

RNA-LEGO: COMBINATORIAL DESIGN OF PSEUDOKNOT RNA

EMMA Y. JIN AND CHRISTIAN M. REIDYS*

ABSTRACT. In this paper we enumerate k -noncrossing RNA pseudoknot structures with given minimum stack-length, σ . One main result of the paper is the asymptotic formula for their numbers: $T_{k,\sigma}(n) \sim n^{-((k-1)^2+(k-1)/2)} \gamma_{k,\sigma}^{-n}$, where $\gamma_{k,\sigma}$ is explicitly known. Our results show that the numbers of k -noncrossing structures without isolated base pairs are significantly smaller than the numbers of all k -noncrossing structures. In particular we prove that, for large n , the number of 3- and 4-noncrossing RNA structures with stack-length ≥ 2 is given by $311.2470 \frac{4!}{n(n-1)\dots(n-4)} 2.5881^n$ and $1.217 \cdot 10^7 n^{-\frac{21}{2}} 3.0382^n$, respectively. Our results are of importance for prediction algorithms for pseudoknot-RNA and provide evidence for the existence of neutral networks of RNA pseudoknot structures.

1. INTRODUCTION

An RNA structure is the helical configuration of an RNA sequence, described by its primary sequence of nucleotides **A**, **G**, **U** and **C** together with the Watson-Crick (**A-U**, **G-C**) and (**U-G**) base pairing rules. Subject to these single stranded RNA forms helical structures. Since the function of many RNA sequences is oftentimes tantamount to their structures, it is of central importance to understand RNA structure in the context of studying the function of biological RNA, as well as in the design process of artificial RNA. In the following we use a coarse grained notion of structure by concentrating on the pairs of nucleotide positions corresponding to the chemical bonds and ignoring any spatial embedding. There are several ways to represent these RNA structures [21, 31]. We choose the diagram representation [22] which is particularly well suited for displaying the crossings of the Watson-Crick base pairs. A diagram is a labeled graph over the vertex set $[n] = \{1, \dots, n\}$ with vertex degrees ≤ 1 , represented by drawing its vertices

Date: May, 2008.

Key words and phrases. k -noncrossing diagram, generating function, transfer theorem, stack, core-structure, singularity analysis.

$1, \dots, n$ in a horizontal line and its arcs (i, j) , where $i < j$, in the upper half-plane. The vertices and arcs correspond to nucleotides and Watson-Crick (**A-U**, **G-C**) and (**U-G**) base pairs, respectively. We categorize diagrams according to the 3 parameters (k, λ, σ) : 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 $((i, j), (i + 1, j - 1), \dots, (i + (\sigma - 1), j - (\sigma - 1)))$, see Figure 1. We call an arc of length λ a λ -arc.

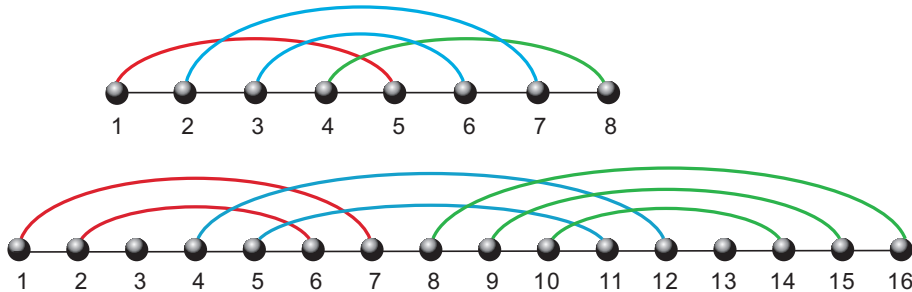


FIGURE 1. k -noncrossing diagrams: in the upper diagram the arcs red/blue/green mutually cross, the arc with minimum length 3 is $(3, 6)$ and the arc $(1, 5)$ and $(4, 8)$ are isolated. Hence this is a 4-noncrossing, $\lambda = 3$, $\sigma = 1$ diagram without isolated vertices. Analogously, below we have a 3-noncrossing (no red/green cross), $\lambda = 4$, $\sigma = 2$ diagram with isolated vertices 3, 13.

In the following we call a k -noncrossing diagram with arc-length ≥ 2 and stack-length $\geq \sigma$ a k -noncrossing RNA structure (of type (k, σ)). We denote the set (number) of k -noncrossing RNA structures of type (k, σ) over $[n]$ 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. Intuitively, a higher number of pairwise crossing arcs is tantamount to higher structural complexity and crossing bonds are reality [20]. These pseudoknot bonds [32] occur in functional RNA (RNaseP [14]), ribosomal RNA [13] and are conserved in the catalytic core of group I introns, see Figure 2, where we show the diagram representation of the catalytic core region of the group I self-splicing intron [3]. For $k = 2$, $\sigma = 1$ we have RNA structures with no 2 crossing arcs, i.e., the well-known RNA secondary structures, whose combinatorics was pioneered by Waterman *et al.* [17, 28, 29, 30, 31].

There are many reasons why pseudoknot structures are fascinating. First, compared to secondary structures their “mathematical” properties are much more intriguing [10, 11, 12]. Their enumeration employs the nontrivial concepts of vacillating tableaux [4, 5] and singular expansions [11, 12]. Second, the recurrence relation for the numbers of 3-noncrossing RNA [10] is, in contrast to that

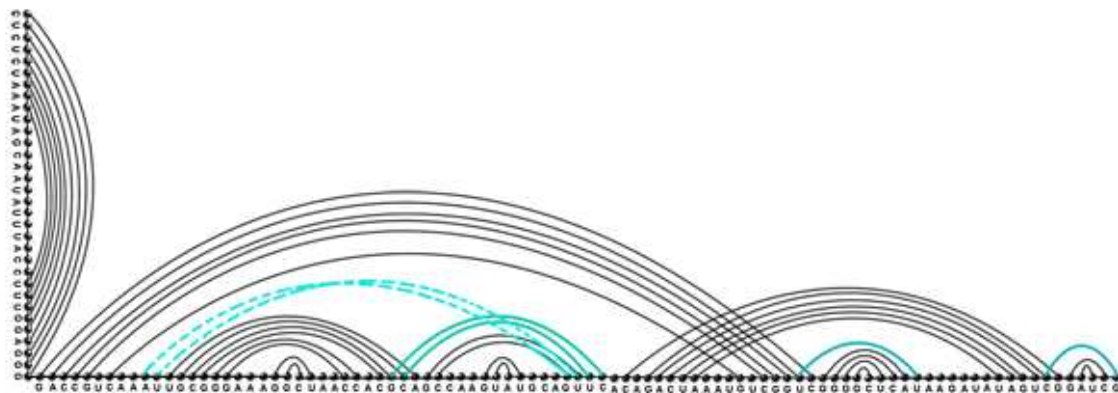


FIGURE 2. Diagram representation of the catalytic core region of the group I self-splicing intron [3]. Six tertiary interactions shown in green lines. The gaps after **G**54, **U**72, **G**103 and **A**112 indicate that some nucleotides are omitted which are involved in an unrelated structural motif.

k	2	3	4	5	6	7	8	9	10
$\gamma_{k,1}^{-1}$	2.6180	4.7913	6.8540	8.8873	10.9087	12.9232	14.9321	16.9405	18.9466

