, including the choice of rate model.

Section III introduces our Arrhenius rule, which considers the almost closed intermediates, while maintaining detailed balance. It can be used in kinetic Monte Carlo simulations as an alternative to the Kawasaki and Metropolis rules in order to add the effects of the almost closed state without cluttering the simulated state space with states whose properties are not an object of study and whose inclusion may be undesirable due to their effect on the equilibrium ensemble. Note that there exist ways to account for the intermediates other than through the Arrhenius rule, such as by inserting the missing states into the system or through a more general temperature- and context-dependent rate constant $\kappa_{ij}(T)$.

This research provides a framework for calibration of κ, as well as a modification to the Kawasaki or Metropolis transition rate models, which improves agreement between simulation and experiment overall in kinetic Monte Carlo simulations on the space of secondary structure. Here, the kinfold simulator is used from the ViennaRNA 2.0 [11] package, with DNA energy models and an empirical salt correction by SantaLucia [20] as implemented in UNAFold (mfold) [14]. The applicability of the Kawasaki and Metropolis rules is studied and the Arrhenius rule is developed as an option for modeling transition rates between nucleic acid states. In particular, it is shown here that existing free-energy models cannot explain experiments on DNA hairpin formation kinetics using standard Kawasaki or Metropolis rate models, which assign a common time scale κ to all processes, owing to their lack of a context-dependent rate-limiting intermediate.

II. ALMOST CLOSED STATE

The kinetic intermediate for formation of a hairpin is an almost closed state, in which a pair of nucleotides is brought into proximity but not yet bound. We separate the free energy of the intermediate into two contributions, "loop" and "stack,"

$$G^I = G^{\mathrm{loop}} + G^{\mathrm{stack}}. \tag{5}$$

The loop contribution includes the entropic cost of bringing the ends into proximity, plus any associated enthalpic cost of bending the single-stranded structure. The stack contribution includes the stacking and dangling-end free energies of preexisting nucleotide pairs, but does not include the pair about to form. Dangling-end contributions are just stacking of single nucleotides onto adjacent helices. Both stacking within helices and dangling-end contributions are usually stabilizing.

We evaluate G_{ji}^{loop} based on a structure comprising the union of all pairs of the initial and final structures (i.e., all pairs from both structures i and j). Should this union produce nucleotides with two pairing partners, an additional nucleotide is inserted adjacent to the duplicated nucleotide and allowed to pair to the second partner. Since G^{loop} omits the energy of double-stranded stacks, duplication of these nucleotides allows standard secondary structure algorithms to perform the calculation without affecting the resulting value. We evaluate G_{ji}^{stack} based on a structure comprising only the intersection of the pairs of the initial and final structures (i.e., the pairs shared in common by structures i and j). Because of the simplicity of these set-theoretic rules for producing the intermediate, our definition extends beyond hairpins to cover any pair of secondary structures.

Figure 2 shows the particular case of an open chain nucleating a single base pair, an example that includes free-energy contributions from the loop, but no stacking energies, since the intersection of pairs is the empty set. All energies are calculated using the DNA energy parameters and salt corrections of SantaLucia [20] as implemented in UNAfold [14], but translated into ViennaRNA format for compatibility with the kinfold kinetic Monte Carlo program. These parameters allow the user to provide Na^+ and Mg^{2+} concentrations to determine the appropriate salt correction. The free-energy model as

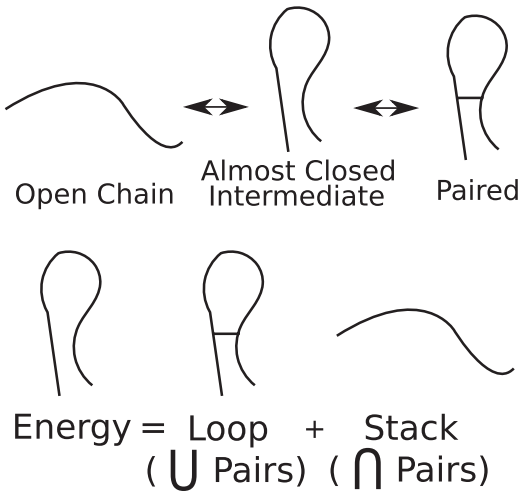

FIG. 2. Diagram of the true barrier state between the open chain and a nucleic acid chain with a single base pair. Loop and stack energy contributions to the intermediate state are diagramed at the bottom. Note that the intersection of base pairs here is the empty set, so there is no double-stranded or dangling-end stacking contribution to the free energy of this intermediate, which was selected for being the overall rate-limiting step in the systems studied here.

defined by SantaLucia goes asymptotically to a logarithmic dependence on loop length as in the Jacobson-Stockmayer formula [24] and hence our almost closed intermediate barrier state is dominated by entropy.

