JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 12, Number 6, 2005

Pp. 862–881

Using Motion Planning to Study RNA Folding Kinetics

XINYU TANG,[1] BONNIE KIRKPATRICK,[2] SHAWNA THOMAS,[1] GUANG SONG,[3] and NANCY M. AMATO[1]

ABSTRACT

We propose a novel, motion planning based approach to approximately map the energy landscape of an RNA molecule. A key feature of our method is that it provides a sparse map that captures the main features of the energy landscape which can be analyzed to compute folding kinetics. Our method is based on probabilistic roadmap motion planners that we have previously successfully applied to protein folding. In this paper, we provide evidence that this approach is also well suited to RNA. We compute population kinetics and transition rates on our roadmaps using the master equation for a few moderately sized RNA and show that our results compare favorably with results of other existing methods.

Key words: RNA, folding kinetics, motion planning.

1. INTRODUCTION

Ribonucleic acid (RNA) molecules perform diverse and important functions such as synthesizing proteins, catalyzing reactions, splicing introns, and regulating cellular activities (Tinoco and Bustamante, 1999). An RNA's nucleotide sequence and the three-dimensional structure of its energetically stable conformations determine how the RNA functions and interacts with its environment. The process by which an RNA molecule (re)configures itself into an energetically stable conformation is called *folding*.

There are two general, but related, issues for RNA folding: structure prediction and folding kinetics. The structure prediction problem is to predict the structure of the native conformation given the RNA's nucleotide sequence. Unlike for the related protein folding problem, efficient algorithms do exist for some forms of RNA structure prediction (Walter *et al.*, 1994; Zuker *et al.*, 1999). However, they do not provide insight into the folding process or the "energy landscape" which determines the folding kinetics.

Each RNA conformation is associated with an energy; the lower its energy, the more stable it is. The energy landscape can be thought of as adding this energy as another dimension to the other parameters specifying the conformation. As will be described in detail later, the energy landscape encodes information about folding pathways, transition rates, intermediate states, and population kinetics. The size of the landscape grows exponentially with the sequence length, so it is infeasible to compute the complete landscape. To reduce this complexity, researchers focus on RNA secondary structures that have planar

[1]Parasol Lab, Dept. of Computer Science, Texas A&M University, College Station, TX 77843-3112.
[2]Dept. of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720.
[3]Baker Center for Bioinformatics and Biological Statistics, Iowa State University, Ames, IA 50011.

representations instead of three-dimensional conformations. Although this dramatically reduces the size of the landscape, it remains impractical to compute the complete landscape for sequences longer than about 40 nucleotides (Cupal *et al.*, 1997).

There are at least three important reasons to study RNA energy landscapes and folding kinetics. First, a better understanding of the folding process will aid the development of more efficient structure prediction algorithms. Second, it has recently been discovered that catalytic RNA often fluctuate away from their native conformation to interact with other RNA, proteins, and ligands (Tinoco and Bustamante, 1999), and we cannot model or predict these fluctuations without studying energy landscapes. Third, we must study energy landscapes and folding kinetics to understand how and why RNA molecules misfold.

In this paper, we propose a novel, motion planning based approach to approximately map the RNA's energy landscape. In particular, we develop a *probabilistic roadmap* (PRM) (Kavraki *et al.*, 1996) based approach that first samples RNA configurations and then connects them together to form a graph, or *roadmap*. The key advantage of our method is that it provides a sparse representation of the landscape that captures its main features which can be analyzed to compute folding kinetics. We have previously applied this strategy to protein folding with considerable success (Song and Amato, 2001; Amato and Song, 2002; Amato *et al.*, 2003); e.g., our method predicted the subtle folding differences between the structurally similar proteins G and L (Song *et al.*, 2003). In this paper, we provide evidence that this approach is also well suited to RNA. In particular, we present results such as population kinetics and transition rates obtained using the master equation (Section 4.1) for a few moderately sized RNA and show that our results match results obtained with other existing methods quite well (Zhang and Chen, 2002).

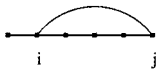
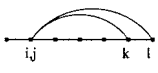
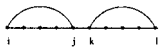
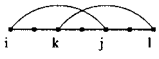
2. PRELIMINARIES AND RELATED WORK

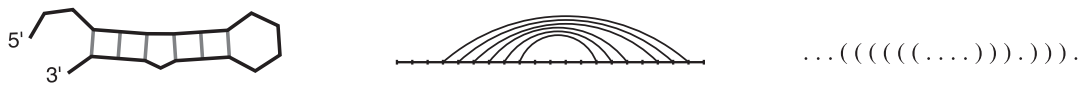
2.1. RNA primer

An RNA molecule is a sequence of nucleotides which differs from other RNA molecules in its bases. There are four types of bases: adenine (A), cytosine (C), guanine (G), and uracil (U). The complementary Watson–Crick bases, C-G and A-U, form stable, hydrogen bonds (*base pairs*) when they form a contact. The wobble pair G-U constitutes another strong base pair. These are the three most commonly considered base pairings (Zuker and Sankoff, 1984; Walter *et al.*, 1994; Hofacker, 1998) and are also what we consider in our model.

RNA structure. *Tertiary structure* is a 3D spatial RNA conformation with a set of base pairs. *Secondary structure* is a planar representation of an RNA conformation. Although there are slightly differing definitions (Hofacker, 1998; Chen and Dill, 2000), secondary structure is usually considered to be a planar subset of the base pair contacts present (see Table 1, Case 3). Nonplanar contacts, often called *pseudoknots*, are usually considered tertiary interactions and not allowed in secondary structure. Many definitions of secondary structure, including the one we adopt, eliminate other types of contacts that are not physically favored. Contacts considered invalid in our secondary structure are defined in Table 1; this definition is

TABLE 1. DEFINITION OF VALID SECONDARY STRUCTURE FOR ANY TWO CONTACTS $[i, j]$ AND $[k, l]$ WITH $i < j$ AND $k < l$

Description	*Valid contact*	*Invalid contact*
Case 1: (Separation) Bases of each pair must be separated by at least 3 other residues, i.e., $\|i - j\| > 3$		
Case 2: (Multiplicity) Each base can be paired to only one other, i.e., $i = k$ if and only if $j = l$		
Case 3: (Planarity) The contacts must be planar (no pseudoknots), i.e., if $i < k < j$, then $i < k < l < j$		



(a) (b) (c)

FIG. 1. Three representations of the same secondary structure for the sequence GGCGUAAGGAUUACCUAUGCC which denote contact pairs with bonds (**a**), arcs (**b**), and pairs of brackets (**c**).

also used by Hofacker (1998). Three common representations for RNA secondary structure are shown in Fig. 1 (Zuker and Sankoff, 1984).

The tertiary structure gives the most complete representation of RNA structure. However, the secondary structure is commonly used (Zuker and Sankoff, 1984; Hofacker, 1998; Zuker *et al.*, 1999) because in many cases it provides sufficient information to study many aspects of folding while dramatically reducing the size of the RNA conformation space to explore. One justification for this simplification is that research has shown that the RNA folding process is hierarchical, i.e., that secondary structure forms before tertiary structure (Zuker and Sankoff, 1984; Tinoco and Bustamante, 1999). In this work, we focus on the first stage, the formation of secondary structure.

Energy calculations. To represent the RNA folding energy landscape, we must be able to calculate the energy of any conformation. One commonly used energy function is the Turner or nearest neighbor rules (Zuker *et al.*, 1999). This method involves determining the types of loops that exist in the molecule and looking up their free energy in a table of experimentally determined values. Intuitively, more contacts, especially adjacent contacts, typically yield more stable structures with lower energy. Much work has been done to make these rules more detailed and accurate.

RNA (secondary structure) conformation space. For a given RNA nucleotide sequence, an RNA (secondary structure) conformation is a planar set of valid base pairs.[1] The secondary structure conformation space, $\mathcal{C}$, of an RNA sequence contains all sets of base pairs that meet the criteria in Table 1. The size of $\mathcal{C}$, $|\mathcal{C}|$, grows exponentially as sequence length increases (Zuker and Sankoff, 1984; Cupal *et al.*, 1996). Knowledge of $|\mathcal{C}|$ is used to determine the feasibility of enumerating all conformations or whether some sampling will be needed. Not only does $|\mathcal{C}|$ depend on the RNA sequence length but also on the sequence itself. Since exact computation of $|\mathcal{C}|$ requires enumerating $\mathcal{C}$, it should be estimated.

Zuker and Sankoff (1984) developed a close estimation of $|\mathcal{C}|$ using a stochastic approach to account for the effect of the specific sequence. Given an RNA sequence of length n, they calculate the probabilities p_A, p_C, p_G, and p_U of the occurrence of each nucleotide, i.e., the percentage of that nucleotide in the sequence. They then use $p = 2(p_A p_U + p_C p_G)$ as the probability of two bases making a contact and obtain the approximation $|\mathcal{C}| \approx h n^{\frac{3}{2}} \alpha^n$, where $\alpha = (\frac{1+\sqrt{1+4\sqrt{p}}}{2})^2$ and $h = \frac{\alpha(1+4\sqrt{p})^{1/4}}{2\sqrt{\pi}\, p^{3/4}}$.

Unfortunately, however, the Zuker and Sankoff estimate does not fit our model because they do not consider the wobble pair G-U or the restriction of the minimal hairpin size to 5. We modified this formula to fit our model by including the wobble pair in the probability $p' = 2(p_A p_U + p_C p_G + p_U p_G)$, and then scaling the probability p' to $p = p' \cdot (n-3)(n-4)/n^2$ to restrict the minimal hairpin size to 5. Our revised estimate results from substituting the new p in the equations for α and h.