TABLE 1. The exponential growth rates of k -noncrossing RNA structures ($\gamma_{k,1}^{-1}$).

for secondary structures, “enumerative” but not “constructive”. This indicates that prediction of pseudoknot RNA is much more involved compared to the dynamic programming routine used for secondary structures. Nevertheless, there exist several prediction algorithms for RNA pseudoknot structures [1, 15, 19, 26] that are able to express certain “types” of pseudoknots. In this context the notion of the “language of RNA” has been tossed [27]. The combinatorial analysis in [10, 11, 12] shows that 3-noncrossing RNA structures ($T_{3,1}(n)$) exhibit an exponential growth rate of $\frac{5+\sqrt{21}}{2} \approx 4.7913$ and even when considering only structures with minimum arc-length 3 the rate is approximately given by 4.5492. This is bad news, since this rate exceeds already for $k = 3$ the number of sequences over the natural alphabet. Therefore, *a priori*, not all 3-noncrossing structures can be folded by sequences. The situation becomes worse for higher k : in Table 1 we display the results of [11, 12], on the exponential growth rates $\gamma_{k,1}^{-1}$ for k -noncrossing RNA structures ¹.

¹here $\gamma_{k,1}$ denotes the dominant real singularity of the generating function

k	2	3	4	5	6	7	8	9	10
$\gamma_{k,1}^{-1}$	2.6180	4.7913	6.8540	8.8873	10.9087	12.9232	14.9321	16.9405	18.9466
$\gamma_{k,2}^{-1}$	1.9680	2.5881	3.0382	3.4138	3.7439	4.0420	4.3159	4.5714	4.8114

TABLE 2. The exponential growth rates of arbitrary ($\gamma_{k,1}^{-1}$) versus canonical ($\gamma_{k,2}^{-1}$) k -noncrossing RNA structures.

These data give rise to the question: can we identify and analyze those k -noncrossing structures that do “occur”? To this end, let us consider this question in the biophysical context: RNA structures are formed by Watson-Crick (**A-U**, **G-C**) and (**U-G**) base pairs and, due to the specific chemistry of the latter, parallel bonds are thermodynamically more stable. This fact is well-known and has led to the notion of “canonical” structures [23], i.e., structures in which there exist *no* isolated base pair, i.e., $\sigma = 2$ see Figure 3. The question then is, do canonical k -noncrossing structures exhibit

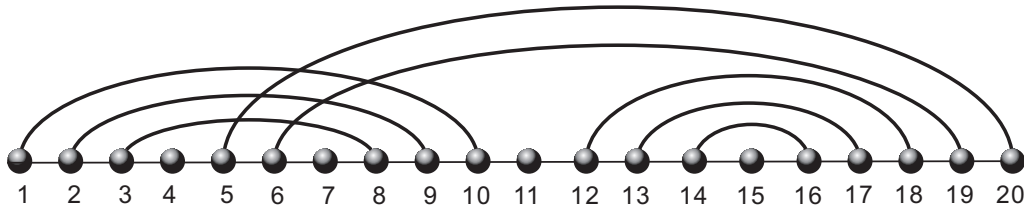


FIGURE 3. A canonical structure

significantly smaller growth rates? In the following we will develop a combinatorial framework which allows us to enumerate any RNA structure class of type (k, σ) , for any k, σ . We then can report good news: as we can show in Table 2 there is a significant drop in exponential growth rates when passing from k -noncrossing RNA structures to their canonical counterparts for $k \geq 3$.

where the case $k = 2$ is due to [6], which is independently confirmed by our approach. In particular, for 3-noncrossing RNA structures, we have a drop in exponential growth rate from 4.7913 to 2.5881, more than 45% and for $k = 10$ there is a drop of more than 74%. As a result, the number of canonical 3-noncrossing RNA structures is very close to that of arbitrary secondary structures.

These small growth rates imply the existence of neutral networks in analogy to the situation for mappings into RNA secondary structures [18].

Intuitively this makes perfect sense since canonicity implies *parallel* arcs which severely limits crossings and it can be expected to have dramatic effect on k -noncrossing RNA for large k . In other words, the biophysical constraints (thermodynamic stability) counteract the combinatorial variety, see Figure 4.

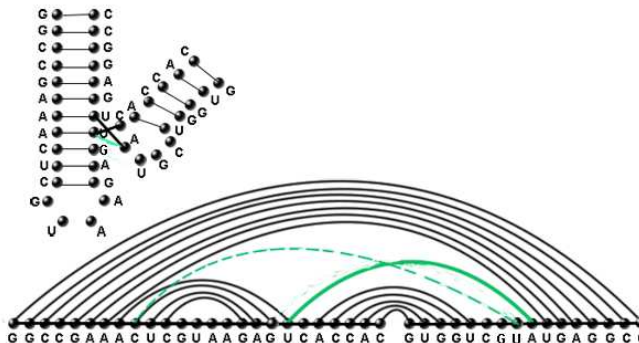


FIGURE 4. Biophysical constraints inducing parallel arcs: the hammerhead ribozyme [2]. Its two tertiary interactions are shown in green lines, The gap after C25 indicates that some nucleotides are omitted, which are involved in an unrelated structural motif.

For our analysis of k -noncrossing canonical structures we consider a new type of k -noncrossing structure. The latter being somewhat “dual” to canonical structures. We consider k -noncrossing structures in which there exists *no* two arcs of the form $(i, j), (i + 1, j - 1)$. These structures are called k -noncrossing core-structures and we use $C_k(n)$ and $C_k(n)$ to denote their set and their number, respectively. We are grateful to a referee pointing out that the idea of core-structures is not new: M.E. Nebel used core-structures in [16] in order to analyze RNA secondary structures of order p . Furthermore, similar ideas appeared in a different context in [7]. The key observation with respect to core-structures is the following: any structure has a unique core obtained by identifying all arcs contained in stacks by a single arc and keeping isolated vertices. In addition, the number of all structures is a sum of the number of the corresponding core-structures with *positive* integer coefficients. In Figure 5 we illustrate the idea of how a core-structure is obtained. It is of particular interest to note that Figure 5 shows that deriving the core-structure can reduce

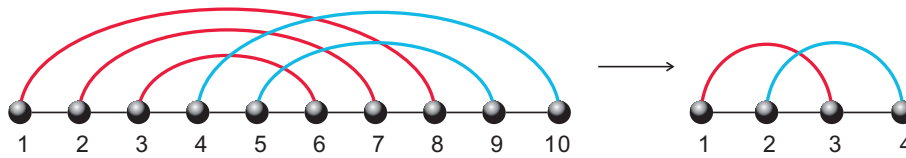


FIGURE 5. Core-structures. Each sequence of stacked arcs in the 3-noncrossing (canonical) structure (lhs) is replaced by its unique arc with minimal length (rhs). The so derived core-structure is unique.

k	2	3	4	5	6	7	8	9	10
$\gamma_{k,1}^{-1}$	2.6180	4.7913	6.8540	8.8873	10.9087	12.9232	14.9321	16.9405	18.9466
κ_k^{-1}	2.5152	4.7097	6.7921	8.8378	10.8672	12.8866	14.9031	16.9119	18.9215

TABLE 3. The exponential growth rates of k -noncrossing RNA structures ($\gamma_{k,1}^{-1}$) versus k -noncrossing core-structures (κ_k^{-1}).

the minimum arc-length, but cannot produce arcs of the form $(i, i + 1)$. In Theorem 4 we derive the generating function for core-structures which shows that “most” k -noncrossing structures are in fact core-structures, see Table 3.

