Loops in canonical RNA pseudoknot structures

Markus E. Nebel, Christian M. Reidys⋆ and Rita R. Wang

TU Kaiserslautern

67663 Kaiserslautern

Germany

nebel@cs.uni-kl.de

Center for Combinatorics, LPMC-TJKLC

Nankai University

Tianjin 300071

P.R. China

Phone: 86-22-2350-6800

Fax: 86-22-2350-9272

⋆ duck@santafe.edu

wangrui@cfc.nankai.edu.cn

Abstract

In this paper we compute the limit distributions of the numbers of hairpin-loops, interior-loops and bulges in k-noncrossing RNA structures. The latter are coarse grained RNA structures allowing for cross-serial interactions, subject to the constraint that there are at most $k-1$ mutually crossing arcs in the diagram representation of the molecule. We prove central limit theorems by means of studying the corresponding bivariate generating functions. These generating functions are obtained by symbolic inflation of $\mathsf{lv}_k^{\mathsf{5}}$-shapes introduced by Reidys and Wang (2009).

Keywords: k-noncrossing τ-canonical RNA structure, bivariate generating function, singularity analysis, central limit theorem, loops

1. Introduction

An RNA molecule is a sequence of the four nucleotides **A**, **G**, **U**, **C** together with the Watson-Crick (**A-U**, **G-C**) and **U-G** base pairing rules. The sequence of bases is called the primary structure of the RNA molecule. Two bases in the primary structure which are not adjacent may form hydrogen bonds following the Watson-Crick base pairing rules. Three decades ago Waterman *et al.* [Kleitman 1970, Nussinov et al. 1978, Waterman 1978] analyzed RNA secondary structures. Secondary structures are coarse grained RNA contact structures. They can be represented as diagrams and planar graphs, see Fig. 1. Diagrams are labeled graphs over the vertex set $[n] = \{1, \dots, n\}$ with vertex degrees ≤ 1, represented by drawing its vertices on a horizontal line and its arcs (i, j) $(i < j)$, in the upper half-plane, see Fig. 1 and Fig. 2. Here, vertices and arcs correspond to the nucleotides **A**, **G**, **U**, **C** and Watson-Crick (**A-U**, **G-C**) and (**U-G**) base pairs, respectively. In a diagram two arcs (i_1, j_1) and (i_2, j_2) are called crossing if $i_1 < i_2 < j_1 < j_2$ holds. Accordingly, a k-crossing is a sequence of arcs $(i_1, j_1), \dots, (i_k, j_k)$ such that

$i_1 < i_2 < \cdots < i_k < j_1 < j_2 < \cdots < j_k$, see Fig. 2. We call diagrams containing at most $(k-1)$-crossings, k-noncrossing diagrams (k-noncrossing partial matchings). The length of an arc (i,j) is given by $j-i$, characterizing the minimal length of a hairpin loop. A stack of length τ is a maximal sequence of "parallel" arcs of the form

$$((i,j),(i+1,j-1),\dots,(i+(\tau-1),j-(\tau-1))), \tag{1.1}$$

and we denote it by $S^{\tau}_{i,j}$. We call an arc of length one a 1-arc. A k-noncrossing, τ-canonical RNA structure is a k-noncrossing diagram without 1-arcs, having a minimum stack-size of τ, see Fig. 2. Let $\mathcal{T}_{k,\tau}(n)$ denote the set of k-noncrossing, τ-canonical RNA structures of length n and let $\mathsf{T}_{k,\tau}(n)$ denote their number.

We next introduce the following structural elements of k-noncrossing, τ-canonical RNA structures, see Fig. 3 and Fig. 4.

Let $[i,j]$ denote an interval, i.e. a sequence of consecutive isolated vertices $(i,i+1,\dots,j-1,j)$. We consider, see Fig. 4

(1) a *hairpin-loop* is a pair

$$((i,j),[i+1,j-1]).$$

(2) an *interior-loop* is a sequence

$$((i_1, j_1), [i_1 + 1, i_2 - 1], (i_2, j_2), [j_2 + 1, j_1 - 1]),$$

where (i_2, j_2) is nested in (i_1, j_1).

(3) a *bulge* is a sequence

$$((i_1, j_1), [i_1 + 1, i_2 - 1], (i_2, j_1 - 1)) \quad \text{or} \quad ((i_1, j_1), (i_1 + 1, j_2), [j_2 + 1, j_1 - 1]).$$

(4) a *stem* of size s is a sequence of stacks

$$\left(S^{\tau_1}_{i_1,j_1}, S^{\tau_2}_{i_2,j_2}, \dots, S^{\tau_s}_{i_s,j_s}\right)$$

where the stack $S^{\tau_m}_{i_m,j_m}$ is nested in $S^{\tau_{m-1}}_{i_{m-1},j_{m-1}}$ such that any arc nested in $S^{\tau_{m-1}}_{i_{m-1},j_{m-1}}$ is either contained or nested in $S^{\tau_m}_{i_m,j_m}$, for $2 \le m \le s$.

In this paper, we derive the limit distributions of the numbers of hairpin-loops, interior-loops and bulges in k-noncrossing τ-canonical RNA structures,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\mathbb{X} - \mu_{k,\tau,\mathbb{X}} n}{\sqrt{n\, \sigma^2_{k,\tau,\mathbb{X}}}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt, \tag{1.2}$$

where $\mathbb{X}$ is a random variable counting the number of hairpin-loops, interior-loops or bulges of k-noncrossing, τ-canonical structures of length n, see Fig. 5.

2. Preliminaries

Let $f_k(n,\ell)$ denote the number of k-noncrossing diagrams on n vertices having exactly ℓ isolated vertices. A diagram without isolated points is called a matching. The exponential generating function of k-noncrossing matchings satisfies the following identity [Chen et al. 2007, Grabiner and Magyar 1993, Jin et al. 2008a]

$$(2.1)\qquad \mathbf{H}_k(z)=\sum_{n\geq 0} f_k(2n,0)\cdot\frac{z^{2n}}{(2n)!}=\det[I_{i-j}(2z)-I_{i+j}(2z)]|_{i,j=1}^{k-1}$$

where $I_r(2z)=\sum_{j\geq 0}\frac{z^{2j+r}}{j!(j+r)!}$ is the hyperbolic Bessel function of the first kind of order r. Eq. (2.1) allows us to conclude that the ordinary generating function

$$\mathbf{F}_k(z)=\sum_{n\geq 0} f_k(2n,0)z^n$$

is D-finite [Stanley 1980]. This follows from the fact that $I_r(2z)$ is D-finite and D-finite power series form an algebra [Stanley 1980]. Consequently, there exists some $e\in\mathbb{N}$ such that

$$(2.2)\qquad q_{0,k}(z)\frac{d^e}{dz^e}\mathbf{F}_k(z)+q_{1,k}(z)\frac{d^{e-1}}{dz^{e-1}}\mathbf{F}_k(z)+\cdots+q_{e,k}(z)\mathbf{F}_k(z)=0,$$

where $q_{j,k}(z)$ are polynomials and $q_{0,k}(z)\neq 0$. The ordinary differential equations (ODE) for $\mathbf{F}_k(z)$, where $2\leq k\leq 7$ are obtained by the MAPLE package `GFUN` from the exact data of $f_k(2n,0)$. They are verified by first deriving the corresponding P-recursions [Stanley 1980] for $f_k(2n,0)$ second transforming these P-recursions into

P-recursions of $f_k(2n,0)/(2n)!$ and third deriving the corresponding ODEs for $\mathbf{H}_k(z)$ and verifying that the RHS of eq. (2.1) is a solution. The key point is that any singularity of $\mathbf{F}_k(z)$ is contained in the set of roots of $q_{0,k}(z)$ [Stanley 1980], which we denote by R_k. For $2 \leq k \leq 7$, we give the polynomials $q_{0,k}(z)$ and their roots in Table 1.