III. ARRHENIUS RATE MODEL

One contribution of this work is an Arrhenius rate model for Monte Carlo in systems with significant intermediates. In particular, the Arrhenius rate model is applied to nucleic acid secondary structure. Figure 1 shows the true intermediate between an open chain and a single nucleotide pair. We call this intermediate an almost closed chain, a situation not described within the space of nucleic acid secondary structure. The Arrhenius rate model uses the free energy of this activated complex to set the rate for transitions between two adjacent states,

$$R_{i\to j}^{\mathrm{Arr}} = \kappa e^{-\Delta G^B/RT}, \tag{6}$$

where

$$\Delta G^B = G_{ji}^B - G_i \tag{7}$$

is the free-energy difference between the initial state i and the barrier state separating i from the final state j. Note that detailed balance is obeyed provided the free energy of the barrier state G_{ji}^B is symmetric between i and j. A given nucleic acid sequence has a fixed number N of possible secondary structures, hence on the order of N^2 distinct barriers B_{ji}. By utilizing the almost closed states to define time scales, while excluding them from the state space, we allow simulations to remain at an appropriately coarse-grained level for the set of secondary structures.

Ideally, the free energy of the barrier state G_{ji}^B should be greater than both G_i and G_j, but our predicted intermediate-state free energy G_{ji}^I sometimes lies below either the initial or the final-state free energy. Hence we allow for a Kawasaki- or Metropolis-type cutoff to relate the barrier free energy G^B to the intermediate-state free energy G^I,

$$G_{ij}^B = G_{ji}^B = \begin{cases} G_{ji}^I & \text{(pure Arrhenius)}, \\ \max\left(G_{ji}^I, \frac{G_i+G_j}{2}\right) & \text{(Kawasaki cutoff)}, \\ \max\left(G_{ji}^I, G_i, G_j\right) & \text{(Metropolis cutoff)}. \end{cases} \tag{8}$$

In these contingencies, an intermediate energy that would result in a faster transition rate than the Kawasaki or Metropolis rate is instead discarded in favor of the slower transition rate. These cutoffs are physically plausible because, in these cases, the true barrier to transition, state i or j, is used to calculate the transition rate.

In the case of DNA hairpin closing, the Kawasaki cutoff provided better agreement with experiment than Metropolis. This may be a function of the Kawasaki cutoff having better dynamic range, allowing transitions to barriers with $\max(G_i, G_j) > G^B > \frac{G_i+G_j}{2}$ instead of truncating these to $\max(G_i, G_j)$.

For the situation of nucleic acid folding explored here, the move set of kinfold [16] is used in order to study conformation changes in DNA secondary structure. The allowed space of conformations does not include pseudoknots, and each conformation is uniquely defined by its secondary structure. In the kinfold move set, there are three allowed types of conformation change that determine neighboring states: to form a new base pair between two unpaired bases; to open a base pair from existing paired bases; for an already paired base to shift the nucleotide with which it is paired. For convenience, the rate constant κ is assigned a fixed common value among all types of move, an assumption that is plausible given the elementary nature of the moves in question. These moves are diagramed in Fig. 3. During any Monte Carlo step, the statistical weight for selecting any neighboring state depends on the transition rate model [Kawasaki (3), Metropolis (4), or Arrhenius (6)]. Given the rates from a given transition rule, the

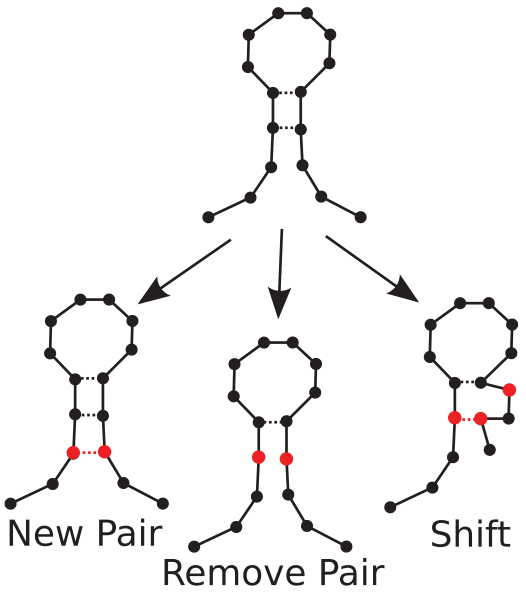

FIG. 3. (Color online) Three types of move allowed by kinfold [16].

dwell time at the initial state is selected from an exponential distribution with mean dwell time $(\sum_j R_{i\to j})^{-1}$.

