MDPI

Article

Markov Chain-Based Sampling for Exploring RNA Secondary Structure under the Nearest Neighbor Thermodynamic Model

Anna Kirkpatrick, Kalen Patton, Prasad Tetali and Cassie Mitchell *

Georgia Institute of Technology, Atlanta, GA 30318, USA

* Correspondence: cassie.mitchell@bme.gatech.edu

Version September 1, 2020 submitted to Math. Comput. Appl.

Abstract: Ribonucleic acid (RNA) secondary structures and branching properties are important for determining functional ramifications in biology. While energy minimization of the Nearest Neighbor Thermodynamic Model (NNTM) is commonly used to identify such properties (number of hairpins, maximum ladder distance, etc.), it is difficult to know whether the resultant values fall within expected dispersion thresholds for a given energy function. The goal of this study was to construct a Markov chain capable of examining the dispersion of RNA secondary structures and branching properties obtained from NNTM energy function minimization independent of a specific nucleotide sequence. Plane trees are studied as a model for RNA secondary structure, with energy assigned to each tree based on the NNTM, and a corresponding Gibbs distribution is defined on the trees. Through a bijection between plane trees and 2-Motzkin paths, a Markov chain converging to the Gibbs distribution is constructed, and fast mixing time is established by estimating the spectral gap of the chain. The spectral gap estimate is obtained through a series of decompositions of the chain and also by building on known mixing time results for other chains on Dyck paths. The resulting algorithm can be used as a tool for exploring the branching structure of RNA, especially for long sequences, and to examine branching structure dependence on energy model parameters.

Keywords: Markov chain Monte Carlo, RNA secondary structure, Nearest Neighbor Thermodynamic Model, Markov chain convergence

1. Introduction

Computational and mathematical applications play a critical role in the analysis of the structure and function of biological molecules, including ribonucleic acid (RNA). RNA is an essential biological polymer with many roles including information transfer, regulation of gene expression, and catalysis of chemical reactions. The *primary structure* of an RNA molecule may be understood as a sequence of amino acids: arginine, urasil, guanine, and cytosine. As is standard, we frequently abbreviate these as A, U, G, and C, respectively. RNA molecules are single-stranded and may therefore interact with themselves, forming A-U, G-U, and G-C bonds. The *secondary structure* of an RNA molecule is a set of such bonds.

The determination of secondary structure is an important step to understanding an RNA molecule's full shape and therefore its function [1,2]. Accordingly, secondary structure information is commonly used in tertiary structure prediction algorithms, see e.g. [3–6]. Identifying the secondary structure of RNA is crucial to understanding its function and mechanism in a cell [7]. Thus, the structure of RNA is critical to the development of biological and pharmaceutical therapeutics. Biologists use inexpensive and expedient means to sequence RNA, but experimental determination of structure is more difficult and time-consuming. Therefore, computational methods are the primary means to determine possible RNA secondary structures.

For decades, one of the main computational approaches for examining RNA structure and branching properties has been thermodynamic free energy minimization using Nearest Neighbor Thermodynamics Modeling (NNTM) [8–10]. This free energy is in turn used in algorithms to predict secondary structure given an RNA sequence, see, e.g., [11–13]. Under the NNTM, the free energy of a structure is computed as the sum of the free energy of its various substructures. Many common programs (e.g. mFold, RNAFold, RNA Structure, sFold, Vienna RNA, etc.) intake a single sequence to produce secondary structures based on NNTM energy minimizations performed via dynamic programming. Nearest neighbor parameter sets include both a set of rules, referred to as equations or features, and a set of parameter values used by the equations. Separate rules exist for predicting stabilities of helices, hairpin loops, small internal loops, large internal loops, bulge loops, multi-branch loops, and exterior loops. Other branching properties of interest include, but are not limited to, average ladder distance, maximum ladder distance, maximum branching degree, average contact distance, average branching degree, degree of branching at the exterior loop, number of multi-loops with n braches, etc. The online nearest neighbor database (NNDB) archives and stores complete nearest neighbor sets, including rules and corresponding parameter values [14].

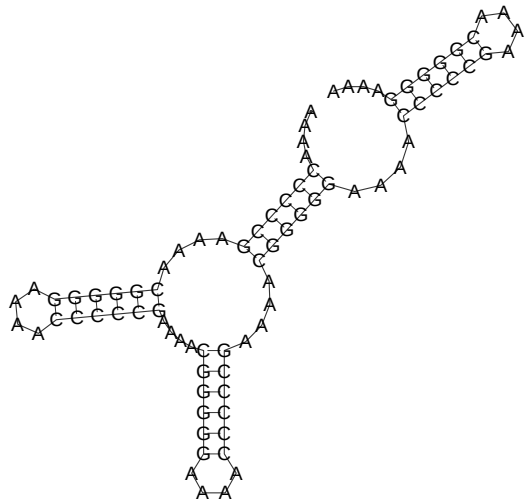
A common challenge is inferring whether the predicted results of NNTM for a set of RNA structural features or branching properties are within expected dispersion thresholds for a given energy model. For example, is the number of hairpins more than 2-3 standard deviations greater than the expected mean for a given energy model? This challenge is particularly vexing if the sequence is relatively long (greater than 1000 nucleotides). If structural features or branching properties are determined to exceed expected energy model dispersion thresholds, it relays potential scientific and/or mechanistic insight. Continuing with our hairpin example, what if an NNTM model produces a result where the number of hairpins seems rather large for the given sequence length? If the number of hairpins exceeds the expected dispersion of the NNTM model, it might be inferred that the greater number of hairpins is evidence of natural selection.

The primary objective of the present study is to enable mathematical determination of the dispersion of RNA secondary structural features for a given sequence length. We present a Markov-based algorithm to provide samples of the branching structure under the NNTM and Gibbs distribution, but without reference to a particular sequence of nucleotides. The algorithm enables the determination of where the predicted feature or branching property for an actual sequence falls within this distribution, which in turn enables the determination of whether the predicted NNTM feature or branching property is within expected dispersion limits.

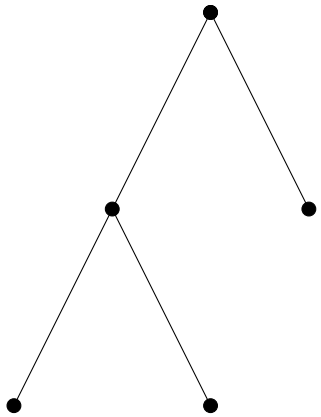
In particular, this work investigates RNA substructures called *multi-loops*, the places where three or more helices join. Though multi-loops are crucial to the overall shape of a secondary structure, the models used to predict them algorithmically do not produce accurate results [15]. This investigation builds on an existing model of RNA branching [16] and provides a theoretical grounding for a Markov chain which may be used to algorithmically investigate branching properties of secondary structure models. The investigational foundation is a model for RNA secondary structure developed by Hower and Heitsch [16], in which secondary structures are in bijection with plane trees and the minimal energy structures of the model have been previously characterized. The present study characterizes the full Gibbs distribution of possible structures. Notably, Bakhtin and Heitsch [17] analyzed a very similar model and determined degree sequence properties of the distribution of plane trees asymptotically. However, the present study utilizes a Markov chain-based sampling algorithm to investigate the Gibbs distribution in the finite case. A full explanation of the plane tree model as well as the derivation of the energy functions is provided in Section 2.1.

2. Methods

The methods are divided into an overview of the RNA secondary structure NNTM plane tree model and energy functions (Section 2.1) and an all-encompassing explanation of the mathematical



(a) A maximally-paired secondary structure for $A^4(C^5GA^4CG^5A^4)^4$ has 4 helices.



(b) The corresponding plane tree has 4 edges and encodes the branching pattern seen in the secondary structure.

Figure 1. An RNA secondary structure for one of the combinatorial RNA sequences used in this work and its corresponding plane tree. The ordering of the edges in the plane tree is derived from the 3′ to 5′ ordering of the RNA sequence. Note that the exterior loop corresponds to the root of the plane tree. The diagram in Figure 1a was generated by ViennaRNA [19].

preliminaries that lay the foundation for the derived results and corresponding algorithms (Section 2.2).

2.1. Derivation of Energy Functions

The energy function studied here is derived from the Nearest Neighbor Thermodynamic Model (NNTM). The numerical parameters from the NNTM can be found in the NNDB [14]. In calculating energy functions for the sequences, we consider thermodynamic parameter values published by Turner in 1989 [8], 1999 [9], and 2004 [10].

The plane trees that we study in this paper come from two combinatorial RNA sequences, both of the form $A^4(Y^5ZA^4YZ^5A^4)^n$. The sequences of interest have $(Y, Z) = (C, G)$ or $(Y, Z) = (G, C)$. For both of these sequences, the set of maximally-paired secondary structures is in bijection with the set of plane trees of size n [18]. Figure 1 shows one example of a secondary structure and corresponding plane tree.

These specific combinatorial sequences are chosen because they allow for the study of the relationship between NNTM multiloop parameters and the branching behavior of secondary structures without interference from the energy contributions have specific base pairing combinations.. In particular, the only places where the free energy differs between different secondary structures (for the same sequence) is in the type and number of multi-loops, the branching at the exterior loop, the number of hairpins, and the number of internal nodes. All of these energies directly relate to branching, not to specific base pairs. This simplification achieved by focusing only on multi-loops and branching both creates a model that is more amenable to theoretical analysis and speeds computation.

Note that these secondary structures should not be considered representative of naturally occurring secondary structures. Instead, the only properties of interest in these structures are branching-related properties.

Three constants determine the free energy contribution of multiloops under NNTM, a, b, and c. The value of a encodes the energy penalty per multiloop. The constant b specifies the energy penalty per single-stranded nucleotide in a multiloop. The value of c gives the energy penalty for each helix branching from a multiloop.

In addition to the multiloop parameters a, b, c discussed above, we must account for the energy contributions of stacking base pairs, hairpins, interior loops, and dangling energy contributions. The energy of one helix is given by h. The energy associated with a hairpin is f, and the energy contribution of an interior loop is i. Finally, the parameter g encodes the dangling energy contributions. All of these values can be computed directly from the parameters found in the NNTM.

We wish to compute the energy of the structure corresponding to plane tree t having (down) degree sequence $d_0, d_1, \ldots, d_{n-1}$ and root degree r. Note that the *down degree* of a node x is equal to the number of children of x, and, in the *down degree sequence*, d_i is the number of non-root nodes with exactly i children. The energy contribution of all hairpin loops will be $d_0 f$, and similarly the total energy of all interior loops will be $d_1 i$. For a multi-loop having down degree j, the energy contribution will be $a + 4b(j+1) + c(j+1) + (j+1)g$, and so the contribution of all multi-loops is given by $\sum_{j=2}^{n} d_j(a + 4b(j+1) + c(j+1) + g(j+1))$. The root vertex of the tree corresponds to the exterior loop and has energy contribution gr. Finally, our structure has n helices, each with energy h. Summing all of these components gives the total energy.

$$d_0 f + d_1 i + \sum_{j=2}^{n} d_j(a + 4b(j+1) + c(j+1) + g(j+1)) + nh + gr \tag{1}$$

$$= (f - a - 4b - c - g)d_0 + (i - a - 8b - 2c - 2g)d_1 + (-4b - c)r + (a + 8b + 2c + h + 2g)n, \tag{2}$$

where we have used the facts $\sum_{k=0}^{n-1} d_k = n$ and $\sum_{k=0}^{n-1} k d_k = n - r$.

Set $\alpha = f - a - 4b - c - g$, $\beta = i - a - 8b - 2c - 2g$, $\gamma = -4b - c$, and $\delta = a + 8b + 2c + h + 2g$. Then, the energy function is $\alpha d_0 + \beta d_1 + \gamma r + \delta n$. Since n will be fixed, we disregard the term δn, giving

$$E(t) = \alpha d_0 + \beta d_1 + \gamma r. \tag{3}$$

Though we study these energy functions for arbitrary values of (α, β, γ), numerical values for both the input energy parameters from NNTM and the resulting energy function coefficients are given in Table 1.