In [Jin et al. 2008b] we showed that for arbitrary k

$$f_k(2n,0) \sim \widetilde{c}_k \, n^{-((k-1)^2+(k-1)/2)} \, (2(k-1))^{2n}, \qquad \widetilde{c}_k > 0 \tag{2.3}$$

in accordance with the fact that $\mathbf{F}_k(z)$ has the unique dominant singularity ρ_k^2, where $\rho_k = 1/(2k-2)$.

We next introduce a central limit theorem due to Bender [Bender 1973]. It is proved by analyzing the characteristic function by the Lévy-Cramér Theorem (Theorem IX.4 in [Flajolet and Sedgewick 2009]).

Theorem 1. *Suppose we are given the bivariate generating function*

$$f(z,u) = \sum_{n,t \geq 0} f(n,t)\, z^n \, u^t, \tag{2.4}$$

where $f(n,t) \geq 0$ *and* $f(n) = \sum_t f(n,t)$. *Let* $\mathbb{X}_n$ *be a r.v. such that* $\mathbb{P}(\mathbb{X}_n = t) = f(n,t)/f(n)$. *Suppose*

$$[z^n] f(z, e^s) \sim c(s)\, n^{\alpha}\, \gamma(s)^{-n} \tag{2.5}$$

uniformly in s in a neighborhood of 0*, where $c(s)$ is continuous and nonzero near* 0*, α is a constant, and $\gamma(s)$ is analytic near* 0*. Then there exists a pair (μ, σ) such that the normalized random variable*

$$\mathbb{X}_n^* = \frac{\mathbb{X}_n - \mu\, n}{\sqrt{n\, \sigma^2}} \tag{2.6}$$

has asymptotically normal distribution with parameter $(0, 1)$*. That is we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\mathbb{X}_n^* < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}c^2} dc \tag{2.7}$$

where μ and σ^2 are given by

$$\mu = -\frac{\gamma'(0)}{\gamma(0)} \quad \textit{and} \quad \sigma^2 = \left(\frac{\gamma'(0)}{\gamma(0)}\right)^2 - \frac{\gamma''(0)}{\gamma(0)}. \tag{2.8}$$

The crucial points for applying Theorem 1 are (a) eq. (2.5)

$$[z^n] f(z, e^s) \sim c(s)\, n^{\alpha}\, \gamma(s)^{-n},$$

uniformly in s in a neighborhood of 0, where $c(s)$ is continuous and nonzero near 0 and α is a constant and (b) the analyticity of $\gamma(s)$ in s near 0. In the following, we have generating functions of the form $\mathbf{F}_k(\psi(z, s))$. In this situation, Theorem 2 below guarantees under specific conditions

$$[z^n]\mathbf{F}_k(\psi(z, s)) \sim A(s)\, n^{-((k-1)^2+(k-1)/2)} \left(\frac{1}{\gamma(s)}\right)^n, \quad A(s) \text{ continuous},$$

for $2 \leq k \leq 7$. The analyticity of $\gamma(s)$ is guaranteed by the analytic implicit function theorem [Flajolet and Sedgewick 2009].

Theorem 2. [Jin and Reidys 2010] *Suppose* $2 \leq k \leq 7$. *Let* $\psi(z,s)$ *be an analytic function in a domain*

$$\mathcal{D} = \{(z,s) | |z| \leq r, |s| < \epsilon\} \tag{2.9}$$

such that $\psi(0,s) = 0$. *In addition suppose* $\gamma(s)$ *is the unique dominant singularity of* $\mathbf{F}_k(\psi(z,s))$ *and analytic solution of* $\psi(\gamma(s),s) = \rho_k^2$, $|\gamma(s)| \leq r$, $\partial_z \psi(\gamma(s),s) \neq 0$ *for* $|s| < \epsilon$. *Then* $\mathbf{F}_k(\psi(z,s))$ *has a singular expansion and*

$$[z^n]\mathbf{F}_k(\psi(z,s)) \sim A(s)\, n^{-((k-1)^2+(k-1)/2)} \left(\frac{1}{\gamma(s)}\right)^n \quad \text{for some continuous } A(s) \in \mathbb{C}, \tag{2.10}$$

uniformly in s *contained in a small neighborhood of* 0.

To keep the paper selfcontained we give a direct proof of Theorem 2 in Section 5. This avoids calling upon generic results, such as the uniformity Lemma of singularity analysis [Flajolet and Sedgewick 2009].

3. The generating function

In this section we compute the bivariate generating functions of hairpin-loops, interior-loops and bulges. Let $h_{k,\tau}(n,t)$, $i_{k,\tau}(n,t)$ and $b_{k,\tau}(n,t)$ denote the numbers of k-noncrossing, τ-canonical RNA structures of length n with t hairpin-loops, interior-loops and bulges. We set

$$\mathbf{H}_{k,\tau}(z,u_1) = \sum_{n\geq 0}\sum_{t\geq 0} h_{k,\tau}(n,t) z^n\, u_1^t, \tag{3.1}$$

$$\mathbf{I}_{k,\tau}(z,u_2) = \sum_{n\geq 0}\sum_{t\geq 0} i_{k,\tau}(n,t) z^n\, u_2^t, \tag{3.2}$$

$$\mathbf{B}_{k,\tau}(z,u_3) = \sum_{n\geq 0}\sum_{t\geq 0} b_{k,\tau}(n,t) z^n\, u_3^t. \tag{3.3}$$

In order to derive the above generating functions we use symbolic enumeration [Flajolet and Sedgewick 2009]. A combinatorial class is a set of finite size with the definition of size function of its elements, whose elements are all finite size and the number of certain size elements is finite. Suppose $\mathcal{C}$ be a combinatorial class and $c \in \mathcal{C}$. We denote the size of c by $|c|$. There are two special combinatorial classes $\mathcal{E}$ and $\mathcal{Z}$ which respectively contains only an element of size 0 and an element of size 1. The subset of $\mathcal{C}$ which contains all the elements of size n in $\mathcal{C}$ is denoted by $\mathcal{C}_n$. Then the generating function of a combinatorial class $\mathcal{C}$ is

$$\mathbf{C}(z) = \sum_{c\in\mathcal{C}} z^{|c|} = \sum_{n\geq 0} C_n\, z^n, \tag{3.4}$$

where $\mathcal{C}_n \subset \mathcal{C}$ and $C_n = |\mathcal{C}_n|$. In particular the generating functions of $\mathcal{E}$ and $\mathcal{Z}$ are given by $\mathbf{E}(z) = 1$ and $\mathbf{Z}(z) = z$. For any two combinatorial classes $\mathcal{C}$, $\mathcal{D}$, we have the following operations:

- $\mathcal{C} + \mathcal{D} := \mathcal{C} \cup \mathcal{D}$, if $\mathcal{C} \cap \mathcal{D} = \varnothing$
- $\mathcal{C} \times \mathcal{D} := \{(c,d) | c \in \mathcal{C}, d \in \mathcal{D}\}$ and $\mathcal{C}^m := \prod_{i=1}^m \mathcal{C}$
- $\textsc{Seq}(\mathcal{C}) = \mathcal{E} + \mathcal{C} + \mathcal{C}^2 + \cdots$.

