

RNA pseudoknot structures with arc-length ≥ 3 and stack-length $\geq \sigma$

Emma Y. Jin and Christian M. Reidys*

Center for Combinatorics, LPMC-TJKLC
Nankai University
Tianjin 300071
P.R. China
Phone: *86-22-2350-6800
Fax: *86-22-2350-9272
email: duck@santafe.edu

Abstract

In this paper, we enumerate k -noncrossing RNA pseudoknot structures with given minimum arc- and stack-length. That is, we study the numbers of RNA pseudoknot structures with arc-length ≥ 3 , stack-length $\geq \sigma$ and in which there are at most $k - 1$ mutually crossing bonds, denoted by $\mathbb{T}_{k,\sigma}^{[3]}(n)$. We prove that the numbers of k -noncrossing RNA structures with arc-length ≥ 3 and stack-length ≥ 2 satisfy $\mathbb{T}_{k,2}^{[3]}(n) \sim C_k n^{-(k-1)^2 - \frac{k-1}{2}} (\gamma_{k,2}^{[3]})^{-n}$. In the case $k = 3, 4, 5$, we derive $\mathbb{T}_{3,2}^{[3]}(n) \sim C_3 n^{-5} 2.5721^n$, $\mathbb{T}_{4,2}^{[3]}(n) \sim C_4 n^{-\frac{21}{2}} 3.0306^n$, and $\mathbb{T}_{5,2}^{[3]}(n) \sim C_5 n^{-18} 3.4092^n$, respectively, where C_3, C_4, C_5 are constants. Our results are of importance for prediction algorithms for RNA pseudoknot structures.

Keywords: RNA secondary structure, pseudoknot, enumeration, generating function.

Date: June 2009

1. INTRODUCTION

One central problem in structural biology is that of predicting the spatial configuration of a molecule. For RNA, this means to understand the configuration of the primary sequence composed by the four nucleotides **A**, **G**, **U** and **C**. These nucleotides can form Watson-Crick (**A-U**, **G-C**) and (**U-G**) base pairs, as well as hydrogen bonds. The formation of these bonds stabilizes the

molecule by lowering its free energy. A well-studied class of RNA conformations is that of RNA secondary structures, pioneered by M. Waterman [22, 21, 7, 23]. An RNA secondary structure is a contact-structure, identified with a set of Watson-Crick (**A-U**, **G-C**) and (**U-G**) base pairs without considering any notion of spatial embedding. The key property of secondary structures is, that they, when represented as diagrams (see Section 2), only contain noncrossing arcs. In other words, secondary structures are specific graphs over n nucleotides whose arcs are the base-pairs, see Figure 1. The biological relevance of secondary structures lies in the fact that their bonds make the main contribution to the free energy of the molecule [4]. Secondary structures can be

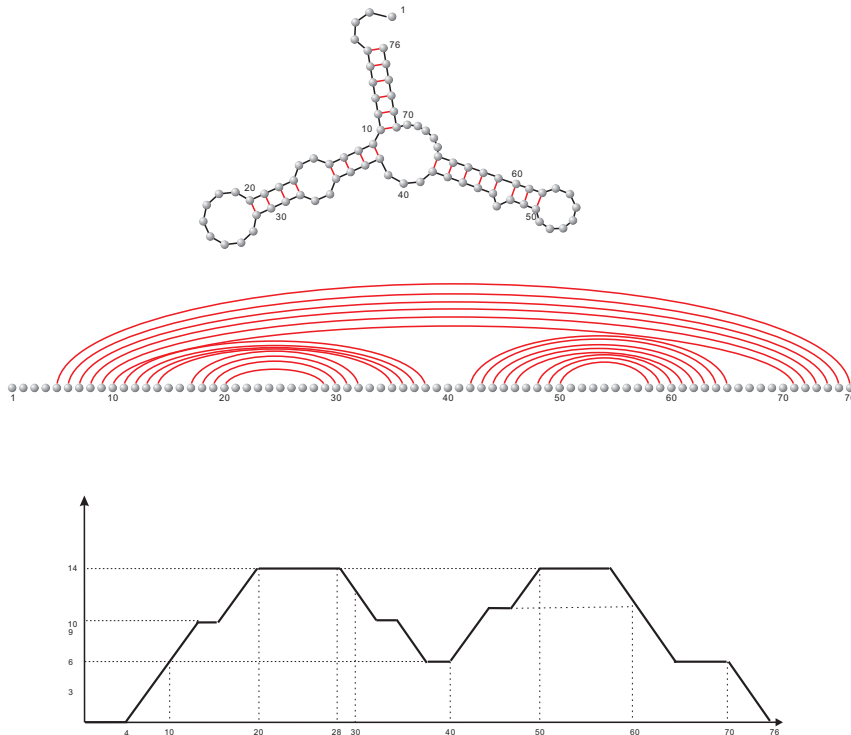


FIGURE 1. The phenylalanine tRNA secondary structure as folded by the *ab initio* folding algorithm **cross** [8], represented as planar graph, diagram and Motzkin-path.

represented as planar graphs, diagrams and Motzkin-paths, see Figure 1. In all representations, vertices and arcs correspond to nucleotides and Watson-Crick (**A-U**, **G-C**) and (**U-G**) base pairs, respectively. In addition to having noncrossing bonds, RNA secondary structures satisfy particular constraints common to all RNA structures. They satisfy specific minimum arc-length and stack-length conditions. These stem from the fact that chemical bonds are subject to specific steric as

well as biophysical constraints: the folded structure has to exhibit a low minimum free energy and key energy contributions come from the “stacking” of base pairs.

RNA structures serve a variety of biochemical functions: RNA acts as a messenger linking DNA and proteins. The discovery of catalytic RNAs, or ribosomes, in 1981 proved that RNA could catalyze reactions just as proteins. RNA plays a central role within living cells and all its functions are closely connected to the structure of the RNA molecules. An increasing number of experimental findings imply that there exist additional types of interactions between RNA nucleotides [24]. These bonds are called pseudoknots and occur in functional RNA like for instance RNaseP [14] as well as ribosomal RNA [13]. RNA pseudoknots are conserved also in the catalytic core of group I introns. In plant viral RNAs, pseudoknots mimic tRNA structure and in *in vitro* RNA evolution [20], experiments have produced families of RNA structures with pseudoknot motifs, when binding HIV-1 reverse transcriptase. In addition, important mechanisms like ribosomal frame shifting [1] also involve pseudoknot interactions. As a result, RNA pseudoknot structures have drawn over the last years a lot of attention [16]. A recent categorization of RNA pseudoknot structures has been obtained in [9, 10]. There the notion of k -noncrossing RNA structures has been introduced, that are structures with at most $k - 1$ *mutually* crossing arcs. This concept naturally generalizes that of 2-noncrossing, i.e. secondary structures and captures the complexity of pseudoknots, since Watson-Crick base pairs are constrained in the way they cross [6, 24]. The number of mutually crossing arcs, $k - 1$, the minimum arc-length and the minimum stack-length are key parameters of the molecular structure, see Figure 2. As in the case RNA secondary structures, due to steric and biophysical constraints, arcs in RNA pseudoknot structures are subject to minimal length conditions and the main energy contributions stem from the stacking of arcs.