Y	Z	Turner	a	b	c	h	f	i	g	α	β	γ
C	G	89	4.6	0.4	0.1	-10.9	3.8	3.0	-1.6	-0.9	-1.8	-1.7
G	C	89	4.6	0.4	0.1	-16.5	3.5	3.0	-1.9	-0.9	-1.2	-1.7
C	G	99	3.4	0	0.4	-12.9	4.5	2.3	-1.6	2.3	1.3	-0.4
G	C	99	3.4	0	0.4	-16.9	4.1	2.3	-1.9	2.2	1.9	-0.4
C	G	04	9.3	0	-0.9	-12.9	4.5	2.3	-1.1	-2.8	-3.0	0.9
G	C	04	9.3	0	-0.9	-16.9	4.1	2.3	-1.5	-2.8	-2.2	0.9

Table 1. NNTM parameters and resulting energy functions. Energy functions are of the form $\alpha d_0 + \beta d_1 + \gamma r$.

2.2. Mathematical Preliminaries

Section 2.2 of this manuscript provides the necessary mathematical background, including a formal introduction of combinatorial objects and a review of the relevant Markov chain mixing results used to construct our resultant sampling Markov chain and corresponding mixing time proof in Section 3.

2.2.1. Combinatorial Objects

A *plane tree* is a rooted, ordered tree. We will use $\mathfrak{T}_n$ to denote the set of plane trees with n edges. It is known that $|\mathfrak{T}_n|$ is given by the nth Catalan number $C_n = \frac{1}{n+1}\binom{2n}{n}$. In a plane tree, a *leaf* is a node with down degree 0, and an *internal node* is a non-root node with down degree 1. For a given plane tree t, we will use $d_0(t)$ to denote the number of leaves and $d_1(t)$ to denote the number of internal nodes.

For a plane tree t, the *energy* of the tree is given by

$$E(t) = \alpha d_0(t) + \beta d_1(t), \tag{4}$$

where α and β are real parameters of the energy function. Note that this function is a simplification of the model due to Hower and Heitsch [16] discussed in Section 2.1. Making this simplification effectively disregards the energy contribution of the exterior loop, which is small in comparison to the total energy of a structure, especially for the longer sequences that are of interest to us. Other authors have made similar simplifications, e.g. [17].

For our purposes, we consider α and β to be arbitrary but fixed. We will consider a Gibbs distribution $\mathbf{g}$ on the set $\mathfrak{T}_n$, where the weight of each tree t is given by

$$\mathbf{g}(t) = \frac{e^{-E(t)}}{Z}, \tag{5}$$

where $Z = \sum_{y \in \mathfrak{T}_n} e^{-E(y)}$ is a normalizing constant.

A *Motzkin path* of length n is a lattice path from $(0,0)$ to $(n,0)$, which consists of steps along the vectors $U = (1,1)$, $H = (1,0)$, and $D = (1,-1)$ and never crosses below the x-axis. We can also represent Motzkin paths as strings from the alphabet $\{U, H, D\}$ where, in any prefix, the number of Us is greater than or equal to the number of Ds. The number of Motzkin paths of length n is given by the Motzkin numbers M_n where

$$M_n = \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k} C_k. \tag{6}$$

Motzkin numbers and Motzkin paths have been well-studied in the combinatorics literature, see e.g. [20–24].

A *Dyck path* is a Motzkin path with no H steps. It is easy to see that a Dyck path must have even length, so we will use $\mathfrak{D}_n$ to denote the set of Dyck paths on length $2n$. It is well known that $|\mathfrak{D}_n| = C_n$ (see, e.g. [25]).

A *2-Motzkin path* is a Motzkin path in which $(1,0)$ steps are given one of two distinguishable colors. Let $\mathfrak{M}_m^2$ be the set of all 2-Motzkin paths of length m. We can also represent 2-Motzkin paths as strings from the alphabet $\{U, H, I, D\}$, where as before, the number of Ds never exceeds the number of Us in any prefix. In a such a string x, we denote by $|x|_a$ the number of times the symbol a appears in x, where $a \in \{U, H, I, D\}$. Notice that we always have $|x|_U = |x|_D$. For any $x \in \mathfrak{M}_n^2$ and $k \in \{1, \cdots, n\}$, let $x(k)$ denote the symbol at index k in the string representation of x. Additionally, the *skeleton* of a 2-Motzkin path x is the Dyck path of Us and Ds which results from removing all Hs and Is from x. We will denote the skeleton of x by $\sigma(x)$.

2.2.2. A Bijection Between $\mathfrak{T}_n$ and $\mathfrak{M}_{n-1}^2$

We will use the particular bijection $\Phi\colon \mathfrak{T}_n \to \mathfrak{M}_{n-1}^2$ between plane trees and 2-Motzkin paths from Deutsch [26], which neatly encodes information about d_0 and d_1. For clarity, we will overview the bijection here.

For a given plane tree t with n edges, assign a label from the set $\{U, H, I, D\}$ to each edge e according to the following rules:

- If e is the leftmost edge off a non-root node of down degree at least 2, assign the label U.
- If e is the rightmost edge off a non-root node of down degree at least 2, assign the label D.
- If e is the only edge off a non-root node of degree 1, assign the label I.
- If e is an edge off the root node, or if e is neither the leftmost nor the rightmost edge off its parent node, assign the label H.

Now, if we traverse t in preorder reading off these labels, we get a 2-Motzkin path of length n. However, this path will always begin with H, so we define $\Phi(t)$ to be the 2-Motzkin path of length $n-1$ after

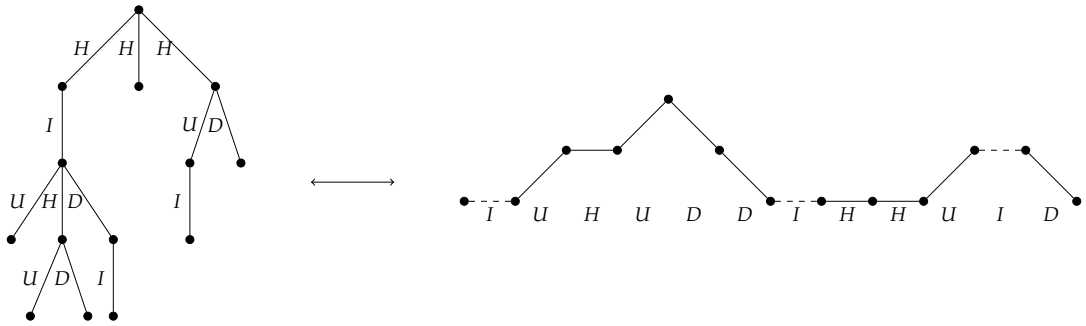


Figure 2. A plane tree with edges labeled according to the bijection Φ, along with its corresponding 2-Motzkin path.

this initial H is removed. Figure 2 gives an example of this labeling process. From Deutsch, we know not only that Φ is a bijection, but also that if $x = \Phi(t)$ then $|x|_I = d_1(t)$ and $|x|_U + |x|_H + 1 = d_0(t)$.

Using this bijection, it is natural to extend our energy function to 2-Motzkin paths. We define the energy of a 2-Motzkin path x to be

$$E(x) = \alpha(|x|_U + |x|_H + 1) + \beta|x|_I, \tag{7}$$

and we extend our definition of the distribution $\mathbf{g}$ to $\mathfrak{M}_n^2$ accordingly. We note that, while this energy function does not capture all possible weightings on 2-Motzkin paths, it does capture all weightings possible under our simplification of the model due to Hower and Heitsch [16] after applying the bijection due to Deutsch [26].

2.2.3. Markov Chains

A *Markov chain* $\mathcal{M}$ is a sequence of random variables $X_0, X_1, X_2, \cdots$ taking values in a state space Ω subject to the condition that

$$\Pr(X_{t+1} = y \mid X_t = x,\ X_{t-1} = s_{t-1},\ \cdots,\ X_0 = s_0) = \Pr(X_{t+1} = y \mid X_t = x). \tag{8}$$

All Markov chains that we consider will be implicitly *time-homogeneous* (meaning $\Pr(X_{t+1} = y \mid X_t = x)$ does not depend on t) and *finite* (meaning $|\Omega| < \infty$). The *transition matrix* of a time-homogeneous Markov chain is the matrix $P\colon \Omega \times \Omega \to [0,1]$ given by

$$P(x,y) = \Pr(X_{t+1} = y \mid X_t = x). \tag{9}$$

It is easy to see that if X_0 has distribution vector $\mathbf{x}$, then X_t has distribution vector $P^t\mathbf{x}$.

A finite Markov chain with transition matrix P is said to be *ergodic* if it has the following two properties.

1. *Irreducibility:* For any $x, y \in \Omega$, there is some integer $t \in \mathbb{N}$ for which $P^t(x,y) > 0$.
2. *Aperiodicity:* For any state $x \in \Omega$, we have $\gcd\{t \in \mathbb{N}\colon P^t(x,x) > 0\} = 1$.

It is well known that if $\mathcal{M}$ is ergodic, then there exists a unique distribution vector π, the *stationary distribution*, such that $P\pi = \pi$, and $\lim_{t\to\infty} P^t(x,y) = \pi(y)$ for any states $x, y \in \Omega$. Additionally, we call $\mathcal{M}$ *reversible* if for all states $x, y \in \Omega$, we have $\pi(x)P(x,y) = \pi(y)P(y,x)$.

For $\epsilon > 0$, the *mixing time* $\tau(\epsilon)$ of $\mathcal{M}$ is given by

$$\tau(\epsilon) = \min\left\{t \in \mathbb{N}\colon \forall s \geq t,\ \max_{x\in\Omega}\left(\frac{1}{2}\sum_{y\in\Omega}|P^s(x,y) - \pi(y)|\right) < \epsilon\right\}. \tag{10}$$

Intuitively, the mixing time gives a measure of the number of steps required for $\mathcal{M}$ to get sufficiently close to its stationary distribution from any starting state.

Let $\mathcal{M}$ be a finite ergodic Markov chain over a state space Ω with transition matrix P. Let the eigenvalues of P be $\lambda_0, \lambda_1, \ldots, \lambda_{|\Omega|-1}$ such that $1 = \lambda_0 > |\lambda_1| \geq \ldots \geq |\lambda_{|\Omega|-1}|$. The *spectral gap* of $\mathcal{M}$ is given by $\text{Gap}(\mathcal{M}) = 1 - |\lambda_1|$. As is standard, it will be convenient to denote the inverse of the spectral gap by *relaxation time* $\tau_{rel}(\mathcal{M}) := 1/\text{Gap}(\mathcal{M})$.

Additionally, the spectral gap is given by the following functional definition [27].

$$\text{Gap}(\mathcal{M}) = \inf_f \frac{\sum_{x,y\in\Omega} |f(x) - f(y)|^2 \pi(x) P(x,y)}{\sum_{x,y\in\Omega} |f(x) - f(y)|^2 \pi(x)\pi(y)}, \tag{11}$$

where the infimum is taken over all non-constant functions $f : \Omega \to \mathbb{R}$. A direct consequence of this definition of the spectral gap is the following lemma.

Lemma 1. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be ergodic Markov chains over Ω with the same stationary distribution. Let P_1 and P_2 be the transition matrices of $\mathcal{M}_1$ and $\mathcal{M}_2$ respectively. If for all $x, y \in \Omega$ and for some constant $c > 0$ we have $P_1(x,y) \leq cP_2(x,y)$, then $\text{Gap}(\mathcal{M}_1) \leq c\text{Gap}(\mathcal{M}_2)$.*

Additionally, spectral gap is related to the mixing time by the following lemma [28].