We have the following relations between the operations of combinatorial classes and the operations of their generating functions:

$$\mathcal{A} = \mathcal{C} + \mathcal{D} \quad \Rightarrow \quad \mathbf{A}(z) = \mathbf{C}(z) + \mathbf{D}(z) \tag{3.5}$$

$$\mathcal{A} = \mathcal{C} \times \mathcal{D} \quad \Rightarrow \quad \mathbf{A}(z) = \mathbf{C}(z) \cdot \mathbf{D}(z) \tag{3.6}$$

$$\mathcal{A} = \textsc{Seq}(\mathcal{C}) \quad \Rightarrow \quad \mathbf{A}(z) = (1 - \mathbf{C}(z))^{-1}, \tag{3.7}$$

where $\mathbf{A}(z)$, $\mathbf{C}(z)$, $\mathbf{D}(z)$ is the generating function of $\mathcal{A}$, $\mathcal{C}$ and $\mathcal{D}$.

Given a k-noncrossing, τ-canonical RNA structure δ, its lv_k^5-shape, $\mathsf{lv}_k^5(\delta)$ [Reidys and Wang 2009], is obtained by first removing all isolated vertices and second collapsing any stack into a single arc, see Fig.6. By construction, lv_k^5-shapes do not preserve stack-lengths, interior loops and unpaired regions. In the following, we shall refer to lv_k^5-shape simply as shape. Let $\mathcal{T}_{k,\tau}$ denote the set of k-noncrossing, τ-canonical structures and $\mathcal{I}_k$ the set of all k-noncrossing shapes and $\mathcal{I}_k(m)$ those having m 1-arcs, see Figure 6. Each stem of a k-noncrossing, τ-canonical RNA structure is mapped into an arc in its corresponding shape and all hairpin-loops are mapped into 1-arcs. Therefore we have the surjective map,

$$\varphi : \ \mathcal{T}_{k,\tau} \to \mathcal{I}_k. \tag{3.8}$$

Indeed, for a given shape γ in $\mathcal{I}_k$, we can derive a k-noncrossing, τ-canonical structure having arc-length≥ 2, we can add arcs to each arc contained in the shape such that every resulting stack has τ arcs and insert one isolated vertex in each 1-arc. Let $\mathcal{I}_k(s,m)$ and $i_k(s,m)$ denote the set and number of the Iv_k^5-shapes of length $2s$ with m 1-arcs and

$$\mathbf{I}_k(x,y) = \sum_{s\geq 0}\sum_{m=0}^{s} i_k(s,m)x^s y^m \tag{3.9}$$

be the bivariate generating function. Furthermore, let $\mathcal{I}_k(m)$ denote the set of shapes γ having m 1-arcs. Let k,s,m be natural numbers where $k\geq 2$, then the generating function $\mathbf{I}_k(x,y)$ [Reidys and Wang 2009] is given by

$$\mathbf{I}_k(x,y) = \frac{1+x}{1+2x-xy}\mathbf{F}_k\left(\frac{x(1+x)}{(1+2x-xy)^2}\right) \tag{3.10}$$

Theorem 3. *Suppose $k,\tau\in\mathbb{N}, k\geq 2, \tau\geq 1$. Then*

$$\begin{aligned}\mathbf{H}_{k,\tau}(z,u_1) =& \frac{(1-z)(1-z^2+z^{2\tau})}{(1-z)^2(1-z^2+z^{2\tau})+z^{2\tau}-z^{2\tau+1}u_1}\\ &\mathbf{F}_k\left(\frac{z^{2\tau}(1-z)^2(1-z^2+z^{2\tau})}{((1-z)^2(1-z^2+z^{2\tau})+z^{2\tau}-z^{2\tau+1}u_1)^2}\right),\end{aligned} \tag{3.11}$$

$$\mathbf{I}_{k,\tau}(z,u_2) = \frac{(1-z^2)(1-z)^2 - u_2 z^{2\tau+2} + (2z^2-2z+1)z^{2\tau}}{(1-z)\left((1-z^2)(1-z)^2 - u_2 z^{2\tau+2} + (2z^2-3z+2)z^{2\tau}\right)} \mathbf{F}_k\left(\frac{z^{2\tau}\left((1-z^2)(1-z)^2 - u_2 z^{2\tau+2} + (2z^2-2z+1)z^{2\tau}\right)}{\left((1-z^2)(1-z)^2 - u_2 z^{2\tau+2} + (2z^2-3z+2)z^{2\tau}\right)^2}\right), \tag{3.12}$$

$$\mathbf{B}_{k,\tau}(z,u_3) = \frac{(1-z^2)(1-z) - 2u_3 z^{2\tau+1} + (z+1)z^{2\tau}}{(1-z)\left((1-z^2)(1-z) - 2u_3 z^{2\tau+1} + (z+2)z^{2\tau}\right)} \mathbf{F}_k\left(\frac{z^{2\tau}\left((1-z^2)(1-z) - 2u_3 z^{2\tau+1} + (z+1)z^{2\tau}\right)}{(1-z)\left((1-z^2)(1-z) - 2u_3 z^{2\tau+1} + (z+2)z^{2\tau}\right)^2}\right). \tag{3.13}$$

Proof. We prove the theorem via symbolic enumeration representing a k-noncrossing, τ-canonical structure as the inflation of a shape, γ. Since a structure inflated from $\gamma \in \mathcal{I}_k(s,m)$ has exactly s stems, $(2s+1)$ (possibly empty) intervals of isolated vertices and m nonempty such intervals we rewrite the generating functions as

$$\begin{aligned}
\mathbf{H}_{k,\tau}(z,u_1) &= \sum_{m\geq 0}\sum_{\gamma\in\mathcal{I}_k(m)} \mathbf{T}_\gamma(z,u_1,1,1),\\
\mathbf{I}_{k,\tau}(z,u_2) &= \sum_{m\geq 0}\sum_{\gamma\in\mathcal{I}_k(m)} \mathbf{T}_\gamma(z,1,u_2,1),\\
\mathbf{B}_{k,\tau}(z,u_3) &= \sum_{m\geq 0}\sum_{\gamma\in\mathcal{I}_k(m)} \mathbf{T}_\gamma(z,1,1,u_3).
\end{aligned}$$

where $\mathbf{T}_\gamma(z,u_1,u_2,u_3)$ is the generating function of all k-noncrossing, τ-canonical structures with shape γ and $u_i (i=1,2,3)$ are variables associated with the number of hairpin-loops, interior-loops and bulges. In order to compute the latter we consider the inflation process: we inflate $\gamma \in \mathcal{I}_k(m)$ having s arcs, where $s \geq m$, to a structure as follows:

- we inflate each arc of the shape to a stem of stacks of minimum size τ. Any isolated vertices inserted during this first inflation step separate the added stacks.
- we insert isolated vertices at the remaining $(2s+1)$ positions.

We inflate any shape-arc to a stack of size at least τ and subsequently add additional stacks. The latter are called induced stacks and have to be separated by means of inserting isolated vertices, see Fig. 7. Note that during this first inflation step no intervals of isolated vertices, other than those necessary for separating the nested stacks are inserted. After the first inflation step we proceed inflating further by inserting only additional isolated vertices at the remaining $(2s+1)$ positions in which such insertions are possible. For each 1-arc at least one such isolated vertex is necessarily inserted, see Fig. 8.