In Theorem 5 we derive a functional identity for the generating function for k -noncrossing RNA structures with stack-length $\geq \sigma$, which allows us to obtain exact and asymptotic results on $\mathbb{T}_{k,\sigma}(n)$, i.e., the number of all k -noncrossing RNA structures with stack-length $\geq \sigma$. In its proof the number of k -noncrossing core-structures plays a central role.

2. SOME BASIC FACTS

In this section we provide the basic facts needed for proving Theorem 4 in section 3 and Theorem 5 in section 4. For background on crossings and nestings in diagrams and partitions we recommend the paper of Chen *et al.* [4] and for analytic combinatorics and asymptotic analysis the book of Flajolet [8]. Our results are based on the generating function of k -noncrossing RNA structures

[10], and asymptotic analysis of k -noncrossing RNA structures [11, 12], summarized in Theorem 1 below.

First let us fix some notation. $T_{k,\sigma}(n)$ denotes the set of k -noncrossing RNA structures with minimum stack-length σ and $\mathbb{T}_{k,\sigma}(n)$ denotes their number. That is, $T_{k,\sigma}(n)$ can be identified with the set of diagrams with vertex degrees ≤ 1 , represented by drawing the vertices $1, \dots, n$ in a horizontal line and its arcs (i, j) , where $i < j$, in the upper half-plane with arc-length ≥ 2 and stack-length $\geq \sigma$, in which the maximum number of mutually crossing arcs is $k - 1$. Furthermore let $T_{k,\sigma}(n, h)$ denote the set of k -noncrossing RNA structures with stack-length $\geq \sigma$ having h arcs over $[n]$ and let $\mathbb{T}_{k,\sigma}(n, h)$ denote their number. We denote by $f_k(n, \ell)$ the number of k -noncrossing diagrams with arbitrary arc-length and ℓ isolated vertices. In Figure 6 we display the various types of diagrams involved.

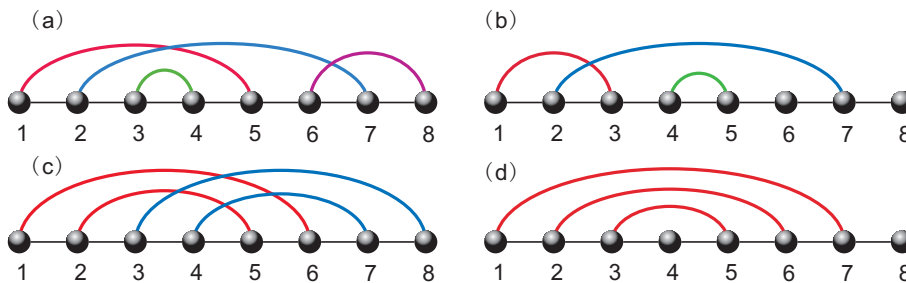


FIGURE 6. Basic diagram types: (a) matching ($f_3(8,0)$), (b) partial matching with 1-arc $(4,5)$ and isolated vertices $6,8$ ($f_3(8,2)$), (c) structure (minimum arc-length ≥ 2) with minimum stack-length 2 and no isolated vertices ($\mathbb{T}_{3,2}(8)$) and (d) structure with minimum arc-length 2 and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

The following identities are due to Grabiner and Magyar [9]

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

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

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. The second formula results from a determinant formula enumerating pairs of nonintersecting Dyck-paths. In view of

$$f_k(n, \ell) = \binom{n}{\ell} f_k(n - \ell, 0)$$

everything can be reduced to matchings.

As for the generating function and asymptotics of k -noncrossing RNA structures we have the following result

Theorem 1. [10, 11] *Let $k \in \mathbb{N}$, $k \geq 2$. Then the number of k -noncrossing RNA structures with $\binom{n-\ell}{2}$ arcs, $\mathsf{T}_{k,1}(n, \frac{n-\ell}{2})$ and the number of k -noncrossing RNA structures, $\mathsf{T}_{k,1}(n)$ are given by*

$$(2.4) \quad \mathsf{T}_{k,1}(n, \frac{n-\ell}{2}) = \sum_{b=0}^{\lfloor n/2 \rfloor} (-1)^b \binom{n-b}{b} f_k(n-2b, \ell)$$

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

where $\{\sum_{\ell=0}^{n-2b} f_k(n-2b, \ell)\}$ is the number of k -noncrossing partial matchings and given via eq. (2.2). Furthermore we have

$$\mathsf{T}_{3,1}(n) \sim \frac{10.4724 \cdot 4!}{n(n-1) \cdots (n-4)} \left(\frac{5 + \sqrt{21}}{2} \right)^n.$$

The following functional identity is due to [11] and relates the bivariate generating function for $\mathsf{T}_{k,1}(n, h)$, the number of k -noncrossing RNA pseudoknot structures with h arcs to the generating function of k -noncrossing matchings.

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

$$(2.6) \quad \sum_{n \geq 0} \sum_{h \leq n/2} \mathsf{T}_{k,1}(n, h) u^{2h} z^n = \frac{1}{u^2 z^2 - z + 1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{uz}{u^2 z^2 - z + 1} \right)^{2n}.$$

In particular we have for $u = 1$,

$$(2.7) \quad \sum_{n \geq 0} \mathbb{T}_{k,1}(n) z^n = \frac{1}{z^2 - z + 1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{z}{z^2 - z + 1} \right)^{2n} .$$

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 [8]. 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. Details on the method can be found in [8] and its application to 3-noncrossing RNA in [11]. To be precise, 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

$$(2.8) \quad \Delta_\rho(\phi, R) = \{z \mid |z| < R, z \neq \rho, |\text{Arg}(z - \rho)| > \phi\}$$

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

$$(2.9) \quad (f(z) = O(g(z)) \text{ as } z \rightarrow \rho) \iff (f(z)/g(z) \text{ is bounded as } z \rightarrow \rho)$$

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. A powerseries $A(z) = \sum_{n \geq 0} a(n)z^n$ is called D -finite if $A(z)$ satisfies on ODE with polynomial coefficients

$$(2.10) \quad q_0(z) \frac{d^e}{dz^e} A(z) + q_1(z) \frac{d^{e-1}}{dz^{e-1}} A(z) + \dots + q_e(z) A(z) = 0, \quad \text{where } e \in \mathbb{N}.$$

Theorem 2. [8] *Let $f(z), g(z)$ be D -finite, Δ_ρ -analytic functions with unique dominant singularity ρ and suppose*

$$(2.11) \quad f(z) = O(g(z)) \quad \text{for } z \rightarrow \rho .$$

Then we have

$$(2.12) \quad [z^n]f(z) = K \left(1 - O\left(\frac{1}{n}\right) \right) [z^n]g(z)$$

where K is some constant.