As can be seen in Table 2, our estimate can be a significantly better estimate of $|\mathcal{C}|$ for our model than the estimate used by Zuker and Sankoff (1984). Our exact enumeration results match Cupal (Hofacker, 1998). It can also be seen that $|\mathcal{C}|$ grows exponentially with sequence length, and hence it becomes impractical to enumerate all conformations when the sequence length exceeds 40 nucleotides (Cupal *et al.*, 1997), and thus some type of sampling must be used instead.

2.2. PRMs and protein folding

Our approach to RNA folding is based on the *probabilistic roadmap* (PRM) technique for motion planning (Kavraki *et al.*, 1996). Motion planning determines valid paths to move objects from one conformation

[1] As we consider only secondary structure in our method, we will usually omit this qualification when referring to conformations and conformation space.

TABLE 2. ESTIMATED AND ACTUAL SIZES OF C-SPACE FOR SEVERAL RNA SEQUENCES

Sequence	*# nucl*	*Exact $\|\mathcal{C}\|$*	*Zuker's estimation (Zuker and Sankoff, 1984)*	*Our estimation*
$(ACGU)_2$	8	5	22	6
$(ACGU)_3$	12	35	206	47
ACUGAUCGUAGUCAC	15	1.4×10^2	1.0×10^3	2.4×10^2
GGCGUAAGGAUUACCUAUGCC	21	8.6×10^3	6.2×10^5	1.3×10^4
$(ACGU)_{10}$	40	1.7×10^8	1.6×10^{10}	3.3×10^9

to another. PRMs build graphs (roadmaps) that ideally approximate the topology of the feasible planning space, and can be used to answer many, varied queries quickly. Briefly, PRMs work by sampling points from the movable object's conformation space (C-space) and retaining those that satisfy feasibility requirements (e.g., collision-free). The movable object's C-space is the set of all positions and orientations of the movable object, feasible or not (Latombe, 1991). Next, the retained points are connected to form a graph, or roadmap, using some simple local planning method (e.g., a straight line) to connect nearby points. During query processing, paths connecting the start and goal are extracted from the roadmap using standard graph search techniques (see Fig. 2).

In previous work, we used PRMs to study protein folding when the native structure is known (Song and Amato, 2001; Amato and Song, 2002; Amato *et al.*, 2003; Song *et al.*, 2003). Here, the moving object is the protein, and the main difference from the usual PRM application is that the collision-detection feasibility test is replaced by a preference for low energy conformations. We have obtained promising results that were validated with experimental data for several moderately sized proteins; e.g., we were able to observe the subtle folding differences in the structurally similar proteins G and L (Song *et al.*, 2003).

2.3. Related work

Research on RNA folding falls into two categories: structure prediction and the study of folding kinetics. Structure prediction is commonly solved with dynamic programming. Nussinov introduced a dynamic programming solution to find the conformation with the maximum number of base pairs (Nussinov *et al.*, 1972). Zuker and Stiegler (1981) formulated an algorithm to address the minimum energy problem. Today, Zuker's MFOLD algorithm is widely used for structure prediction (Walter *et al.*, 1994). McCaskill's algorithm (McCaskill, 1990) uses dynamic programming to calculate the partition function $Q = \sum_s exp(-\Delta G(s)/kT)$ over all secondary structures s, while Chen (Chen and Dill, 2000) uses matrices for approximation. As described by Chen and Dill (2000), the partition function is "the sum of Boltzmann factors over all possible branching patterns in which the chain can be arranged into helices and intervening regions." As we will see, the partition function can also be used to study folding kinetics. Wuchty (1998) extended the algorithm to compute the density of states at a predefined energy resolution. The ViennaRNA package, based on above work, implements Zuker and McCaskill's algorithms as well as some energy functions (Hofacker, 1998).

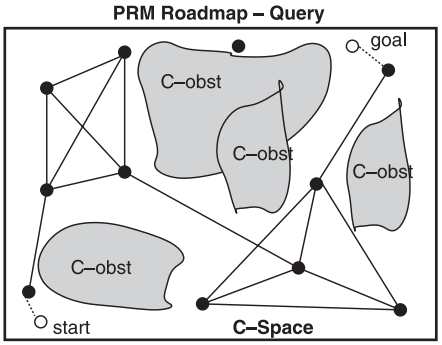


FIG. 2. A PRM roadmap in C-space and a query.

Several approaches other than thermodynamics have been used to investigate RNA kinetics, e.g., Gillespie (1977) used a Monte Carlo algorithm to find folding pathways, and Shapiro *et al.* (2001) used a genetic algorithm to study RNA folding pathways.

Several methods have been proposed that involve computations on the folding landscape. Dill (Chen and Dill, 2000) used matrices to compute the partition function over all possible structures. Complete folding landscapes are approximated by this method. Wuchty (1998) modified Zuker's algorithm to generate all the secondary structures within some given energy range of the native structure. Flammy and Wolfinger (Flamm, 1998; Wolfinger, 2001) extended this algorithm to find local minima within some energy difference of the native state, then to connect them via energy barriers. Their algorithm, Kinfold, uses an energy barrier tree to represent the energy landscape.

The master equation can be used to compute the population kinetics of the folding landscape. It uses a matrix of differential equations to represent the probability of transition between different conformations. Once solved, the dominant modes of the solution describe the general folding kinetics (Kampen, 1992; Zhang and Chen, 2002; Ozkan *et al.*, 2002, 2003).

3. RNA FOLDING WITH PRMs

In this section, we discuss how to apply PRMs to study RNA folding. There are two main steps in our approach: constructing the roadmap and analyzing it. Constructing the roadmap requires sampling a set of RNA conformations and computing their energies. Next, we determine which conformations we should attempt to connect and try to connect them using a "local planner," i.e., a simple method to find transitions between RNA conformations. One difference from the protein folding application is that our C-space is not continuous but discrete, and hence our options for making the local connections are more restricted. The local planner also assigns weights to the transitions to reflect their energetic feasibility. This results in a roadmap (graph) of conformations (nodes) connected by transitions (edges) that represents the energy landscape and where each pathway is a sequence of conformational changes the RNA molecule goes through as it transforms from one conformation to another.

After the roadmap is built, we perform some analysis on it to study the population kinetics and provide insight into the folding process. We can identify transitional conformations where the folding process could be trapped or delayed, the folding rate, and representative folding pathways. In this work, we analyze the landscape via folding pathways and the master equation.

Energy computations are required to measure conformation feasibility and to calculate the roadmap edge weights (as discussed in Section 3.1.2). Our current implementation uses a third-party energy function relying on the Turner rules to determine the validity of a point in C-space. This energy function is part of the ViennaRNA package (Hofacker, 1998).

3.1. Roadmap construction

The goal of roadmap construction is to build an approximation of the energy landscape that captures its important features. The approximation quality depends on our node sampling and connection methods.

3.1.1. Node generation. Our framework currently has three methods for generating RNA conformations: complete base-pair enumeration (for small RNA), stack-pair enumeration, and maximal-contact sampling.

Complete base-pair enumeration (BPE). Our discrete RNA C-space makes it possible to enumerate all the conformations for small RNA molecules. However, it is not feasible for molecules with more than 40 nucleotides (Cupal *et al.*, 1997). Let $\mathcal{S}$ be the set of all possible base-pair contacts. To generate a valid conformation, we first select one contact in $\mathcal{S}$. Then we remove all contacts from $\mathcal{S}$ that would violate the criteria in Table 1 if combined with already selected contacts. The process of selecting a valid contact from $\mathcal{S}$ and then removing invalid contacts from $\mathcal{S}$ continues until $\mathcal{S}$ is empty. Each time we select a new contact, we define a new secondary structure. To enumerate the entire space, we enumerate all possible combinations of a valid set of contacts from $\mathcal{S}$ as above. Figure 3 shows the complete enumeration for the RNA sequence ACGUCACGU.

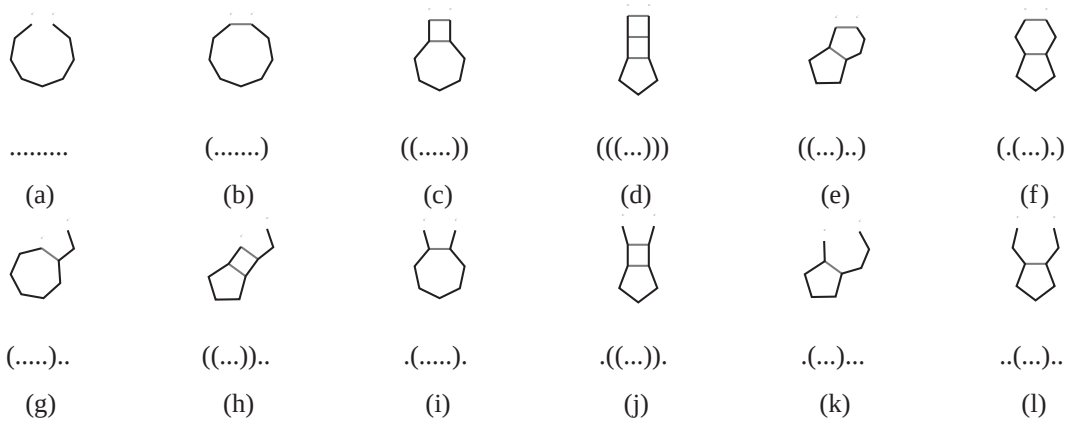


FIG. 3. Complete enumeration of all conformations for RNA sequence ACGUCACGU. Conformations (**a**), (**c**), (**d**), (**h**), and (**j**) are stack-pair conformations. Conformations (**a**), (**d**), (**e**), (**g**), and (**h**) are maximal-contact conformations.