In this paper, we study RNA pseudoknot structures that are subject to such biophysical constraints. We analyze k -noncrossing RNA structures having arc-length ≥ 3 and stack-length $\geq \sigma$, where $\sigma \geq 2$, the number of which are denoted by $T_{k,\sigma}^{[3]}(n)$. In order to construct this structure class, the main idea is to consider a certain subset of k -noncrossing core-structures [11]. We recall that a core-structure has no stacks of size > 1 , i.e. there exists no two arcs of the form $(i, j), (i + 1, j - 1)$ and no arcs of the form $(i, i + 1)$ (1-arcs). To be explicit, we consider the subset of core-structures having minimum arc-length 2 and having no arcs of the form $(i, i + 2)$, where $i + 1$ is isolated. We prove a bijection between this subset of core-structures with multiplicities and k -noncrossing RNA structures with arc-length ≥ 3 and stack-length $\geq \sigma$, where $\sigma \geq 2$. Subsequently, we derive several functional equations between generating functions, based on which transfer theorems imply our asymptotic formulas. In Figure 3, we show that the results derived here produce an approximation

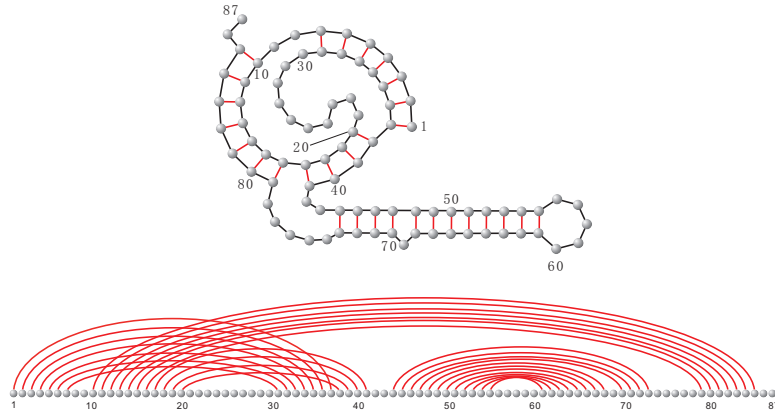


FIGURE 2. The Hepatitis Delta Virus (HDV)-pseudoknot structure and its diagram representation: we display the structure as folded by *ab initio* algorithm **cross** with minimum stack size three [8] (top) and its diagram representation (bottom).

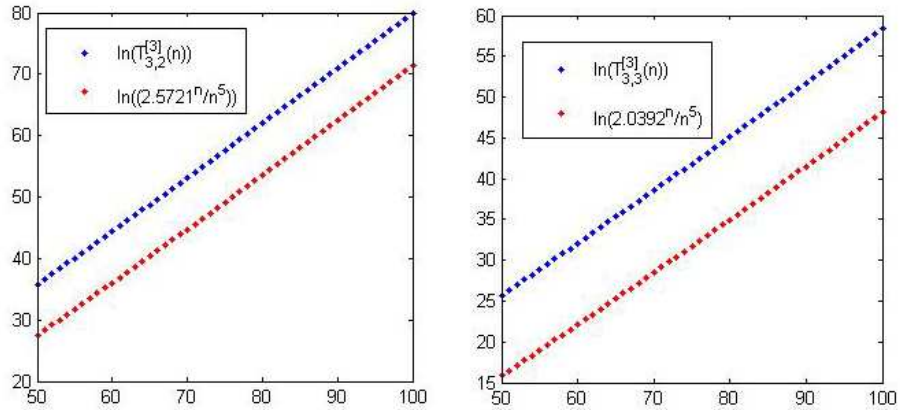


FIGURE 3. Exact enumeration data versus asymptotic formula. We plot the number of 3-noncrossing RNA structures with arc-length ≥ 3 and stack-length ≥ 2 , denoted by $T_{3,2}^{[3]}(n)$ and its asymptotic formula $C_{3,2} n^{-5} 2.5721^n$ (lhs) and the number of 3-noncrossing RNA structures with arc-length ≥ 3 and stack-length ≥ 3 , denoted by $T_{3,3}^{[3]}(n)$ versus $C_{3,3} n^{-5} 2.0392^n$. For the purpose of representation we set $C_{3,2} = C_{3,3} = 1$.

that works well already for relatively small sequence length. Figure 3 compares the exact values $T_{3,2}^{[3]}(n)$, i.e., the number of 3-noncrossing RNA structures with arc-length ≥ 3 and stack-length ≥ 2 and $T_{3,3}^{[3]}(n)$, i.e., the number of 3-noncrossing RNA structures with arc-length ≥ 3 and stack-length

≥ 3 , with the asymptotic formulas given in Theorem 5:

$$\Gamma_{3,2}^{[3]}(n) \sim C_{3,2} n^{-5} 2.5721^n \quad \text{and} \quad \Gamma_{3,3}^{[3]}(n) \sim C_{3,3} n^{-5} 2.0392^n.$$

The paper is relevant for *ab initio* prediction algorithms of pseudoknot RNA, since it proves that the numbers of k -noncrossing RNA structures with arc-length ≥ 3 and stack-length $\geq \sigma$ exhibit small exponential growth rates. The results suggest a novel strategy for RNA pseudoknot prediction and have led to novel folding strategies [8].

2. DIAGRAMS, MATCHINGS AND STRUCTURES

A diagram is labeled graph over the vertex set $[n] = \{1, \dots, n\}$ with degree ≤ 1 , represented by drawing its vertices $1, \dots, n$ in a horizontal line and its arcs (i, j) , where $i < j$, in the upper half plane. Here the degree of i refers to the number of non-horizontal arcs incident to i , i.e. the backbone of the primary sequence is not accounted for. The vertices and arcs correspond to nucleotides and Watson-Crick (**A-U**, **G-C**) and (**U-G**) base pairs, respectively. We categorize

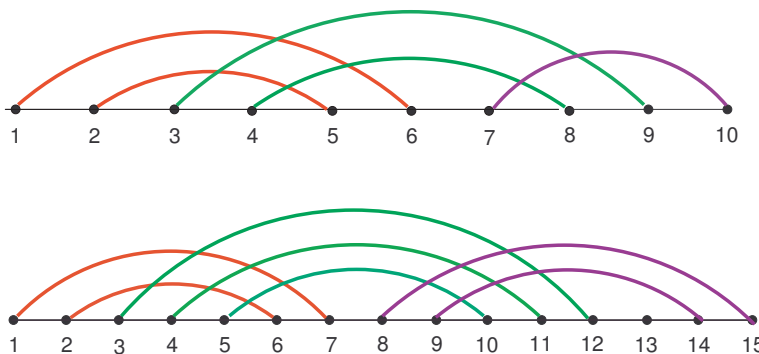


FIGURE 4. k -noncrossing diagrams. Top: 3-noncrossing diagram (no red/purple cross) with arc-length ≥ 3 , $(2, 5)$, $(7, 10)$, the arc $(7, 10)$ being isolated. Hence we have a 3-noncrossing, $\lambda = 3$, $\sigma = 1$ diagram without isolated vertices. Bottom: 3-noncrossing, $\lambda = 4$, $\sigma = 2$ diagram with isolated vertices 13.

diagrams according to the maximum number of mutually crossing arcs, $k - 1$, the minimum arc-length, λ , and the minimum stack-length, σ . Here the length of an arc (i, j) is $j - i$ and a stack of

length σ is a sequence of “parallel” arcs of the form

$$((i, j), (i + 1, j - 1), \dots, (i + (\sigma - 1), j - (\sigma - 1))).$$

In the following, we call a k -noncrossing diagram with arc-length ≥ 2 and stack-length $\geq \sigma$ a k -noncrossing RNA structure. We denote the set (number) of k -noncrossing RNA structures with stack-size $\geq \sigma$ by $T_{k,\sigma}(n)$ ($\mathbb{T}_{k,\sigma}(n)$) and refer to k -noncrossing RNA structures for $k \geq 3$ as pseudoknot RNA structures. A k -noncrossing core-structure is a k -noncrossing RNA structures in which there exists *no* two arcs of the form $(i, j), (i + 1, j - 1)$. The set (number) of RNA structures with arc-length ≥ 3 , is denoted by $T_{k,\sigma}^{[3]}(n)$ ($\mathbb{T}_{k,\sigma}^{[3]}(n)$). For $k = 2$ and $\sigma = 1$ we have RNA structures with no two crossing arcs, i.e. the well-known RNA secondary structures, as mentioned above. RNA secondary structures are $T_{2,1}(n)$ -structures. We denote by $f_k(n, \ell)$ the number of k -noncrossing diagrams with arbitrary arc-length and ℓ isolated vertices over n vertices. In Figure 5 we display various types of diagrams involved. In light of the bijection between k -noncrossing