By default, the kinfold program [16] uses the ViennaRNA energy models [11], which are limited in that they do not allow the salt concentrations to be continuously varied. Instead, ViennaRNA relies on the user to provide energy parameters suitable to the system to be simulated.

IV. RESULTS

We utilize our Arrhenius rate model in order to model experimental measurements of DNA hairpin loop opening and closing kinetics [19]. In particular, the sequences A_{21}, T_{12}, T_{16}, T_{21}, and T_{30} are studied, where sequence X_n is given by CCCAA(X^n)TTGGG. An example of minimum free-energy conformations of the T_{12} hairpin is given in Fig. 1. In all sequences, the closed conformation of each is a hairpin structure with the outermost five nucleotides bound, and the open conformation has no nucleotides bound. By averaging the first passage times from repeated kinfold simulations, opening (k_{open}) and closing (k_{close}) rates are determined which can be compared to the experimental results as illustrated in Fig. 4.

Experimental and simulated opening and closing rates vary almost linearly in a logarithmic plot of rate versus $1/T$. This is qualitatively understood taking the logarithm of a generic Arrhenius rate model,

$$\ln R = \ln\kappa - \frac{\Delta G^B}{RT} = \ln\kappa + \frac{\Delta S^B}{R} - \frac{\Delta H^B}{RT}. \tag{9}$$

Assuming ΔS^B and ΔH^B are temperature independent predicts the observed linear variations. Notice that the slope is proportional to the negative of the enthalpy barrier, $-\Delta H^B/R$. Comparing experiment with the Kawasaki rate model simulation shows reasonable agreement in hairpin opening rates, but the closing rates have slopes of opposite signs, with the simulation indicating a negative ΔH^B, while the experiments indicate it should be positive.

The positive slopes of the simulated closing rates are an inevitable result of the Kawasaki rate model utilizing the SantaLucia free-energy model in the space of secondary structures. It turns out that, within traditional secondary structure, the activated state for both the opening and closing processes consists of a state with only a single nucleotide pair, as indicated in Fig. 1 for the sequence T_{12}. The free-energy model assigns an entropic cost $\Delta S < 0$ for closing a loop, leading to a free-energy increase $-T\Delta S = 4.8$ kcal/mol, but there is an enthalpic benefit $\Delta H < 0$ kcal/mol for pairing the nucleotides, leading to an apparent barrier of $+2.93$ kcal/mol. This negative ΔH leads to the positive slope in closing rate versus $1/T$, for all sequences studied. On the other hand, the process of opening a closed hairpin requires a large positive enthalpic barrier to be crossed, opening four nucleotide pairs before the final activated complex can be considered. Thus the slope for the opening process is always negative.

The equilibrium constant

$$K = k_{\text{open}}/k_{\text{close}} = e^{-\Delta G_{ji}/RT} \tag{10}$$

depends only on the energy model, with no dependence on the dynamics. Inspecting Fig. 4, we see qualitative but not quantitative agreement between experiment and theory. In the case of poly(T) sequences, the model agrees with experiment on average, but has an incorrect slope with regard to temperature (higher temperatures are to the right in each graph of K). poly(A) has an additional offset with calculated values of K systematically lying below the experimental values. Naturally, an improved energy model that matches the experimental equilibrium constants would result in agreement of the relative slopes of k_{open} and k_{close}. However, this is not sufficient to give the individual slopes properly.

It is the disagreement in slope for closing rates that motivated our development of our Arrhenius rate and intermediate-energy models. The favorable negative ΔH associated with base pairing should not belong to the true rate-limiting intermediate barrier state for hairpin loop closing. Rather, we believe the true barrier is the almost closed state. Thus we carry out simulations utilizing our Arrhenius rate model [Eq. (6)] with a Kawasaki-like cutoff [Eq. (8)], with results shown in Fig. 5 for T_n sequences and in Fig. 6 for A_{21} sequences. In both cases, the change replaces the positive slope for closing rates with a nearly temperature-independent rate. This occurs because the negative ΔH of base pairing is now hidden behind an almost closed intermediate state with a high barrier. The closing rate is nearly temperature-independent because the barrier is dominated by loop closing entropy. Because we modify only the elementary step transition rates, and maintain detailed balance among all secondary structures, the equilibrium constants of the transitions are not affected.