Stack-pair enumeration (SPE). This enumeration contains only those conformations containing stack-pair contacts. A *stack-pair contact* is a set of adjacent base-pair contacts; i.e., no contacts are isolated from the others. More formally, if a stack-pair contact has a contact $[i, j]$, where $i < j$, then it must also have at least one of the contacts $[i - 1, j + 1]$ or $[i + 1, j - 1]$. For example, the contacts in Fig. 3(c) form a stack, but the contacts in Fig. 3(f) do not because they are not adjacent. A conformation is a valid *stack-pair conformation* if it has only stack-pair contacts, i.e., if there are no isolated base pairs. The conformations in Figs. 3(a), (c), (d), (h), and (j) are the enumeration of stack-pair conformations for RNA sequence ACGUCACGU. We favor these conformations because isolated base pairs are unstable. This simplification has been used by Zhang and Chen (2002). We can study larger RNA molecules with this method than is possible with complete enumeration because we can enumerate all stack-pair conformations without enumerating all conformations. The stack-pair enumeration is implemented similarly to the base-pair enumeration except that $\mathcal{S}$ contains stacks instead of base-contact pairs.

Maximal-contact sampling (MCS). In this method, nodes are generated in a more "random" fashion. To get lower energy conformations, we generate only conformations with maximal contacts; i.e., no more contacts can be added to those conformations without causing a violation (Table 1). First, we create a conformation c without any contacts. Then, single contacts are successively added until it is not possible to add a contact and maintain a valid conformation. This method biases the node distribution toward the areas of C-space with more contacts. Since more contacts usually means more stability for the conformation, the energy of these conformations is usually lower. Each time a contact is added, it is randomly selected from all currently feasible contacts, and the set of feasible contacts is updated. This continues until no more contacts can possibly be added. In Figs. 3(a), (d), (e), (g), and (h) are the maximal-contact conformations.

3.1.2. Node connection. After node generation, it would be expensive, and generally not necessary, to make all $\theta(n^2)$ connections. Here, we restrict our attention to connecting nearby conformations. This requires distance metrics to identify nearby conformations for connection and techniques for connecting them.

Distance metrics. The distance metric defines which nodes are close to each other and which are far apart. Here we use base-pair distance (the number of contact pairs that differ between two conformations). This denotes the number of base pairs that have to be opened or closed to transform one conformation into another. Our approach can utilize other distance metrics such as string edit distance or tree edit distance (Sankoff and Kruskal, 1983), but we found that base-pair distances perform the best on the RNA we have studied.

Identifying nodes for connection. Neighboring roadmap nodes are connected using a local planner. We use two different strategies for determining neighbors. One strategy attempts to connect a node with the k closest nodes, and the other attempts to connect a node with all nodes within a fixed radius r.

Generating transitional conformations. Once the neighbors are determined, the local planner connects each pair of nodes by generating transitions between them. To generate a transition from conformation

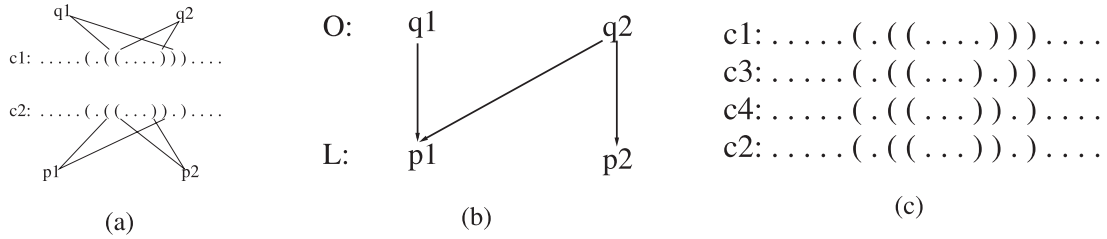


FIG. 4. Transitional node generation. (**a**) Start and goal conformations and contact pairs to be opened and closed: q_1, q_2 are in $\mathcal{O}$; p_1, p_2 are in $\mathcal{L}$. (**b**) Conflict graph: q_1 and q_2 conflict with p_1, q_2 conflicts with p_2. (**c**) Sequences generated: First open q_2 and close p_2, then open q_1 and close p_1. Conformations c_3 and c_4 are the two transitional conformations to connect c_1 and c_2; here, c_4 happens to be identical to c_2.

c_1 to conformation c_2, we first identify the set $\mathcal{O}$ of contacts to be opened (i.e., contacts in c_1 but not in c_2) and the set $\mathcal{L}$ of contacts to be closed (i.e., contacts in c_2 but not in c_1). See Fig. 4(a): contacts q_1 and q_2 are in $\mathcal{O}$; contacts p_1 and p_2 are in $\mathcal{L}$. To ensure that transitional conformations do not violate our planarity constraint, we construct a *conflict graph* G between $\mathcal{O}$ and $\mathcal{L}$. Graph G describes which contact pairs cannot exist together in a valid conformation. If one contact $p \in \mathcal{L}$ conflicts with another contact $q \in \mathcal{O}$, then p cannot be closed until q is opened, and we have an edge from q to p in G. See Fig. 4(b): q_1 and q_2 conflict with p_1; q_2 conflicts with p_2. A valid transition is a sequence of transitional conformations that doesn't violate G.

Our framework can use any strategy to determine the order to open contacts in $\mathcal{O}$ and close contacts in $\mathcal{L}$. The most naive method is to first open all the contacts in $\mathcal{O}$ and then to close all the contacts in $\mathcal{L}$. This does not violate G, but it produces high energy transitional conformations. To find low energy transitions, we want to produce conformations with as many contacts as possible since they usually have lower energy. So, once we open a contact, we close all contacts in $\mathcal{L}$ that do not violate G.

We use a greedy strategy to determine the order for opening the contacts. In particular, we sort the contacts in $\mathcal{L}$ according to the number of contacts in $\mathcal{O}$ they conflict with (given by their indegree in $\mathcal{L}$). We select the contact in $\mathcal{L}$ with the smallest number of conflicts and open all the contacts in $\mathcal{O}$ that conflict with it. We then close all the contacts in $\mathcal{L}$ that have no conflicts. See Fig. 4(c): c_3, c_4 are the two transitional conformations generated for the connection. This is repeated until both $\mathcal{O}$ and $\mathcal{L}$ are empty. This strategy works well for the RNA we have studied.

Edge weights. Edge weights are assigned to reflect the transition rate between neighboring conformations, i.e., the probability the molecule folds from one conformation to the other. Thus, the edge weight reflects the energetic feasibility for the folding process on the edge.

When an edge (q_1, q_2) is added to the roadmap, it is assigned a weight that depends on the sequence of transitional conformations $\{q_1 = c_0, c_1, c_2, \ldots, c_{n-1}, c_n = q_2\}$ determined by the local planner. For each pair of consecutive conformations c_i and c_{i+1}, the probability P_i of moving from c_i to c_{i+1} is

$$P_i = \begin{cases} e^{\frac{-\Delta E_i}{kT}} & \text{if } \Delta E_i > 0 \\ 1 & \text{if } \Delta E_i \leq 0 \end{cases} \tag{1}$$

where $\Delta E_i = E(c_{i+1}) - E(c_i)$, k is the Boltzmann constant, and T is the temperature of folding. For a detailed discussion of different rules to calculate the transition probabilities, please refer to Dill and Chan (1997). The edge weight $w(q_1, q_2)$ is calculated as

$$w(q_1, q_2) = \sum_{i=0}^{n-1} -log(P_i). \tag{2}$$

(Negative logs are used since each $0 \leq P_i \leq 1$.) By assigning the weights in this manner, we can find the most energetically feasible path in our roadmap when performing queries. This is the same method used in our previous work on protein folding.

4. ROADMAP ANALYSIS

The roadmap is an approximation of the folding landscape, and it can be used to study individual folding pathways as well as the global folding kinetics.

A folding pathway is a sequence of transitional conformations the RNA molecules goes through during the folding process from any unfolded conformation to the native conformation. As with our previous work on protein folding (Amato *et al.*, 2003), we can extract folding pathways and compute the free-energy profile, energy barriers, and important states of the folding process. From all the folding pathways to the native conformation, we extract the pathway with minimum total weight because this corresponds to the most energetically feasible path *in our roadmap*. Individual pathway results are provided for one case study in Section 5.4.

Beyond the study of specific folding pathways, we are interested in the global properties of the energy landscape, for example, how the population of conformations in the landscape vary as a function of time, i.e., the population kinetics. Folding rates and transition states are also of great interest. These can all be studied using the master equation.

4.1. Folding kinetics and the master equation

Master equation formalism has been developed for folding kinetics in a number of earlier studies (Kampen, 1992; Zhang and Chen, 2002). The stochastic process of folding is represented as a set of transitions among all n conformations (states). The time evolution of the population of each state, $P_i(t)$, can be described by the following differential equation:

$$dP_i(t)/dt = \sum_{i \neq j}^{n} (k_{ji} P_j(t) - k_{ij} P_i(t)) \tag{3}$$

where k_{ij} denotes the transition rate from state i to state j. Thus the change in population $P_i(t)$ is the difference between transitions *to* state i and transitions *from* state i. The transition rates are computed from the roadmap's edge weight: $k_{ij} = K_0 e^{-w_{ij}}$ where K_0 is a constant adjusted according to experimental results.

If we use an n-dimensional column vector $\mathbf{p}(t) = (P_1(t), P_2(t), \ldots, P_n(t))'$ to denote the population of all n conformational states, then we can construct an $n \times n$ matrix M to represent the transitions, where

$$\begin{cases} M_{ij} = k_{ji} & i \neq j \\ M_{ii} = -\sum_{i \neq j} k_{ij} & i \neq j \end{cases}. \tag{4}$$

The master equation can be represented in matrix form:

$$d\mathbf{p}(t)/dt = M\mathbf{p}(t). \tag{5}$$

The solution to the master equation is

$$P_i(t) = \sum_k \sum_j N_{ik} e^{\lambda_k t} N_{kj}^{-1} P_j(0) \tag{6}$$

where N is the matrix of eigenvectors N_i for the matrix M in Equation (4), Λ is the diagonal matrix of its eigenvalues λ_i, and $P_j(0)$ is the initial population of conformation j.