We proceed by expressing the above two inflations in terms of symbolic enumeration. For this purpose we introduce the combinatorial classes $\mathcal{T}_\gamma$ (k-noncrossing, τ-canonical RNA structures with shape γ), $\mathcal{M}$ (stems), $\mathcal{K}^\tau$ (stacks), $\mathcal{N}^\tau$ (induced stacks), $\mathcal{L}$ (isolated vertices), $\mathcal{R}$ (arcs) and $\mathcal{Z}$ (vertices), where $\mathbf{Z}(z) = z$ and $\mathbf{R}(z) = z^2$. Let μ_1, μ_2 and μ_3 be the labels for hairpin-loops, interior-loops and bulges, respectively. Then

$$
\begin{aligned}
&(3.14) \qquad & \mathcal{T}_\gamma &= (\mathcal{M})^s \times \mathcal{L}^{2s+1-m} \times ([\mathcal{Z}\times\mathcal{L}]_{\mu_1})^m\,, \\
&(3.15) \qquad & \mathcal{M} &= \mathcal{K}^\tau \times \textsc{Seq}\,(\mathcal{N}^\tau)\,, \\
&(3.16) \qquad & \mathcal{N}^\tau &= \mathcal{K}^\tau \times \big([\mathcal{Z}\times\mathcal{L}]_{\mu_3} + [\mathcal{Z}\times\mathcal{L}]_{\mu_3} + [(\mathcal{Z}\times\mathcal{L})^2]_{\mu_2}\big)\,, \\
&(3.17) \qquad & \mathcal{K}^\tau &= \mathcal{R}^\tau \times \textsc{Seq}\,(\mathcal{R})\,, \\
&(3.18) \qquad & \mathcal{L} &= \textsc{Seq}\,(\mathcal{Z})\,.
\end{aligned}
$$

and consequently, translating the above relations into generating functions the generating function $\mathbf{T}_\gamma(z,u_1,u_2,u_3)$ is given by

$$
\begin{aligned}
&\left(\frac{\frac{z^{2\tau}}{1-z^2}}{1-\frac{z^{2\tau}}{1-z^2}\left(2\frac{u_3\,z}{1-z}+u_2\left(\frac{z}{1-z}\right)^2\right)}\right)^s \left(\frac{1}{1-z}\right)^{2s+1-m}\left(\frac{u_1\,z}{1-z}\right)^m \\
&= (1-z)^{-1}\left(\frac{z^{2\tau}}{(1-z^2)(1-z)^2-(2u_3\,z\,(1-z)+u_2\,z^2)z^{2\tau}}\right)^s (u_1\,z)^m,
\end{aligned}
$$

where the indeterminants u_i $(i=1,2,3)$ correspond to the labels μ_i, i.e. the occurrences of hairpin-loops, interior-loops and bulges. Accordingly, for any two shapes $\gamma_1,\gamma_2 \in \mathcal{I}_k(m)$ having s arcs, we have

$$(3.19) \qquad \mathbf{T}_{\gamma_1}(z,u_1,u_2,u_3) = \mathbf{T}_{\gamma_2}(z,u_1,u_2,u_3).$$

We set

$$(3.20) \qquad \eta(u_2,u_3) = \frac{z^{2\tau}}{(1-z^2)(1-z)^2-(2u_3\,z\,(1-z)+u_2\,z^2)z^{2\tau}}.$$

and accordingly derive

$$\begin{aligned}
\mathbf{H}_{k,\tau}(z,u_1) &= \sum_{m\geq 0}\sum_{\gamma\in \mathcal{I}_k(m)} \mathbf{T}_\gamma(z,u_1,1,1) = \sum_{s\geq 0}\sum_{m=0}^{s} i_k(s,m)\mathbf{T}_\gamma(z,u_1,1,1),\\
\mathbf{I}_{k,\tau}(z,u_2) &= \sum_{m\geq 0}\sum_{\gamma\in \mathcal{I}_k(m)} \mathbf{T}_\gamma(z,1,u_2,1) = \sum_{s\geq 0}\sum_{m=0}^{s} i_k(s,m)\mathbf{T}_\gamma(z,1,u_2,1),\\
\mathbf{B}_{k,\tau}(z,u_3) &= \sum_{m\geq 0}\sum_{\gamma\in \mathcal{I}_k(m)} \mathbf{T}_\gamma(z,1,1,u_3) = \sum_{s\geq 0}\sum_{m=0}^{s} i_k(s,m)\mathbf{T}_\gamma(z,1,1,u_3).
\end{aligned}$$

It now remains to observe

$$\sum_{s\geq 0}\sum_{m=0}^{s} i_k(s,m)\, x^s\, y^m = \frac{1+x}{1+2x-xy}\mathbf{F}_k\left(\frac{x(1+x)}{(1+2x-xy)^2}\right).$$

and to subsequently substitute $x=\eta(1,1)$ and $y=u_1\,z$ for deriving $\mathbf{H}_{k,\tau}(z,u_1)$. Substituting $x=\eta(u_2,1)$ and $y=z$ in we obtain $\mathbf{I}_{k,\tau}(z,u_2)$ and finally $x=\eta(1,u_3)$ and $y=z$ produce the expression for $\mathbf{B}_{k,\tau}(z,u_3)$, whence the theorem. $\square$

4. The central limit theorem

For fixed k-noncrossing, τ-canonical structure, S, let $\mathbb{H}_{n,k,\tau}(S)$, $\mathbb{I}_{n,k,\tau}(S)$ and $\mathbb{B}_{n,k,\tau}(S)$ denote the number of hairpin-loops, interior-loops and bulges in S. Then we have the r.v.s

- $\mathbb{H}_{n,k,\tau}$, where $\mathbb{P}\left(\mathbb{H}_{n,k,\tau} = t\right) = \frac{h_{k,\tau}(n,t)}{\mathsf{T}_{k,\tau}(n)}$
- $\mathbb{I}_{n,k,\tau}$, where $\mathbb{P}\left(\mathbb{I}_{n,k,\tau} = t\right) = \frac{i_{k,\tau}(n,t)}{\mathsf{T}_{k,\tau}(n)}$
- $\mathbb{B}_{n,k,\tau}$, where $\mathbb{P}\left(\mathbb{B}_{n,k,\tau} = t\right) = \frac{b_{k,\tau}(n,t)}{\mathsf{T}_{k,\tau}(n)}$.

Here $h_{k,\tau}(n,t)$, $i_{k,\tau}(n,t)$ and $b_{k,\tau}(n,t)$ are the numbers of k-noncrossing, τ-canonical structures of length n with t hairpin-loops, interior-loops and bulges. The key for computing the distributions of the above r.v.s are the bivariate generating functions derived in Theorem 3:

$$\mathbf{H}_{k,\tau}(z,u_1) = \sum_{n\geq 0}\sum_{t\geq 0} h_{k,\tau}(n,t) z^n\, u_1^t, \tag{4.1}$$

$$\mathbf{I}_{k,\tau}(z,u_2) = \sum_{n\geq 0}\sum_{t\geq 0} i_{k,\tau}(n,t) z^n\, u_2^t, \tag{4.2}$$

$$\mathbf{B}_{k,\tau}(z,u_3) = \sum_{n\geq 0}\sum_{t\geq 0} b_{k,\tau}(n,t) z^n\, u_3^t. \tag{4.3}$$

The following proposition is based on Theorem 2 and facilitates the application of Theorem 1.