Lemma 2. *Let $\mathcal{M}$ be an ergodic Markov chain with state space Ω, and let $\lambda_0, \lambda_1, \ldots, \lambda_{|\Omega|-1}$ be the eigenvalues of the transition matrix P as defined above. Then, for all $\epsilon > 0$ and $x \in \Omega$, we have*

$$\frac{|\lambda_1|}{\text{Gap}(\mathcal{M})} \log\left(\frac{1}{2\epsilon}\right) \leq \tau(\epsilon) \leq \frac{1}{\text{Gap}(\mathcal{M})} \log\left(\frac{1}{\pi(x)\epsilon}\right). \tag{12}$$

We say that a Markov chain $\mathcal{M}$, whose state space depends on a variable $n \in \mathbb{N}$, is *rapidly mixing* if $\tau(\epsilon)$ is bounded above by some polynomial in n and $\log(\epsilon^{-1})$. For the specific chains studied in this manuscript, we will show that $\tau(\epsilon)(\mathcal{M})$ is bounded by a polynomial in n and $\log(\epsilon^{-1})$ if and only if $\tau_{rel}(\mathcal{M})$ is bounded by a polynomial in n and $log(\epsilon^{-1})$. Our next lemma presents sufficient conditions.

Lemma 3. *Let $\mathcal{M}$ be an ergodic Markov chain with state space Ω and let $\lambda_0, \lambda_1, \ldots, \lambda_{|\Omega|-1}$ be the eigenvalues of its transition matrix. Let $\epsilon > 0$. If $\tau(\epsilon)$ is bounded by a polynomial in n and $log(\epsilon^{-1})$, then τ_{rel} is also bounded by a polynomial in n and $\log(\epsilon^{-1})$. Further, suppose we have $\log(1/\pi(x))$ bounded by some polynomial $q(n)$ for all $x \in \Omega$. Then, $\tau_{rel}(\mathcal{M})$ being bounded by a polynomial in n and $\log(\epsilon^{-1})$ implies that $\tau(\epsilon)$ is also bounded by some polynomial in n and $\log(\epsilon^{-1})$.*

Proof. Suppose that $\tau(\epsilon) \leq p(n, \log(\epsilon^{-1}))$, where p is a polynomial. Beginning with the left hand side of Lemma 2, note that

$$\frac{|\lambda_1|}{1 - |\lambda_1|} \log\left(\frac{1}{2\epsilon}\right) = (\tau_{rel}(\mathcal{M}) - 1) \log\left(\frac{1}{2\epsilon}\right).$$

Then, applying Lemma 2 and the bound on $\tau(\epsilon)$,

$$\tau_{rel}(\mathcal{M}) \leq \frac{\tau(\epsilon)}{\log((2\epsilon)^{-1})} + 1 \leq \frac{p(n, \log(\epsilon^{-1}))}{\log((2\epsilon)^{-1})} + 1 \leq p'(n, \log(\epsilon^{-1})),$$

where p' is again a polynomial in n and $\log(\epsilon^{-1})$.

Turning now to converse, suppose that we have $\tau_{rel} \leq p(n, \log(\epsilon^{-1}))$, for some polynomial p. Additionally suppose $\log(1/\pi(x)) \leq q(n)$ for all $x \in \Omega$, for some polynomial q.

Applying Lemma 2,

$$\tau(\epsilon) \leq \tau_{rel}(\mathcal{M}) \log\left(\frac{1}{\pi(x)\epsilon}\right) \leq p(n, \log(\epsilon^{-1})) \log(\epsilon^{-1}) q(n) \leq p'(n, \log(\epsilon^{-1})),$$

where p' is some polynomial. □

2.2.4. Coupling

A *coupling* of a Markov chain $\mathcal{M}$ on Ω is a chain $(X_t, Y_t)_{t=0}^{\infty}$ on $\Omega \times \Omega$ for which the following properties hold.

1. Each chain $(X_t)_{t=0}^{\infty}$ and $(Y_t)_{t=0}^{\infty}$, when viewed in isolation, is a copy of $\mathcal{M}$, given initial states $X_0 = x$ and $Y_0 = y$.
2. Whenever $X_t = Y_t$, we have $X_{t+1} = Y_{t+1}$.

Formally, the first item above requires that the joint distribution of (X_t, Y_t) (given (X_{t-1}, Y_{t-1})) should satisfy the property that the marginal of X_t (and also Y_t) is consistent with the probability transitions of $\mathcal{M}$. We define the *coupling time* T to be

$$T = \max_{x,y \in \Omega} \mathbb{E}\left[\min\{t \colon X_t = Y_t \mid X_0 = x,\, Y_0 = y\}\right] \tag{13}$$

The following lemma [29] is useful in bounding the coupling time T.

Lemma 4. *Suppose that $(X_t, Y_t)_{t=0}^{\infty}$ is a coupling of a Markov chain M. Let φ be an integer-valued distance function on $\Omega \times \Omega$ taking values in the range $[0, B]$, and suppose that $\varphi(x, y) = 0$ if and only if $x = y$. Let $\varphi(t) = \varphi(x_t, y_t)$. Suppose that the coupling satisfies $E\left(\varphi(t+1) - \varphi(t) | X_t, Y_t\right) \leq 0$. Additionally, suppose that whenever $\varphi(t) > 0$, $E\left(|\varphi(t+1) - \varphi(t)|^2 | X_t, Y_t\right) \geq V$. Then, the expected coupling time satisfies $E\left(T^{x,y}\right) \leq \varphi(0)(2B - \varphi(0))/V$.*

Coupling time and mixing time are then related by the following theorem [28].

Theorem 1. *A Markov chain M with coupling time T has mixing time $\tau(\epsilon)$ bounded by*

$$\tau(\epsilon) \leq \lceil Te \log \epsilon^{-1} \rceil. \tag{14}$$

2.2.5. Decomposition

We use two disjoint decomposition methods for bounding the spectral gap, one developed by Martin and Randall [30], and a very recent one given by Hermon and Salez [31], building on the work by Jerrum, Son, Tetali and Vigoda [32]. We use both theorems because, while the latter gives better bounds, the former has more relaxed conditions, which is necessary in one of our applications. The setup for both methods is the same.

Let $\mathcal{M}$ be an ergodic, reversible Markov chain over a state space Ω with transition matrix P and stationary distribution π. Suppose Ω can be partitioned into disjoint subsets $\Omega_1, \ldots, \Omega_m$. For each $i \in [m]$, let $\mathcal{M}_i$ be the *restriction* of $\mathcal{M}$ to Ω_i, which is obtained by rejecting any transition that would leave Ω_i. Let P_i be the transition matrix of $\mathcal{M}_i$ Additionally, we define $\overline{\mathcal{M}}$ to be the *projection chain* of $\mathcal{M}$ over the state space $[m]$ as follows. Let the transition matrix $\overline{P}$ of $\overline{\mathcal{M}}$ be given by

$$\overline{P}(i, j) = \frac{1}{\pi(\Omega_i)} \sum_{\substack{x \in \Omega_i \\ y \in \Omega_j}} \pi(x) P(x, y). \tag{15}$$

One can check that $\overline{\mathcal{M}}$ is reversible and has stationary distribution

$$\overline{\pi}(i) = \pi(\Omega_i),$$

while each $\mathcal{M}_i$ has stationary distribution

$$\pi_i(x) = \frac{\pi(x)}{\overline{\pi}(i)}.$$

With this notation, we have the following theorem by Martin and Randall [30].

Theorem 2. *Defining $\mathcal{M}_i$ and $\overline{\mathcal{M}}$ as above, we have*

$$\text{Gap}(\mathcal{M}) \geq \frac{1}{2}\text{Gap}(\overline{\mathcal{M}}) \min_{i\in[m]} \text{Gap}(\mathcal{M}_i). \tag{16}$$

The theorem due to Hermon and Salez obtains better bounds if, for each pair $(i,j) \in [m] \times [m]$ with $\overline{P}(i,j) > 0$, we can find an effective joint distribution (often referred to as a "coupling") $\kappa_{ij} \colon \Omega_i \times \Omega_j \to [0,1]$ of the distributions π_i and π_j. In other words, we must have

$$\forall x \in \Omega_i, \qquad \sum_{y\in\Omega_j} \kappa_{ij}(x,y) = \pi_i(x), \tag{17}$$

$$\forall y \in \Omega_j, \qquad \sum_{x\in\Omega_i} \kappa_{ij}(x,y) = \pi_j(y). \tag{18}$$

The *quality* of the joint distribution κ is defined as

$$\chi := \chi(\kappa) := \min\left\{\frac{\pi(x)P(x,y)}{\overline{\pi}(i)\overline{P}(i,j)\kappa_{ij}(x,y)}\right\}, \tag{19}$$

where the minimum is taken over all (x,y,i,j) with $x \in \Omega_i, y \in \Omega_j$ for which $\overline{P}(i,j) > 0$ and $\kappa_{ij}(x,y) > 0$. Hermon and Salez [31] prove the following.

Theorem 3. *With P, $\bar{P}$, P_i, and χ defined as above,*

$$\text{Gap}(\mathcal{M}) \geq \min\left\{\chi\text{Gap}(\overline{\mathcal{M}}), \min_{i\in[m]} \text{Gap}(\mathcal{M}_i)\right\}. \tag{20}$$

The utility of these decomposition theorems is that they allow us to break down a more complicated Markov chain into pieces that are easier to analyze. If we can show that the pieces rapidly mix, and the projection chain rapidly mixes, then we may conclude that the original chain rapidly mixes as well.

Additionally, to aid with the analysis of some projection chains, we will need another lemma from [30].

Let $\mathcal{M}_M$ be the Markov chain on $[m]$ with Metropolis transitions $P_M(i,j) = \frac{1}{2\Delta}\min\{1, \frac{\pi(\Omega_j)}{\pi(\Omega_i)}\}$ whenever $\overline{P}(i,j) > 0$, where Δ is the maximum degree of vertices in the transition graph of $\overline{M}$ Let $\partial_i(\Omega_j) = \{y \in \Omega_j \colon \exists x \in \Omega_i \text{ with } P(x,y) > 0\}$. Then we have the following

Lemma 5. *With $\mathcal{M}_M$ as defined above, suppose there exist constants $a > 0$ and $b > 0$ with*

1. *$P(x,y) \geq a$ for all x, y such that $P(x,y) > 0$.*
2. *$\pi(\partial_i(\Omega_j)) \geq b\pi(\Omega_j)$ for all i, j with $\overline{P}(i,j) > 0$.*

Then $\text{Gap}(\overline{\mathcal{M}}) \geq ab \cdot \text{Gap}(\mathcal{M}_M)$.

In order to help analyze the mixing time of $\mathcal{M}_M$, we will also require the following two lemmas. Note that Lemma 6 is used only in the proof of Lemma 7.

Lemma 6. *Let $(a_i)_{i=1}^m$ be a log concave sequence, with $a_i > 0$ for all $1 \leq i \leq m$. Then,*

$$\frac{a_{i+1}}{a_i} \geq \frac{a_{j+1}}{a_j} \tag{21}$$

for all $1 \leq i \leq j \leq m$.

Proof. In order to use induction, we will slightly reframe the statement. We will prove

$$\frac{a_{i+1}}{a_i} \geq \frac{a_{i+1+k}}{a_{i+k}}$$

for all $i + k \leq n$.

We now proceed by induction on k. The base case, $k = 0$, is trivial.