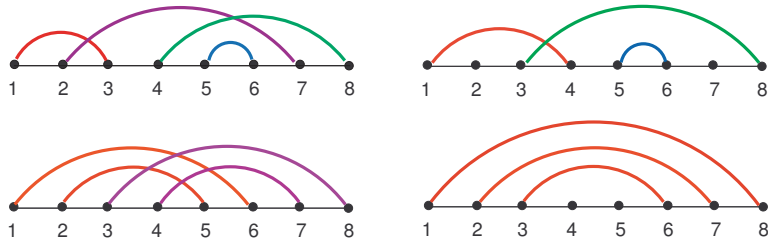


FIGURE 5. Basic diagram types: (a) (perfect) matching ($f_3(8, 0)$), (b) partial matching with 1-arc (5, 6) and isolated points 2, 7 ($f_3(8, 2)$), (c) structure with arc-length ≥ 3 and stack-length ≥ 2 and no isolated points ($\mathbb{T}_{3,2}^{[3]}(8)$) and (d) structure with arc-length ≥ 3 , stack-length ≥ 3 and isolated points 4, 5 ($\mathbb{T}_{2,3}^{[3]}(8)$).

matchings on $[2n]$ and the oscillating tableaux of empty shape and length $2n$ due to Stanley [2] and Sundaram [18], and the interpretation of an oscillating tableau as a path with elementary moves $\pm e_i$ inside a Weyl Chamber [2], we come to conclude that the number of k -noncrossing matchings on $[2n]$ equals the number of walks from η to itself that stay inside the Weyl Chamber $x_1 > x_2 > \dots > x_{k-1} > 0$ with steps $\pm e_i$, $1 \leq i \leq k - 1$, given by Grabiner *et al.* [5]. It is exactly the situation $\eta = \lambda = (k - 1, k - 2, \dots, 1)$ of equation (38) in [5]. As shown in detail in

[9], Lemma 2, it follows

$$(2.1) \quad \sum_{n \geq 0} f_k(n, 0) \cdot \frac{x^n}{n!} = \det[I_{i-j}(2x) - I_{i+j}(2x)]_{i,j=1}^{k-1}$$

$$(2.2) \quad \sum_{n \geq 0} \left\{ \sum_{\ell=0}^n f_k(n, \ell) \right\} \cdot \frac{x^n}{n!} = e^x \det[I_{i-j}(2x) - I_{i+j}(2x)]_{i,j=1}^{k-1},$$

where $I_r(2x) = \sum_{j \geq 0} \frac{x^{2j+r}}{j!(r+j)!}$ denotes the hyperbolic Bessel function of the first kind of order r . Eq. (2.1) and (2.2) allow “in principle” for explicit computation of the numbers $f_k(n, \ell)$. In particular for $k = 2$ and $k = 3$ we have the formulas

$$(2.3) \quad f_2(n, \ell) = \binom{n}{\ell} C_{(n-\ell)/2} \quad \text{and} \quad f_3(n, \ell) = \binom{n}{\ell} \left[C_{\frac{n-\ell}{2}+2} C_{\frac{n-\ell}{2}} - C_{\frac{n-\ell}{2}+1}^2 \right],$$

where C_m denotes the m -th Catalan number. $f_3(n, \ell)$ results from a determinant formula enumerating pairs of noncrossing Dyck-paths. In view of $f_k(n, \ell) = \binom{n}{\ell} f_k(n - \ell, 0)$ everything can be reduced to (perfect) matchings, where we have the following situation: there exists an asymptotic approximation of the determinant of hyperbolic Bessel function for general order k due to [12] and employing the subtraction of singularities-principle [15] one can prove [12]

$$(2.4) \quad \forall k \in \mathbb{N}; \quad f_k(2n, 0) \sim c_k n^{-((k-1)^2+(k-1)/2)} (2(k-1))^{2n}, \quad c_k > 0.$$

Let $\mathbf{F}_k(z) = \sum_{n \geq 0} f_k(2n, 0) z^{2n}$, the generating function of k -noncrossing matchings. As for the generating function and asymptotics of k -noncrossing RNA structures, we have the following results from [9, 10]. First the number of k -noncrossing RNA structures with $\binom{n-\ell}{2}$ arcs, $\mathbb{T}_{k,1}(n, \frac{n-\ell}{2})$, and the number of k -noncrossing RNA structures, $\mathbb{T}_{k,1}(n)$, is given by

$$(2.5) \quad \mathbb{T}_{k,1}(n, \frac{n-\ell}{2}) = \sum_{b=0}^{\lfloor n/2 \rfloor} (-1)^b \binom{n-b}{b} f_k(n-2b, \ell)$$

$$(2.6) \quad \mathbb{T}_{k,1}(n) = \sum_{b=0}^{\lfloor n/2 \rfloor} (-1)^b \binom{n-b}{b} \left\{ \sum_{\ell=0}^{n-2b} f_k(n-2b, \ell) \right\},$$

where $\{\sum_{\ell=0}^{n-2b} f_k(n-2b, \ell)\}$ is given via eq. (2.2). Secondly we have

$$(2.7) \quad \mathbb{T}_{3,1}(n) \sim \frac{1.9572 \cdot 4!}{n(n-1) \cdots (n-4)} \left(\frac{5 + \sqrt{21}}{2} \right)^n$$

$$(2.8) \quad \mathbb{T}_{3,1}^{[3]}(n) \sim \frac{6.1117 \cdot 4!}{n(n-1) \cdots (n-4)} 4.54920^n.$$

The particular class of k -noncrossing core-structures, i.e. structures in which there exists no two arcs of the form $(i, j), (i+1, j-1)$ with minimal arc-length 2, will play a central role in the following:

Theorem 1. (Core-structures) *Suppose $k \in \mathbb{N}$, $k \geq 2$, x is an indeterminate, ρ_k is the dominant, positive real singularity of $\mathbf{F}_k(x)$ and $u_1(x) = \frac{1}{1+x^2}$. Then for $h \geq 1$, the number of k -noncrossing core-structures, $C_k(n, h)$ is given by*

$$(2.9) \quad C_k(n, h) = \sum_{b=0}^{h-1} (-1)^{h-b-1} \binom{h-1}{b} \mathsf{T}_{k,1}(n-2h+2b+2, b+1).$$

Furthermore we have the functional equation

$$(2.10) \quad \sum_{n \geq 0} C_k(n) x^n = \frac{1}{u_1(x)x^2 - x + 1} \mathbf{F}_k \left(\frac{\sqrt{u_1(x)}x}{u_1(x)x^2 - x + 1} \right)$$

and the asymptotic formula

$$(2.11) \quad C_k(n) \sim n^{-((k-1)^2+(k-1)/2)} \left(\frac{1}{\kappa_k} \right)^n, \quad \text{for } k = 3, 4, \dots, 7$$

where κ_k is the dominant positive real singularity of $\sum_{n \geq 0} C_k(n)x^n$ and the minimal positive real solution of the equation $\frac{\sqrt{u_1}x}{u_1x^2-x+1} = \rho_k$ for $k = 3, 4, \dots, 7$.

Next we present a functional identity [10] which relates the bivariate generating function for $\mathsf{T}_{k,1}(n, h)$, the number of k -noncrossing RNA pseudoknot structures with h arcs, and the generating function of k -noncrossing matchings.

Lemma 1. *Let $k \in \mathbb{N}$, $k \geq 2$ and z, u be indeterminates. Then we have*

$$(2.12) \quad \sum_{n \geq 0} \sum_{h \leq n/2} \mathsf{T}_{k,1}(n, h) u^{2h} z^n = \frac{1}{u^2 z^2 - z + 1} \mathbf{F}_k \left(\frac{uz}{u^2 z^2 - z + 1} \right).$$

In particular we have for $u = 1$,

$$(2.13) \quad \sum_{n \geq 0} \mathsf{T}_{k,1}(n) z^n = \frac{1}{z^2 - z + 1} \mathbf{F}_k \left(\frac{z}{z^2 - z + 1} \right).$$

In view of Lemma 1, it is of interest to deduce relations between the coefficients from the equality of generating functions. The class of theorems that deal with this deduction are called transfer-theorems [3]. One key ingredient in this framework is a specific domain, in which the functions in question are analytic, which is “slightly” bigger than their respective radius of convergence. It

is tailored for extracting the coefficients via Cauchy's integral formula: given two numbers ϕ, R , where $R > 1$ and $0 < \phi < \frac{\pi}{2}$ and $\rho \in \mathbb{R}$, the open domain $\Delta_\rho(\phi, R)$ is defined as

$$(2.14) \quad \Delta_\rho(\phi, R) = \{z \mid |z| < R, z \neq \rho, |\text{Arg}(z - \rho)| > \phi\}.$$

A domain is a Δ_ρ -domain if it is of the form $\Delta_\rho(\phi, R)$ for some R and ϕ . A function is Δ_ρ -analytic if it is analytic in some Δ_ρ -domain. We use the notation

$$(2.15) \quad (f(z) = O(g(z)) \text{ as } z \rightarrow \rho) \iff (f(z)/g(z) \text{ is bounded as } z \rightarrow \rho)$$

and if we write $f(z) = O(g(z))$, it is implicitly assumed that z tends to a (unique) singularity. $[z^n]f(z)$ denotes the coefficient of z^n in the power series expansion of $f(z)$ around 0.