Proposition 1. *Suppose* $2 \le k \le 7$, $1 \le \tau \le 10$. *There exists a unique dominant* $\mathbf{H}_{k,\tau}(z, e^s)$*-singularity,* $\gamma_{k,\tau}(s)$*, such that for* $|s| < \epsilon$*, where* $\epsilon > 0$*:*

(1) $\gamma_{k,\tau}(s)$ *is analytic,*

(2) $\gamma_{k,\tau}(s)$ *is the solution of minimal modulus of*

$$\frac{z^{2\tau}(1-z)^2(1-z^2+z^{2\tau})}{\left((1-z)^2(1-z^2+z^{2\tau})+z^{2\tau}-z^{2\tau+1}e^s\right)^2} - \rho_k^2 = 0. \tag{4.4}$$

and

$$[z^n]\mathbf{H}_{k,\tau}(z, e^s) \sim C(s)\, n^{-((k-1)^2+\frac{k-1}{2})} \left(\frac{1}{\gamma_{k,\tau}(s)}\right)^n, \tag{4.5}$$

uniformly in s *in a neighborhood of* 0 *and continuous* $C(s)$*.*

Proof. The first step is to establish the existence and uniqueness of the dominant singularity $\gamma_{k,\tau}(s)$.

We denote

$$\vartheta(z,s) = (1-z)^2(1-z^2+z^{2\tau}) + z^{2\tau} - z^{2\tau+1}e^s, \tag{4.6}$$

$$\psi_\tau(z,s) = z^{2\tau}(1-z)^2(1-z^2+z^{2\tau})\vartheta(z,s)^{-2}, \tag{4.7}$$

$$\omega_\tau(z,s) = (1-z)(1-z^2+z^{2\tau})\vartheta(z,s)^{-1}, \tag{4.8}$$

and consider the equations

$$\forall\, 2 \le i \le k; \qquad F_{i,\tau}(z,s) = \psi_\tau(z,s) - \rho_i^2, \tag{4.9}$$

where $\rho_i = 1/(2i-2)$. Theorem 3 and Table 1 imply that the singularities of $\mathbf{H}_{k,\tau}(z, e^s)$ are are contained in the set of roots of

$$F_{i,\tau}(z, s) = 0 \quad \text{and} \quad \vartheta(z, s) = 0 \tag{4.10}$$

where $i \leq k$. Let $r_{i,\tau}$ denote the solution of minimal modulus of

$$F_{i,\tau}(z, 0) = \psi_\tau(z, 0) - \rho_i^2 = 0. \tag{4.11}$$

We next verify that, for sufficiently small $\epsilon_i > 0$, $|z - r_{i,\tau}| < \epsilon_i$, $|s| < \epsilon_i$, the following assertions hold

- $\frac{\partial}{\partial z} F_{i,\tau}(r_{i,\tau}, 0) \neq 0$
- $\frac{\partial}{\partial z} F_{i,\tau}(z, s)$ and $\frac{\partial}{\partial s} F_{i,\tau}(z, s)$ are continuous.

The analytic implicit function theorem, guarantees the existence of a unique analytic function $\gamma_{i,\tau}(s)$ such that, for $|s| < \epsilon_i$,

$$F_{i,\tau}(\gamma_{i,\tau}(s), s) = 0 \quad \text{and} \quad \gamma_{i,\tau}(0) = r_{i,\tau}. \tag{4.12}$$

Analogously, we obtain the unique analytic function $\delta(s)$ satisfying $\vartheta(z, s) = 0$ where $\delta(0)$ is the minimal solution of $\vartheta(z, 0) = 0$ for $|s| < \epsilon_\delta$, for some $\epsilon_\delta > 0$. We next verify that the unique dominant singularity of $\mathbf{H}_{k,\tau}(z, 1)$ is the minimal positive solution $r_{k,\tau}$ of $F_{k,\tau}(z, 0) = 0$ and subsequently using an continuity argument. Therefore, for sufficiently small ϵ where $\epsilon < \epsilon_i$ and $\epsilon < \epsilon_\delta$, $|s| < \epsilon$, the module of $\gamma_{i,\tau}(s)$, $i < k$ and

$\delta(s)$ are all strictly larger than the modulus of $\gamma_{k,\tau}(s)$. Consequently, $\gamma_{k,\tau}(s)$ is the unique dominant singularity of $\mathbf{H}_{k,\tau}(z, e^s)$ for $|s| < \epsilon$.

Claim. There exists some continuous $C(s)$ such that, uniformly in s, for s in a neighborhood of 0

$$[z^n]\mathbf{H}_{k,\tau}(z, e^s) \sim C(s)\, n^{-((k-1)^2+\frac{k-1}{2})} \left(\frac{1}{\gamma_{k,\tau}(s)}\right)^n .$$

To prove the Claim, let r be some positive real number such that $r_{k,\tau} < r < \delta(0)$. For sufficiently small $\epsilon > 0$ and $|s| < \epsilon$,

$$|\gamma_{k,\tau}(s)| \le r \quad \text{and} \quad |\delta(s)| > r.$$

Then $\psi_\tau(z, s)$ and $\omega_\tau(z, s)$ are all analytic in $\mathcal{D} = \{(z, s)||z| \le r, |s| < \epsilon\}$ and $\psi_\tau(0, s) = 0$. Since $\gamma_{k,\tau}(s)$ is the unique dominant singularity of

$$\mathbf{H}_{k,\tau}(z, e^s) = \omega_\tau(z, s)\, \mathbf{F}_k(\psi_\tau(z, s)),$$

satisfying

$$\psi_\tau(\gamma_{k,\tau}(s), s) = \rho_k^2 \quad \text{and} \quad |\gamma_{k,\tau}(s)| \le r, \tag{4.13}$$

for $|s| < \epsilon$. For sufficiently small $\epsilon > 0$, $\frac{\partial}{\partial z} F_{k,\tau}(z, s)$ is continuous and $\frac{\partial}{\partial z} F_{k,\tau}(r_{k,\tau}, 0) \neq 0$. Thus there exists some $\epsilon > 0$, such that for $|s| < \epsilon$, $\frac{\partial}{\partial z} F_{k,\tau}(\gamma_{k,\tau}(s), s) \neq 0$. According to Theorem 2, we therefore derive

$$[z^n]\mathbf{H}_{k,\tau}(z, e^s) \sim C(s)\, n^{-((k-1)^2+\frac{k-1}{2})} \left(\frac{1}{\gamma_{k,\tau}(s)}\right)^n , \tag{4.14}$$

uniformly in s in a neighborhood of 0 and continuous $C(s)$. $\square$

After establishing the analogues of Proposition 1 for $\mathbf{I}_{k,\tau}(z,u)$ and $\mathbf{B}_{k,\tau}(z,u)$, see the Supplemental Materials, Theorem 1 implies the following central limit theorem for the distributions of hairpin-loops, interior-loops and bulges in k-noncrossing structures.

Theorem 4. *Let $k,\tau \in \mathbb{N}$, $2 \le k \le 7$, $1 \le \tau \le 10$ and suppose the random variable $\mathbb{X}$ denotes either $\mathbb{H}_{n,k,\tau}$, $\mathbb{I}_{n,k,\tau}$ or $\mathbb{B}_{n,k,\tau}$. Then there exists a pair*

$$(\mu_{k,\tau,\mathbb{X}}, \sigma^2_{k,\tau,\mathbb{X}})$$

such that the normalized random variable $\mathbb{X}^*$ has asymptotically normal distribution with parameter $(0,1)$, where $\mu_{k,\tau,\mathbb{X}}$ and $\sigma^2_{k,\tau,\mathbb{X}}$ are given by

$$\mu_{k,\tau,\mathbb{X}} = -\frac{\gamma'_{k,\tau,\mathbb{X}}(0)}{\gamma_{k,\tau,\mathbb{X}}(0)}, \qquad \sigma^2_{k,\tau,\mathbb{X}} = \left(\frac{\gamma'_{k,\tau,\mathbb{X}}(0)}{\gamma_{k,\tau,\mathbb{X}}(0)}\right)^2 - \frac{\gamma''_{k,\tau,\mathbb{X}}(0)}{\gamma_{k,\tau,\mathbb{X}}(0)}, \tag{4.15}$$

where $\gamma_{k,\tau,\mathbb{X}}(s)$ represents the unique dominant singularity of $\mathbf{H}_{k,\tau}(z,e^s)$, $\mathbf{I}_{k,\tau}(z,e^s)$, and $\mathbf{B}_{k,\tau}(z,e^s)$, respectively.