Now fix $l > 0$ and suppose that the induction hypothesis is true for $k = l - 1$, that is,

$$\frac{a_{i+1}}{a_i} \geq \frac{a_{i+l}}{a_{i+l-1}}.$$

By log concavity $a_{i+l}^2 \geq a_{i+l-1}a_{i+l+1}$, or, equivalently,

$$\frac{a_{i+l}}{a_{i+l-1}} \geq \frac{a_{i+l+1}}{a_{i+l}}.$$

Therefore,

$$\frac{a_{i+1}}{a_i} \geq \frac{a_{i+l}}{a_{i+l-1}} \geq \frac{a_{i+l+1}}{a_{i+l}},$$

where the first inequality follows from the induction hypothesis, and the second inequality follows from log concavity. □

Lemma 7. *Let π be a probability distribution on $[m]$. Let $\mathcal{M}$ be a Markov chain on $[m]$ with the transition probabilities*

$$P(i,j) = \begin{cases} \frac{1}{4}\min\left\{1, \frac{\pi(j)}{\pi(i)}\right\} & \text{if } |i-j| = 1 \\ 0 & \text{if } |i-j| > 1 \end{cases} \tag{22}$$

and the appropriate self-loop probabilities $P(i,i)$. If $\pi(i)$ is log concave in i, then $\mathcal{M}$ has mixing time (and hence also relaxation time) $O(m^2)$.

Proof. We define a coupling (X_t, Y_t) on $\mathcal{M}$ as follows. If $X_t \neq Y_t$, then at time step $t+1$, flip a fair coin.

- If heads, set $Y_{t+1} = Y_t$. Let l be either 1 or -1, each with probability $1/2$. If possible, let $X_{t+1} = X_t + l$ with probability $\frac{1}{2}\min\left\{1, \frac{\pi(X_t+l)}{\pi(X_t)}\right\}$. Otherwise, let $X_{t+1} = X_t$.
- If tails, set $X_{t+1} = X_t$, and update Y_{t+1} the same way as we did for X_{t+1} in the previous case.

Now, suppose that for some t we have $X_t = i$ and $Y_t = j$ for $i \neq j$. WLOG, assume that $i < j$. Let $\varphi(t) = \varphi(X_t, Y_t) = j - i$, and let $\Delta\varphi(t) = \varphi(t) - \varphi(t-1)$. Note that we have two moves, with probabilities $P(i, i-1)$ and $P(j, j+1)$, which will increase the distance φ by 1 and similarly two moves, with probabilities $P(i, i+1)$ and $P(j, j+1)$, will decrease the distance by 1. Then we have

$$\mathbb{E}(\Delta\varphi(t)) = -P(i,i+1) + P(i,i-1) + P(j,j+1) - P(j,j-1).$$

By the log-concavity of $\pi(i)$ and Lemma 6, we have $P(i,i+1) \geq P(j,j+1)$ and $P(i,i-1) \leq P(j,j-1)$. Therefore, the expected change in $\varphi(t)$ is always non-positive. We also have

$$\begin{aligned}\mathbb{E}\left((\Delta\varphi(t))^2 \,|X_t, Y_t\right) &= P(j,j+1) + P(i,i+1) + P(j,j-1) + P(i,i-1) \\ &= \frac{1}{4}\left(\min\left\{1, \frac{\pi(j+1)}{\pi(j)}\right\} + \min\left\{1, \frac{\pi(i+1)}{\pi(i)}\right\} + \min\left\{1, \frac{\pi(j-1)}{\pi(j)}\right\} + \min\left\{1, \frac{\pi(i-1)}{\pi(i)}\right\}\right).\end{aligned}$$

We claim that $E\left((\Delta\varphi)^2 \,|X_t, Y_t\right) \geq \frac{1}{4}$. Suppose, for contradiction, that the expectation is less than $\frac{1}{4}$. Then, for each of the minimum functions in the above expression, 1 must be the larger argument. Equivalently, $\pi(i-1) < \pi(i)$, $\pi(i) > \pi(i+1)$, $\pi(j-1) < \pi(j)$, and $\pi(j) > \pi(j+1)$. Therefore, $\pi(i)$ is not unimodal in i and is therefore also not log concave in i, contradicting our hypothesis. Therefore we have $E\left((\Delta\varphi)^2 \,|X_t, Y_t\right) \geq \frac{1}{4}$, as desired. □

3. Results

Here we present the constructed Markov chain and corresponding algorithms devised for the sampling task and the proof of an upper bound on the relaxation time - that the chain mixes rapidly. Collectively, the results illustrate an analytical approach to calculate dispersion of the secondary structure and corresponding branching properties of RNA based on the NNTM energy function minimization and without reference to a specific nucleotide sequence.

3.1. Our Markov Chain on $\mathfrak{M}_m^2$

We define a Markov chain $\mathcal{M} = X_0, X_1, X_2, \cdots$ on $\mathfrak{M}_m^2$ to sample 2-Motzkin paths as a representation of plane trees. Here, we use $m = n-1$ to denote the length of the 2-Motzkin paths corresponding to plane trees with n edges.

We define each step of $\mathcal{M}$ as follows. First, pick a random element l uniformly from $\{1,2,3,4\}$. Now choose y as follows.

- If $l = 1$, pick a random pair of consecutive symbols in X_t, and call this pair s. If s is UD or HH, let s' be either UD or HH with probabilities $\frac{1}{1+e^{-\alpha}}$ and $\frac{e^{-\alpha}}{1+e^{-\alpha}}$ respectively. Let y be the string X_t with s replaced by s'. Otherwise, let $y = X_t$.
- If $l = 2$, pick i uniformly from $\{1, \cdots, m\}$. If $X_t(i)$ is H or I, choose a symbol c to be either H or I with probabilities $\frac{e^{-\alpha}}{e^{-\alpha}+e^{-\beta}}$ and $\frac{e^{-\beta}}{e^{-\alpha}+e^{-\beta}}$ respectively. Let y be the 2-Motzkin path given by changing the symbol in $X_t(j)$ to c. Otherwise, we let $y = X_t$.
- If $l = 3$, pick i and j each uniformly from $\{1, \cdots, m\}$. If each of $X_t(i)$ and $X_t(j)$ are either U or D, let y be the string X_t with the symbols at indices i and j swapped. Otherwise, let $y = X_t$.
- If $l = 4$, pick a random pair of consecutive symbols in X_t, and call this pair s. If s is of the form ab or ba for some $a \in \{U, D\}$ and $b \in \{H, I\}$, let s' be the reverse of s, and let y be the string X_t with s replaced by s'. Otherwise, let $y = X_t$.

If y is a valid 2-Motzkin path, set $X_{t+1} = y$ with probability $\frac{1}{2}$. Otherwise, set $X_{t+1} = X_t$.

One can see that $\mathcal{M}$ is irreducible by noting that every path can be transformed to the path consisting of all H's. To make this transformation, first use the $l = 4$ rule to move all H's and I's to the end of the path. If there are any U's in the path, we must now have at least one consecutive pair UD. Use the $l = 1$ rule to convert the UD to a HH. From here we can repeat, again moving all H's to the end and replacing UD with HH, until only H's and I's remain. Finally, we can use the $l = 2$ rule to convert all I's to H's. Since all of these steps can also be taken in reverse, this gives a procedure to move between two arbitrary paths, demonstrating irreducibility. We can also conclude that $\mathcal{M}$ is aperiodic, due to the existence of self loops. Combined with irreducibility, this establishes that $\mathcal{M}$ is ergodic.

We claim that $\mathcal{M}$ is reversible with respect to the stationary distribution $\pi(x) = \frac{e^{-E(x)}}{Z}$, where $Z = \sum_{y \in \mathfrak{M}_m^2} e^{-E(y)}$. This can be easily verified by considering the 4 move types listed above. For

example, for the first move type given above (transforming UD to HH and vice versa), let x and y be the states of interest. Suppose that y has the consecutive symbols HH where x contains UD. Then,

$$\begin{aligned}\pi(x)P(x,y) &= \frac{e^{-\alpha(|x|_U+|x|_H+1)-\beta|x|_I}}{Z}\cdot\frac{e^{-\alpha}}{1+e^{-\alpha}}\\ &= \frac{e^{-\alpha\left((|y|_U+1)+(|y|_H-2)+1\right)-\beta|y|_I}}{Z}\cdot\frac{e^{-\alpha}}{1+e^{-\alpha}}\\ &= \frac{e^{-\alpha\left(|y|_U+|y|_H+1\right)-\beta|y|_I}}{Z}\cdot\frac{1}{1+e^{-\alpha}}\\ &= \pi(y)P(y,x).\end{aligned}$$

One can verify that similar computations hold for the remaining 3 types of moves. Therefore, we conclude that the chain $\mathcal{M}$ has stationary distribution $\pi(x) = \frac{e^{-E(x)}}{Z}$.

The Markov chain $\mathcal{M}$ can be implemented in pseudocode as in Table 2. Here, the $\text{Ber}(p)$ function returns true with probability p, and false otherwise. We also use addition of strings to denote concatenation.

Table 2. The main Markov chain algorithm. This pseudocode calculates X_t given X_0.

Require: X_0 is a valid 2-Motzkin path of length m.
  $x \leftarrow X_0$
  for $s = 1 \to t$ **do**
    $y \leftarrow x$
    $l \leftarrow \text{randInt}(1, 4)$
    if $l = 1$ **then**
      $i \leftarrow \text{randInt}(1, m-1)$
      if $x[i:i+1] = UD$ **and** $\text{Ber}\left(\frac{e^{-\alpha}}{2(1+e^{-\alpha})}\right)$ **then**
        $y[i:i+1] \leftarrow HH$
      else if $x[i:i+1] = HH$ **and** $\text{Ber}\left(\frac{1}{2(1+e^{-\alpha})}\right)$ **then**
        $y[i:i+1] \leftarrow UD$
    else if $l = 2$ **then**
      $i \leftarrow \text{randInt}(1, m)$
      if $x[i] = I$ **and** $\text{Ber}\left(\frac{e^{-\alpha}}{2(e^{-\alpha}+e^{-\beta})}\right)$ **then**
        $y[i] \leftarrow H$
      else if $x(i) = H$ **and** $\text{Ber}\left(\frac{e^{-\beta}}{2(e^{-\alpha}+e^{-\beta})}\right)$ **then**
        $y[i] \leftarrow I$
    else if $l = 3$ **then**
      $i \leftarrow \text{randInt}(1, m)$
      $j \leftarrow \text{randInt}(1, m)$
      if $(x[i] \in \{U, D\}$ **and** $x[j] \in \{U, D\})$ **and** $\text{Ber}\left(\frac{1}{2}\right)$ **then**
        $y[i] \leftarrow x[j]$
        $y[j] \leftarrow x[i]$
        if y is not a valid 2-Motzkin path **then**
          $y \leftarrow x$
    else if $l = 4$ **then**
      $i \leftarrow \text{randInt}(1, m-1)$
      if $(x[i] \in \{U, D\}$ **and** $x[j+1] \in \{H, I\})$ **or** $(x[i] \in \{H, I\}$ **and** $x[j+1] \in \{U, D\})$ **and** $\text{Ber}\left(\frac{1}{2}\right)$ **then**
        $y[i:i+1] \leftarrow x[j+1] + x[j]$
    $x \leftarrow y$
  return x

Additionally, in order to convert the 2-Motzkin path X_t into a plane tree, we use the algorithm in Table 3, which assumes the existence of a Node object with children and parent attributes.

Table 3. Algorithm to convert a sampled 2-Motzkin path to a plan tree. The pseudocode calculates $\Phi^{-1}(x)$.