Note that the slopes of closing rates k_{close} are not exactly zero. This is due primarily to the complexity of true barrier crossing processes [25]. The actual transitions pass through an ensemble of barrier states. For example, any of the first five nucleotides in Fig. 1 could close first, or even a nucleotide pair not seen in the ultimate closed state.

Although we have successfully removed the strong positive slopes in the closing rates, we still have not achieved negative slopes as seen in the experiment. We believe these reflect a rigidity arising from stacking interactions within a single-stranded molecule, especially in the case of poly(A) sequences, where a rigidity of 0.5 kcal/mol/nt bending enthalpy was measured [26] in $0.25M$ NaCl, compared with no significant effect observed in poly(T). The SantaLucia energy model explicitly sets the loop bending enthalpy to zero. In Fig. 4, we show that inclusion of this bending enthalpy dramatically improves the calculated equilibrium constant of the A_{21} sequence in $0.20M$ NaCl. Although the enthalpic cost of bending single strands arises from stacking interactions, for the purpose of defining the free energy of the intermediate, we include it in G^{loop}. That is, we pay the cost at the time of bending, as a hairpin closes, prior to the formation of the first bound nucleotide pair. Notice that this bending enthalpy could be considered as a contribution to nucleation enthalpy ΔH_{nuc}. While the SantaLucia energy model assigns $\Delta H_{\text{nuc}} < 0$, inclusion of bending brings us to a positive value.

A recently formulated free-energy model for RNA [27] does include positive loop bending enthalpies of roughly the necessary order of magnitude, though these are sequence-independent [i.e., they do not distinguish between poly(A) and poly(T)]. We have confirmed that simulations utilizing this new free-energy model qualitatively reproduce our results for DNA shown in Fig. 6. In particular, using the Kawasaki