Let $F_k(z) = \sum_n f_k(2n, 0)z^{2n}$, the ordinary generating function of k -noncrossing matchings. It follows from eq. (2.1) that the power series $F_k(z)$ is D -finite, i.e. there exists some $e \in \mathbb{N}$ such that

$$(2.13) \quad q_{0,k}(z) \frac{d^e}{dz^e} F_k(z) + q_{1,k}(z) \frac{d^{e-1}}{dz^{e-1}} F_k(z) + \dots + q_{e,k}(z) F_k(z) = 0 ,$$

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

where $q_{j,k}(z)$ are polynomials. The key point is that any dominant singularity of $F_k(z)$ is contained in the set of roots of $q_{0,k}(z)$ [24], which we denote by M_k . The polynomials $q_{0,k}(z)$ and their sets of roots for $k = 3, \dots, 7$ are given in Table 4.

Hence $F_k(z)$ has singularities $\pm\rho_k$, where $\rho_k = (2(k-1))^{-1}$. As a consequence of Theorem 2 we have the following result tailored for our functional equations.

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 unique solution with minimal modulus $< \delta$ of the two equations $\vartheta_\sigma(x) = \rho_k$ and $\vartheta_\sigma(x) = -\rho_k$. Then*

$$(2.14) \quad [z^n] F_k(\vartheta_\sigma(z)) \sim c_k n^{-((k-1)^2 + (k-1)/2)} \left(\gamma_{k,\sigma}^{-1} \right)^n.$$

Since $\vartheta_\sigma(z)$ is algebraic over $K(z)$ and satisfies $\vartheta_\sigma(0) = 0$ we can conclude that the composition $F_k(\vartheta_\sigma(z))$ is D -finite [24]. In particular $F_k(\vartheta_\sigma(z))$ has a singular expansion. Since $F_k(z)$ has the two dominant singularities $\pm\rho_k$ and $\gamma_{k,\sigma}$ is the unique solution with minimal modulus of the two equations $\vartheta_\sigma(x) = \rho_k$ and $\vartheta_\sigma(x) = -\rho_k$ we can conclude that $\gamma_{k,\sigma}$ is the unique dominant singularity of $F_k(\vartheta_\sigma(z))$. We proceed by studying the singular expansion in general. According to Theorem 2 in [27] we have

$$(2.15) \quad f_k(n) \sim \tilde{c}_k n^{-((k-1)^2 + (k-1)/2)} (2(k-1))^{2n}$$

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

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

in accordance with basic structure theorems for singularities of solutions of $|\vartheta_\sigma(z)| = \rho_k$ [8], p. 499. Since $\vartheta_\sigma(z)$ is regular at $\gamma_{k,\sigma}$ we are given the supercritical case of singularity analysis [8]. In the supercritical case the subexponential factors of the compositum, $F_k(\vartheta_\sigma(z))$ coincide with those of the outer function, $F_k(z)$. Consequently we have

$$(2.16) \quad [z^n] F_k(\vartheta_\sigma(z)) \sim c_k n^{-((k-1)^2+(k-1)/2)} \left(\gamma_{k,\sigma}^{-1} \right)^n,$$

whence Theorem 3.

3. CORE-STRUCTURES

As discussed in the introduction, a core-structure is a k -noncrossing structure with no stacked base pairs. We denote the set and number of core-structures over $[n]$ by $C_k(n)$ and $\mathbf{C}_k(n)$, respectively. Analogously $C_k(n, h)$ and $\mathbf{C}_k(n, h)$ denote the set and the number of core-structures having h arcs. In Lemma 2 below we establish that the number of all k -noncrossing structures with stack-length $\geq \sigma$ is a sum of the number of k -noncrossing cores with positive integer coefficients.

Lemma 2. (Core-lemma) *For $k, h, \sigma \in \mathbb{N}$, $k \geq 2$, $1 \leq h \leq n/2$ we have*

$$(3.1) \quad \mathbf{T}_{k,\sigma}(n, h) = \sum_{b=\sigma-1}^{h-1} \binom{b + (2-\sigma)(h-b) - 1}{h-b-1} \mathbf{C}_k(n-2b, h-b).$$

Remark 1. Lemma 2 cannot be used in order to enumerate diagrams with arc-length $\geq \lambda$, where $\lambda > 2$ and stack-length σ . Basically, k -noncrossing structures with arc-length $\geq \lambda$ have core-structures with arc-length 2, see Figure 7. The enumeration of k -noncrossing RNA structures with arc-length ≥ 3 and stack-length ≥ 2 is work in progress.

Proof. First, there exists a mapping from k -noncrossing structures with h arcs and minimum stack size σ over $[n]$ into core-structures:

$$(3.2) \quad c: T_{k,\sigma}(n, h) \rightarrow \bigcup_{0 \leq b \leq h-1} C_k(n-2b, h-b), \quad \delta \mapsto c(\delta)$$

where the core-structure $c(\delta)$ is obtained in two steps: first we map arcs and isolated vertices as follows

$$(3.3) \quad \forall \ell \geq \sigma - 1; \quad ((i - \ell, j + \ell), \dots, (i, j)) \mapsto (i, j) \quad \text{and} \quad j \mapsto j \quad \text{if } j \text{ is isolated.}$$

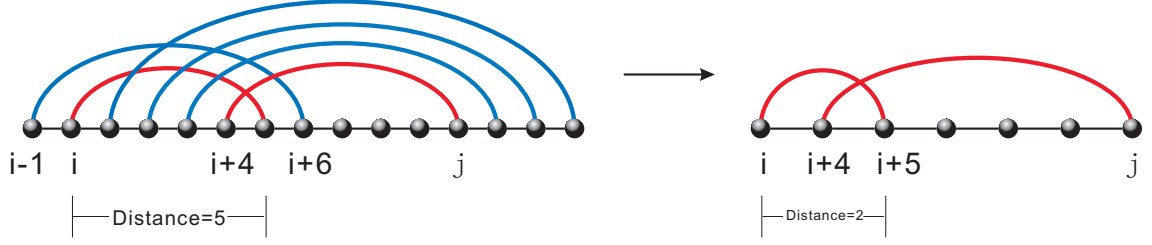


FIGURE 7. Core-structures will in general have 2-arcs: the structure $\delta \in T_{3,2}(19)$ (lhs) is mapped into its core $c(\delta)$ (rhs). Clearly δ has arc-length ≥ 5 and as a consequence of the collapse of the stack $((i+1, j+3), \dots, (i+4, j))$ (the blue arcs are being removed) into the arc $(i+4, j)$, $c(\delta)$ contains the 2-arc $(i, i+5)$.