Theorem 2. [3] *Let $f(z), g(z)$ be D-finite, Δ_ρ -analytic functions with unique dominant singularity ρ and suppose*

$$(2.16) \quad f(z) = O(g(z)) \quad \text{for } z \rightarrow \rho.$$

Then we have

$$(2.17) \quad [z^n]f(z) = K \left(1 - O\left(\frac{1}{n}\right) \right) [z^n]g(z),$$

where K is some constant.

As a consequence of Theorem 2, eq. (2.4) and the so called supercritical case of singularity analysis [3], VI.9., p. 411, we give the following result tailored for our functional equations [11]. Let ρ_k denote the dominant positive real singularity of $\mathbf{F}_k(z)$.

Theorem 3. *Suppose $\vartheta_\sigma(z)$ is algebraic over $K(z)$, analytic for $|z| < \delta$ and satisfies $\vartheta_\sigma(0) = 0$. Suppose further $\gamma_{k,\sigma}$ is the real unique solution with minimal modulus $< \delta$ of the two equations $\vartheta_\sigma(z) = \rho_k$ and $\vartheta_\sigma(z) = -\rho_k$. Then*

$$(2.18) \quad [z^n] \mathbf{F}_k(\vartheta_\sigma(z)) \sim c_k n^{-((k-1)^2+(k-1)/2)} \left(\gamma_{k,\sigma}^{-1} \right)^n.$$

Since $\vartheta_\sigma(z)$ is algebraic over $K(z)$ and satisfies $\vartheta_\sigma(0) = 0$, we can conclude that the composition $\mathbf{F}_k(\vartheta_\sigma(z))$ is D-finite [17]. In particular, $\mathbf{F}_k(\vartheta_\sigma(z))$ has a singular expansion. Since $\mathbf{F}_k(z)$ has the two dominant singularities $\pm\rho_k$ and $\gamma_{k,\sigma}$ is the unique solution with minimal modulus $< \delta$ of the two equations $\vartheta_\sigma(z) = \rho_k$ and $\vartheta_\sigma(z) = -\rho_k$, we can conclude that $\gamma_{k,\sigma}$ is the unique dominant singularity of $\mathbf{F}_k(\vartheta_\sigma(z))$. We proceed by studying the singular expansion in more details. According to Theorem 2 in [12], we have

$$(2.19) \quad f_k(2n, 0) \sim c_k n^{-((k-1)^2+(k-1)/2)} (2(k-1))^{2n}$$

for some $c_k > 0$. In combination with Theorem 2, this allows us to conclude

$$\mathbf{F}_k(z) = \begin{cases} O((z - \rho_k)^{(k-1)^2 + (k-1)/2 - 1} \ln(z - \rho_k)) & \text{for } k \text{ odd, } z \rightarrow \rho_k \\ O((z - \rho_k)^{(k-1)^2 + (k-1)/2 - 1}) & \text{for } k \text{ even, } z \rightarrow \rho_k, \end{cases}$$

in accordance with basic structure theorems for singular expansions of D -finite functions [3]. Since $\vartheta_\sigma(z)$ is regular at $\gamma_{k,\sigma}$, we are given the supercritical case of singularity analysis [3]. In the supercritical case, the subexponential factors of the composition, $\mathbf{F}_k(\vartheta_\sigma(z))$ coincide with those of the outer function, $\mathbf{F}_k(z)$. Consequently we have

$$[z^n] \mathbf{F}_k(\vartheta_\sigma(z)) \sim c_k n^{-((k-1)^2 + (k-1)/2)} \left(\gamma_{k,\sigma}^{-1} \right)^n,$$

whence the theorem 3.

3. EXACT ENUMERATION

In Theorem 4 below, we enumerate k -noncrossing RNA structures with arc-length ≥ 3 and stack-length $\geq \sigma$. The structure of the formula is analogous to the Möbius inversion of eq. (2.9) proved in [11], which relates the numbers of all structures and the numbers of core-structures: $\mathbb{T}_{k,\sigma}(n, h) = \sum_{b=\sigma-1}^{h-1} \binom{b+(2-\sigma)(h-b)-1}{h-b-1} \mathbf{C}_k(n-2b, h-b)$. The latter cannot be used in order to enumerate k -noncrossing structures with arc-length ≥ 3 , see Figure 6. The sets of structures

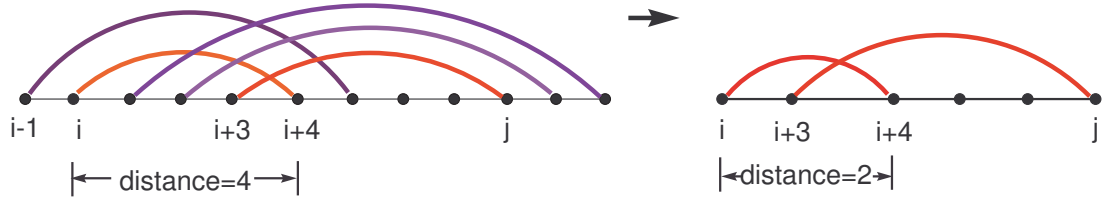


FIGURE 6. Core-structures will in general have 2-arcs: the structure $\delta \in T_{3,2}(12)$ (lhs) is mapped into its core $c(\delta)$ (rhs). Clearly δ has arc-length ≥ 4 and as a consequence of the collapse of the stack $((i+1, j+2), (i+2, j+1), (i+3, j))$ (the purple arcs are being removed) into the arc $(i+3, j)$, $c(\delta)$ contains the arc $(i, i+4)$, which is, after relabeling, a 2-arc.

$$(3.1) \quad C_k^*(n, h) = \{ \delta \mid \delta \in C_k(n, h); \exists (i, i+2); i+1 \text{ is an isolated vertex} \}$$

$$(3.2) \quad T_{k,1}^*(n, h) = \{ \delta \mid \delta \in T_{k,1}(n, h); \exists (i, i+2); i+1 \text{ is an isolated vertex} \}$$

turn out to be the key. Here the cardinality of the sets $C_k^*(n, h)$, $T_{k,1}^*(n, h)$ and $T_{k,\sigma}^{[3]}(n, h)$ are denoted by $C_k^*(n, h)$, $T_{k,1}^*(n, h)$ and $T_{k,\sigma}^{[3]}(n, h)$, respectively. Note that $C_k^*(n, 0) = 1$, for $n \geq 0$.