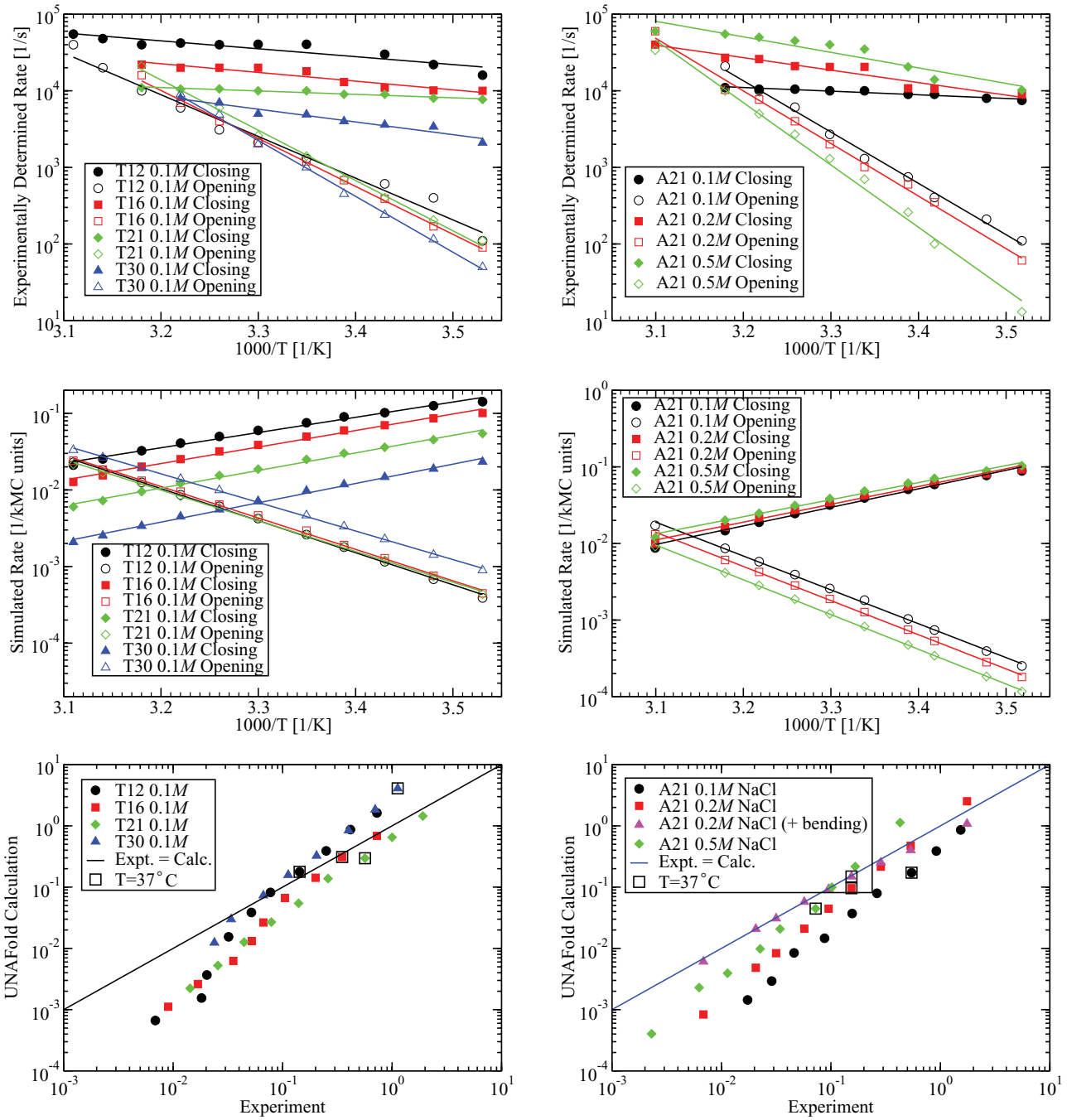

FIG. 4. (Color online) Comparison of experiment [19] and calculated rates. All lines shown are a linear regression of the correspondingly colored data. The left-hand column shows sequences T_n for various n. The right-hand column shows sequence A_{21} at various salt concentrations. (Top row) Experimental opening (k_{open}) and closing (k_{close}) rates. (Middle row) Simulations utilizing the Kawasaki rate model. (Bottom row) Equilibrium constants $K = k_{\text{open}}/k_{\text{close}}$ compared between experiment and calculation.

rate model results in a positive slope of the closing rates for both the A21 sequence and the U_n sequences (U_n rather than T_n because we treat RNA), while introducing our Arrhenius rates changes these positive slopes to negative.

Other options for improving the agreement between the kinetic Monte Carlo simulation and experiment over a broader temperature range were considered and discarded. First, consider a time-dependent rate to attempt a move corresponding to the DNA polymer colliding with itself. This could not provide the exponential dependence on $1/T$ observed, as the characteristic relaxation time τ_r given by the Rouse model has $\tau_r \propto 1/T$ [28]. Alternatively, suppose there were some cost for nucleating [29] a DNA hairpin $\Delta G_{\text{nuc}} = \Delta H_{\text{nuc}} - T\Delta S_{\text{nuc}}$. Such a cost, if it exists, should in principle be included in the free-energy model. However, the model assigns a negative $\Delta H_{\text{nuc}} < 0$ for the first pair, even including the adjacent dangling nucleotides. An enthalpy increase of $\Delta H_{\text{nuc}} = +11$ kcal/mol would be needed to reproduce the experimental