Second we relabel the vertices of the resulting diagram from left to right in increasing order. That is we replace each stack by a single arc and keep isolated vertices and then relabel, see Figure 8. We have to prove that $c: T_{k,\sigma}(n, h) \rightarrow \dot{\bigcup}_{0 \leq b \leq h-1} C_k(n-2b, h-b)$ is well-defined, i.e., that c

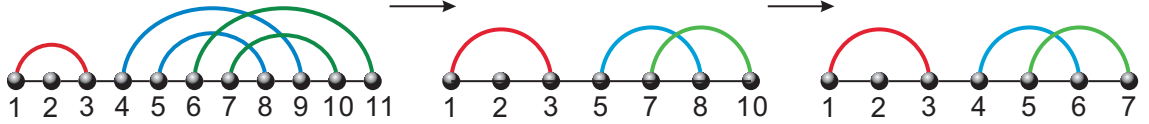


FIGURE 8. The mapping $c: T_{k,\sigma}(n, h) \rightarrow \dot{\bigcup}_{0 \leq b \leq h-1} C_k(n-2b, h-b)$ is obtained in two steps: first contraction of the stacks and secondly relabeling of the resulting diagram.

cannot produce 1-arcs. Indeed, since $\delta \in T_{k,\sigma}(n, h)$, δ does not contain 1-arcs we can conclude that $c(\delta)$ has by construction arcs of length ≥ 2 . c is by construction surjective. Keeping track of multiplicities gives rise to the map

$$(3.4) \quad f_{k,\sigma}: T_{k,\sigma}(n, h) \rightarrow \dot{\bigcup}_{0 \leq b \leq h-1} \left[C_k(n-2b, h-b) \times \left\{ (a_j)_{1 \leq j \leq h-b} \mid \sum_{j=1}^{h-b} a_j = b, a_j \geq \sigma - 1 \right\} \right],$$

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 computing the multiplicities of the resulting core-structures:

claim.

$$(3.5) \quad |\{(a_j)_{1 \leq j \leq h-b} \mid \sum_{j=1}^{h-b} a_j = b; a_j \geq \sigma - 1\}| = \binom{b + (2 - \sigma)(h - b) - 1}{h - b - 1}.$$

Clearly, $a_j \geq \sigma - 1$ is equivalent to $\mu_j = a_j - \sigma + 2 \geq 1$ and we have

$$\sum_{j=1}^{h-b} \mu_j = \sum_{j=1}^{h-b} (a_j - \sigma + 2) = b + (2 - \sigma)(h - b) .$$

We next show that

$$(3.6) \quad |\{(\mu_j)_{1 \leq j \leq h-b}\} \mid \sum_{j=1}^{h-b} \mu_j = b + (2 - \sigma)(h - b); \mu_j \geq 1|$$

is equal the number of $(h - b - 1)$ -subset in $\{1, 2, \dots, b + (2 - \sigma)(h - b) - 1\}$. Consider the set

$$(3.7) \quad \{\mu_1, \mu_1 + \mu_2, \dots, \mu_1 + \mu_2 + \dots + \mu_{h-b-1}\}$$

consisting of $h - b - 1$ distinct elements of $[b + (2 - \sigma)(h - b) - 1] = \{1, 2, \dots, b + (2 - \sigma)(h - b) - 1\}$. Therefore $\{\mu_1, \mu_1 + \mu_2, \dots, \mu_1 + \mu_2 + \dots + \mu_{h-b-1}\}$ is a $(h - b - 1)$ -subset of $[b + (2 - \sigma)(h - b) - 1]$. Given any $(h - b - 1)$ -subset of $[b + (2 - \sigma)(h - b) - 1]$, we can arrange its elements in linear order and retrieve the sequence $\{\mu_i \mid 1 \leq i \leq h - b\}$ of positive integers with sum $b + (2 - \sigma)(h - b)$. Therefore the above assignment is a bijection. Since the number of $(h - b - 1)$ -subset of $[b + (2 - \sigma)(h - b) - 1]$ is given by $\binom{b + (2 - \sigma)(h - b) - 1}{h - b - 1}$ the claim follows.

We can conclude from the claim and eq. (3.4) that

$$(3.8) \quad \mathbb{T}_{k,\sigma}(n, h) = \sum_{b=\sigma-1}^{h-1} \binom{b + (2 - \sigma)(h - b) - 1}{h - b - 1} \mathbb{C}_k(n - 2b, h - b)$$

holds and the lemma follows. \square

Next, we prove a functional identity between the bivariate generating functions of $\mathbb{T}_{k,\sigma}(n, h)$ and $\mathbb{C}_k(n, h)$. This identity plays a central role in proving Theorem 4 and Theorem 5 in section 4.

Lemma 3. *Let $k, \sigma \in \mathbb{N}$, $k \geq 2$ and let u, x be indeterminants. Then we have the functional relation*

$$(3.9) \quad \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathbb{T}_{k,\sigma}(n, h) u^h x^n = \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathbb{C}_k(n, h) \left(\frac{u \cdot (ux^2)^{\sigma-1}}{1 - ux^2} \right)^h x^n$$

and in particular, for $u = 1$

$$(3.10) \quad \sum_{n \geq 0} \mathbb{T}_{k,\sigma}(n) x^n = \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathbb{C}_k(n, h) \left(\frac{(x^2)^{\sigma-1}}{1 - x^2} \right)^h x^n .$$

Proof. We set $\sum_{n \geq 0} \sum_{1 \leq h \leq \frac{n}{2}} \mathbf{C}_k(n, h) u^h x^n = \sum_{n \geq 2} \sum_{1 \leq h \leq \frac{n}{2}} \mathbf{C}_k(n, h) u^h x^n = \sum_{h \geq 1} \varphi_h(x) u^h$ and proceed by deducing a functional equation for $\sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathbf{T}_{k, \sigma}(n, h) u^h x^n$ via Lemma 2. For this purpose we note that for $h = 0$ the Binomial coefficient

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

is zero, while the term $\mathbf{T}_{k, \sigma}(n, 0) = 1$ for $n \geq 1$. Clearly, $\mathbf{T}_{k, \sigma}(n, 0)$ counts for each $n \geq 1$ the “open” structure consisting of isolated vertices. We accordingly have to extend the identity of Lemma 2

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

to the case $h = 0, n \geq 1$ which gives rise to the term $\sum_{n \geq 1} x^n = \frac{x}{1 - x}$. Accordingly, we derive

$$(3.11) \quad \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathbf{T}_{k, \sigma}(n, h) u^h x^n = \sum_{n \geq 2} \sum_{1 \leq h \leq \frac{n}{2}} \sum_{b \leq h - 1} \mathbf{C}_k(n - 2b, h - b) \binom{b + (2 - \sigma)(h - b) - 1}{h - b - 1} u^h x^n + \sum_{n \geq 0} x^n.$$

We proceed by rewriting the (rhs) of eq. (3.11)

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

Setting $m = h - b$ (note that $m \geq 1$) and subsequently interchanging the summation indices we arrive at

$$(3.12) \quad \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathbf{T}_{k, \sigma}(n, h) u^h x^n = \sum_{b \geq 0} \sum_{m \geq 1} \varphi_m(x) \binom{b + (2 - \sigma)m - 1}{m - 1} u^m (ux^2)^b + \frac{1}{1 - x}.$$

Since $\binom{b + (2 - \sigma)m - 1}{m - 1}$ is only nonzero if and only if $b \geq (\sigma - 1)m$ we obtain

$$(3.13) \quad \sum_{b \geq m(\sigma - 1)} \binom{b + (2 - \sigma)m - 1}{m - 1} (ux^2)^b = (ux^2)^{(\sigma - 2)m + 1} \sum_{\omega \geq m - 1} \binom{\omega}{m - 1} (ux^2)^\omega,$$

where $\omega = b + (2 - \sigma)m - 1$. Using $\sum_{a \geq j} \binom{a}{j} x^a = \frac{x^j}{(1 - x)^{j + 1}}$ we find

$$\sum_{b \geq m(\sigma - 1)} \binom{b + (2 - \sigma)m - 1}{m - 1} (ux^2)^b = (ux^2)^{(\sigma - 2)m + 1} \frac{(ux^2)^{m - 1}}{(1 - ux^2)^m} = \left(\frac{(ux^2)^{\sigma - 1}}{1 - ux^2} \right)^m.$$