Theorem 4. *Suppose we have $k, h, \sigma \in \mathbb{N}$, $k \geq 2$, $h \leq n/2$ and $\sigma \geq 2$. Then the following assertions hold:*

(a) *The numbers of k -noncrossing RNA structures with arc-length ≥ 3 and stack-length $\geq \sigma$ having h arcs are given by*

$$(3.3) \quad T_{k,\sigma}^{[3]}(n, h) = \sum_{b=\sigma-1}^{h-1} \binom{b + (2-\sigma)(h-b) - 1}{h-b-1} C_k^*(n-2b, h-b).$$

(b) *The numbers $C_k^*(n, h)$ and $T_{k,1}^*(n, h)$ are given by*

$$(3.4) \quad C_k^*(n, h) = \sum_{b=0}^{h-1} (-1)^{h-b-1} \binom{h-1}{b} T_{k,1}^*(n-2h+2b+2, b+1) \quad \text{for } h \geq 1$$

$$(3.5) \quad T_{k,1}^*(n, h) = \sum_{0 \leq j_1 + j_2 \leq h} (-1)^{j_1 + j_2} \lambda(n, j_1, j_2) f_k(n-2j_1-3j_2, n-2h-j_2),$$

where

$$\lambda(n, j_1, j_2) = \binom{n-j_1-2j_2}{j_1, j_2, n-2j_1-3j_2}.$$

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$T_{3,2}^{[3]}(n)$	1	1	1	1	1	2	4	9	19	40	82	166	334	678	1394	2905	6111	12961
$T_{3,3}^{[3]}(n)$	1	1	1	1	1	1	1	2	4	8	14	24	40	68	118	209	371	653

Proof. We observe that there exists a mapping from k -noncrossing structures with h arcs with arc-length ≥ 3 and stack-length $\sigma \geq 2$ over $[n]$ into $\dot{\bigcup}_{\sigma-1 \leq b \leq h-1} C_k^*(n-2b, h-b)$:

$$(3.6) \quad c: T_{k,\sigma}^{[3]}(n, h) \rightarrow \dot{\bigcup}_{\sigma-1 \leq b \leq h-1} C_k^*(n-2b, h-b), \quad \delta \mapsto c(\delta)$$

which is obtained in two steps: first induce $c(\delta)$ by mapping arcs and isolated vertices as follows:

$$(3.7) \quad \forall \ell \geq \sigma - 1; \quad ((i - \ell, j + \ell), \dots, (i, j)) \mapsto (i, j) \quad \text{and} \quad j \mapsto j \quad \text{if } j \text{ is an isolated vertex}$$

and secondly relabel the resulting diagram from left to right in increasing order, see Figure 7.

Claim 1. $c: T_{k,\sigma}^{[3]}(n, h) \rightarrow \dot{\bigcup}_{\sigma-1 \leq b \leq h-1} C_k^*(n-2b, h-b)$ is well-defined and surjective.

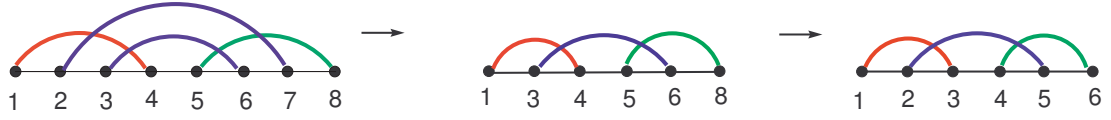


FIGURE 7. The mapping $c: T_{k,\sigma}^{[3]}(n, h) \rightarrow \dot{\bigcup}_{\sigma-1 \leq b \leq h-1} C_k^*(n-2b, h-b)$ is obtained in two steps: first contraction of the stacks and secondly relabeling of the resulting diagram.

By construction, c does not change the crossing number. Since $T_{k,\sigma}^{[3]}(n)$ contains only arcs of length ≥ 3 , we derive $c(T_{k,\sigma}^{[3]}(n)) \subset C_k^*(n-2b, h-b)$. Therefore c is well-defined. It remains to show that c is surjective. For this purpose, let $\delta \in C_k^*(n-2b, h-b)$ and set $a = b - (\sigma-1)(h-b)$. We proceed by constructing a k -noncrossing structure $\tilde{\delta}$ in three steps:

Step 1. replace each label i by r_i , where $r_i \leq r_s$ if and only if $i \leq s$.

Step 2. replace the leftmost arc (r_p, r_q) by the sequence of arcs

$$(3.8) \quad ((\tau_p - ([\sigma-1] + a), \tau_q + ([\sigma-1] + a)), \dots, (\tau_p, \tau_q)),$$

replace any other arc (r_p, r_q) by the sequence

$$(3.9) \quad ((\tau_p - [\sigma-1], \tau_q + [\sigma-1]), \dots, (\tau_p, \tau_q))$$

and each isolated vertex r_s by τ_s .

Step 3. Set for $x, y \in \mathbb{Z}$, $\tau_b + y \leq \tau_c + x$ if and only if $(b < c)$ or $(b = c$ and $y \leq x)$. By construction, \leq is a linear order over

$$n - 2b + 2(h-b)(\sigma-1) + 2a = n - 2b + 2(h-b)(\sigma-1) + 2(b - (\sigma-1)(h-b)) = n$$

elements, which we then label from 1 to n (left to right) in increasing order. It is straightforward to verify that $c(\tilde{\delta}) = \delta$ holds. It remains to show that $\tilde{\delta} \in T_{k,\sigma}^{[3]}(n)$. Suppose *a contrario* $\tilde{\delta}$ contains an arc $(i, i+2)$. Since $\sigma \geq 2$, we can then conclude that $i+1$ is necessarily isolated. The arc $(i, i+2)$ is mapped by c into $(j, j+2)$ with isolated point $j+1$, which is impossible by definition of $C_k^*(n', h')$ and Claim 1 follows.

Labeling the h arcs of $\delta \in T_{k,\sigma}^{[3]}(n, h)$ from left to right and keeping track of multiplicities gives rise to the map

$$(3.10) \quad f_{k,\sigma}: T_{k,\sigma}^{[3]}(n, h) \rightarrow \dot{\bigcup}_{\sigma-1 \leq b \leq h-1} \left[C_k^*(n-2b, h-b) \times \left\{ (a_j)_{1 \leq j \leq h-b} \mid \sum_{j=1}^{h-b} a_j = b, a_j \geq \sigma-1 \right\} \right],$$

given by $f_{k,\sigma}(\delta) = (c(\delta), (a_j)_{1 \leq j \leq h-b})$. We can conclude that $f_{k,\sigma}$ is well-defined and a bijection. We proceed by computing the multiplicities of the resulting core-structures [11]:

$$(3.11) \quad |\{(a_j)_{1 \leq j \leq h-b} \mid \sum_{j=1}^{h-b} a_j = b; a_j \geq \sigma - 1\}| = \binom{b + (2 - \sigma)(h - b) - 1}{h - b - 1}.$$

Eq. (3.11) and eq. (3.10) imply

$$\mathbb{T}_{k,\sigma}^{[3]}(n, h) = \sum_{b=\sigma-1}^{h-1} \binom{b + (2 - \sigma)(h - b) - 1}{h - b - 1} \mathbb{C}_k^*(n - 2b, h - b),$$

whence eq. (3.3). Next we consider the map

$$(3.12) \quad c^* : T_{k,1}^*(n, h) \rightarrow \bigcup_{0 \leq b \leq h-1} C_k^*(n - 2b, h - b), \quad \delta \mapsto c^*(\delta)$$

In the following, we call the 2-arc $(i, i + 2)$ with isolated $i + 1$ a bad-arc. In analogy to the above, each diagram in $T_{k,1}^*(n, h)$ without 1-arcs and bad-arcs can be reduced into a core-structure without 1-arcs and bad arcs in $C_k^*(n', h')$. That gives rise to

$$\mathbb{T}_{k,1}^*(n, h) = \sum_{b=0}^{h-1} \binom{h-1}{b} \mathbb{C}_k^*(n - 2b, h - b).$$

Then, via Möbius-inversion, we obtain eq. (3.4). Recall that $\mathbb{T}_{k,1}^*(n, h)$ counts the number of k -noncrossing partial matchings without 1-arcs and bad-arcs. It is straightforward to show there are $\lambda(n, j_1, j_2) = \binom{n-j_1-2j_2}{j_1, j_2, n-2j_1-3j_2}$ ways to select j_1 1-arc and j_2 bad-arcs over $[n]$. Since removing j_1 1-arc and j_2 bad-arcs by construction removes $2j_1 + 3j_2$ vertices, we observe the number of configurations of at least j_1 1-arc and j_2 bad-arcs is given by $\lambda(n, j_1, j_2) f_k(n - 2j_1 - 3j_2, n - 2h - j_2)$. Via inclusion-exclusion principle, we arrive at

$$\mathbb{T}_{k,1}^*(n, h) = \sum_{0 \leq j_1 + j_2 \leq h} (-1)^{j_1 + j_2} \lambda(n, j_1, j_2) f_k(n - 2j_1 - 3j_2, n - 2h - j_2),$$

whence Theorem 4. □

Remark 1. As for the case of minimal arc length four, the ideas in Theorem 4 work for the enumeration of k -noncrossing RNA structures with minimal arc length 4 and stack size $\sigma \geq 3$, denoted by $\mathbb{T}_{k,\sigma}^{[4]}(n)$. However, in the case $\sigma = 2$, the above strategy fails: the subset of core-structures for generating $T_{k,2}^{[4]}$ -structures is readily identified to have no 1-arcs, 2-arcs and no isolated 3-arcs, i.e., no arcs of form $(i, i + 3)$ where $i, i + 1$ and $i + 2$ as isolated vertices. While the elimination of 1-arcs, bad-arcs and isolated 3-arcs can be dealt with, the difficulty lies in considering 2-arcs of the form $(i, i + 2)$ with non-isolated vertex $i + 1$. When these arcs are being considered, the inclusion-exclusion principle no longer works.