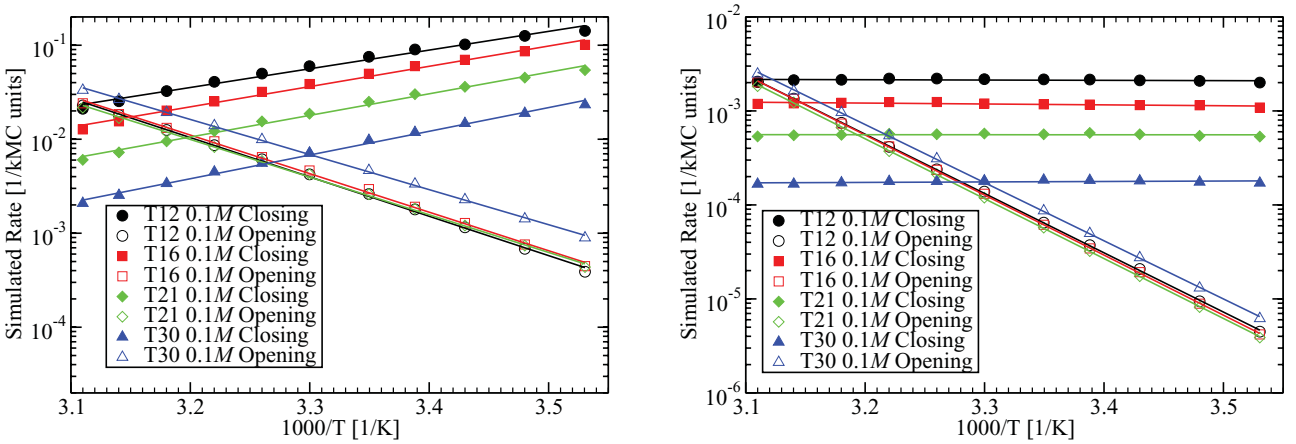

FIG. 5. (Color online) Simulated opening and closing rates of T_n sequences. (Left) Kawasaki rate model as previously shown in Fig. 4. (Right) Arrhenius rate model with a Kawasaki cutoff. Lines shown are linear regressions of the corresponding colored data.

data for the T12 sequence, for example, which is presumably far outside the realm of experimental error, given that SantaLucia reports $\Delta H_{\text{nuc}} < 0$. In our model, a sequence-dependent rigidity could serve as the source of a positive ΔH_{nuc}, since it would be provided by the loop in the almost-closed state, and combined with the intermediate-state energy model above to remove the favorable double stranded stacking and dangling end contributions, would create an overall positive ΔH_{nuc} as required for agreement with experiment. Indeed, the nucleation and growth model [29] assumes that ΔG_{nuc} is dominated by entropy, consistent with our Arrhenius rate model for the large hairpin loops such as we consider here.