Require: x is a valid 2-Motzkin path of length m.
root ← new Node()
// u will be where a new node will be added for an H or D symbol
$u \leftarrow$ root
// v will be always the last node added
$v \leftarrow$ new Node()
// the stack will keep track of previous values of u
stack = new Stack()
root.children.append(v)
for $i = 1 \rightarrow m$ **do**
  node ← new Node()
  if $x[i] = U$ **then**
    v.children.append(node)
    stack.push(u)
    $u \leftarrow v$
  else if $x[i] = I$ **then**
    v.children.append(node)
  else if $x[i] = H$ **then**
    u.children.append(node)
  else if $x[i] = D$ **then**
    u.children.append(node)
    $u \leftarrow$ stack.pop()
  $v \leftarrow$ node
return root

3.2. Mixing Time Results

Our main result is to prove the rapid mixing of the Markov chain defined in Section 3.1. An upper bound on the relaxation time is achieved by bounding the spectral gap from below. A spectral gap bound for the complex chain at hand is obtained through the use of multiple decomposition theorems, which give bounds on the spectral gap of the complex chain in terms of the spectral gaps of multiple simpler chains. The disjoint decomposition theorem due to Martin and Randall [30] provides a flexible approach to decomposition of Markov chains. Very recent work by Hermon and Salez [31], building on the work of Jerrum, Son, Tetali, and Vigoda [32], proves a decomposition theorem with tighter bounds but stronger hypotheses.

Since this proof involves multiple decomposition steps, we provide an overview here. The primary tools used in this proof are the two decomposition theorems presented in Section 2.2.5. We first partition the state space of all 2-Motzkin paths by the number of Us in the path. The projection chain from this first decomposition is linear and is proved to be rapidly mixing using a result of Martin and Randall [30] (Lemma 8). Each of the restriction chains are decomposed again, this time by the pattern of H and I symbols. The projection chains for this second decomposition are shown to be rapidly mixing by coupling (Lemma 9). The restriction chains are decomposed a third time, this time according to the skeleton of U and D steps. The projection chains for this third decomposition are shown to be rapidly mixing by comparison to the classic mountain valley moves chain on Dyck paths (Lemma 10). This last set of restriction chains are found to be rapidly mixing by isomorphism to the chain consisting of adjacent transpositions on binary strings (Lemma 11). Finally, starting from the

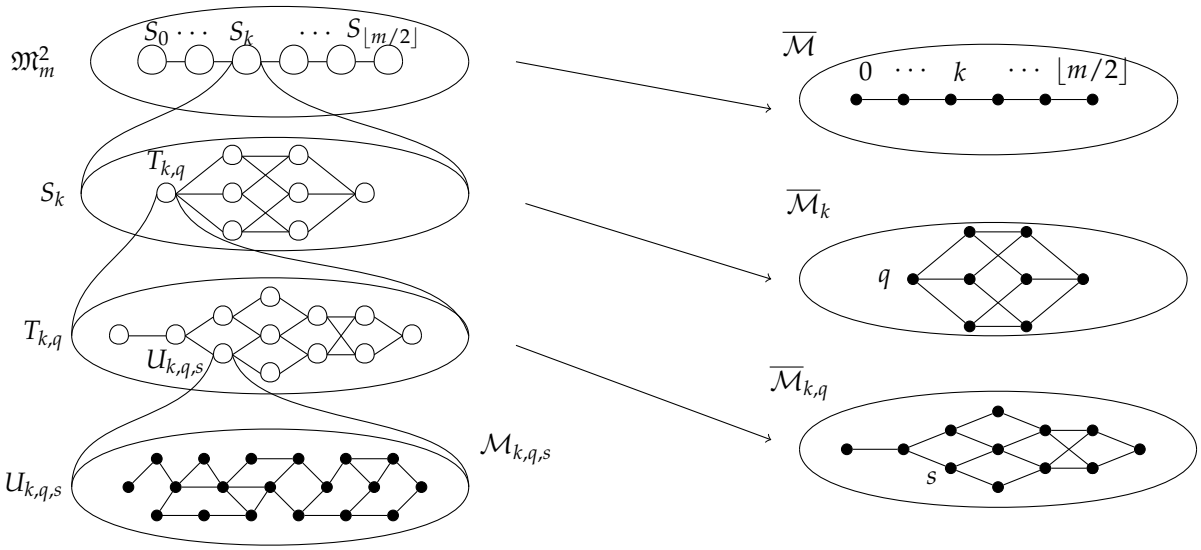


Figure 3. The four level decomposition of $\mathfrak{M}_m^2$ (left), and the projection chains corresponding to each decomposition (right).

most restricted chains, we use the decomposition theorems to obtain a bound on the spectral gap of the original chain (Theorem 4).

We now proceed with a formal presentation. We will use a series of decompositions of $\mathcal{M}$. We will first decompose our state space $\mathfrak{M}_m^2$ into $S_0, \cdots, S_{\lfloor m/2 \rfloor}$, where

$$S_k = \{x \in \mathfrak{M}_m^2 \colon |x|_U = k\}.$$

Let $\mathcal{M}_k$ denote the Markov chain $\mathcal{M}$ restricted to the set S_k, and let $\overline{\mathcal{M}}$ be the projection chain over this decomposition as outlined for Theorem 2.

Additionally, we will decompose each S_k into the sets $\{T_{k,q} \colon q \in (H+I)^{m-2k}\}$, where $(H+I)^{m-2k}$ denotes the set of strings with length $m-2k$ from the alphabet $\{H, I\}$. We define $T_{k,q}$ to be the set of 2-Motzkin paths $x \in S_k$ such that the substring of H and I symbols in x is q. Let $\mathcal{M}_{k,q}$ denote the chain $\mathcal{M}_k$ restricted to $T_{k,q}$, and let $\overline{M}_k$ be the projection chain of $\mathcal{M}_k$ over this decomposition.

Finally, we decompose each $T_{k,q}$ into the partition $\{U_{k,q,s} \colon s \in \mathfrak{D}_k\}$ based on the skeletons of the 2-Motzkin paths. For each $s \in \mathfrak{D}_k$, we define

$$U_{k,q,s} = \{x \in T_{k,q} \mid \sigma(x) = s\}.$$

As before, we let $\mathcal{M}_{k,q,s}$ be the Markov chain $\mathcal{M}_{k,q}$ restricted to $U_{k,q,s}$, and let $\overline{\mathcal{M}}_{k,q}$ be the appropriate projection chain. For clarity, this four-level decomposition is summarized in Figure 3.

Lemma 8. *$\overline{\mathcal{M}}$ has relaxation time $\tau_{rel}(\overline{\mathcal{M}}) = O(m^4)$.*

Proof. The chain $\overline{\mathcal{M}}$ is a linear chain with states k in $\{0, \ldots, \lfloor m/2 \rfloor\}$, and with stationary distribution

$$\begin{aligned}\overline{\pi}(k) = \pi(S_k) &= \frac{C_k}{Z_m} \cdot \sum_{i=0}^{m-2k} \binom{m}{2k} \binom{m-2k}{i} e^{-\alpha(k+i+1)-\beta(m-2k-i)} \\ &= \frac{e^{-\alpha(k+1)}}{Z_m} \binom{m}{2k} C_k \cdot (e^{-\alpha} + e^{-\beta})^{m-2k},\end{aligned}$$

where $\overline{\pi}$ is defined as in Section 2.2.5. Notice that transitions in $\mathcal{M}$ which move between the S_k sets are those which change a HH substring into a UD or DU substring, or vise versa. Thus, the transitions in

$\overline{\mathcal{M}}$ only increase or decrease k by at most 1. We seek to apply Lemma 5. To choose a, notice that for $x \in S_k$ and $y \in S_{k\pm1}$ with $P(x,y) > 0$, we have

$$P(x,y) = \frac{1}{4(m-1)}\frac{1}{1+e^{\alpha}} \text{ or } P(x,y) = \frac{1}{4(m-1)}\frac{1}{1+e^{-\alpha}}.$$

Note that the factor $1/4$ comes from the choice $l = 4$, and the factor $1/(m-1)$ comes from the fact that there are $m-1$ adjacent pairs to pick from. Then,

$$P(x,y) \geq \frac{1}{4(m-1)(1+e^{-|\alpha|})}.$$

Thus, we pick $a = \frac{1}{4(m-1)(1+e^{-|\alpha|})}$.

To pick b, we let

$$\partial_-(S_k) = \{y \in S_k \colon \exists x \in S_{k-1}, P(x,y) > 0\}$$

for $k \in \{1, \cdots, \lfloor m/2 \rfloor\}$, and we let

$$\partial_+(S_k) = \{y \in S_k \colon \exists x \in S_{k+1}, P(x,y) > 0\}$$

for $k \in \{0, \cdots, \lfloor m/2 \rfloor - 1\}$.

Additionally, let A_k for $k \in \{1, \cdots, \lfloor m/2 \rfloor\}$ be the subset of S_k consisting of the 2-Motzkin paths in which the first D symbol appears immediately after a U. Let B_k for $k \in \{0, \cdots, \lfloor m/2 \rfloor - 1\}$ be the subset of S_k consisting of the 2-Motzkin paths in which a pair of adjacent H symbols occurs before all other H or I symbols. It is easy to see that $A_k \subset \partial_-(S_k)$ and $B_k \subset \partial_+(S_k)$. We have

$$\pi(A_k) = \frac{C_k e^{-\alpha(k+1)}}{Z_m}\binom{m-1}{2k-1}(e^{-\alpha}+e^{-\beta})^{m-2k},$$

as there are C_k ways to arrange the U and D symbols and $\binom{m-1}{2k-1}$ ways to insert $m-2k$ H or I symbols (treating H and I as being identical for now) without placing anything between the first D and the U immediately before it. The energy contribution of the U and D symbols is given by $e^{-\alpha(k+1)}$, and the energy contribution of the H and I symbols is $\left(e^{-\alpha}+e^{-\beta}\right)^{m-2k}$. The required normalizing constant is Z_n. Similarly, we also get

$$\pi(B_k) = \frac{C_k e^{-\alpha(k+3)}e^{-2\beta}}{Z_m}\binom{m-1}{2k}(e^{-\alpha}+e^{-\beta})^{m-2k-2}$$

because there are C_k ways to arrange the U and D symbols and $\binom{m-1}{2k}$ ways insert $m-2k-1$ H or I symbols (treating the initial pair of H's as a single symbol gives us only $m-2k-1$ symbols to insert). The energy contribution of the U's, D's, and the initial two H's is given by $e^{-\alpha(k+3)}e^{-2\beta}$, and the energy contribution of the remaining H's and I's is $(e^{-\alpha}+e^{-\beta})^{m-2k-2}$. Finally, Z_m is again a normalizing constant.

Hence combining these two results, we have

$$\frac{\pi(\partial_-(S_k))}{\pi(S_k)} \geq \frac{\pi(A_k)}{\pi(S_k)} = \frac{2k}{m}$$

and

$$\frac{\pi(\partial_+(S_k))}{\pi(S_k)} \geq \frac{\pi(B_k)}{\pi(S_k)} = \frac{m-2k}{m}\left(\frac{e^{-\alpha}e^{-\beta}}{e^{-\alpha}+e^{-\beta}}\right)^2.$$

Thus, we may let $b = \frac{1}{m}\left(\frac{e^{-\alpha}e^{-\beta}}{e^{-\alpha}+e^{-\beta}}\right)^2$.

Applying Lemma 5, we get that $\text{Gap}(\overline{\mathcal{M}}) \geq \frac{\text{Gap}(\mathcal{M}_M)}{O(m^2)}$. Additionally, one can check that $\overline{\pi}(i)$ is log concave in i. Hence, using Lemma 7, we get $\tau_{rel}(\mathcal{M}_M) = O(m^2)$, and in turn $\tau_{rel}(\overline{\mathcal{M}}) = O(m^4)$, as claimed. □

Lemma 9. *$\overline{\mathcal{M}}_k$ has mixing time $\tau(\overline{\mathcal{M}}_k) = O(m \log m)$, for all k.*

Proof. Notice that $\overline{\mathcal{M}}_k$ appears as a chain with states q in the set $Q = (H+I)^{m-2k}$. Additionally, transitions in $\overline{\mathcal{M}}_k$ only occur between strings in Q that differ at only one index. The stationary distribution of $\overline{\mathcal{M}}_k$ is given by $\overline{\pi}_k(q) \propto e^{(\beta-\alpha)|q|_H}$, where we have intentionally used the constant of proportionality to remove all dependence on k, which we consider in this context to be fixed.