We proceed by proving a functional identity between the bivariate generating functions of $\mathbb{T}_{k,\sigma}^{[3]}(n, h)$ and $\mathbb{C}_k^*(n, h)$. This identity is based on Theorem 4 and is crucial for proving Theorem 5 in Section 4. Its proof is analogous to Lemma 3 in [11].

Lemma 2. *Let $k, \sigma \in \mathbb{N}$, $k \geq 2$ and let u, x be indeterminates. Suppose we have*

$$(3.13) \quad \forall h \geq 1; \quad \mathbb{A}_{k,\sigma}(n, h) = \sum_{b=\sigma-1}^{h-1} \binom{b + (2-\sigma)(h-b) - 1}{h-b-1} \mathbb{B}_k(n-2b, h-b)$$

and $\mathbb{A}_{k,\sigma}(n, 0) = 1$, $\mathbb{B}_k(n, 0) = 1$ for $n \geq 0$. Then we have the functional relation

$$(3.14) \quad \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \mathbb{A}_{k,\sigma}(n, h) u^h x^n = \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \mathbb{B}_k(n, h) \left(\frac{u \cdot (ux^2)^{\sigma-1}}{1-ux^2} \right)^h x^n.$$

Proof. We set $\sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \mathbb{B}_k(n, h) u^h x^n = \sum_{h \geq 0} \varphi_h(x) u^h$ and compute in view of eq. (3.13)

$$(3.15) \quad \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \mathbb{A}_{k,\sigma}(n, h) u^h x^n = \sum_{n \geq 2} \sum_{\substack{h \leq \frac{n}{2} \\ h \geq 1}} \sum_{b \leq h-1} \mathbb{B}_k(n-2b, h-b) \binom{b + (2-\sigma)(h-b) - 1}{h-b-1} u^h x^n + \sum_{n \geq 0} x^n$$

where the term $\sum_{n \geq 0} x^n = \frac{1}{1-x}$ comes from the fact that for $h = 0$ the binomial

$$\binom{b + (2-\sigma)(h-b) - 1}{h-b-1}$$

is zero, while for any $n \geq 0$ the (lhs) counts $\mathbb{A}_{k,\sigma}(n, 0) = 1$. We proceed by computing

$$\begin{aligned} &= \sum_{h \geq 1} \sum_{b \leq h-1} \sum_{n \geq 2h} \mathbb{B}_k(n-2b, h-b) x^{n-2b} \binom{b + (2-\sigma)(h-b) - 1}{h-b-1} u^h x^{2b} + \frac{1}{1-x} \\ &= \sum_{b \geq 0} \sum_{b < h} \varphi_{h-b}(x) \binom{b + (2-\sigma)(h-b) - 1}{h-b-1} u^h x^{2b} + \frac{1}{1-x}. \end{aligned}$$

Setting $m = h - b$ and subsequently interchanging the summation indices, we arrive at

$$\begin{aligned}
\sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} A_{k,\sigma}(n, h) u^h x^n &= \sum_{b \geq 0} \sum_{1 \leq m} \varphi_m(x) \binom{b + (2 - \sigma)m - 1}{m - 1} u^m (ux^2)^b + \frac{1}{1 - x} \\
&= \sum_{m \geq 1} \varphi_m(x) \left(\frac{u \cdot (ux^2)^{\sigma-1}}{1 - ux^2} \right)^m + \frac{1}{1 - x} \\
&= \sum_{n \geq 2} \sum_{\substack{h \leq \frac{n}{2} \\ h \geq 1}} B_k(n, h) \left(\frac{u \cdot (ux^2)^{\sigma-1}}{1 - ux^2} \right)^h x^n + \frac{1}{1 - x} \\
&= \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} B_k(n, h) \left(\frac{u \cdot (ux^2)^{\sigma-1}}{1 - ux^2} \right)^h x^n,
\end{aligned}$$

whence Lemma 2. □

According to Lemma 2 and eq. (3.3) we have

$$(3.16) \quad \mathbb{T}_{k,\sigma}(n, h) = \sum_{b=\sigma-1}^{h-1} \binom{b + (2 - \sigma)(h - b) - 1}{h - b - 1} C_k(n - 2b, h - b)$$

$$(3.17) \quad \mathbb{T}_{k,\sigma}^{[3]}(n, h) = \sum_{b=\sigma-1}^{h-1} \binom{b + (2 - \sigma)(h - b) - 1}{h - b - 1} C_k^*(n - 2b, h - b)$$

$$(3.18) \quad \mathbb{T}_{k,1}^*(n, h) = \sum_{b=0}^{h-1} \binom{h-1}{b} C_k^*(n - 2b, h - b)$$

and Lemma 2 implies the following three functional identities, which are instrumental for the proof of Theorem 5 in Section 4.

$$(3.19) \quad \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \mathbb{T}_{k,\sigma}(n, h) u^h x^n = \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} C_k(n, h) \left(\frac{u \cdot (ux^2)^{\sigma-1}}{1 - ux^2} \right)^h x^n$$

$$(3.20) \quad \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \mathbb{T}_{k,1}^*(n, h) u^h x^n = \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} C_k^*(n, h) \left(\frac{u}{1 - ux^2} \right)^h x^n$$

$$(3.21) \quad \sum_{n \geq 0} \mathbb{T}_{k,\sigma}^{[3]}(n) x^n = \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} C_k^*(n, h) \left(\frac{(x^2)^{\sigma-1}}{1 - x^2} \right)^h x^n \quad \text{for } \sigma \geq 2.$$

4. ASYMPTOTIC ENUMERATION

In this Section, we study the asymptotics of k -noncrossing RNA pseudoknot structures with arc-length ≥ 3 and minimum stack length σ . We are particularly interested in deriving simple formulas, that can be used assessing the complexity of prediction algorithms for k -noncrossing RNA structures. In order to state Theorem 5, below we introduce the following rational function

$$(4.1) \quad w_0(x) = \frac{x^{2\sigma-2}}{1-x^2+x^{2\sigma}}.$$

Theorem 5. *Let $k, \sigma \in \mathbb{N}$, $k, \sigma \geq 2$, x be an indeterminate and ρ_k the dominant, positive real singularity of $\mathbf{F}_k(z)$. Then $\mathsf{T}_{k,\sigma}^{[3]}(n)$, the number of RNA structures with arc-length ≥ 3 and stack-length $\geq \sigma$ satisfies the following identity*

$$(4.2) \quad \sum_{n \geq 0} \mathsf{T}_{k,\sigma}^{[3]}(n)x^n = \frac{1}{1-x+w_0(x)x^2+w_0(x)x^3} \mathbf{F}_k \left(\frac{\sqrt{w_0(x)}x}{1-x+w_0(x)x^2+w_0(x)x^3} \right),$$

where $w_0(x)$ is given by eq. (4.1). Furthermore

$$(4.3) \quad \mathsf{T}_{k,\sigma}^{[3]}(n) \sim n^{-(k-1)^2 - \frac{k-1}{2}} \left(\frac{1}{\gamma_{k,\sigma}^{[3]}} \right)^n, \quad \text{for } k = 3, 4, \dots, 7$$

holds, where $\gamma_{k,\sigma}^{[3]}$ is the positive real dominant singularity of $\sum_{n \geq 0} \mathsf{T}_{k,\sigma}^{[3]}(n)z^n$ and minimal real solution of the equation

$$(4.4) \quad \frac{\sqrt{w_0(x)}x}{1-x+w_0(x)x^2+w_0(x)x^3} = \rho_k = \frac{1}{2(k-1)}$$

and $f_k(2n, 0) \sim n^{-(k-1)^2 - \frac{k-1}{2}} \left(\frac{1}{\rho_k} \right)^{2n}$, see eq. (2.4).