Accordingly, we can express the (rhs) of eq. (3.12) as follows

$$\begin{aligned}
\sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathsf{T}_{k,\sigma}(n, h) u^h x^n &= \sum_{m \geq 1} \varphi_m(x) \left(\frac{u \cdot (ux^2)^{\sigma-1}}{1 - ux^2} \right)^m + \frac{1}{1-x} \\
&= \sum_{n \geq 0} \sum_{1 \leq h \leq \frac{n}{2}} \mathsf{C}_k(n, h) \left(\frac{u \cdot (ux^2)^{\sigma-1}}{1 - ux^2} \right)^h x^n + \frac{1}{1-x} \\
&= \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathsf{C}_k(n, h) \left(\frac{u \cdot (ux^2)^{\sigma-1}}{1 - ux^2} \right)^h x^n,
\end{aligned}$$

whence Lemma 3. □

We next enumerate core-structures. Theorem 4 has two main parts, the first being the “inversion” of Lemma 2. It allows us to express core-structures via all structures and follows by Möbius inversion. The second part deals with the asymptotics of core-structures. The asymptotic formula follows by applying transfer Theorems (the super-critical case) [8] to some version of the functional identity of Lemma 1.

Theorem 4. (Core-structures) *Suppose $k \in \mathbb{N}$, $k \geq 2$, let x be an indeterminate, ρ_k the dominant, positive real singularity of $\sum_{n \geq 0} f_k(2n, 0) z^{2n}$ and $u_1(x) = \frac{1}{1+x^2}$. Then for $h \geq 1$, the numbers of k -noncrossing core-structures, $\mathsf{C}_k(n, h)$ are given by*

$$(3.14) \quad \mathsf{C}_k(n, h) = \sum_{b=0}^{h-1} (-1)^{h-b-1} \binom{h-1}{b} \mathsf{T}_{k,1}(n-2h+2b+2, b+1).$$

Furthermore we have the functional equation

$$(3.15) \quad \sum_{n \geq 0} \mathsf{C}_k(n) x^n = \frac{1}{u_1 x^2 - x + 1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{\sqrt{u_1} x}{u_1 x^2 - x + 1} \right)^{2n}$$

and the asymptotic formula

$$(3.16) \quad \mathsf{C}_k(n) \sim n^{-((k-1)^2 + (k-1)/2)} \left(\frac{1}{\kappa_k} \right)^n, \quad \text{for } k = 3, 4, \dots, 7$$

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

In the following, we present the number of 3- and 4-noncrossing core-structures for $n = 1, \dots, 15$:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$C_3(n)$	1	1	2	5	12	31	88	263	814	2604	8575	28936	99726	350151	1249865
$C_4(n)$	1	1	2	5	12	32	95	301	1001	3495	12708	47932	186581	747619	3073207

Proof. We set

$$\begin{aligned} \forall 0 \leq i \leq h-1; \quad a(i) &= C_k(n-2(h-1-i), i+1) \\ \forall 0 \leq i \leq h-1; \quad b(i) &= T_{k,1}(n-2(h-1-i), i+1). \end{aligned}$$

We first employ Lemma 2 for $\sigma = 1$:

$$T_k(n, h) = \sum_{b=0}^{h-1} \binom{h-1}{b} C_k(n-2b, h-b) \iff b(h-1) = \sum_{i=0}^{h-1} \binom{h-1}{i} a(i).$$

Via Möbius-inversion we arrive at $a(h-1) = \sum_{i=0}^{h-1} (-1)^{h-1-i} \binom{h-1}{i} b(i)$, which is equivalent to

$$(3.17) \quad C_k(n, h) = \sum_{b=0}^{h-1} (-1)^{h-b-1} \binom{h-1}{b} T_{k,1}(n-2h+2b+2, b+1),$$

whence eq. (3.14). We proceed by proving eq. (3.15). First Lemma 3 implies:

$$(3.18) \quad \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} T_{k,1}(n, h) u^h x^n = \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} C_k(n, h) \left(\frac{u}{1-ux^2} \right)^h x^n$$

and we inspect that $u_1(x) = \frac{1}{1+x^2}$ is the unique solution for $\frac{u}{1-ux^2} = 1$. Accordingly, we obtain

$$\sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} T_{k,1}(n, h) u_1^h x^n = \sum_{n \geq 0} C_k(n) x^n.$$

Second, setting $u = \sqrt{u_1}$ in Lemma 1, provides an interpretation of the lhs of eq. (3.18):

$$(3.19) \quad \sum_{n \geq 0} \sum_{h \leq n/2} T_{k,1}(n, h) u_1^h x^n = \frac{1}{u_1 x^2 - x + 1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{\sqrt{u_1} x}{u_1 x^2 - x + 1} \right)^{2n}$$

and we can conclude

$$\begin{aligned} \sum_{n \geq 0} C_k(n) x^n &= \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} T_{k,1}(n, h) u_1^h x^n \\ &= \frac{1}{u_1 x^2 - x + 1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{\sqrt{u_1} x}{u_1 x^2 - x + 1} \right)^{2n}, \end{aligned}$$

whence eq. (3.15). As for eq. (3.16) we consider the functional equation

$$\sum_{n \geq 0} \mathbf{C}_k(n) x^n = \frac{1}{u_1 x^2 - x + 1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{\sqrt{u_1} x}{u_1 x^2 - x + 1} \right)^{2n}.$$

claim. All dominant singularities of $\sum_{n \geq 0} \mathbf{C}_k(n) x^n$ are dominant singularities of

$$W_k(x) = \sum_{n \geq 0} f_k(2n, 0) \left(\frac{\sqrt{u_1} x}{u_1 x^2 - x + 1} \right)^{2n}.$$

To prove the claim we observe that a dominant singularity of

$$\frac{1}{u_1 x^2 - x + 1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{\sqrt{u_1} x}{u_1 x^2 - x + 1} \right)^{2n}$$

is either a singularity of $W_k(x)$ or $\frac{1}{u_1 x^2 - x + 1}$. Suppose there exists some singularity $\zeta \in \mathbb{C}$ which is a root of $u_1 x^2 - x + 1$. By construction $\zeta \neq 0$ and ζ is necessarily a singularity of $W_k(x)$. Pringsheim's Theorem [25] guarantees that $\sum_{n \geq 0} \mathbf{C}_k(n) x^n$ has a dominant real positive singularity κ_k . Suppose $|\zeta| \leq \kappa_k$ then we arrive at the contradiction $|W_k(\zeta)| > W_k(\kappa_k)$ since $W_k(\zeta)$ is not finite and $W_k(\kappa_k) = \sum_{n \geq 0} f_k(2n, 0) \rho_k^{2n} < \infty$. Therefore all dominant singularities of $\sum_{n \geq 0} \mathbf{C}_k(n) x^n$ are dominant singularities of $W_k(x)$ and the claim follows.