Additionally, for $q_1, q_2 \in Q$ which differ at exactly one index, we have the transition probability

$$\overline{P}_k(q_1, q_2) = \begin{cases} \frac{(m-2k)e^{-\alpha}}{4m(e^{-\alpha}+e^{-\beta})} & \text{if } |q_2|_H = |q_1|_H + 1 \\ \frac{(m-2k)e^{-\beta}}{4m(e^{-\alpha}+e^{-\beta})} & \text{if } |q_2|_H = |q_1|_H - 1 \end{cases}.$$

We may show that $\overline{\mathcal{M}}_k$ rapidly mixes by a simple coupling argument. Let $(X_t, Y_t)_{t=0}^{\infty}$ be our coupled Markov chain on $Q \times Q$. We define one step in this coupled chain as follows.

1. With probability $1 - \frac{m-2k}{4m}$, set $(X_{t+1}, Y_{t+1}) = (X_t, Y_t)$.
2. Otherwise, pick a random index $j \in [m-2k]$. Let $a \in \{H, I\}$ be a random symbol such that $\Pr(a = H) = \frac{e^{-\alpha}}{e^{-\alpha}+e^{-\beta}}$ and $\Pr(a = I) = \frac{e^{-\beta}}{e^{-\alpha}+e^{-\beta}}$. Now let X_{t+1} and Y_{t+1} be X_t and Y_t respectively, each with the jth symbol changed to a.

One can check that each of $(X_t)_t$ and $(Y_t)_t$ are indeed copies of $\overline{\mathcal{M}}_k$. Additionally, notice that we will have $X_t = Y_t$ after all $m-2k$ possible indices j have been updated. By the Coupon Collector Theorem, we have the coupling time of this chain to be $T_{\overline{\mathcal{M}}_k} = \frac{4m}{m-2k} \cdot O((m-2k)\log(m-2k)) = O(m \log m)$. Thus, using Theorem 1, we have the mixing time (and the relaxation time) also $O(m \log m)$. □

Lemma 10. *$\overline{\mathcal{M}}_{k,q}$ has relaxation time $\tau_{rel}(\overline{\mathcal{M}}_{k,q}) = O(m^2)$, for all pairs (k, q).*

Proof. Notice that all $x \in T_{k,q}$ have equal energy, and that $|U_{k,q,s}| = \binom{m}{2k}$ for all s. Thus, $\overline{\mathcal{M}}_{k,q}$ has a uniform stationary distribution. If we represent each set $U_{k,q,s}$ by the Dyck path s, we can think of $\overline{\mathcal{M}}_{k,q}$ as a chain over $\mathfrak{D}_k$. Since all the transitions in $\mathcal{M}_{k,q}$ that move between the $U_{k,q,s}$ sets are moves that exchange the positions of a U and a D, the transitions in $\overline{\mathcal{M}}_{k,q}$ are simply the moves on elements of $\mathfrak{D}_k$ which exchange a U with a D. We call these moves on the elements of $\mathfrak{D}_k$, *transposition moves*.

For each $s_1, s_2 \in \mathfrak{D}_k$ that differ by a transposition move, the transition probabilities in our projection chain are given by

$$\begin{aligned} \overline{P}_{k,q}(s_1, s_2) &= \frac{1}{\pi(U_{k,q,s_1})} \sum_{\substack{x \in U_{k,q,s_1} \\ y \in U_{k,q,s_2}}} \pi(x) P(x,y) = \frac{1}{|U_{k,q,s}|} \sum_{\substack{x \in U_{k,q,s_1} \\ y \in U_{k,q,s_2}}} P(x,y) \\ &= \frac{1}{\binom{m}{2k}} \sum_{\substack{x,y \\ P(x,y)>0}} \frac{1}{4m^2} = \frac{1}{4m^2}. \end{aligned}$$

The last equality above relies on counting the number of terms in the sum. Notice that for each $x \in U_{k,q,s_1}$, there is a unique $y \in U_{k,q,s_2}$ for which $P(x,y) > 0$. Therefore, the number of terms is simply $|U_{k,q,s_1}| = \binom{m}{2k}$. Compare this chain to the traditional mountain valley Markov chain on $\mathfrak{D}_k$, which we will denote by $\mathcal{M}'$. The transition probabilities of $\mathcal{M}'$ are given by $P'(s_1, s_2) = \frac{1}{k^2}$ for each pair (s_1, s_2) which differ by a mountain-valley move. It is known from Cohen [33] that $\text{Gap}(\mathcal{M}') = \frac{1}{O(k^2)}$. Thus, applying Lemma 1 to $\overline{\mathcal{M}}_{k,q}$ and $\mathcal{M}'$, we see that $\text{Gap}(\overline{\mathcal{M}}_{k,q}) = \frac{1}{O(m^2)}$. □

Lemma 11. *$\mathcal{M}_{k,q,s}$ has relaxation time $\tau_{rel}(\mathcal{M}_{k,q,s}) = O(m^3)$, for all valid triples (k,q,s).*

Proof. Notice that transitions in $\mathcal{M}_{k,q,s}$ consist only of moves which involve swapping an H or an I with an adjacent U or D. Additionally, all 2-Motzkin paths in $U_{k,q,s}$ have equal energy, so for all $x, y \in U_{k,q,s}$ such that $P(x,y) > 0$, we have $P(x,y) = \frac{1}{8(m-1)}$.

To determine the mixing time of $\mathcal{M}_{k,q,s}$, consider an isomorphic chain. Let U' be the set of all binary strings of length m with $2k$ zeros and $m-2k$ ones. Let $\mathcal{M}'$ be the Markov chain on U' where each step does nothing with probability $7/8$ and swaps a random pair of adjacent (potentially identical) digits with probability $1/8$. From Wilson [34], we know that the spectral gap of $\mathcal{M}'$ is $\frac{1}{O(m^3)}$. □

Finally, we can combine our bounds on the spectral gaps of all of these chains to prove our main result.

Theorem 4. *The Markov chain $\mathcal{M}$ has relaxation time $\tau_{rel}(\mathcal{M}) = O(m^7)$, for all $\alpha, \beta \in \mathbb{R}$.*

Proof. We use Lemmas 11 and 10 with Theorem 3 to obtain a bound on $\text{Gap}(\mathcal{M}_{k,q})$. We define a coupling κ_{s_1,s_2} for each pair $(s_1,s_2) \in \mathfrak{D}_k \times \mathfrak{D}_k$ with $\overline{P}_{k,q}(s_1,s_2) > 0$. For each such pair, notice that the set of pairs $(x,y) \in U_{k,q,s_1} \times U_{k,q,s_2}$ with $P(x,y) > 0$ is a perfect matching. Thus, we may set

$$\kappa_{s_1,s_2}(x,y) = \begin{cases} \frac{1}{\binom{m}{2k}} & \text{if } P(x,y) > 0 \\ 0 & P(x,y) = 0 \end{cases}.$$

To compute χ, we begin by observing $\pi(x) = \pi(y)$ for all $x, y \in \mathcal{M}_{k,q}$. Also note $\left|U_{k,q,s}\right| = \binom{m}{2k}$ for all skeletons s of length $2k$. Before computing χ, we start by finding $\overline{P}(s_1,s_2)$.

$$\begin{aligned} \overline{P}(s_1,s_2) &= \frac{1}{\pi(U_{k,q,s_1})} \sum_{x \in U_{k,q,s_1}, y \in U_{k,q,s_2}} \pi(x)P(x,y) \\ &= \frac{1}{\pi(U_{k,q,s_1})} \sum_{x \in U_{k,q,s_1}, y \in U_{k,q,s_2}} \frac{\pi(x)}{\frac{1}{4}\binom{m}{2}} \\ &= \frac{1}{\pi(U_{k,q,s_1})} \left|U_{k,q,s_1}\right| \frac{4\pi(x)}{\binom{m}{2}} \\ &= \frac{4}{\binom{m}{2}}. \end{aligned}$$

We now proceed with the calculation of χ. Recall that the minimum is taken over all tuples x, y, s_1, s_2 where $\overline{P}(s_1,s_2) > 0$ and $\kappa_{0_1,s_2}(x,y) > 0$.

$$\begin{aligned} \chi &= \min\left\{ \frac{\pi(x)P(x,y)}{\overline{\pi}(s_1)\overline{P}(s_1,s_2)\kappa_{s_1,s_2}(x,y)} \right\} \\ &= \min\left\{ \frac{\pi(x)\frac{4}{\binom{m}{2}}}{\pi(U_{k,q,s_1})\frac{4}{\binom{m}{2}}\frac{1}{\binom{m}{2k}}} \right\} \\ &= \frac{\binom{m}{2k}}{\binom{m}{2k}} = 1. \end{aligned}$$

Theorem 3 then gives

$$\begin{aligned}\text{Gap}(\mathcal{M}_{k,q}) &\geq \min\left\{\chi\text{Gap}(\overline{\mathcal{M}}_{k,q}), \min_s \text{Gap}(\mathcal{M}_{k,q,s})\right\} \\ &\min\left\{\frac{1}{O(m^2)}, \frac{1}{O(m^3)}\right\} \\ &= \frac{1}{O(m^3)}.\end{aligned}$$

Similarly, we define a coupling κ_{q_1,q_2} for each pair $(q_1, q_2) \in (H+I)^{m-2k} \times (H+I)^{m-2k}$ with $\overline{P}_k(q_1, q_2) > 0$ to apply Theorem 3 to $\overline{M}_k$. Notice that once again, the set of pairs $(x, y) \in T_{k,q_1} \times T_{k,q_2}$ for which $P(x, y) > 0$ forms a perfect matching. Thus, we take

$$\kappa_{q_1,q_2}(x,y) = \begin{cases} \frac{1}{\binom{m}{2k}C_k} & \text{if } P(x,y) > 0 \\ 0 & P(x,y) = 0 \end{cases}.$$

To compute χ for this coupling, we again begin with a few preliminary computations. In all of the following, let $x \in T_{k,q_1}, y \in T_{k,q_2}$ with $\overline{P}(q_1, q_2) > 0$. Note that q_1 and q_2 have the same length and differ at only one index. We will show the computations for the case where q_1 has a I where q_2 has a H. The computations for the other case are nearly identical.