Theorem 5 implies the following growth rates for 3-, 4- and 5-noncrossing RNA structures with arc-length ≥ 3 and stack-length $\geq 2, 3$:

$$\begin{array}{lll} (\gamma_{3,2}^{[3]})^{-1} = 2.5721 & (\gamma_{4,2}^{[3]})^{-1} = 3.0306 & (\gamma_{5,2}^{[3]})^{-1} = 3.4092 \\ (\gamma_{3,3}^{[3]})^{-1} = 2.0392 & (\gamma_{4,3}^{[3]})^{-1} = 2.2663 & (\gamma_{5,3}^{[3]})^{-1} = 2.4442. \end{array}$$

Proof. In the following, we will use the notation $w_0(x)$, eq. (4.1), for short without specifying the variable x . The first step consists in deriving a functional equation relating the bivariate generating functions of $T_k^*(n, h)$ and $f_k(2h', 0)$. For this purpose, we use eq. (3.5).

Claim 1.

$$(4.5) \quad \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \mathsf{T}_{k,1}^*(n, h) w^h x^n = \frac{1}{1 - x + wx^2 + wx^3} \mathbf{F}_k \left(\frac{\sqrt{wx}}{1 - x + wx^2 + wx^3} \right).$$

Set $\varphi_m(w) = \sum_{h \leq \frac{m}{2}} \binom{m}{2h} f_k(2h, 0) w^h$. In order to prove Claim 1, we compute

$$\begin{aligned} & \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \mathsf{T}_{k,1}^*(n, h) w^h x^n \\ &= \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \sum_{0 \leq j_1 + j_2 \leq h} (-1)^{j_1 + j_2} \lambda(n, j_1, j_2) f_k(n - 2j_1 - 3j_2, n - 2h - j_2) w^h x^n \\ &= \sum_{n \geq 0} \sum_{j_1 + j_2 \leq \frac{n}{2}} (-1)^{j_1 + j_2} \lambda(n, j_1, j_2) x^n \sum_{h \geq j_1 + j_2} f_k(n - 2j_1 - 3j_2, n - 2h - j_2) w^h \\ &= \sum_{n \geq 0} \sum_{j_1 + j_2 \leq \frac{n}{2}} (-1)^{j_1 + j_2} \lambda(n, j_1, j_2) w^{j_1 + j_2} \varphi_{n - 2j_1 - 3j_2}(w) x^n. \end{aligned}$$

We interchange the summation over $j_1 + j_2$ and n and arrive at

$$\begin{aligned} & \sum_{j_1 + j_2 \geq 0} \sum_{n \geq 2j_1 + 2j_2} (-1)^{j_1 + j_2} \binom{n - j_1 - 2j_2}{j_1, j_2, n - 2j_1 - 3j_2} w^{j_1 + j_2} \varphi_{n - 2j_1 - 3j_2}(w) x^n \\ &= \sum_{j_1 + j_2 \geq 0} \frac{(-w)^{j_1 + j_2}}{j_1! j_2!} \sum_{n \geq 2j_1 + 3j_2} \frac{(n - j_1 - 2j_2)!}{(n - 2j_1 - 3j_2)!} \varphi_{n - 2j_1 - 3j_2}(w) x^n. \end{aligned}$$

Setting $m = n - 2j_1 - 3j_2$, this becomes

$$\begin{aligned} &= \sum_{j_1 + j_2 \geq 0} \frac{(-w)^{j_1 + j_2}}{j_1! j_2!} x^{2j_1 + 3j_2} \sum_{m \geq 0} \frac{(m + j_1 + j_2)!}{m!} \varphi_m(w) x^m \\ &= \sum_{m \geq 0} \left[\sum_{j_1 + j_2 \geq 0} \binom{m + j_1 + j_2}{m, j_1, j_2} (-wx^2)^{j_1} (-wx^3)^{j_2} \right] \varphi_m(w) x^m \\ &= \sum_{m \geq 0} \varphi_m(w) x^m \left(\frac{1}{1 + wx^2 + wx^3} \right)^{m+1} \\ &= \frac{1}{1 + wx^2 + wx^3} \sum_{m \geq 0} \varphi_m(w) \left(\frac{x}{1 + wx^2 + wx^3} \right)^m. \end{aligned}$$

Next we compute

$$\begin{aligned}
\sum_{m \geq 0} \varphi_m(w)y^m &= \int_0^\infty \sum_{m \geq 0} \varphi_m(w) \frac{(xy)^m}{m!} e^{-x} dx \\
&= \int_0^\infty \det(2\sqrt{wy}x) e^{-(1-y)x} dx \\
&= \sum_{n \geq 0} f_k(2n, 0) \int_0^\infty e^{-(1-y)x} \frac{(\sqrt{wy}x)^{2n}}{(2n)!} d((1-y)x) \\
&= \sum_{n \geq 0} f_k(2n, 0) \frac{(\sqrt{wy})^{2n}}{(2n)!} \frac{\int_0^\infty e^{-(1-y)x} ((1-y)x)^{2n} d((1-y)x)}{(1-y)^{2n+1}} \\
&= \frac{1}{1-y} \mathbf{F}_k \left(\frac{\sqrt{wy}}{1-y} \right).
\end{aligned}$$

Therefore the bivariate generating function can be written as

$$\sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \mathsf{T}_{k,1}^*(n, h) w^h x^n = \frac{1}{1-x+wx^2+wx^3} \mathbf{F}_k \left(\frac{\sqrt{wx}}{1-x+wx^2+wx^3} \right),$$

which is immediately identified as an identity of power series, whence Claim 1. In view of eq. (3.20) we arrive at

$$(4.6) \quad \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \mathsf{C}_k^*(n, h) \left(\frac{w}{1-wx^2} \right)^h x^n$$

$$(4.7) \quad = \frac{1}{1-x+wx^2+wx^3} \mathbf{F}_k \left(\frac{\sqrt{wx}}{1-x+wx^2+wx^3} \right).$$

According to eq. (3.21), we have

$$(4.8) \quad \sum_{n \geq 0} \mathsf{T}_{k,\sigma}^{[3]}(n) x^n = \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \mathsf{C}_k^*(n, h) \left(\frac{(x^2)^{\sigma-1}}{1-x^2} \right)^h x^n$$

and Claim 1 provides, setting

$$(4.9) \quad w_0 = \frac{(x^2)^{\sigma-1}}{1-x^2+x^{2\sigma}},$$

the following interpretation of the (rhs) of eq. (4.6):

$$\begin{aligned}
&\sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \mathsf{C}_k^*(n, h) \left(\frac{(x^2)^{\sigma-1}}{1-x^2} \right)^h x^n \\
&= \frac{1}{1-x+w_0x^2+w_0x^3} \mathbf{F}_k \left(\frac{\sqrt{w_0}x}{1-x+w_0x^2+w_0x^3} \right).
\end{aligned}$$

According to Lemma 1 and eq. (4.8), we have

$$(4.10) \quad \sum_{n \geq 0} \mathsf{T}_{k,\sigma}^{[3]}(n)z^n = \frac{1}{1-z+w_0z^2+w_0z^3} \mathbf{F}_k \left(\frac{\sqrt{w_0}z}{1-z+w_0z^2+w_0z^3} \right).$$

Let us denote $V_k(z) = \sum_{n \geq 0} f_k(2n, 0) \left(\frac{\sqrt{w_0}z}{1-z+w_0z^2+w_0z^3} \right)^{2n}$.

Claim 2. All dominant singularities of $\sum_{n \geq 0} \mathsf{T}_{k,\sigma}^{[3]}(n)z^n$ are singularities of $V_k(z)$. Furthermore the unique, minimal, positive, real solution of

$$(4.11) \quad \vartheta_\sigma(z) = \frac{\sqrt{w_0}z}{1-z+w_0z^2+w_0z^3} = \rho_k = \frac{1}{2(k-1)}, \quad \text{for } k = 3, 4, \dots, 7$$

denoted by $\gamma_{k,\sigma}^{[3]}$ is a dominant singularity of $\sum_{n \geq 0} \mathsf{T}_{k,\sigma}^{[3]}(n)z^n$.