From Equation (6), we see that the eigenvalue spectrum is composed of n modes. If sorted by magnitude in ascending order, the eigenvalues include $\lambda_0 = 0$ and several small magnitude eigenvalues. Since all the eigenvalues are negative, the population kinetics will stabilize over time. The population distribution $\mathbf{p}(t)$ will converge to the equilibrium Boltzmann distribution, and no mode other than the mode with the zero eigenvalue will contribute to the equilibrium. Thus the eigenmode with eigenvalue $\lambda_0 = 0$ corresponds to the stable distribution, and its eigenvector corresponds to the Boltzmann distribution of all conformations

in equilibrium. To validate our implementation, we compared our master equation results to the Boltzmann distribution, and they match exactly.

For the same reason, we see that the large magnitude eigenvalues correspond to the fast-folding modes, that is, those modes which fold in a burst. Their contribution to the population will die away quickly. Similarly, the smaller the magnitude of the eigenvalue is, the more influence its corresponding eigenvector has on the global folding process. Thus, the global folding rates are determined by the slow modes.

For some folders (i.e., 2-state folders), their folding rate is dominated by only one nonzero slowest mode. If we sort the eigen spectrum by ascending magnitude, there will be one other eigenvalue λ_1 in addition to eigenvalue λ_0 that is significantly smaller in magnitude than all other eigenvalues. This λ_1 corresponds to the folding mode which determines the global folding rate. We will refer it as the *master folding mode*. Its corresponding eigenvector denotes its contribution to the population of each state. Hence, the large magnitude components of the eigenvector correspond to the states whose populations are most impacted by the master folding mode. These states are the transition states (Ozkan *et al.*, 2002, 2003).

5. RESULTS AND DISCUSSION

With our kinetics analysis tools, we are able to evaluate our roadmap-based approximation of the energy landscape. Generally, the best way to evaluate an approximation is to compare it to the exact method. Thus, ideally, we should compare the "exact" full base-pair enumeration (BPE) roadmap with the "approximate" stack-pair enumeration (SPE) and the maximal contact sampling (MCS) roadmaps. As previously mentioned, we can afford to do a full enumeration for RNA with up to approximately 40 nucleotides (Cupal *et al.*, 1997). Also, there is currently a limit on the size of the master equation we can accurately handle (due to a limitation in our present implementation). For these reasons, we performed the full comparison on a 15 nucleotide sequence, an 18-nucleotide sequence, two 21-nucleotide sequences, and a 22-nucleotide sequence. As shown in Table 3, we used all three strategies to generate nodes for the roadmap. For the maximal-contact sampling, we tried to generate at least twice as many nodes generated with the stack-pair enumeration.

Each of the roadmaps was connected using both the k-closest and the radius connection strategies described in Section 3.1.2. The parameters for these methods need to be carefully selected. The roadmap generated using complete base-pair enumeration and radius connections with $r = 1$ corresponds to a fine mesh on the energy landscape. Recall that our distance metric is the base-pair distance, therefore setting $r = 1$ creates transitions between all pairs of nodes differing by a single contact. This roadmap is used as a basis for comparison to determine appropriate k and r values for the other node generation methods. If we increase k or r, the connections will be more complete and more expensive. We want these parameters to be as small as possible yet still large enough to capture the important transitions.

For the complete base-pair enumeration, we tested $k = 5$, 10, 15, 20, 30, and 40. We found the smallest values that closely matched the complete landscape for each RNA, respectively. For the other sampling strategies, the distance between conformations is usually greater than 1, thus, we must use larger values for r and k. To determine the appropriate parameters, we compared the kinetics results using $r = 1$, 2, 5, 10, and 20, and $k = 5$, 10, 15, 20, 30, and 40. For the 21-nucleotide RNA, $k = 40$ always generated a close approximation to the complete energy landscape.

Table 3 gives the roadmap sizes, parameters used in the detailed results comparison, and running times for each RNA sequence studied. Note that the number of nodes for MCS differs due to different random number generator seeds. We show results only for the k-closest-connected roadmaps for three of the RNA sequences studied. However, for every RNA in Table 3, we found the population kinetics, eigenvalues, and eigenvectors of the BPE roadmap and the SPE roadmap to match very well.

We also compared our results with a modified version of the Kinfold algorithm (Wolfinger, 2001). Kinfold simulates RNA folding as a Markov process in the RNA's conformation space. Moves consist of opening, closing, and shifting individual base pairs. Kinfold uses the same energy evaluation function as our method (Hofacker, 1998), although the energy constants and transition rate constants may be different, which may cause some discrepancies in results. Kinfold, however, makes two assumptions not found in our approach. (1) Kinfold assumes that the native state is "ground-state absorbing," i.e., that conformations may transit to the native state but may not transit out of it. (2) At each time step, Kinfold considers

TABLE 3. COMPARISON BETWEEN DIFFERENT ROADMAP CONSTRUCTION STRATEGIES[a]

Name	*Sequence*	*# nucleotides*	*Generation method*	*# nodes*	*Connection method*	*# edges*	*Running time (sec)*
RNA0	ACUGAUCGUAGUCAC	15	BPE	142	k-closest 10	865	1.00
					radius 1	307	0.28
			SPE	15	k-closest 10	80	0.07
					radius 10	105	0.04
			MCS	33	k-closest 10	186	1.02
				33	radius 20	406	0.24
RNA1	CGCGCUACUCCUAGAGCU	18	BPE	876	k-closest 20	11,006	36.43
					radius 1	2,498	8.41
			SPE	22	k-closest 20	225	0.23
					radius 20	231	0.18
			MCS	157	k-closest 20	1,928	3.02
				161	radius 40	12,880	10.45
RNA2	UAUAUAUCGACACGAUAUAUA	21	BPE	5,353	k-closest 40	133,165	1,548.58
					radius 1	121,656	504.08
			SPE	250	k-closest 40	5,890	15.49
					radius 20	30,001	14.36
			MCS	612	k-closest 40	14,565	113.07
				578	radius 40	166,753	208.27
RNA3	GGCGUAAGGAUUACCUAUGCC	21	BPE	8,622	k-closest 40	216,033	4,017.87
					radius 1	205,064	1,308.50
			SPE	167	k-closest 40	4,110	9.28
					radius 20	13,262	6.23
			MCS	573	k-closest 40	13,832	47.22
				548	radius 40	159,895	182.14
1K2G	CAGACUUCGGUCGCAGAGAUGG	22	BPE	12,137	k-closest 40	304,595	8,230.84
					radius 1	45,706	2,155.03
			SPE	71	k-closest 40	1,736	3.33
					radius 20	2,485	2.08
			MCS	641	k-closest 40	15,739	56.93
				654	radius 40	213,531	276.11
28nt	GGCGUCAGGUCCGGAAGGAAGCAGCGCC	28	BPE	132,596	k-closest 40	4,612,854	984,280.00
			SPE	246	k-closest 40	6,288	23.47
					radius 20	30,135	43.95
			MCS	946	k-closest 40	23,162	99.75
				944	radius 40	445,096	1,041.38

[a]BPE, SPE, and MCS denote base-pair enumeration, stack-pair enumeration, and maximal contact sampling.

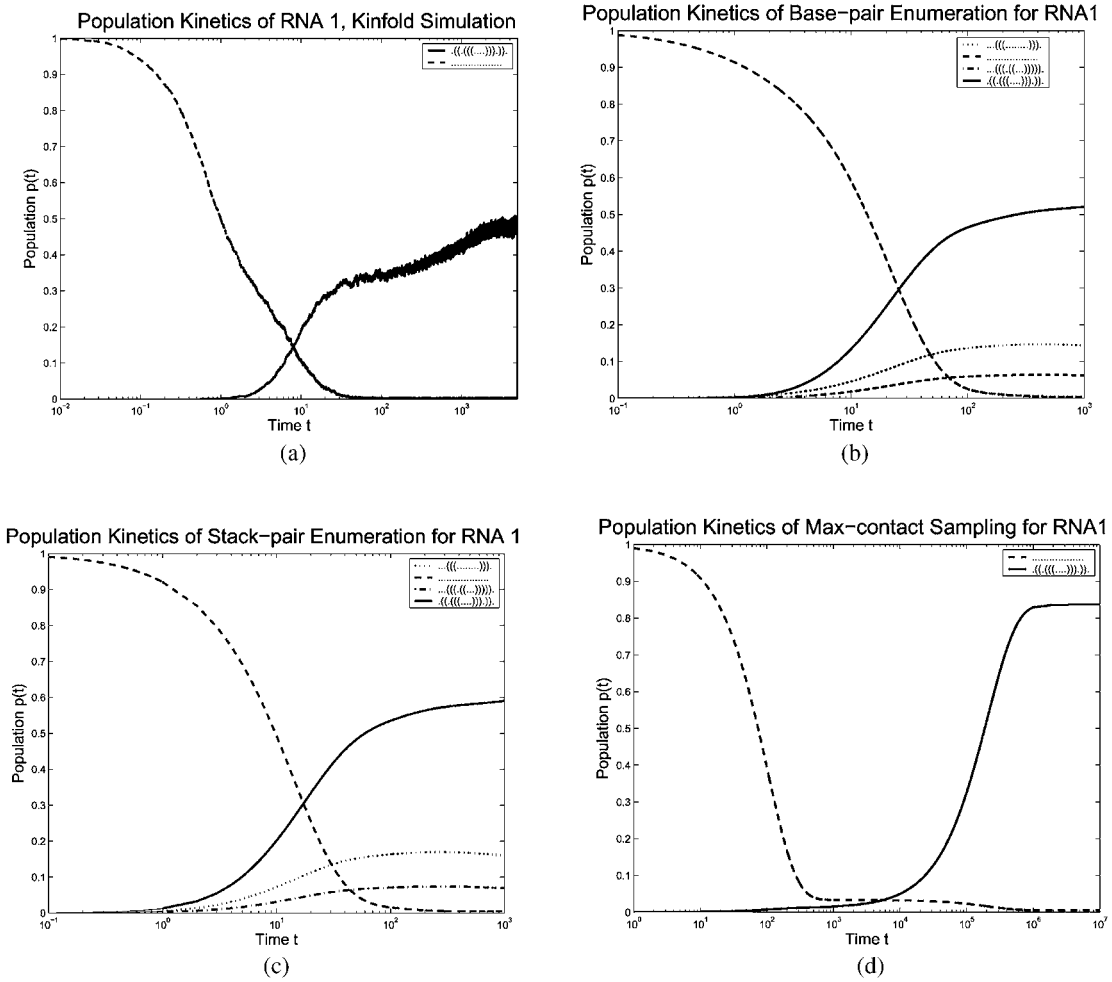


FIG. 5. The population kinetics of the 18 nucleotide hairpin sequence CGCGCUACUCCUAGAGCU with the native structure .((.(((....))).)). Figure (**a**) gives the Kinfold folding kinetics of the native state and open chain conformations. Figures (**b**), (**c**), and (**d**) give a comparison the folding kinetics of the base-pair enumeration roadmap (876 conformations) to the stack-pair enumeration roadmap (22 conformations) and the maximal-contact random sampling roadmap (161 conformations).