Note that $P(x, y) = \frac{e^{-\alpha}}{e^{-\alpha}+a^{-\beta}}$. Also note

$$\overline{\pi}(q_1) = \pi(T_{k,q_1}) = \pi(x)\left|T_{k,q_1}\right| = \pi(x)C_k\binom{m}{2k}$$

and

$$\begin{aligned}\overline{P}(q_1,q_2) &= \frac{1}{\pi(T_{k,q_1})} \sum_{x' \in T_{k,q_1}, y' \in T_{k,q_2}} P(x',y') \\ &= \frac{1}{\left|T_{s,q_1}\right|} \cdot \frac{e^{-\alpha}}{e^{-\alpha}+e^{-\beta}} \left|T_{s,q_1}\right| \\ &= \frac{e^{-\alpha}}{e^{-\alpha}+e^{-\beta}}.\end{aligned}$$

Now we can compute

$$\begin{aligned}\chi &= \min\left\{\frac{\pi(x)P(x,y)}{\overline{\pi}(q_1)\overline{P}}(q_1,q_2)\kappa_{q_1,q_2}(x,y)\right\} \\ &= \min\left\{\frac{\pi(x)\frac{e^{-\alpha}}{e^{-\alpha}+e^{-\beta}}}{\pi(x)C_k\binom{m}{2k}\frac{e^{-\alpha}}{e^{-\alpha}+e^{-\beta}} \cdot \frac{1}{C_k\binom{m}{2k}}}\right\} \\ &= 1.\end{aligned}$$

Applying Theorem 3 then gives

$$\begin{aligned}\text{Gap}(\mathcal{M}_k) &\geq \min\left\{\chi\text{Gap}(\overline{\mathcal{M}_k}), \min_q \text{Gap}(\mathcal{M}_{k,q})\right\} \\ &= \min\left\{\frac{1}{O(m\log m)}, \frac{1}{O(m^3)}\right\} \\ &= \frac{1}{O(m^3)}.\end{aligned}$$

Unfortunately, we have not been able to find a useful coupling for $\overline{\mathcal{M}}$, so for the last step of our decomposition, we apply Theorem 2. Since $\text{Gap}(\overline{\mathcal{M}}) = O\left(\frac{1}{m^4}\right)$ and $\text{Gap}(\mathcal{M}_k) = O(\frac{1}{m^3})$ for all k, we have

$$\begin{aligned}\text{Gap}(\mathcal{M}) &\geq \frac{1}{2}\text{Gap}(\overline{\mathcal{M}}) \min_{k\in[m/2]} \text{Gap}(\mathcal{M}_k) \\ &= \frac{1}{2O(m^4)O(m^3)} \\ &= \frac{1}{O(m^7)},\end{aligned}$$

establishing Theorem 4. □

Finally, an application of Lemma 3 allows us to conclude that the mixing time is also polynomially-bounded.

Corollary 1. *$\mathcal{M}$ is rapidly mixing.*

Proof. In order to apply Lemma 3, we need to obtain a polynomial bound on $\log(1/\pi(x))$ for all $x \in \Omega$. Let $t \in \Omega$ have maximum energy among all elements of Ω. For any $x \in \Omega$,

$$\begin{aligned}\log\left(\frac{1}{\pi(x)}\right) &= \log\left(\frac{\sum_{y\in\Omega} e^{-\alpha d_0(y)-\beta d_1(y)}}{e^{-\alpha d_0(x)-\beta d_1(x)}}\right) \\ &\leq \log\left(\frac{C_n e^{-\alpha d_0(t)-\beta d_1(t)}}{e^{-\alpha d_0(x)-\beta d_1(x)}}\right) \\ &\leq \log\left(\frac{C_n e^{-\alpha n-\beta n}}{e^{-\alpha}}\right) \\ &= \log\left(C_n e^{-\alpha(n-1)} e^{-\beta n}\right) \\ &\leq n\log(2n) + \log\left(\frac{1}{n+1}\right) - \alpha(n-1) - \beta n.\end{aligned}$$

This gives us the required polynomial bound, and therefore Lemma 3 implies that $\mathcal{M}$ is rapidly mixing. □

4. Discussion and Conclusions

The goal of this work was to identify a Markov chain and construct a corresponding algorithm by which to examine the non-uniform distribution and dispersion properties of NNTM RNA secondary structures and branching properties independent of a specific nucleotide sequence. This study successfully identifies the existence of a Markov chain, with a provably polynomial mixing time, which generates a Gibbs distribution on plane trees. This stationary probability distribution models branching characteristics of RNA secondary structure under the NNTM. While exploration of sampled structures obtained from this algorithm are beyond the scope of the presented results, pseudocode (see Section 3.1) is provided to facilitate future work in this area. Below we discuss the direct applications and implications of this work to RNA modeling, the possibility of implementing a dynamic programming approach, the possibility of an approach using stochastic context free grammars, other biological applications of this work, contributions of this work towards independent mathematical research interests, and limitations and future directions of the present work.

4.1. Applications to RNA modeling

The most straightforward application of this work is in understanding the background distribution of the branching behavior for secondary structures predicted under the NNTM. While the NNTM is widely used to predict secondary structures from sequence data, little is known about the general branching characteristics of the predicted structures, independent of a specific input sequence. Quantities such as the number of hairpins, the maximum branching in a multiloop, the average branching in a multiloop, and the maximum ladder distance of the structure [7,35] help to characterize the branching behavior and could be computed from samples obtained from this algorithm. These quantities also have been studied in native structures and/or could be easily obtained from databases such as RNA STRAND [36]. The parameter values of α, β, and γ corresponding to various revisions of the NNTM are given in Table 1 in Section 2.1. The Markov chain and corresponding algorithms presented will enable biologists to calculate the dispersion of key branching properties for a specific energy function. As described with the detailed hairpin dispersion example in the Introduction (Section 1), knowing whether branching properties fall within acceptable dispersion limits is crucial for deducing potential functional insight or hypothesizing other scientific ramifications.

Another key application to RNA modeling of the presented algorithms is the ability to explore the parameter space of possible values for α and β. While the various revisions of the NNTM correspond to specific values for these parameters, in principle any real-valued parameters could be used. Finding values for these parameters that approximate reality remains an open question. Yet, determination of how differences in parameter values change the distribution of NNTM branching properties, such as maximum ladder distance, is crucial. Moreover, parameter space exploration is necessary to identify and further explore the phase transitions that exist. The presented Markov chain and corresponding algorithms expedite such future computational experimentation. Therefore, collectively, the presented algorithm enables exploration that will greatly improve understanding of NNTM-basesd RNA secondary structures and branching properties, as well as identify potential limitations or specific branching structures where the NNTM models do not sufficiently emulate reality. For example, NNTM-based free energy minimization algorithms achieved an accuracy of at least 60% in only 9% of 16S secondary structures analyzed by Doshi et. al. [15].

The algorithm presented here can only sample under an energy function of the form $\alpha d_0 + \beta d_1$, and this does not capture the entirety of the model presented in [16], which considers energy functions of the form $\alpha d_0 + \beta d_1 + \gamma r$. However, the missing term, γr, represents the energy contribution of exterior loop, and the exterior loop contributes less of the total free energy as sequence length increases. Therefore, when interested in sequences of at least moderate length, this algorithm may be able to provide insight, as long as information about the exterior loop is not the specific object of study. Also note that other authors have made similar simplifications with respect to the exterior loop, e.g. [17].

4.2. Possibility of a dynamic programming approach

This sampling problem to calculate the dispersion of NNTM RNA secondary structure and properties utilized Markov chain techniques. However, is it possible to utilize a dynamic programming algorithm? It is straightforward to sample Dyck paths under a uniform probability distribution using dynamic programming techniques. However, it is not clear whether a similar technique could be used for the Gibbs distribution we define here, due to the complexity of the energy function. In particular, large numeric computations may be required to handle the variable k, the number of U steps in a path. While Alonso presents a way to sample from the unweighted distribution $\Pr(k = l) \propto \binom{m}{2l} C_l$ in $O(n)$ time without large computations [37], it is unclear if a similar method may be used for the present application.

4.3. Possibility of an SCFG approach

Stochastic context free grammars (SCFGs) have been widely used in the field of RNA secondary structure prediction, e.g. [38–41]. Most commonly, the probabilities for production rules in an SCFG are determined by training on a set of known secondary structures, often including covariance information from homologous structures. These approaches are not immediately applicable to the problem we study here, as they do not give any insight into the NNTM multiloop energy parameters.

However, some authors have constructed SCFGs based on the NNTM. In particular, Nebel and Scheid [38] construct a SCFG with 29 distinct production rules to mirror the NNTM features. They also present a sampling algorithm allowing for sampling structures of a fixed size using the grammar. However, they do not actually compute probabilities for the production rules that would allow one to sample from a Gibbs distribution (with NNTM energy) and instead rely on training on a set of known structures. Indeed, it is not clear from the paper whether such a set of probabilities must exist.

Even in the case of the simplified model we present in this manuscript, it is not clear how to assign probabilities to production rules in an SCFG so that the probability of obtaining a given structure matches the Gibbs probability under the NNTM.

Even if a suitable SCFG could be formulated, the SCFG approach is not necessarily superior. The sampling algorithm presented by Nebel and Scheid has time complexity $O(n^3)$ and space complexity $O(n^2)$. While the algorithm we present does have large time complexity, it only requires linear space, which may be an advantage for some applications.

4.4. Other biological applications

While many Markov chains with nonuniform stationary distributions have been used for biological applications (e.g. [42–45]), theoretical guarantees on the mixing time are generally not known. Instead, researchers must rely on convergence heuristics, and if fact many introductions to Markov chain Monte Carlo written for biologists explain such heuristic techniques [46–49]. Of course, heuristics can be misleading, and rigorous mixing time guarantees would be significantly preferable. The same techniques used in this work might be used to generate algorithms with rigorous mixing time bounds for other biological problems concerning a nonuniform distribution.

4.5. Independent mathematical research interests

The plane trees examined as a model for RNA secondary structure are of independent mathematical interest. As Catalan objects, they have been studied combinatorially (see, for example, [25,50]), and Markov chains on Catalan objects have received significant attention over the years [33,34,51–53], but with very few results providing tight estimates on the corresponding mixing times; most commonly these are discussed in the language of Dyck paths. Cohen's thesis [33] gives an overview of the known mixing time results for chains on Catalan objects. All of the chains surveyed there have uniform distribution over the Catalan-sized state space as their stationary distribution. Among these, essentially the only known chain with tight bounds (upper and lower bounds differing by a small multiplicative constant) is due to Wilson [34] and gives the relaxation time of $O(n^3)$ for the walk consisting of adjacent transpositions on Dyck paths. In comparison, in [51] the chain using *all* (allowed) transpositions has been shown to have relaxation time of $O(n^2)$, and further conjectured to have $O(n)$ as the relaxation time, in analogy with the random transposition shuffle of n cards.

Judging from the lack of progress on several of these chains, it is evident that determining mixing or relaxation time for these chains is typically a challenging problem, even in the case where the stationary distribution is uniform.

In the current work, the RNA secondary structure modeling naturally leads to a state space on Catalan objects with a nonuniform distribution, making the corresponding mixing time analysis even more challenging. Another example where mixing times are estimated for Markov chains on

Catalan objects with nonuniform stationary distribution is the work of Martin and Randall [30], which examines a Gibbs distribution on Dyck paths weighted by the number of returns to the x-axis.

4.6. Limitations and Future Directions

While the mixing time proved here is polynomial, it is almost certainly too large to allow for any practical computational sampling experiments. However, we conjecture the actual mixing time to be much smaller, and future work may provide a better bound. Even without additional theoretical results, interesting work is likely possible using the algorithm we present and heuristic methods for evaluating Markov chain mixing. See [54, Ch. 8] for a discussion of heuristic methods for monitoring Markov chain convergence.

The results of this study provide an important mathematical foundation for examining dispersion of RNA secondary structures and branching properties using a Markov chain. However, more work is necessary to optimize the developed computational application for incorporation into the software utilized by biologists that study RNA. Example questions that strongly compel further investigation include:

1. Can the mixing time bound in our main result be improved?
2. Is there a rapidly mixing chain, with the same stationary distribution studied here, whose transitions correspond naturally to moves on the set plane trees? Mixing time bounds on the chain of matching exchange moves, as defined in [55], would be especially interesting, as such a chain may relate to RNA folding kinetics.
3. Is there a rapidly mixing chain converging to the Gibbs distribution using the full energy function for the utilized NNTM model [16]? The chain presented here uses only the parameters α and β, setting $\gamma = 0$.
4. Is there a stochastic context free grammar which generate secondary structures (in our simplified model or using the full NNTM) according to a Gibbs distribution with NNTM energy?

Author Contributions: Conceptualization, A.K., K.P.; Investigation, A.K., K.P.; Formal Analysis, A.K., K.P.; Funding Acquisition, A.K., K.P., C.M.; Methodology, A.K., K.P.; Supervision, T.P., C.M.; Validation, A.K., K.P., T.P., C.M.; Writing original draft, A.K., K.P., C.M.; Writing - review & editing, A.K., T.P., C.M. All authors have read and approved the published version of the manuscript.