In Tables 2, 3 and 4 we present the values of the pairs $(\mu_{k,\tau,\mathbb{X}}, \sigma^2_{k,\tau,\mathbb{X}})$.

5. Proof of Theorem 2

Proof of Theorem 2. We consider the composite function $\mathbf{F}_k(\psi(z,s))$. In view of $[z^n]f(z,s) = \gamma^n [z^n] f(\frac{z}{\gamma}, s)$ it suffices to analyze the function $\mathbf{F}_k(\psi(\gamma(s)z,s))$ and to subsequently rescale in order to obtain the correct exponential factor. For this purpose we set

$$\widetilde{\psi}(z,s) = \psi(\gamma(s)z,s),$$

where $\psi(z,s)$ is analytic in a domain $\mathcal{D} = \{(z,s) \| z| \le r, |s| < \epsilon\}$. Consequently $\widetilde{\psi}(z,s)$ is analytic in $|z| < \widetilde{r}$ and $|s| < \widetilde{\epsilon}$, for some $1 < \widetilde{r}$, $0 < \widetilde{\epsilon} < \epsilon$, since it's a composition of two analytic functions in $\mathcal{D}$. Taking its Taylor expansion at $z = 1$,

$$\widetilde{\psi}(z,s) = \sum_{n \ge 0} \widetilde{\psi}_n(s)(1-z)^n, \tag{5.1}$$

where $\widetilde{\psi}_n(s)$ is analytic in $|s| < \widetilde{\epsilon}$. The singular expansion of $\mathbf{F}_k(z)$, $2 \le k \le 7$, for $z \to \rho_k^2$, follows from the ODEs, see eq. (2.2), and is given by

$$\mathbf{F}_k(z) = \begin{cases} P_k(z-\rho_k^2) + c_k'(z-\rho_k^2)^{((k-1)^2+(k-1)/2)-1} \log(z-\rho_k^2)\,(1+o(1)) \\ P_k(z-\rho_k^2) + c_k'(z-\rho_k^2)^{((k-1)^2+(k-1)/2)-1}\,(1+o(1)) \end{cases} \tag{5.2}$$

depending on whether k is odd or even and where $P_k(z)$ are polynomials of degree $\le (k-1)^2 + (k-1)/2 - 1$, c_k' is some constant, and $\rho_k = 1/2(k-1)$. By assumption, $\gamma(s)$ is the unique analytic solution of $\psi(\gamma(s),s) = \rho_k^2$ and by construction

$\mathbf{F}_k(\psi(\gamma(s)z,s)) = \mathbf{F}_k(\widetilde{\psi}(z,s))$. In view of eq. (5.1), we have for $z \to 1$ the expansion

$$(5.3) \qquad \widetilde{\psi}(z,s) - \rho_k^2 = \sum_{n \geq 1} \widetilde{\psi}_n(s)(1-z)^n = \widetilde{\psi}_1(s)(1-z)(1+o(1)),$$

that is uniform in s since $\widetilde{\psi}_n(s)$ is analytic for $|s| < \widetilde{\epsilon}$ and $\widetilde{\psi}_0(s) = \psi(\gamma(s), s) = \rho_k^2$. As for the singular expansion of $\mathbf{F}_k(\widetilde{\psi}(z,s))$ we derive, substituting the eq. (5.3) into the singular expansion of $\mathbf{F}_k(z)$, for $z \to 1$,

$$(5.4) \qquad \begin{cases} \widetilde{P}_k(z,s) + c_k(s)(1-z)^{((k-1)^2+(k-1)/2)-1} \log(1-z)\,(1+o(1)) & \text{for } k \text{ odd} \\ \widetilde{P}_k(z,s) + c_k(s)(1-z)^{((k-1)^2+(k-1)/2)-1}\,(1+o(1)) & \text{for } k \text{ even} \end{cases}$$

where $\widetilde{P}_k(z,s) = P_k(\widetilde{\psi}(z,s) - \rho_k^2)$ and $c_k(s) = c_k' \widetilde{\psi}_1(s)^{((k-1)^2+(k-1)/2)-1}$ and

$$\widetilde{\psi}_1(s) = \partial_z \widetilde{\psi}(z,s)|_{z=1} = \gamma(s)\partial_z \psi(\gamma(s), s) \neq 0 \quad \text{for } |s| < \epsilon.$$

Furthermore $\widetilde{P}_k(z,s)$ is analytic at $|z| \leq 1$, whence $[z^n]\widetilde{P}_k(z,s)$ is exponentially small compared to 1. Therefore we arrive at

$$(5.5) \quad [z^n]\mathbf{F}_k(\widetilde{\psi}(z,s)) \sim \begin{cases} [z^n]c_k(s)(1-z)^{((k-1)^2+(k-1)/2)-1} \log(1-z)\,(1+o(1)) \\ [z^n]c_k(s)(1-z)^{((k-1)^2+(k-1)/2)-1}\,(1+o(1)) \end{cases}$$

depending on k being odd or even and uniformly in $|s| < \widetilde{\epsilon}$. We observe that $c_k(s)$ is analytic in $|s| < \widetilde{\epsilon}$. Note that a dependency in the parameter s is only given in the coefficients $c_k(s)$, that are analytic in s. Standard transfer theorems

[Flajolet and Sedgewick 2009] imply that

$$[z^n]\mathbf{F}_k(\widetilde{\psi}(z,s)) \sim A(s)\, n^{-((k-1)^2+(k-1)/2)} \quad \text{for some } A(s) \in \mathbb{C}, \tag{5.6}$$

uniformly in s contained in a small neighborhood of 0. Finally, as mention in the beginning of the proof, we use the scaling property of Taylor expansions in order to derive

$$[z^n]\mathbf{F}_k(\psi(z,s)) = (\gamma(s))^{-n}\, [z^n]\mathbf{F}_k(\widetilde{\psi}(z,s)) \tag{5.7}$$

and the proof of the Theorem is complete.

Acknowledgments. This work was supported by the 973 Project, the PCSIRT Project of the Ministry of Education, the Ministry of Science and Technology, and the National Science Foundation of China.

References

[Bender 1973] Bender, E.A. 1973. Central and local limit theorems applied to asymptotic enumeration. *J. Combin. Theory A.* 15, 91–111.

[Chen et al. 2007] Chen, W.Y.C., Deng, E.Y.P., Du, R.R.X., Stanley, R.P., Yan, C.H. 2007. Crossings and nestings of matchings and partitions. *Trans. Amer. Math. Soc.* 359, 1555–1575.

[Chen et al. 2009] Chen, W.Y.C., Han, H.S.W. and Reidys, C.M. 2009. Random k-noncrossing RNA Structures. *Proc. Natl. Acad. Sci. USA* 16, 22061–22066.

[Flajolet and Sedgewick 2009] Flajolet, P. and Sedgewick, R. 2009. *Analytic combinatorics.* Cambridge University Press, New York.