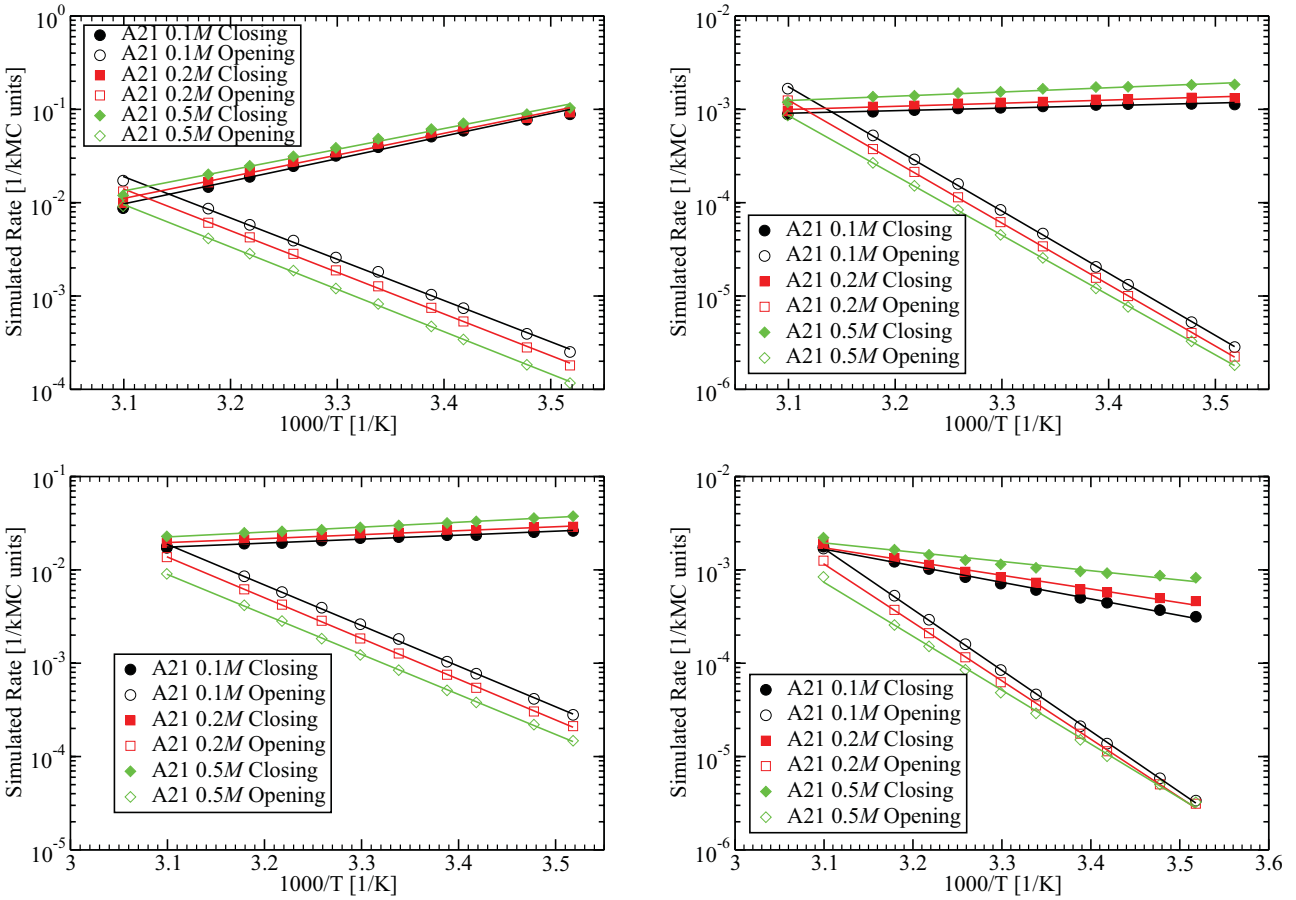

FIG. 6. (Color online) Simulated opening and closing rates of A_{21} sequences. Lines shown are linear regression of corresponding colored data points. (Top left) Conventional Kawasaki as previously shown in Fig. 4. (Top right) Arrhenius rate model with a Kawasaki cutoff. (Bottom left) Kawasaki rate model including stacking rigidity of 0.5 kcal/mol/nt. (Bottom right) Arrhenius rate model including stacking rigidity.

TABLE I. Calibrated time scales for an elementary kinfold conformation change of DNA at the data points nearest 37 °C using the Arrhenius rate model. A kinfold time unit equals the time to make a conformation change that does not affect the free energy of the global structure.

Sequence	NaCl (M)	Process	Expt. rate (1/s)	Calculated rate (1/kMC)	Time scale $1/\kappa$ (ns)
A_{21}	0.1	Closing	1.05×10^4	9.82×10^{-4}	93.5
A_{21}	0.2		2.60×10^4	1.09×10^{-3}	41.9
A_{21}	0.5		5.00×10^4	1.39×10^{-3}	27.8
T_{12}	0.1		4.20×10^4	2.20×10^{-3}	52.3
T_{16}	0.1		2.00×10^4	1.24×10^{-3}	62.0
T_{21}	0.1		1.06×10^4	5.70×10^{-4}	53.7
T_{30}	0.1		7.00×10^3	1.78×10^{-4}	25.4
A_{21}	0.1	Opening	1.01×10^4	2.90×10^{-4}	28.7
A_{21}	0.2		7.70×10^3	2.13×10^{-4}	27.7
A_{21}	0.5		5.00×10^3	1.51×10^{-4}	30.2
T_{12}	0.1		6.00×10^3	4.14×10^{-4}	69.0
T_{16}	0.1		7.00×10^3	4.18×10^{-4}	59.7
T_{21}	0.1		1.05×10^4	3.73×10^{-4}	35.5
T_{30}	0.1		9.00×10^3	5.44×10^{-4}	60.4

Given the improved agreement between simulation with the Arrhenius rule and experiment, we use this experimental data to calibrate the value of κ. Here, the DNA energy parameters from SantaLucia and Hicks are most accurate at 50 °C, but are intended to be used at 37 °C [20], where good agreement between experimental and calculated equilibrium constants K is seen in Fig. 4. Also, 37 °C is a common ambient temperature for biological processes. Therefore, the simulations here are compared to experiment at 37 °C, since experimental data is available in this regime and since it is the most common choice of temperature at which to simulate as well as experiment.

Results of this calibration are shown in Table I. Note the characteristic time scale $1/\kappa$ is of order tens of nanoseconds. Recall that the value of κ is sensitive to the dynamical model in which it will be used. For example, an equivalent time scale that does not explicitly include the intermediate barrier state must be much larger. We find κ of order 1 μs (microsecond) is required to match the experimental opening and closing times with the Kawasaki rate model.