the probability of transiting to each neighboring state; however, it does not consider the probability to remain in the current state. We modified the Kinfold algorithm to eliminate these two assumptions. (1) We stop the algorithm after a user-specified maximum time threshold instead of immediately stopping the algorithm once the conformation reaches the native state. This allows conformations to transit both in and out of the native state. (2) We added a probability of 1 to remain in the current state, following the standard Metropolis criteria (Metropolis *et al.*, 1953). This provides a more accurate comparison between our method and the Kinfold approach.

5.1. RNA1

Here we provide detailed results for RNA1. Figure 5 shows the population kinetics of the four most significant conformations[2] calculated using the base-pair, stack-pair, and maximal-contact roadmaps. These

[2]The four significant conformations are `.((.(((....))).))., ..................., ...(((........)))., and ...(((.((...))))).`. Note that all four are both base-pair and stack-pair conformations, while only the first two are maximal-contact conformations.

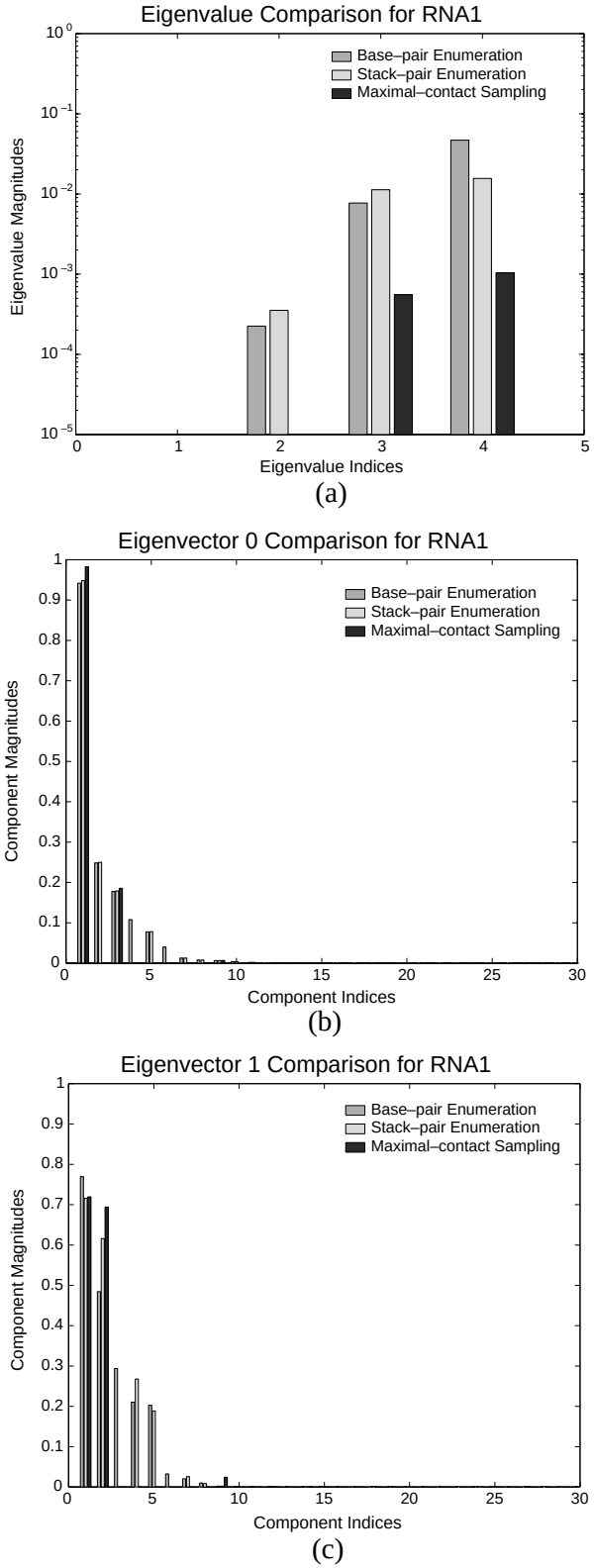


FIG. 6. The folding kinetics of the 18 nucleotide hairpin sequence CGCGCUACUCCUAGAGCU with the native structure .((.(((....))).)). and a C-space of 876 conformations. Figure (**a**) illustrates the differences in the eigenvalues and overall folding rates for base-pair enumeration, stack-pair enumeration, and maximal-contact sampling. Figures (**b**) and (**c**) compare the 40 biggest components of eigenvector N_0 and N_1, respectively.

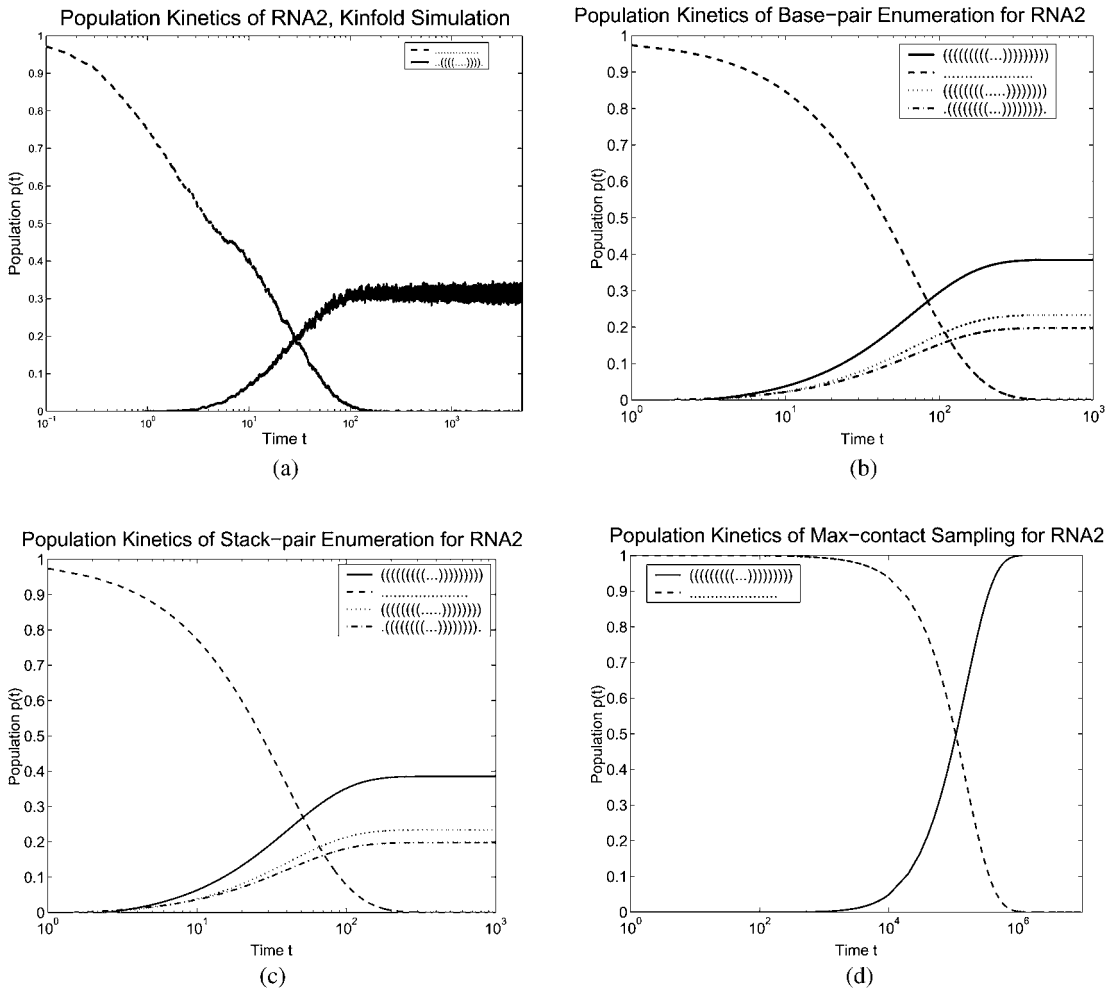


FIG. 7. The population kinetics of the 21 nucleotide hairpin sequence UAUAUAUCGACACGAUAUAUA with the native structure ((((((((((...)))))))))) . Figure (**a**) gives the folding kinetics from Kinfold of the native state and open chain conformations. Figures (**b**), (**c**), and (**d**) give a comparison the folding kinetics of the base-pair enumeration roadmap (5,353 conformations) to the stack-pair enumeration roadmap (250 conformations) and the maximal-contact random sampling roadmap (612 conformations).

have the largest population during or after the folding process, so their existence is more likely to be observed in experiments.