[Grabiner and Magyar 1993] Grabiner, D.J. and Magyar, P. 1993. Random walks in Weyl chambers and the decomposition of tensor powers. *J. Algebr. Comb.* 2, 239–260.

[Jin et al. 2008a] Jin, E.Y., Qin, J. and Reidys, C.M. 2008. Combinatorics of RNA structures with pseudoknots. *Bull. Math. Biol.* 70, 45–67.

[Jin et al. 2008b] Jin, E.Y., Reidys, C.M. and Wang, R.R. 2008. Asympotic analysis of k-noncrossing matchings. *arXiv:0803.0848.*

[Jin and Reidys 2009] Jin, E.Y. and Reidys, C.M. 2009. Combinatorial design of pseudoknot RNA. *Adv. Appl. Math.* 42, 135–151.

[Jin and Reidys 2010] Jin, E.Y. and Reidys, C.M. 2010. RNA pseudoknots structures with arc-length ≥ 3 and stack-length $\geq \sigma$, *Discrete Appl. Math.* 158, 25–36.

[Kleitman 1970] Kleitman, D.J. 1970. Proportions of irreducible diagrams. *Stud. Appl. Math.* 49, 297–299.

[Nussinov et al. 1978] Nussinov, R., Pieczenik, G., Griggs, J.R., Kleitman, D.J. 1978. Algorithms for loop matchings. *SIAM J. Appl. Math.* 35, 68–82.

[Reidys and Wang 2009] Reidys, C.M. and Wang, R.R. 2009. Shapes of RNA pseudoknot structures, submitted.

[Stanley 1980] Stanley, R. 1980. Differentiably finite power series. *Europ. J. Combinatorics* 1, 175–188.

[Waterman 1978] Waterman, M.S. 1978. Combinatorics of RNA hairpins and cloverleafs. *Stud. Appl. Math.* 60, 91–96.

k	$q_{0,k}(z)$	R_k
2	$(4z-1)z$	$\{\frac{1}{4}\}$
3	$(16z-1)z^2$	$\{\frac{1}{16}\}$
4	$(144z^2-40z+1)z^3$	$\{\frac{1}{4},\frac{1}{36}\}$
5	$(1024z^2-80z+1)z^4$	$\{\frac{1}{16},\frac{1}{64}\}$
6	$(14400z^3-4144z^2+140z-1)z^5$	$\{\frac{1}{4},\frac{1}{36},\frac{1}{100}\}$
7	$(147456z^3-12544z^2+224z-1)z^6$	$\{\frac{1}{16},\frac{1}{64},\frac{1}{144}\}$

TABLE 1. We present the polynomials $q_{0,k}(z)$ and their nonzero roots obtained by the MAPLE package `GFUN`.

	$k=2$		$k=3$		$k=4$	
	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$
$\tau=1$	0.105573	0.032260	0.012013	0.011202	0.003715	0.003641
$\tau=2$	0.061281	0.018116	0.009845	0.008879	0.003734	0.003602
$\tau=3$	0.043900	0.012752	0.007966	0.007060	0.003200	0.003060
$\tau=4$	0.034477	0.009896	0.006680	0.005854	0.002757	0.002622
	$k=5$		$k=6$		$k=7$	
	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$
$\tau=1$	0.001626	0.001612	0.000855	0.000852	0.000505	0.000504
$\tau=2$	0.001897	0.001864	0.001123	0.001111	0.000731	0.000726
$\tau=3$	0.001693	0.001655	0.001035	0.001021	0.000692	0.000686
$\tau=4$	0.001486	0.001448	0.000922	0.000907	0.000624	0.000618

TABLE 2. Hairpin-loops: The central limit theorem for the numbers of hairpin-loops in k-noncrossing, τ-canonical structures. We list $\mu_{k,\tau}$ and $\sigma^2_{k,\tau}$ derived from eq. (4.15).

	$k=2$		$k=3$		$k=4$	
	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$
$\tau=1$	0.015403	0.013916	0.001185	0.001176	0.000264	0.000264
$\tau=2$	0.012959	0.011395	0.001823	0.001793	0.000603	0.000599
$\tau=3$	0.011075	0.009570	0.001878	0.001837	0.000693	0.000688
$\tau=4$	0.009682	0.008261	0.001803	0.001755	0.000700	0.000693
	$k=5$		$k=6$		$k=7$	
	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$
$\tau=1$	0.000090	0.000090	0.000039	0.000039	0.000019	0.000019
$\tau=2$	0.000275	0.000274	0.000149	0.000149	0.000090	0.000090
$\tau=3$	0.000343	0.000341	0.000198	0.000198	0.000126	0.000126
$\tau=4$	0.000359	0.000357	0.000214	0.000213	0.000140	0.000140

TABLE 3. Interior-loops: The central limit theorem for the numbers of interior-loops in k-noncrossing,τ-canonical structures. We list $\mu_{k,\tau}$ and $\sigma^2_{k,\tau}$ derived from eq. (4.15).

	$k=2$		$k=3$		$k=4$	
	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$
$\tau=1$	0.049845	0.042310	0.008982	0.008684	0.003094	0.003058
$\tau=2$	0.025088	0.021785	0.005789	0.005597	0.002457	0.002422
$\tau=3$	0.015859	0.013979	0.003936	0.003814	0.001762	0.001737
$\tau=4$	0.011197	0.009980	0.002878	0.002795	0.001318	0.001301
	$k=5$		$k=6$		$k=7$	
	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$	$\mu_{k,\tau}$	$\sigma^2_{k,\tau}$
$\tau=1$	0.001422	0.001414	0.000770	0.000767	0.000463	0.000462
$\tau=2$	0.001326	0.001316	0.000817	0.000813	0.000547	0.000546
$\tau=3$	0.000991	0.000984	0.000632	0.000629	0.000436	0.000435
$\tau=4$	0.000755	0.000750	0.000489	0.000486	0.000342	0.000341

TABLE 4. Bulges: The central limit theorems for the numbers of bulges in k-noncrossing, τ-canonical structures. We list $\mu_{k,\tau}$ and $\sigma^2_{k,\tau}$ derived from eq. (4.15).

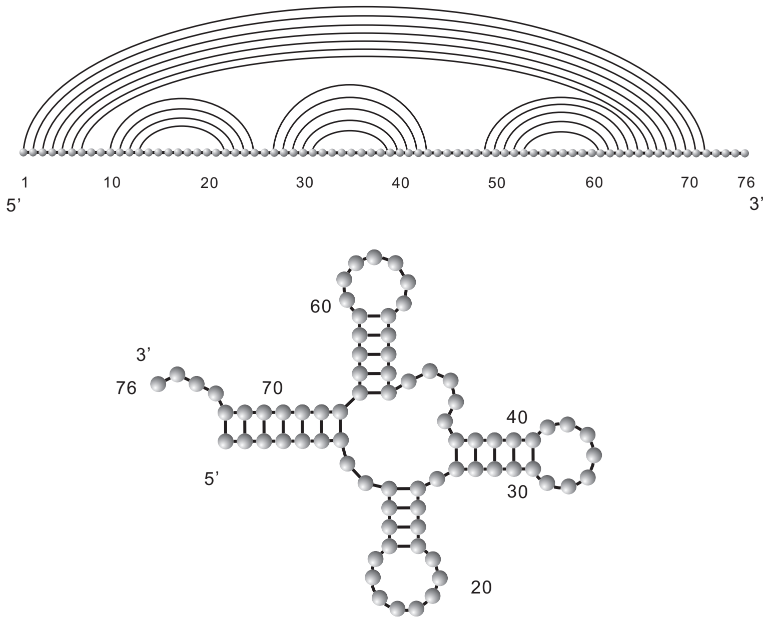

FIGURE 1. The phenylalanine tRNA secondary structure represented as 2-noncrossing diagram (top) and planar graph (bottom).