The claim immediately implies that the exponential growth rate is the inverse of κ_k , the minimal positive real solution of the equation $\frac{\sqrt{u_1} x}{u_1 x^2 - x + 1} = \rho_k$. Using Table 4 of section 2 it is straightforward to verify that Theorem 3 applies for $k = 3, 4, \dots, 7$. I.e. we have

$$(3.20) \quad \mathbf{C}_k(n) \sim n^{-((k-1)^2 + (k-1)/2)} (\kappa_k^{-1})^n.$$

This completes the proof of Theorem 4. □

4. PSEUDOKNOT RNA WITH STACK-LENGTH $\geq \sigma$

In this section we combine Lemma 1 and Lemma 3 in order to derive the generating function of k -noncrossing RNA pseudoknot structures with minimum stack-size σ . Core-structures are only implicit (via Lemma 3) in its proof: all expressions and relations are based on $\mathbf{T}_{k,1}(n', h')$ and $\mathbf{T}_{k,1}(n)$, respectively. The latter are given by Theorem 1. Our main result reads

Theorem 5. Let $k, \sigma \in \mathbb{N}$, $k \geq 2$, let x be an indeterminate and ρ_k the dominant, positive real singularity of $\sum_{n \geq 0} f_k(2n, 0)z^{2n}$. Then

$$\begin{aligned} \mathbb{T}_{k,\sigma}(n, h) &= \sum_{b=\sigma-1}^{h-1} \sum_{j=0}^{(h-b)-1} \binom{b+(2-\sigma)(h-b)-1}{h-b-1} (-1)^{(h-b)-j-1} \\ &\quad \times \binom{(h-b)-1}{j} \mathbb{T}_{k,1}(n-2h+2j+2, j+1). \end{aligned}$$

Furthermore, $\mathbb{T}_{k,\sigma}(n)$, satisfies the following identity

$$(4.1) \quad \sum_{n \geq 0} \mathbb{T}_{k,\sigma}(n)x^n = \frac{1}{u_0x^2 - x + 1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{\sqrt{u_0}x}{u_0x^2 - x + 1} \right)^{2n},$$

where $u_0 = \frac{(x^2)^{\sigma-1}}{(x^2)^{\sigma-x^2+1}}$. We have the asymptotic formula

$$(4.2) \quad \mathbb{T}_{k,\sigma}(n) \sim n^{-((k-1)^2+(k-1)/2)} \left(\frac{1}{\gamma_{k,\sigma}} \right)^n \quad \text{for } 3 \leq k \leq 7 \text{ and } 1 \leq \sigma \leq 10$$

where $\gamma_{k,\sigma}$ is the dominant real singularity of $\sum_{n \geq 0} \mathbb{T}_{k,\sigma}(n)x^n$ and the minimal positive real solution of the equation

$$(4.3) \quad \frac{\sqrt{\frac{(x^2)^{\sigma-1}}{(x^2)^{\sigma-x^2+1}}} x}{\left(\frac{(x^2)^{\sigma-1}}{(x^2)^{\sigma-x^2+1}} \right) x^2 - x + 1} = \rho_k$$

The first 18 numbers of $\mathbb{T}_{3,2}(n)$, $\mathbb{T}_{3,3}(n)$, $\mathbb{T}_{4,2}(n)$ and $\mathbb{T}_{4,3}(n)$ are given by

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$\mathbb{T}_{3,2}(n)$	1	1	1	1	2	4	8	15	28	55	110	222	448	913	1890	3964	8385	17846
$\mathbb{T}_{3,3}(n)$	1	1	1	1	1	1	2	4	8	14	23	36	56	91	155	275	491	869
n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$\mathbb{T}_{4,2}(n)$	1	1	1	1	2	4	8	15	28	55	110	223	455	944	1995	4274	9244	20182
$\mathbb{T}_{4,3}(n)$	1	1	1	1	1	1	2	4	8	14	23	36	56	91	155	275	491	870

Proof. The first assertion follows from Lemma 2 and eq. (3.17), which allows us to express the terms $C_k(n-2b, h-b)$ via $\mathbb{T}_{k,1}(n', h')$. In order to prove eq. (4.2) we apply Lemma 3 twice. First,

Lemma 3 implies for arbitrary σ and $u = 1$

$$(4.4) \quad \sum_{n \geq 0} \mathbb{T}_{k,\sigma}(n)x^n = \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathbb{C}_k(n, h) \left(\frac{(x^2)^{\sigma-1}}{1-x^2} \right)^h x^n$$

and second, it guarantees for arbitrary $u \in \mathbb{C}$ and $\sigma = 1$

$$(4.5) \quad \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \mathbb{T}_{k,1}(n, h)u^h x^n = \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathbb{C}_k(n, h) \left(\frac{u}{1-ux^2} \right)^h x^n.$$

Here, the key observation (the ‘‘bridge’’) is the relation between σ and u via the terms $\frac{(x^2)^{\sigma-1}}{1-x^2}$ and $\frac{u}{1-ux^2}$. It is clear that for any $\sigma \in \mathbb{N}$ there exists a unique solution u_0 for

$$(4.6) \quad \frac{(x^2)^{\sigma-1}}{1-x^2} = \frac{u}{1-ux^2}$$

given by $u_0 = \frac{(x^2)^{\sigma-1}}{(x^2)^{\sigma-x^2+1}}$. This allows us to express

$$\sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathbb{C}_k(n, h) \left(\frac{(x^2)^{\sigma-1}}{1-x^2} \right)^h x^n$$

for any σ via the bivariate generating function $\sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \mathbb{T}_{k,1}(n, h) u^h x^n$. Now we employ Lemma 1, which provides an interpretation of the latter as follows:

$$(4.7) \quad \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \mathbb{T}_{k,1}(n, h)u^h x^n = \frac{1}{ux^2 - x + 1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{\sqrt{ux}}{ux^2 - x + 1} \right)^{2n}.$$

We accordingly obtain

$$\begin{aligned} \sum_{n \geq 0} \mathbb{T}_{k,\sigma}(n)x^n &= \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathbb{C}_k(n, h) \left(\frac{(x^2)^{\sigma-1}}{1-x^2} \right)^h x^n \\ &= \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathbb{C}_k(n, h) \left(\frac{u_0}{1-u_0x^2} \right)^h x^n \\ &= \frac{1}{u_0x^2 - x + 1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{\sqrt{u_0x}}{u_0x^2 - x + 1} \right)^{2n} \end{aligned}$$

and eq. (4.1) follows. We set $V_k(z) = \sum_{n \geq 0} f_k(2n, 0) \left(\frac{\sqrt{u_0z}}{u_0z^2 - z + 1} \right)^{2n}$.

claim. All dominant singularities of $\sum_{n \geq 0} \mathbb{T}_{k,\sigma}(n)x^n$ are singularities of $V_k(x)$ and $\gamma_{k,\sigma}$ is a dominant singularity for both.