As illustrated in Figs. 5(b) and (c), the population kinetics calculated from the base-pair and stack-pair roadmaps are very similar to each other during the folding process. Hence, the stack-pair roadmap is a good approximation of the complete energy landscape. It preserves the main characteristics of the energy landscape while using notably fewer nodes (22 vs. 876). In addition, both the base-pair and stack-pair roadmaps yield population kinetics similar to those generated by Kinfold, Fig. 5(a). Minor discrepancies are caused by different energy and transition rate constants. Note that the folding kinetics of the maximal-contact sampling method, Fig. 5(d), are farther from the completely enumerated kinetics than the stack-pair kinetics are. We expected this because the stack-pair method encourages the formation of energetically stable conformations with stacks. The maximal-contact sampling is more random than the stack-pair method and does not attempt to capture the stability inherent in stacking pairs.

Figures 6(a), (b), and (c) demonstrate the similarities of the eigenvalues and eigenvectors between the three maps. Most significant is the discovery that the eigenvalues for the base-pair enumeration and the stack-pair enumeration are approximately the same (Fig. 6(a)). In addition, the components of the

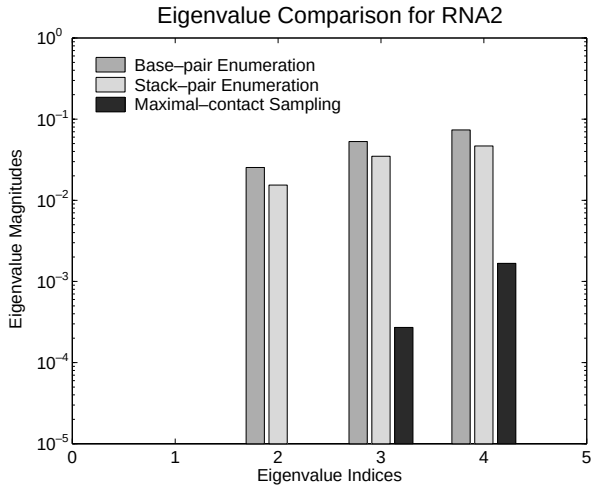


FIG. 8. Comparison of the eigenvalues of the 21 nucleotide hairpin sequence UAUAUAUCGACACGAUAUAUA and a C-space of 5,353 conformations. It illustrates the differences in the eigenvalues and overall folding rates for base-pair enumeration, the stack-pair enumeration, and the maximal-contact sampling.

eigenvectors (Figs. 6(b) and (c)) are close. Figure 6(b) shows the equilibrium solutions of the three folding landscapes. They all match very well with the Boltzmann distribution for this molecule. Figure 6(c) illustrates the small differences in magnitude of the components of the second eigenvector for all three folding landscapes. Although the maximal-contact sampling varies more from the complete base-pair enumeration in eigenvector N_1 than in N_0, the differences in magnitude are still relatively small. These results indicate that given some specific conformations, it is possible to examine the folding kinetics by computing the folding landscape of that set combined with some additional random sampling. This combination will approximate the slow mode eigenvectors.

5.2. RNA2

Figures 7(b), (c), and (d) show the population kinetics of the four most significant conformations[3] calculated using the base-pair, stack-pair and maximal-contact roadmaps. The population kinetics calculated from the base-pair and stack-pair roadmaps are very similar to each other during the folding process. Again, the stack-pair roadmap is a good approximation of the complete energy landscape while using notably fewer nodes (250 vs. 5,353). Also, the folding kinetics for the base-pair and stack-pair roadmaps compare well to the Kinfold folding kinetics, Fig. 7(a).

Figure 8 shows the similarities between the four smallest eigenvalues of the three roadmaps. In contrast to the results on RNA1, we found that except for the zero eigenvalue, the other three eigenvalues are comparable to each other. This means its folding behavior is different from RNA1 examined above. As described in Section 4.1, all three small modes have a nonnegligible influence on the global folding rate. Hence, the contributions of their corresponding eigenvectors to the transition states should not be ignored.

In Fig. 9, we compare the corresponding eigenvectors of the four small eigenvalues for the three roadmaps. Figure 9(a) shows the equilibrium of distributions, while (b), (c), and (d) show the contribution of the three eigenmodes on the transitional conformations. We found that in Figs. 9(a), (b), and (c), the eigenvectors for base-pair and stack-pair roadmaps are very similar to each other. The random sampling

[3]The four most significant conformations are `(((((((((...)))))))))`, `.....................`, `(((((((.....)))))))`, and `.(((((((((...)))))))))..` Note that all four are both base-pair and stack-pair conformations, while only the first two are maximal-contact conformations.

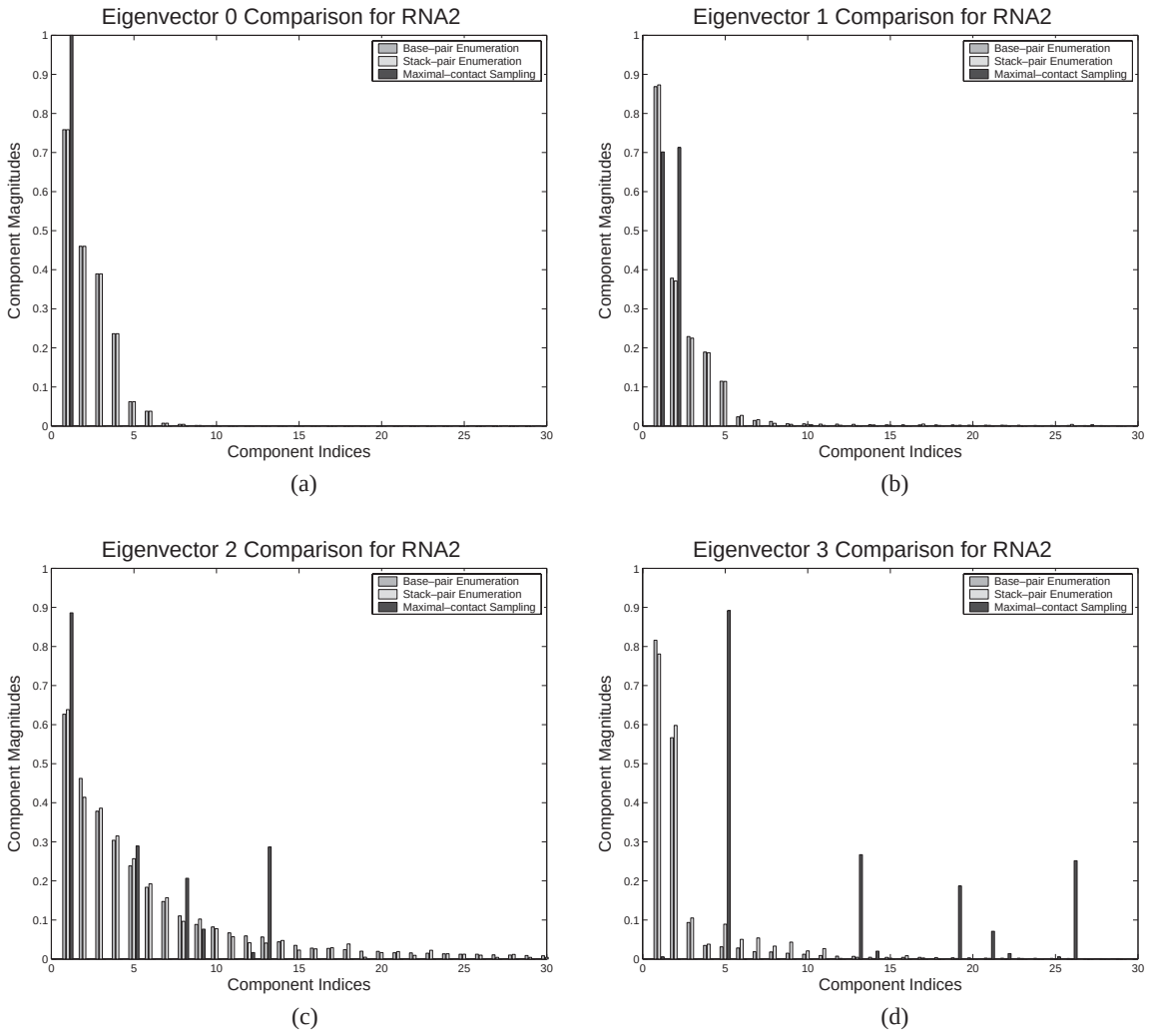


FIG. 9. The folding kinetics of the 21 nucleotide hairpin sequence UAUAUAUCGACACGAUAUAUA with the native structure ((((((((...)))))))) and a C-space of 5,353 conformations. The figures compare the biggest 30 components of eigenvectors (**a**) N_0, (**b**) N_1, (**c**) N_2, and (**d**) N_3 for base-pair enumeration, stack-pair enumeration and maximal-contact sampling.

for the maximal-contact roadmap, however, missed some important conformations. Yet, for those sampled conformations, the values are similar in magnitude to the corresponding components in the stack-pair and base-pair roadmaps. This means that although the random maximal-contact sampling is not accurate enough, it does capture some global properties of the folding process. Also, we found that we can easily increase our approximation accuracy by connecting more conformations.

5.3. 1K2G

Figure 10(b) shows the folding kinetics of the native state and open chain conformations calculated from the stack-pair roadmap. It has the same shape as the Kinfold folding kinetics, Fig. 10(a), but the final distribution of the native state is higher. The stack-pair roadmap contains only 71 conformations from the a C-space of 12,137 conformations. In this case, it misses a few key conformations containing some population (as indicated in Fig. 12) which results in a relatively higher final population of the native state. However, the similar shape of the folding kinetics between Figs. 10(a) and (b) shows that the stack-pair approximation still captures major features of the folding landscape.

