Bulletin of Mathematical Biology (2007)
DOI 10.1007/s11538-007-9265-2

ORIGINAL ARTICLE

Asymptotic Enumeration of RNA Structures with Pseudoknots

Emma Y. Jin, Christian M. Reidys*

Center for Combinatorics, LPMC-TJKLC, Nankai University, Tianjin 300071, People's Republic of China

Received: 21 June 2007 / Accepted: 22 August 2007

Abstract In this paper, we present the asymptotic enumeration of RNA structures with pseudoknots. We develop a general framework for the computation of exponential growth rate and the asymptotic expansion for the numbers of k-noncrossing RNA structures. Our results are based on the generating function for the number of k-noncrossing RNA pseudoknot structures, $\mathsf{S}_k(n)$, derived in *Bull. Math. Biol.* (2007, in press), where $k-1$ denotes the maximal size of sets of mutually intersecting bonds. We prove a functional equation for the generating function $\sum_{n\geq 0}\mathsf{S}_k(n)z^n$ and obtain for $k=2$ and $k=3$, the analytic continuation and singular expansions, respectively. It is implicit in our results that for arbitrary k singular expansions exist and via transfer theorems of analytic combinatorics, we obtain asymptotic expression for the coefficients. We explicitly derive the asymptotic expressions for 2- and 3-noncrossing RNA structures. Our main result is the derivation of the formula $\mathsf{S}_3(n) \sim \frac{10.4724\cdot 4!}{n(n-1)\dots(n-4)}(\frac{5+\sqrt{21}}{2})^n$.

Keywords Asymptotic enumeration · RNA secondary structure · k-noncrossing RNA structure · Pseudoknot · Generating function · Transfer theorem · Hankel contour · Singular expansion

1. Introduction

RNA molecules are particularly fascinating since they represent both: genotypic legislative via their primary sequence and phenotypic executive via their functionality associated to 2D or 3D-structures, respectively. Accordingly, it is believed that RNA may have been instrumental for early evolution–before proteins emerged. The primary sequence of an RNA molecule is the sequence of nucleotides **A**, **G**, **U** and **C** together with the Watson–Crick (**A–U**, **G–C**) and (**U–G**) base pairing rules specifying the pairs of nucleotides can potentially form bonds. Single stranded RNA molecules form helical structures whose bonds satisfy the above base pairing rules and which, in many cases, determine their function. For instance, RNA ribozymes are capable of catalytic activity, cleaving other RNA

*Corresponding author.
E-mail address: reidys@nankai.edu.cn (Christian M. Reidys).

molecules. RNA secondary structure prediction is of polynomial complexity (Waterman, 1978) which is the result from the fact that in secondary structures no two bonds can cross. Leaving the paradigm of RNA secondary structures, i.e., studying RNA structures with crossing bonds, the RNA pseudoknot structures, poses challenging problems for computational biology. Prediction algorithms for RNA pseudoknot structures are much harder to derive since there exists no a priori tree-structure and the subadditivity of local solutions is not guaranteed. RNA pseudoknot structures can be categorized in terms of the maximal size of sets of mutually crossing bonds (Jin et al., 2007). To be precise a k-noncrossing RNA structure has at most $k-1$ mutually crossing bonds and a minimum bond-length of 2, i.e., for any i, the nucleotides i and $i+1$ cannot form a bond. The asymptotics of k-noncrossing RNA structures is of central importance in this context. The key question is how to decompose a k-noncrossing RNA structure into a collection of substructures (which can easily be computed), and what are the properties of this decomposition. Given such a decomposition, we can predict the factors and reassemble the corresponding pseudoknot structure. A first step toward finding such decompositions is to have information about the cardinalities of the respective sets of structures involved. The asymptotic analysis of k-noncrossing RNA structures is based on their generating function, obtained in Jin et al. (2007). The particular formulas are, however, alternating sums, which make even the computation of the exponential growth rate a nontrivial task. In this paper, we develop a framework for the asymptotic enumeration of k-noncrossing RNA structures. Before we go into this in more detail, let us first provide some background on coarse grained RNA structures and put our results into context.

1.1. RNA secondary structures or the universality of the square root

About three decades ago, Waterman et al. pioneered the concept of RNA secondary structures (Penner and Waterman, 1993; Waterman, 1979, 1978; Schmitt and Waterman, 1994; Howell et al., 1980). The key property of secondary structures is best understood, considering a structure as a diagram, which is obtained as follows: one draws the primary sequence of nucleotides horizontally and ignores all chemical bonds of its backbone. Then, one draws all bonds, i.e., nucleotide interactions satisfying the Watson–Crick base pairing rules (and **G–U** pairs) as arcs in the upper halfplane, effectively identifying structure with the set of all arcs. In this representation, RNA secondary structures have then following property: there exist no two arcs (i_1, j_1), (i_2, j_2), where $i_1 < j_1$ and $i_2 < j_2$ with the property $i_1 < i_2 < j_1 < j_2$ and all arcs have at least length 2. Equivalently, there exist no two arcs that cross in the diagram representation of the structure, see Fig. 1. Basically, all combinatorial properties of secondary structures are derived from Waterman's basic recursion (Waterman, 1978)