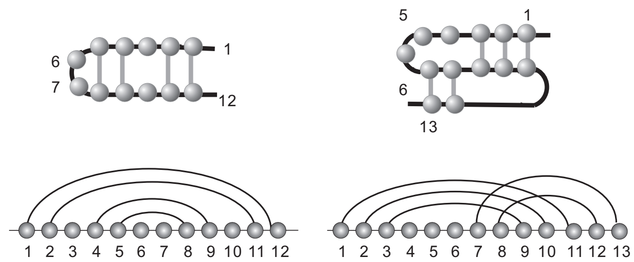

FIGURE 2. A 2-noncrossing, 2-canonical RNA structure (left) and a 3-noncrossing, 2-canonical RNA structure (right) represented as planer graphs (top) and diagrams (bottom).

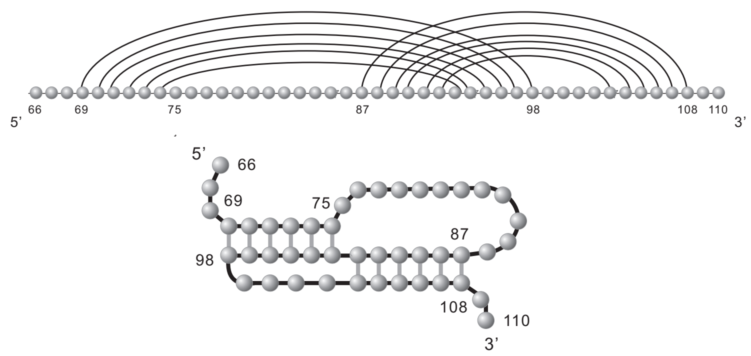

FIGURE 3. 3-noncrossing, 6-canonical structures: the pseudoknot structure of the PrP-encoding mRNA represented as diagrams (top) and planer graphs (bottom).

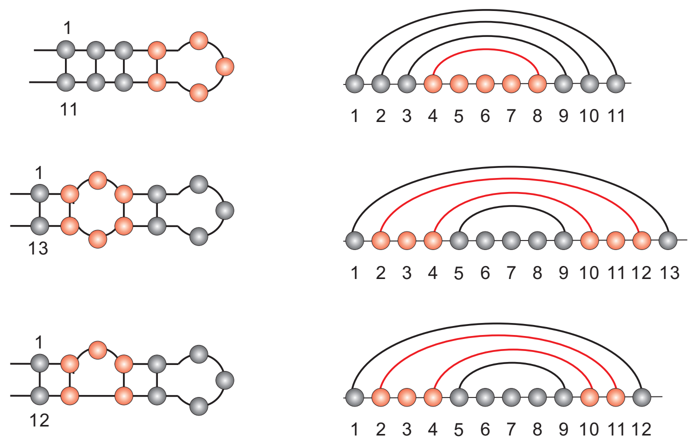

FIGURE 4. The loop-types: hairpin-loop (top), interior-loop (middle) and bulge (bottom).

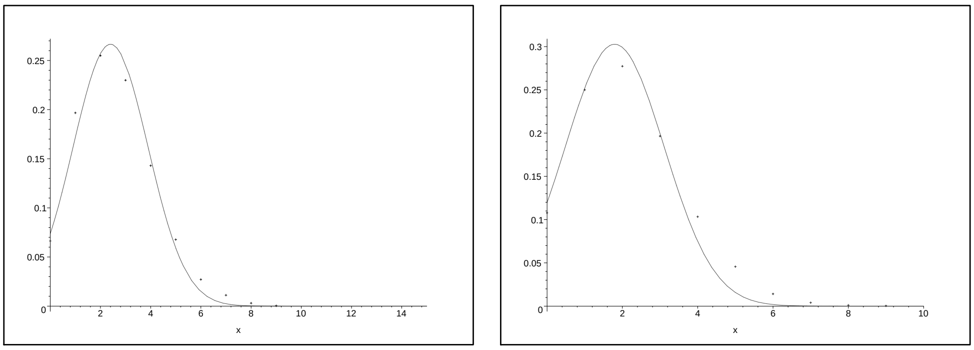

FIGURE 5. The distribution of hairpins (left) and bulges (right) in 3-noncrossing 1-canonical RNA structures of length $n = 200$. The solid curves are derived from the central limit theorem Theorem 4. The data points are obtained by uniformly generating 3-noncrossing structures [Chen et al. 2009].

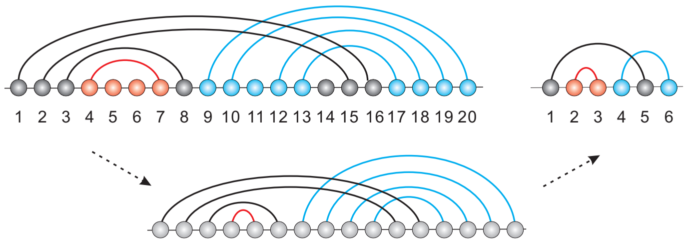

FIGURE 6. A 3-noncrossing, 2-canonical RNA structure (top-left) is mapped into its shape (top-right) in two steps. A stem (blue) is mapped into a single shape-arc (blue). A hairpin-loop (red) is mapped into a shape-1-arc (red).

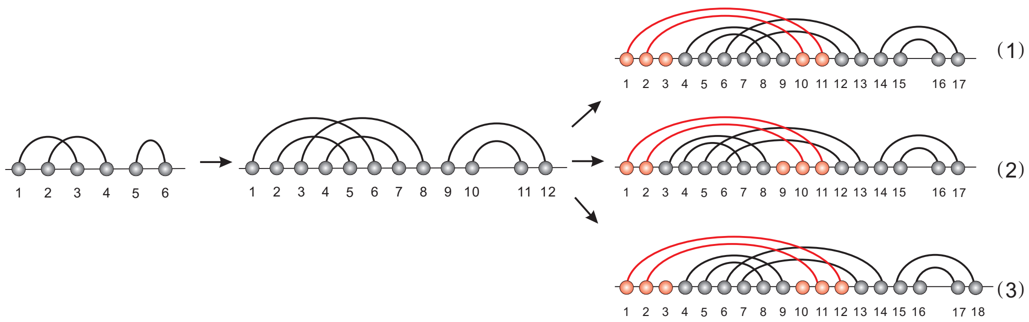

FIGURE 7. The first inflation step a shape (left) is inflated to a 3-noncrossing, 2-canonical structure. First, every arc in the shape is inflated to a stack of size at least two (middle), and then the shape is inflated to a new 3-noncrossing, 2-canonical structure (right) by adding one stack of size two. There are three ways to insert the isolated vertices.

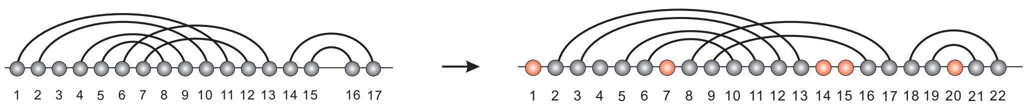

FIGURE 8. The second inflation step: the structure (left) obtained in (1) in Fig. 7 is inflated to a new 3-noncrossing, 2-canonical RNA structures (right) by adding isolated vertices (red).