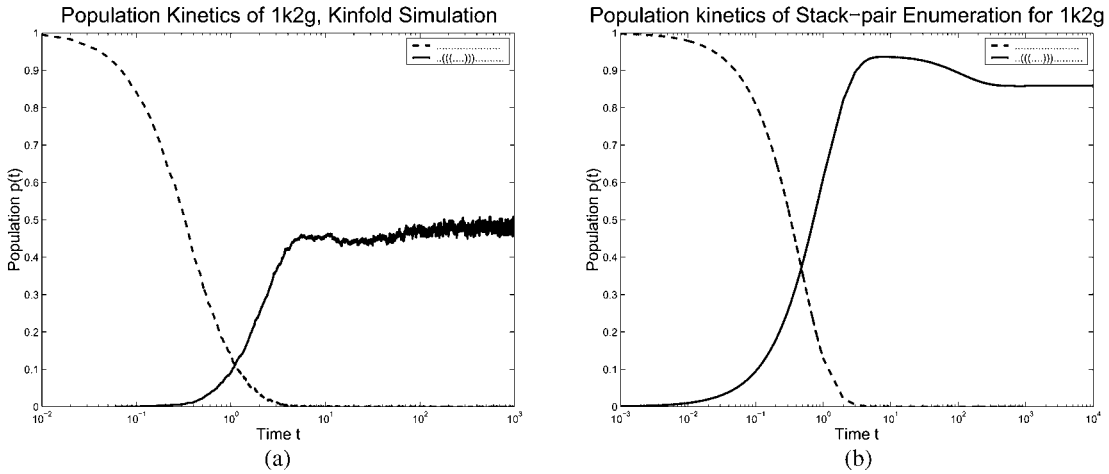


FIG. 10. The population kinetics of 1k2g: CAGACUUCGGUCGCAGAGAUGG with the native structure ..(((....)))..... Figure (**a**) gives the folding kinetics from Kinfold of the native state and open chain conformations. Figure (**b**) gives the folding kinetics from the stack-pair enumeration roadmap (71 conformations).

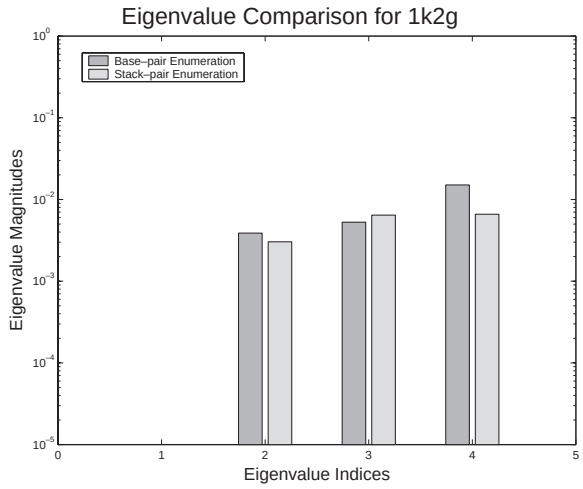


FIG. 11. Comparison of the eigenvalues of the 22 nucleotide hairpin sequence CAGACUUCGGUCGCAGAGAUGG and a C-space of 12,137 conformations. It illustrates the differences in the eigenvalues and overall folding rates for base-pair enumeration and stack-pair enumeration.

Figure 11 shows the comparison of the four smallest eigenvalues of the base-pair and stack-pair roadmaps. All the eigenvalues are similar to each other.

Figure 12 compares the corresponding eigenvectors for both roadmaps. Eigenvectors 0 and 1 match well, while there is a larger difference between the roadmaps for Eigenvectors 2 and 3. Recall that Eigenvector 0 corresponds to the final distribution of each conformation. As explained earlier, a few conformations with some final population are missing from the stack-pair roadmap. This yields an inflated final population of the native state.

Using roadmaps generated by three different strategies, we compared the kinetics analysis of RNA molecules which have different folding behaviors. The roadmap generated by the base-pair enumeration is the most accurate representation. However, it is not feasible to enumerate RNA with more than 40 nucleotides. While the stack-pair roadmap is also generated from an enumeration, it yields a much smaller subset of the entire conformation space that effectively approximates the energy landscape, even for RNA

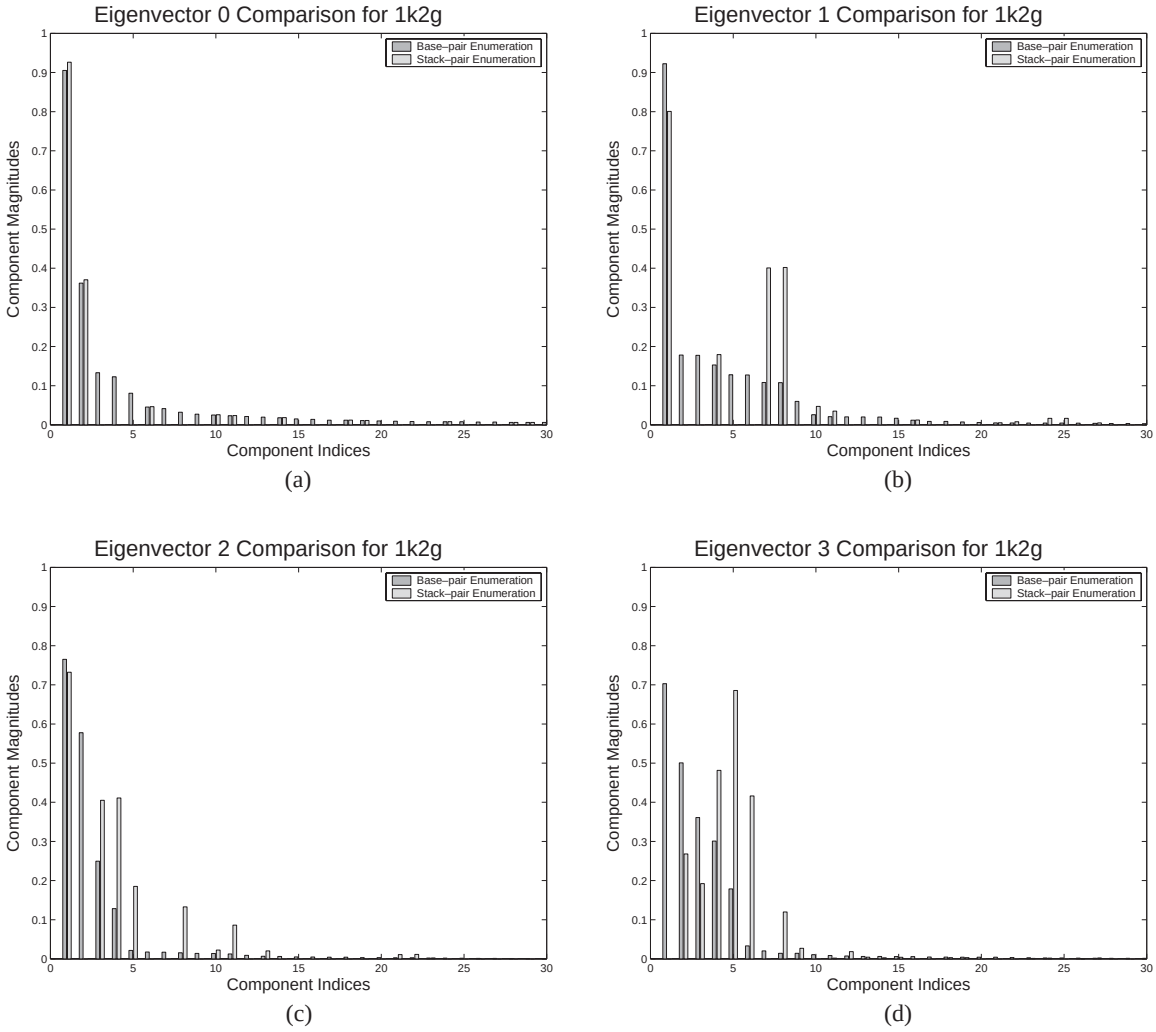


FIG. 12. The folding kinetics of the 22 nucleotide hairpin sequence CAGACUUCGGUCGCAGAGAUGG with the native structure ..(((....))).......... and a C-space of 12,137 conformations. The figures compare the biggest 30 components of eigenvectors (**a**) N_0, (**b**) N_1, (**c**) N_2, and (**d**) N_3 for base-pair enumeration and stack-pair enumeration.

with different folding behaviors. The maximal-contact roadmap does not require enumeration and can be of any size we desire. Although its approximation is slightly inferior to the stack-pair roadmap, our preliminary work indicates that this approximation can be improved. Most important, our work demonstrates that we can effectively characterize the energy landscape using notably fewer conformations than exist in the complete enumeration. These results signify that putting more work into improving our sampling strategies will yield more concise and efficient representations of the energy landscape. Our method is applicable to much longer RNA sequences.

5.4. Folding pathways results

Similarly to our previous work on protein folding (Amato *et al.*, 2003), we extract folding pathways and compute the free-energy profile, energy barriers, and important states of the folding process. From all the folding pathways to the native conformation, we extract the pathway with minimum total weight because this corresponds to the most energetically feasible path *in our roadmap*. For a given pathway, its energy profile shows the energy of each transitional conformation, and it is easy for us to find the local minima and energy barriers on the pathway. These profiles provide an informal visualization of the folding process.