Clearly, a dominant singularity of $\frac{1}{1-z+w_0z^2+w_0z^3} V_k(z)$ is either a singularity of $V_k(z)$ or $\frac{1}{1-z+w_0z^2+w_0z^3}$. Suppose there exists some singularity $\zeta \in \mathbb{C}$ which is a pole of $\frac{1}{1-z+w_0z^2+w_0z^3}$. By construction $\zeta \neq 0$ and ζ is necessarily a non-finite singularity of $V_k(z)$. If $|\zeta| \leq \gamma_{k,\sigma}^{[3]}$, then we arrive at the contradiction $|V_k(\zeta)| > |V_k(\gamma_{k,\sigma}^{[3]})|$, since $V_k(\zeta)$ is not finite and $V_k(\gamma_{k,\sigma}^{[3]}) = \sum_{n \geq 0} f_k(2n, 0) \rho_k^{2n} < \infty$. Therefore all dominant singularities of $\sum_{n \geq 0} \mathsf{T}_{k,\sigma}^{[3]}(n)z^n$ are singularities of $V_k(z)$. According to Pringsheim's Theorem [19], $\sum_{n \geq 0} \mathsf{T}_{k,\sigma}^{[3]}(n)z^n$ has a dominant positive real singularity, which by construction equals $\gamma_{k,\sigma}^{[3]}$ being the minimal positive real solution of eq. (4.11). To prove this, we inspect that for $3 \leq k \leq 7$ (see Remark 2), $\gamma_{k,\sigma}^{[3]}$, has strictly smaller modulus than all solutions of $\vartheta_\sigma(z) = -\rho_k$. Indeed, we observe that, independent of σ , any real positive root of $\vartheta_\sigma(z) = -\rho_k$ must be larger than 1. Accordingly, Theorem 3 applies and we have

$$(4.12) \quad \mathsf{T}_{k,\sigma}^{[3]}(n) \sim K n^{-(k-1)^2 - \frac{k-1}{2}} \left(\frac{1}{\gamma_{k,\sigma}^{[3]}} \right)^n,$$

Hence Claim 2 follows. This completes the proof of Theorem 5. \square

Remark 2. The power series $\mathbf{F}_k(z)$ is D -finite. Accordingly there exists some $e \in \mathbb{N}$ for which $\mathbf{F}_k(z)$ satisfies an ODE of the form

$$(4.13) \quad q_{0,k}(z) \frac{d^e}{dz^e} \mathbf{F}_k(z) + q_{1,k}(z) \frac{d^{e-1}}{dz^{e-1}} \mathbf{F}_k(z) + \dots + q_{e,k}(z) \mathbf{F}_k(z) = 0,$$

where $q_{j,k}(z)$ are polynomials. The key point is that any dominant singularity of $\mathbf{F}_k(z)$ is contained in the set of roots of $q_{0,k}(z)$, which we denote by M_k [17]. In Table 1 we present the polynomials $q_{0,k}(z)$ and their nonzero roots for $k = 3, \dots, 7$. Table 1 validates the reduction to the singularities $\pm \rho_k$. We then verify that for $k = 3, \dots, 7$, $\gamma_{k,\sigma}^{[3]}$ is the unique solution with minimal modulus of $\vartheta_\sigma(x) = \rho_k$ and is strictly smaller than the moduli of the solutions of $\vartheta_\sigma(x) = -\rho_k$.

k	$q_{0,k}(z)$	M_k
3	$(1/4 - 4z^2) z^2$	$\{1/4, -1/4\}$
4	$(144z^4 - 40z^2 + 1) z^6$	$\{1/2, -1/2, 1/6, -1/6\}$
5	$(-80z^2 + 1024z^4 + 1) z^8$	$\{1/4, -1/4, 1/8, -1/8\}$
6	$(-4144z^4 + 140z^2 + 14400z^6 + 1) z^{10}$	$\{1/2, -1/2, 1/6, -1/6, 1/10, -1/10\}$
7	$(-1 - 12544z^4 + 224z^2 + 147456z^6) z^{12}$	$\{1/4, -1/4, 1/8, -1/8, 1/12, -1/12\}$

TABLE 1. The polynomials $q_{0,k}(z)$ and their nonzero roots.

Acknowledgments. We are grateful to Prof. W.Y.C. Chen for stimulating discussions and helpful comments. Particular thanks to the two referees whose suggestions have helped a lot improving the paper. 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

- [1] M. Chamorro, N. Parkin and H.E. Varmus, *An RNA pseudoknots and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA*, Proc. Natl. Acad. Sci. USA, **89**(2)1992, 713–717.
- [2] W.Y.C. Chen, E.Y.P. Deng, R.R.X. Du, R.P. Stanley and C.H. Yan, *Crossings and Nestings of Matchings and Partitions* Trans. Amer. Math. Soc. **359** (2007), No. 4, 1555–1575.
- [3] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, 2008.
- [4] C. Flamm, W. Fontana, I. L. Hofacker and P. Schuster, *RNA folding kinetics at elementary step resolution* RNA, **6**(2000), 325–338.
- [5] D.J. Grabiner and P. Magyar, *Random walks in Weyl chambers and the decomposition of tensor powers* J. Alg. Comb. **2**(1993), 239–260.
- [6] C. Haslinger and P.F. Stadler, *RNA Structures with Pseudo-Knots* Bull. Math. Biol. **61**(1999), 437–467.
- [7] J.A. Howell, T.F. Smith and M.S. Waterman, *Computation of generating functions for biological molecules* SIAM J. Appl. Math. **39**(1980), 119–133.
- [8] F.W.D. Huang, W.W.J. Peng and C. M. Reidys, *Folding 3-noncrossing RNA pseudoknot structures* J. Comp. Biol., accepted, 2008.
- [9] E.Y. Jin, J. Qin and C.M. Reidys, *Combinatorics of RNA structures with pseudoknots* Bull. Math. Biol. **70**(1)(2008), 45–67.
- [10] E.Y. Jin and C.M. Reidys, *Asymptotic enumeration of RNA structures with pseudoknots* Bull. Math. Biol. **70**(4)(2008), 951–970.

- [11] E.Y. Jin and C.M. Reidys, *Combinatorial Design of Pseudoknot RNA*. Adv. Appl. Math. **42**(2009), 135–151.
- [12] E.Y. Jin, C.M. Reidys and R. Wang, *Asymptotic analysis k -noncrossing matchings*, 2008, arXiv:0803.0848
- [13] D.A.M. Konings and R.R. Gutell, *A comparison of thermodynamic foldings with comparatively derived structures of 16S and 16S-like rRNAs* RNA. **1**(1995), 559–574.
- [14] A. Loria and T. Pan, *Domain structure of the ribozyme from eubacterial ribonuclease P* RNA. **2**(1996), 551–563.
- [15] A.M. Odlyzko, *Asymptotic enumeration methods*, Handbook of combinatorics Vol. 2 (1995), 1021–1231, MIT Press.
- [16] *Mapping RNA Form and Function* Science **309**(2005), 1441–1632.
- [17] R.P. Stanley, *Differentiably finite power series* Europ. J. Comb. **1**(1980), 175–188.
- [18] S. Sundaram, *The Cauchy identity for $Sp(2n)$* J. Comb. Theory, Ser. A. **53**(1990), 209–238.
- [19] E.C. Titchmarsh, *The theory of functions* Oxford University Press, London, 1939.
- [20] C. Tuerk, S. MacDougal, and L. Gold, *RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase* Proc. Natl. Acad. Sci. USA, **89**(1992), 6988–6992.
- [21] M.S. Waterman, *Secondary structure of single - stranded nucleic acids* Adv. Math. I (suppl.) **1**(1978), 167–212.
- [22] M.S. Waterman, *Combinatorics of RNA hairpins and cloverleaves* Stud. Appl. Math. **60**(1979), 91–96.
- [23] M. S. Waterman and T. F. Smith, *Rapid dynamic programming algorithms for RNA secondary structure* Adv. Appl. Math., **7**(1986), 455–464.
- [24] E. Westhof and L. Jaeger, *RNA pseudoknots* Curr. Opin. Struct. Biol. **2** (1992), 327–333.