To prove the claim we observe that a dominant singularity of

$$\frac{1}{u_0x^2 - x + 1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{\sqrt{u_0}x}{u_0x^2 - x + 1} \right)^{2n}$$

is either a singularity of $V_k(x)$ or $\frac{1}{u_0x^2 - x + 1}$. Suppose there exists some singularity $\zeta \in \mathbb{C}$ which is a pole of $\frac{1}{u_0x^2 - x + 1}$. By construction $\zeta \neq 0$ and ζ is necessarily a singularity of $V_k(x)$. Suppose $|\zeta| \leq \gamma_{k,\sigma}$ where $\gamma_{k,\sigma}$ denotes the dominant singularity of $\sum_{n \geq 0} \mathbb{T}_{k,\sigma}(n)x^n$. Then we arrive at the contradiction $|V_k(\zeta)| > |V_k(\gamma_{k,\sigma})|$ since $V_k(\zeta)$ is not finite and

$$V_k(\gamma_{k,\sigma}) = \sum_{n \geq 0} f_k(2n, 0) \rho_k^{2n} < \infty .$$

Therefore all dominant singularities of $\sum_{n \geq 0} \mathbb{T}_{k,\sigma}(n)x^n$ are singularities of $V_k(x)$, whence the claim. Set

$$(4.8) \quad \vartheta_\sigma(x) = \frac{\sqrt{u_0}x}{u_0x^2 - x + 1} = \frac{\sqrt{\frac{(x^2)^{\sigma-1}}{(x^2)^\sigma - x^2 + 1}} x}{\left(\frac{(x^2)^{\sigma-1}}{(x^2)^\sigma - x^2 + 1} \right) x^2 - x + 1} .$$

To prove eq. (4.2) we inspect that for $3 \leq k \leq 7$ and $1 \leq \sigma \leq 10$, $\gamma_{k,\sigma}$, has strictly smaller modulus than all solutions of $\vartheta_\sigma(x) = -\rho_k$. Indeed, $|\vartheta_\sigma(x)| = \rho_k$ implies $\frac{x^\sigma \sqrt{x^{2\sigma} - x^2 + 1}}{x^{2\sigma} + (1-x)(x^{2\sigma} - x^2 + 1)} = \pm \rho_k$ and we observe that, independent of σ , any real positive root of $\frac{x^\sigma \sqrt{x^{2\sigma} - x^2 + 1}}{x^{2\sigma} + (1-x)(x^{2\sigma} - x^2 + 1)} = -\rho_k$ must be larger than 1. Since $\vartheta_\sigma(x)$ is algebraic over $K(x)$ and satisfies $\vartheta_\sigma(0) = 0$. Accordingly, Theorem 3 applies and we have

$$(4.9) \quad \mathbb{T}_{k,\sigma}(n) \sim K n^{-((k-1)^2 + (k-1)/2)} (\gamma_{k,\sigma}^{-1})^n .$$

This completes the proof of Theorem 5. □

Acknowledgments. We are grateful to Prof. William Y.C. Chen for helpful comments. Special thanks to Jing Qin for creating Figures 2 and 4. 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] T. Akutsu, *Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots* Discr Appl. Math. **104**(2000), 45-62.

- [2] R.T. Batey, R.P. Rambo and J.A. Doudna, *Tertiary Motifs in RNA Structure and Folding* Angew. Chem. Int. Ed. **38** (1999), 2326-2343.
- [3] M. Chastain and I.Jr. Tinoco, *A base-triple structural domain in RNA* Biochemistry. **31**(1992), 12733-12741.
- [4] 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.
- [5] W.Y.C. Chen, J. Qin and C.M. Reidys, *Crossings and Nestings of tangled-diagrams*, submitted.
- [6] E. Dam, I. Brierley, S. Inglis and C. Pleij, *Identification and analysis of the pseudoknot containing gag-pro ribosomal frameshift signal of simian retrovirus-1* Discrete Appl. Math. **22**(1994), 2304-2310.
- [7] P. Flajolet and H. Prodinger, *Register allocation for unary-binary trees* SIAM J. on Computing. **15**(1986), 629-640.
- [8] P. Flajolet and R. Sedgewick, *Analytic combinatorics* (2007).
- [9] D.J. Grabiner and P. Magyar, *Random walks in Weyl chambers and the decomposition of tensor powers* Discr. Appl. Math. **2**(1993), 239-260.
- [10] E.Y. Jin, J. Qin and C.M. Reidys, *Combinatorics of RNA structures with pseudoknots* Bull. Math. Biol. (2007), PMID: 17896159.
- [11] E.Y. Jin and C.M. Reidys, *Asymptotic enumeration of RNA structures with pseudoknots* Bull. Math. Biol. (2007), to appear.
- [12] E.Y. Jin and C.M. Reidys, *Pseudoknot RNA structures with arc-length ≥ 3 .*, submitted.
- [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] R. Lyngso and C. Pedersen, *Pseudoknots in RNA secondary structures* Physics of Biological Systems: From Molecules to Species Sci. (1996).
- [16] M.E. Nebel, *Combinatorial Properties of RNA Secondary Structures* J. Comput. Biol. **9**(3)(2003), 541-574.
- [17] R.C. Penner and M.S. Waterman, *Spaces of RNA secondary structures* Adv. Math. **101**(1993), 31-49.
- [18] C.M. Reidys, P.F. Stadler and P. Schuster *Generic properties of combinatory maps: neutral networks of RNA secondary structures* Bulletin of Mathematical Biology. **59**(2), 339–397.
- [19] E. Rivas and S. Eddy, *A Dynamic Programming Algorithm for RNA structure prediction including pseudoknots* J. Mol. Biol. **285**(1999), 2053-2068.
- [20] *Mapping RNA Form and Function* Science **2** (2005).
- [21] I.L. Hofacker, P. Schuster and P.F. Stadler, *Combinatorics of RNA Secondary Structures* Discr. Appl. Math. **88**(1998), 207-237.
- [22] P.F. Stadler and C. Haslinger, *RNA Structures with Pseudo-Knots* Bull.Math.Biol. **61** (1999), 437-467.
- [23] P.F. Stadler, private communication.
- [24] R.P. Stanley, *Differentiably finite power series*. Europ. J. Combinatorics **1**, (1980), 175-188.
- [25] E.C. Titchmarsh, *The theory of functions* Oxford University Press, London, 1939.
- [26] Y. Uemura, A. Hasegawa, S. Kobayashi and T. Yokomori, *Tree adjoining grammars for RNA structure prediction* Theoret. Comput. Sci. **210**(1999), 277-303.
- [27] E.Y. Jin, C.M. Reidys and R. Wang, *Asymptotic enumeration of k-noncrossing RNA structures*, in preparation.
- [28] M.S. Waterman, *Combinatorics of RNA hairpins and cloverleaves* Stud. Appl. Math. **60**(1979), 91-96.
- [29] M.S. Waterman, *Secondary structure of single - stranded nucleic acids* Adv. Math.I (suppl.) **1**(1978), 167-212.

- [30] 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.
- [31] M.S. Waterman and W.R. Schmitt, *Linear trees and RNA secondary structure* Discr. Appl. Math. **51** (1994), 317-323.
- [32] E. Westhof and L. Jaeger, *RNA pseudoknots* Current Opinion Struct. Biol. **2** (1992), 327-333.

CENTER FOR COMBINATORICS, LPMC-TJKLC, NANKAI UNIVERSITY, TIANJIN 300071, P.R. CHINA, PHONE: *86-22-2350-6800, FAX: *86-22-2350-9272

E-mail address: reidys@nankai.edu.cn