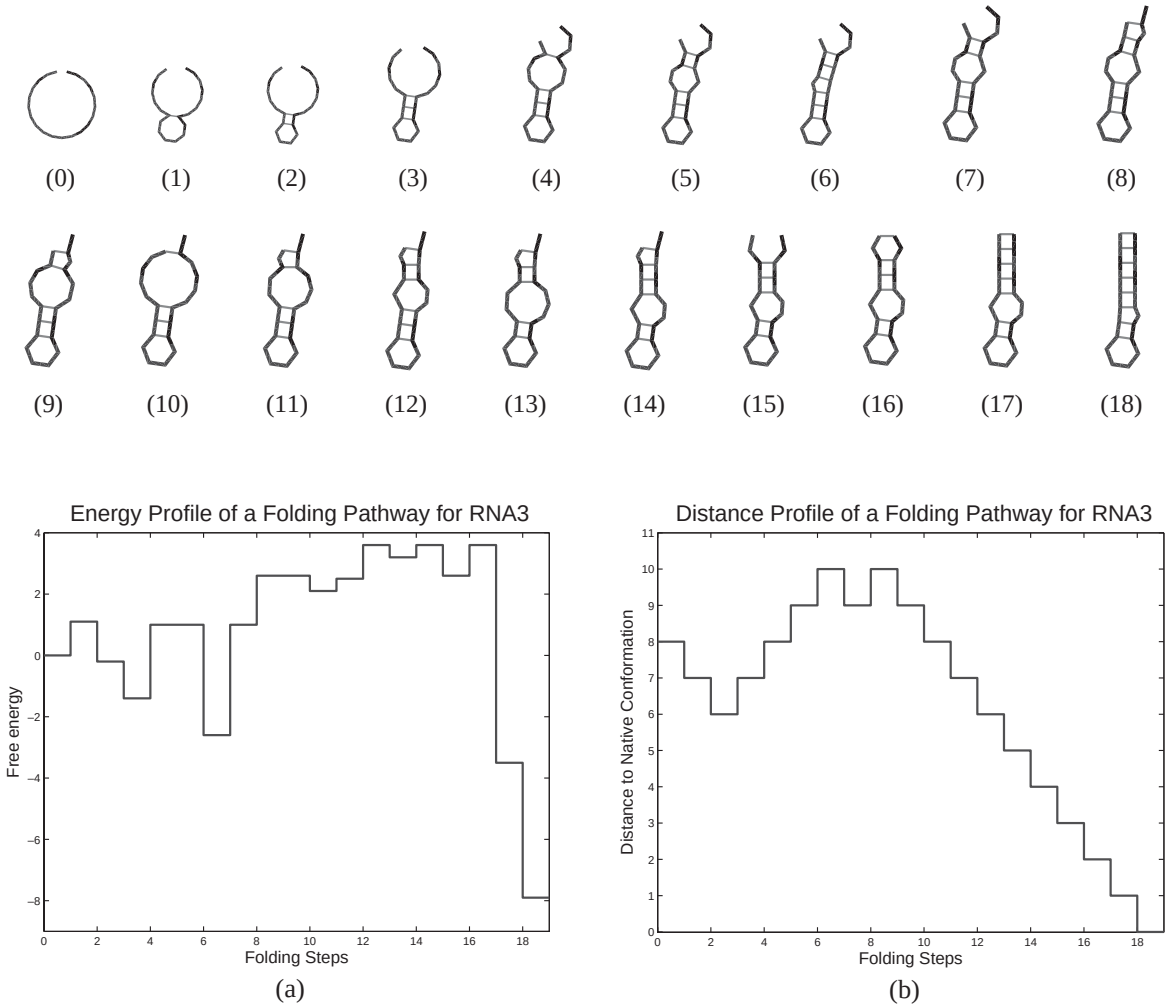


FIG. 13. A folding pathway for RNA sequence GGCGUAAGGAUUACCUAUGCC from a open conformation (0) to a misfolded conformation (6), then to the native conformation (18). Each transitional conformation is numbered according to its position on the pathway. (**a**) The energy profile of transitional conformations. (**b**) The distance of each transitional conformation to the native conformation.

An example is given in Fig. 13. It shows the energy profile and folding pathway for RNA3 (GGCGUAAG-GAUUACCUAUGCC). It first folds into a misfolded conformation and then folds to the native state. From the misfolded conformation, it has to overcome a high energy barrier to reach the native conformation as shown in its energy profile in Fig. 13(a).

6. CONCLUSION

We have demonstrated that the PRM method is a promising technique for studying RNA folding kinetics. PRMs allow us to efficiently characterize the folding landscape using small roadmaps, and moreover, our roadmaps were suitable for computing the folding kinetics for the RNA we have studied so far. Our results also indicate that further work on more sophisticated generation and connection methods will yield better results, and this is the subject of current work.

One key advantage of our method is that it is generic. We can use any method to sample representative conformations and compute transitions to approximate the energy landscape. Thus, with more sophisticated sampling and connection schemes, our method is extensible to much longer RNA molecules. Our preliminary results indicate that our method has the potential to capture the main features of the energy

landscape with a notably smaller number of conformations than the complete enumeration. We believe that our method has huge potential to enable us to analyze folding problems with a global view.

ACKNOWLEDGMENTS

This research supported in part by NSF Grants ACI-9872126, EIA-9975018, EIA-0103742, EIA-9805823, ACR-0081510, ACR-0113971, CCR-0113974, EIA-9810937, EIA-0079874. Song's work performed at the Parasol Lab at Texas A&M and supported in part by an IBM TJ Watson PhD Fellowship. Thomas supported in part by an NSF Graduate Research Fellowship. Kirkpatrick's work performed at the Parasol Lab at Texas A&M during research internships in summer 2002 and 2003 that were supported by the CRA Distributed Mentor Program.

REFERENCES

Amato, N.M., and Song, G. 2002. Using motion planning to study protein folding pathways. *J. Comp. Biol.* 9(2), 149–168.

Amato, N.M., Dill, K.A., and Song, G. 2003. Using motion planning to map protein folding landscapes and analyze folding kinetics of known native structures. *J. Comp. Biol.* 10(3–4), 239–256.

Chen, S.-J., and Dill, K.A. 2000. RNA folding energy landscapes. *Proc. Natl. Acad. Sci. USA* 97, 646–651.

Cupal, J., Flamm, C., Renner, A., and Stadler, P.F. 1997. Density of states, metastable states, and saddle points exploring the energy landscape of an RNA molecule. *Proc. Int. Conf. Intelligent Systems for Molecular Biology (ISMB)*, 88–91.

Cupal, J., Hofacker, I.L., and Stadler, P.F. 1996. Dynamic programming algorithm for the density of states of RNA secondry structures. *Computer Science and Biology* 96, 184–186.

Dill, K.A., and Chan, H.S. 1997. From Levinthalto pathways to funnels: The new view of protein folding kinetics. *Nat. Struct. Biol.* 4, 10–19.

Flamm, C. 1998. *Kinetic Folding of RNA*. Ph.D. Thesis, University of Vienna, Austria.

Gillespie, D.T. 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81, 2340–2361.

Hofacker, I.L. 1998. RNA secondary structures: A tractable model of biopolymer folding. *J. Theor. Biol.* 212, 35–46.

Kampen, N.G. Van. 1992. *Stochastic Processes in Physics and Chemistry*, North-Holland, New York.

Kavraki, L.E., Svestka, P., Latombe, J.C., and Overmars, M.H. 1996. Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *IEEE Trans. Robot. Automat.* 12(4), 566–580.

Latombe, J.-C. 1991. *Robot Motion Planning*, Kluwer Academic Publishers, Boston, MA.

McCaskill, J.S. 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* 29, 1105–1119.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. 1953. Equation of state caculations by fast computing machines. *J. Chem. Phys.* 21, 1087–1092.

Nussinov, R., Piecznik, G., Griggs, J.R., and Kleitman, D.J. 1972. Algorithms for loop matching. *SIAM J. Appl. Math.* 35, 68–82.

Ozkan, S. Banu, Dill, K.A., and Bahar, I. 2002. Fast-folding protein kinetics, hidden intermediates, and the seuential stabilization model. *Protein Sci.* 11, 1958–1970.

Ozkan, S. Banu, Dill, K.A., and Bahar, I. 2003. Computing the transition state population in simple protein models. *Biopolymers* 68, 35–46.

Sankoff, D., and Kruskal, J.B. 1983. *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*, Addison Wesley, London.

Shapiro, B.A., Bengali, D., Kasprzak, W., and Wu, J.C. 2001. RNA folding pathway functional intermediates: Their prediction and analysis. *J. Mol. Biol.* 312, 27–44.

Song, G., and Amato, N.M. 2001. Using motion planning to study protein folding pathways. *Proc. Int. Conf. Comput. Molecular Biology (RECOMB)*, 287–296.

Song, G., Thomas, S.L., Dill, K.A., Scholtz, J.M., and Amato, N.M. 2003. A path planning-based study of protein folding with a case study of hairpin formation in protein G and L. *Proc. Pacific Symposium of Biocomputing (PSB)*, 240–251.

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

Walter, A.E., Turner, D.H., Kim, J., Lyttle, M.H., Muller, P., Mathews, D.H., and Zuker, M. 1994. Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl. Acad. Sci. USA* 91, 9218–9222.

Wolfinger, M. 2001. *The Energy Landscape of RNA Folding*. M.Phil. Thesis, University of Vienna, Austria.
Wuchty, S. 1998. *Suboptimal Secondary Structures of RNA*. M.Phil. Thesis, University of Vienna, Austria.
Zhang, W., and Chen, S. 2002. RNA hairpin-folding kinetics. *Proc. Natl. Acad. Sci. USA* 99, 1931–1936.
Zuker, M., and Sankoff, D. 1984. RNA secondary structure and their prediction. *Bull. Math. Biol.* 46, 591–621.
Zuker, M., Mathews, D.H., and Turner, D.H. 1999. Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide. In Barciszewski, J., and Clark, B.F.C., eds., *RNA Biochemistry and Biotechnology*, NATO ASI Series, Kluwer Academic Publishers, Amsterdam.
Zuker, M., and Stiegler, P. 1981. Optimal computer folding of large Pelf sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* 9, 133–148.

Address correspondence to:
Nancy M. Amato
Parasol Lab
301 Harvey R. Bright Bldg.
3112 Texas A&M University
College Station, TX 77843-3112

E-mail: amato@cs.tamu.edu