Funding: Funding provided by National Science Foundation Graduate Research Fellowship Program to A.K.; Georgia Institute of Technology President's Undergraduate Research Award to K.P.; National Science Foundation grant DMS-1811935 to P.T.; National Science Foundation CAREER award (1944247), National Institute of Health grant R21CA232249, and Alzheimer's Association Research Grant to C.M.

Acknowledgments: The authors would like to thank Christine Heitsch for generous consultation. The authors would also like to think the reviewers for their insightful comments.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

RNA	ribonucleic acid
NNTM	Nearest Neighbor Thermodynamic Model
SCFG	stochastic context free grammar

1. Doudna, J.A. Structural genomics of RNA. *Nature Structural Biology* **2000**, *7*, 954–956.
2. Tinoco Jr, I.; Bustamante, C. How RNA folds. *Journal of molecular biology* **1999**, *293*, 271–281.
3. Massire, C.; Westhof, E. MANIP: an interactive tool for modelling RNA. *Journal of Molecular Graphics and Modelling* **1998**, *16*, 197 – 205. doi:https://doi.org/10.1016/S1093-3263(98)80004-1.
4. Seetin, M.G.; Mathews, D.H. Automated RNA tertiary structure prediction from secondary structure and low-resolution restraints. *Journal of Computational Chemistry* **2011**, *32*, 2232–2244. doi:10.1002/jcc.21806.

5. Zhao, Y.; Gong, Z.; Xiao, Y. Improvements of the Hierarchical Approach for Predicting RNA Tertiary Structure. *Journal of Biomolecular Structure and Dynamics* **2011**, *28*, 815–826. PMID: 21294592, doi:10.1080/07391102.2011.10508609.
6. Zhao, Y.; Huang, Y.; Gong, Z.; Wang, Y.; Man, J.; Xiao, Y. Automated and fast building of three-dimensional RNA structures. *Scientific Reports* **2012**, *2*, 734.
7. Borodavka, A.; Singaram, S.W.; Stockley, P.G.; Gelbart, W.M.; Ben-Shaul, A.; Tuma, R. Sizes of long RNA molecules are determined by the branching patterns of their secondary structures. *Biophysical Journal* **2016**, *111*, 2077–2085.
8. Jaeger, J.A.; Turner, D.H.; Zuker, M. Improved predictions of secondary structures for RNA. *Proceedings of the National Academy of Sciences* **1989**, *86*, 7706–7710.
9. Mathews, D.H.; Sabina, J.; Zuker, M.; Turner, D.H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology* **1999**, *288*, 911–940.
10. Mathews, D.H.; Disney, M.D.; Childs, J.L.; Schroeder, S.J.; Zuker, M.; Turner, D.H. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences* **2004**, *101*, 7287–7292.
11. Ding, Y.; Chan, C.Y.; Lawrence, C.E. Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Research* **2004**, *32*, W135–W141. doi:10.1093/nar/gkh449.
12. Hofacker, I.L.; Fontana, W.; Stadler, P.F.; Bonhoeffer, L.S.; Tacker, M.; Schuster, P. Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie/Chemical Monthly* **1994**, *125*, 167–188.
13. Mathuriya, A.; Bader, D.A.; Heitsch, C.E.; Harvey, S.C. GTfold: a scalable multicore code for RNA secondary structure prediction. Proceedings of the 2009 ACM symposium on Applied Computing, 2009, pp. 981–988.
14. Turner, D.H.; Mathews, D.H. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Research* **2009**, *38*, D280–D282. doi:10.1093/nar/gkp892.
15. Doshi, K.J.; Cannone, J.J.; Cobaugh, C.W.; Gutell, R.R. Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics* **2004**, *5*, 105.
16. Hower, V.; Heitsch, C.E. Parametric analysis of RNA branching configurations. *Bulletin of Mathematical Biology* **2011**, *73*, 754–776.
17. Bakhtin, Y.; Heitsch, C.E. Large deviations for random trees and the branching of RNA secondary structures. *Bulletin of Mathematical Biology* **2009**, *71*, 84–106.
18. Heitsch, C.; Poznanović, S. Combinatorial insights into RNA secondary structure. In *Discrete and topological models in molecular biology*; Springer, 2014; pp. 145–166.
19. Lorenz, R.; Bernhart, S.H.; Zu Siederdissen, C.H.; Tafer, H.; Flamm, C.; Stadler, P.F.; Hofacker, I.L. ViennaRNA Package 2.0. *Algorithms for molecular biology* **2011**, *6*, 26.
20. Donaghey, R.; Shapiro, L.W. Motzkin numbers. *J. Combinatorial Theory Ser. A* **1977**, *23*, 291–301. doi:10.1016/0097-3165(77)90020-6.
21. Bernhart, F.R. Catalan, Motzkin, and Riordan numbers. *Discrete Math.* **1999**, *204*, 73–112. doi:10.1016/S0012-365X(99)00054-0.
22. Eu, S.P.; Fu, T.S.; Hou, J.T.; Hsu, T.W. Standard Young tableaux and colored Motzkin paths. *J. Combin. Theory Ser. A* **2013**, *120*, 1786–1803. doi:10.1016/j.jcta.2013.06.007.
23. Baril, J.L.; Kirgizov, S.; Petrossian, A. Motzkin paths with a restricted first return decomposition. *Integers* **2019**, *19*, Paper No. A46, 19.
24. Fang, W. A partial order on Motzkin paths. *Discrete Math.* **2020**, *343*, 111802, 9. doi:10.1016/j.disc.2019.111802.
25. Stanley, R.P. *Enumerative Combinatorics: Volume 1*, 2nd ed.; Cambridge University Press: New York, NY, USA, 2011.
26. Deutsch, E.; Shapiro, L.W. A bijection between ordered trees and 2-Motzkin paths and its many consequences. *Discrete Mathematics* **2002**, *256*, 655 – 670. LaCIM 2000 Conference on Combinatorics, Computer Science and Applications.
27. Madras, N.; Randall, D. Markov chain decomposition for convergence rate analysis. *Ann. Appl. Probab.* **2002**, *12*, 581–606. doi:10.1214/aoap/1026915617.

28. Randall, D. Rapidly Mixing Markov Chains with Applications in Computer Science and Physics. *Computing in Science and Engg.* **2006**, *8*, 30–41. doi:10.1109/MCSE.2006.30.
29. Luby, M.; Randall, D.; Sinclair, A. Markov Chain Algorithms for Planar Lattice Structures. *SIAM Journal on Computing* **2001**, *31*, 167–192. doi:10.1137/S0097539799360355.
30. Martin, R.; Randall, D. Sampling adsorbing staircase walks using a new Markov chain decomposition method. *Proceedings 41st Annual Symposium on Foundations of Computer Science* **2000**, pp. 492–502.
31. Hermon, J.; Salez, J. Modified log-Sobolev inequalities for strong-Rayleigh measures, 2019, [arXiv:math.PR/1902.02775].
32. Jerrum, M.; Son, J.B.; Tetali, P.; Vigoda, E. Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains. *Ann. Appl. Probab.* **2004**, *14*, 1741–1765. doi:10.1214/105051604000000639.
33. Cohen, E. Problems in catalan mixing and matchings in regular hypergraphs. PhD thesis, Georgia Institute of Technology, 2016.
34. Wilson, D.B. Mixing times of lozenge tiling and card shuffling Markov chains. *Ann. Appl. Probab.* **2004**, *14*, 274–325. doi:10.1214/aoap/1075828054.
35. Yoffe, A.M.; Prinsen, P.; Gopal, A.; Knobler, C.M.; Gelbart, W.M.; Ben-Shaul, A. Predicting the sizes of large RNA molecules. *Proceedings of the National Academy of Sciences* **2008**, *105*, 16153–16158.
36. Andronescu, M.; Bereg, V.; Hoos, H.H.; Condon, A. RNA STRAND: the RNA secondary structure and statistical analysis database. *BMC bioinformatics* **2008**, *9*, 340.
37. Alonso, L. Uniform generation of a Motzkin word. *Theoretical Computer Science* **1994**, *134*, 529 – 536. doi:https://doi.org/10.1016/0304-3975(94)00086-7.
38. Nebel, M.E.; Scheid, A. Evaluation of a sophisticated SCFG design for RNA secondary structure prediction. *Theory in Biosciences* **2011**, *130*, 313–336.
39. Rivas, E.; Eddy, S.R. Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. *Bioinformatics* **2000**, *16*, 583–605.
40. Rivas, E.; Eddy, S.R. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *Journal of molecular biology* **1999**, *285*, 2053–2068.
41. Knudsen, B.; Hein, J. Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic acids research* **2003**, *31*, 3423–3428.
42. Huelsenbeck, J.P.; Larget, B.; Alfaro, M.E. Bayesian Phylogenetic Model Selection Using Reversible Jump Markov Chain Monte Carlo. *Molecular Biology and Evolution* **2004**, *21*, 1123–1133. doi:10.1093/molbev/msh123.
43. Kim, S.; Li, H.; Dougherty, E.R.; Cao, N.; Chen, Y.; Bittner, M.; Suh, E.B. Can Markov chain models mimic biological regulation? *Journal of Biological Systems* **2002**, *10*, 337–357.
44. Lusseau, D. Effects of tour boats on the behavior of bottlenose dolphins: using Markov chains to model anthropogenic impacts. *Conservation Biology* **2003**, *17*, 1785–1793.
45. Said, M.R.; Oppenheim, A.V.; Lauffenburger, D.A. Modeling cellular signal processing using interacting Markov chains. 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)., 2003, Vol. 6, pp. VI–41.
46. Gelman, A.; Rubin, D.B. Markov chain Monte Carlo methods in biostatistics. *Statistical Methods in Medical Research* **1996**, *5*, 339–355.
47. Hamra, G.; MacLehose, R.; Richardson, D. Markov chain Monte Carlo: an introduction for epidemiologists. *International Journal of Epidemiology* **2013**, *42*, 627–634.
48. Nascimento, F.F.; dos Reis, M.; Yang, Z. A biologist's guide to Bayesian phylogenetic analysis. *Nature Ecology & Evolution* **2017**, *1*, 1446–1454.
49. Van Ravenzwaaij, D.; Cassey, P.; Brown, S.D. A simple introduction to Markov Chain Monte–Carlo sampling. *Psychonomic Bulletin & Review* **2018**, *25*, 143–154.
50. Dershowitz, N.; Zaks, S. Ordered trees and non-crossing partitions. *Discrete Mathematics* **1986**, *62*, 215–218.
51. Cohen, E.; Tetali, P.; Yeliussizov, D. Lattice path matroids: negative correlation and fast mixing. *arXiv preprint arXiv:1505.06710* **2015**.
52. McShine, L.; Tetali, P. On the mixing time of the triangulation walk and other Catalan structures. In *Randomization Methods in Algorithm Design (Princeton, NJ, 1997)*; Amer. Math. Soc., Providence, RI, 1999; Vol. 43, *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pp. 147–160.

649 53. Sohoni, M. Rapid mixing of some linear matroids and other combinatorial objects. *Graphs Combin.* **1999**,
650 *15*, 93–107. doi:10.1007/s003730050032.
651 54. Robert, C.P.; Casella, G.; Casella, G. *Introducing Monte Carlo methods with R*; Vol. 18, Springer, 2010.
652 55. Heitsch, C.E.; Tetali, P. Meander graphs. In *23rd International Conference on Formal Power Series and Algebraic*
653 *Combinatorics (FPSAC 2011)*; Discrete Math. Theor. Comput. Sci. Proc., AO, Assoc. Discrete Math. Theor.
654 Comput. Sci., Nancy, 2011; pp. 469–480.