$$\mathsf{S}_2(n) = \mathsf{S}_2(n-1) + \sum_{s=1}^{n-2} \mathsf{S}_2(n-2-s)\mathsf{S}_2(s), \tag{1.1}$$

where $\mathsf{S}_2(n)$ denotes the number of RNA secondary structures. Equation (1.1) is an immediate consequence considering secondary structures as Motzkin paths, i.e., peak-free paths with *up*, *down* and *horizontal* steps that stay in the upper halfplane, starting at the origin and end on the x-axis. The recursion is in particular the key for all asymptotic results since it allows to obtain an implicit function equation for the generating

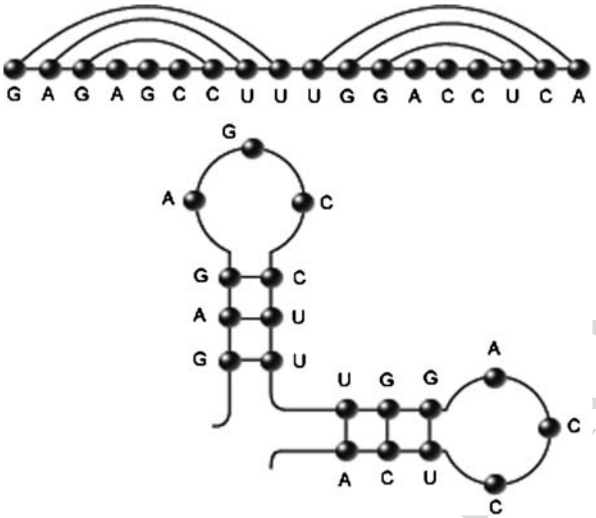

Fig. 1 RNA secondary structures. Diagram representation (top): the primary sequence, **GAGAGCCUUUGGACCUCA**, is drawn horizontally and its backbone bonds are ignored. All bonds are drawn in the upper halfplane and secondary structures have the property that no two arcs intersect and all arcs have minimum length 2. Outer planar graph representation (bottom).

function and subsequent application of Darboux-type theorems (Hofacker et al., 1998; Wong and Wyman, 1974). If specific conditions are being imposed, for instance, minimum loop-size or stack length; it is straightforward to translate these constraints into restricted Motzkin paths, all of which satisfy some variant of Eq. (1.1). As a result, all asymptotic formulae are of the same type: a square root, that is, the asymptotic behavior is determined by an algebraic branch singularity with the factor $n^{-\frac{3}{2}}$. For instance, the number of RNA secondary structures having a minimum hairpin-loop length of 3 and minimum stack-length 2 is asymptotically given by $\mathsf{S}_2(n) \sim 1.4848\, n^{-\frac{3}{2}}\, 1.8488^n$ (Hofacker et al., 1998). The number of RNA secondary structures having exactly ℓ isolated vertices, $\mathsf{S}_2(n,\ell)$, satisfies the two term recursion $(n-\ell)(n-\ell+2)\,\mathsf{S}_2(n,\ell) - (n+\ell)(n+\ell-2)\,\mathsf{S}_2(n-2,\ell) = 0$ (Jin et al., 2007) and Waterman proved in Schmitt and Waterman (1994) the following beautiful formula

$$\mathsf{S}_2(n,\ell) = \frac{2}{n-\ell}\binom{\frac{n+\ell}{2}}{\frac{n-\ell}{2}+1}\binom{\frac{n+\ell}{2}-1}{\frac{n-\ell}{2}-1} \tag{1.2}$$

resulting from a bijection between secondary structure and linear trees. In Waterman and Smith (1986), it is shown that the prediction of secondary structures can be obtained in polynomial time and yet again Eq. (1.1) is central for all folding algorithms (Zuker and Sankoff, 1984; Hofacker et al., 1998; Waterman and Smith, 1986; Tacker et al., 1994, 1996; McCaskill, 1990).