V. CONCLUSIONS

This work advances our ability to simulate nucleic acid hairpin closing in three distinct ways. First, we recognize that the true barrier state for hairpin closing lies outside the state space of the widely used secondary structure models. This helps explain the puzzling slopes in transition rate versus $1/T$, where experiments on DNA showed opposite slopes from secondary structure simulation. Including the effects of the barrier provides better agreement with experiment than the Kawasaki or Metropolis rates in Monte Carlo simulations. Second, we demonstrate how the effects of this barrier state can be efficiently included in kinetic Monte Carlo simulations without losing the convenience of secondary structure models. We avoid adding up to N^2 new states, most of them merely transients, in a system where N secondary structures are the states of principal interest. Our improved kinetic Monte Carlo on the space of secondary structure is able to simulate the behavior of nucleic acids on physically and biologically relevant time scales. Finally, we show that including sequence-dependent rigidity of single-stranded DNA significantly improves the agreement of both free energies and transition rates with experiment. Single-stranded stacking energy parameters must be developed and algorithms adapted in order to accommodate this feature of nucleic acid loops, because this effect could significantly alter some predicted secondary structures.

ACKNOWLEDGMENTS

The authors wish to acknowledge Tanmay K. Mudholkar for implementing the UNAfold DNA parameters and salt corrections in the ViennaRNA 2.0 package. We also acknowledge useful discussions with Albert Libchaber and Oleg Krichevsky. We acknowledge financial support of this research by the DSF Charitable Foundation.

[1] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
[2] M. J. Klein, Phys. Rev. **97**, 1440 (1955).
[3] K. Kawasaki, Phys. Rev. **145**, 224 (1966).
[4] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[5] D. P. Bartel, Cell **116**, 281 (2004).
[6] B. J. Tucker and R. R. Breaker, Curr. Opin. Struct. Biol. **15**, 342 (2005).
[7] K. S. Wilson and P. H. von Hippel, Proc. Natl. Acad. Sci. USA **92**, 8793 (1995).
[8] J. Reeder and R. Geigerich, BMC Bioinformatics **5**, 104 (2004).
[9] J. Ren, B. Rastegari, A. Condon, and H. H. Hoos, RNA **11**, 1494 (2005).
[10] E. Bindewald, T. Kluth, and B. A. Shapiro, Nucl. Acids Res. **38**, W368 (2010).
[11] I. L. Hofacker, W. Fontana, P. F. Stadler, S. Bonhoeffer, M. Tacker, and P. Schuster, Monatsh. Chem. **125**, 167 (1994).
[12] M. Zuker and P. Stiegler, Nucl. Acids Res. **9**, 133 (1981).
[13] J. S. McCaskill, Biopolymers **29**, 1105 (1990).
[14] N. R. Markham and M. Zuker, Methods Mol. Biol. **453**, 3 (2008).
[15] A. Xayaphoummine, V. Viasnoff, S. Harlepp, and H. Isambert, Nucl. Acids Res. **35**, 614 (2007).
[16] C. Flamm, W. Fontana, I. L. Hofacker, and P. Schuster, RNA **6**, 325 (2000).
[17] S. Sharma, F. Ding, and N. V. Dokholyan, Bioinformatics **24**, 1951 (2008).
[18] M. Parisien and F. Major, Nature (London) **452**, 51 (2008).
[19] G. Bonnet, O. Krichevsky, and A. Libchaber, Proc. Natl. Acad. Sci. USA **95**, 8602 (1998).
[20] J. Santalucia Jr., and D. Hicks, Annu. Rev. Biophys. Biomol. Struct. **33**, 415 (2004).

[21] L. V. Danilova, D. D. Pervouchine, A. V. Favorov, and A. A. Mironov, J. Bioinformat. Comput. Biol. **4**, 589 (2006).
[22] F. Liu and Z.-C. Ou-Yang, Biophys. J. **88**, 76 (2005).
[23] G. Alton-Bonnet, A. Libchaber, and O. Krichevsky, Phys. Rev. Lett. **90**, 138101 (2003).
[24] H. Jacobson and W. Stockmayer, J. Chem. Phys. **18**, 1600 (1950).
[25] B. Sauerwine and M. Widom, J. Stat. Phys **142**, 1337 (2011).
[26] N. L. Goddard, G. Bonnet, O. Krichevsky, and A. Libchaber, Phys. Rev. Lett. **85**, 2400 (2000).
[27] Z. J. Lu, D. H. Turner, and D. H. Mathews, Nucl. Acids Res. **34**, 4912 (2006).
[28] P. E. Rouse, J. Chem. Phys **21**, 1272 (1953).
[29] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry* (W. H. Freeman, San Francisco, 1980), Vol. 3.