1.2. Beyond secondary structures

While the concept of secondary structure is of fundamental importance, it is well known that there exist additional types of nucleotide interactions (1). These bonds are called pseudoknots (Westhof and Jaeger, 1992) and occur in functional RNA (RNAseP; Loria and Pan, 1996), ribosomal RNA (Konings and Gutell, 1995) and are conserved in the catalytic core of group I introns. In plant viral RNAs, pseudoknots mimic tRNA structure and in *in vitro* RNA evolution (Tuerk et al., 1992) experiments have produced families of RNA structures with pseudoknot motifs, when binding HIV-1 reverse transcriptase. Important mechanisms like ribosomal frame shifting (Chamorro et al., 1991) also involve pseudoknot interactions. There exist several prediction algorithms for pseudoknot RNA structures (Rivas and Eddy, 1999; Uemura et al., 1999; Akutsu, 2000; Lyngso and Pedersen, 1996), all of which can identify particular respective pseudoknot motifs. Stadler et al. (1999) suggested a classification of RNA pseudoknot-types based on a notion of inconsistency graphs and computed the upper bound of 4.7613 for the exponential growth factor of bi-secondary structures. Bi-secondary structures are "superpositions" of two secondary structures, i.e., they can be drawn as a set of nonintersecting arcs in the upper and lower half plane, respectively. Figure 3 shows how bi-secondary structures naturally arise when passing from outer-planar to planar diagram representations. The concept of k-noncrossing RNA structures generalize both: secondary and bi-secondary structures, respectively. While RNA secondary structures are precisely 2-noncrossing RNA structures, bi-secondary structures correspond to planar 3-noncrossing RNA structures. The key advantage of k-noncrossing RNA structures is that their defining property is intrinsically local. It can be expected that this facilitates fast folding algorithms. In Fig. 4, we contrast all three structural concepts, secondary, bi-secondary and k-noncrossing RNA structures.

1.3. Organization and main results

In Section 2, we provide the necessary background on k-noncrossing RNA structures and the generating function $\sum_{n\geq 0} \mathsf{S}_k(n)z^n$. In Section 3, we derive the exponential factor for k-noncrossing RNA structures, i.e., we compute the base at which k-noncrossing RNA structures asymptotically grow. The exponential factor is *the* key result for all complexity considerations arising in the context of prediction algorithms for RNA pseudoknot structures. To make it easily accessible to a broad readership, we give an elementary proof based on real analysis and transformations of the generating function. Central to our proof is a functional identity (Lemma 1) whose true power is revealed only later in Section 4, where it is put in the context of analytic functions. Remarkably, Stadler's upper bound for bi-secondary structures coincides with the exact exponential factor obtained via Theorem 2 for 3-noncrossing RNA structures up to $O(10^{-2})$. In Section 4, we compute the asymptotics for 2-noncrossing RNA structures and 3-noncrossing RNA structures, respectively. Since the method via implicit functions used for secondary structures (Hofacker et al., 1998) does not work for $k > 2$, we develop a new approach which is based on concepts developed by Flajolet et al. using singular expansions and transfer theorems (Flajolet et al., 1994, 2005; Flajolet, 1999; Popken, 1953; Odlyzko, 1992). The basic strategy is as follows: we first obtain an analytic continuation $f(z)$ generalizing the functional equation of Lemma 1 to complex indeterminant z.

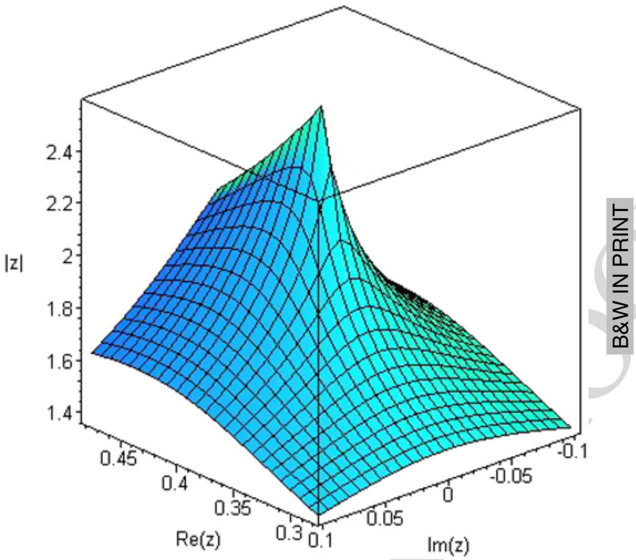

Fig. 2 Universality of the square root. We display the branch-point singularity (here at $\rho_2 = \frac{3-\sqrt{5}}{2}$), i.e. the critical singularity for the asymptotics of RNA secondary structures. *All* singularities arising from enumeration of certain classes of secondary structures produces this type, whence the factor $n^{-\frac{3}{2}}$.

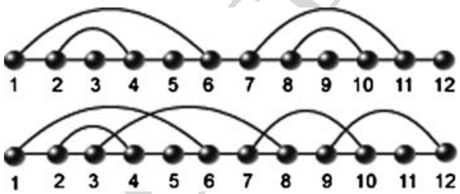

Fig. 3 Beyond secondary structures: an RNA bi-secondary structure as the generalization from outer-planar to planar diagrams. We display a secondary RNA structure (top) and a bi-secondary structure (bottom). Reflecting the arcs (3, 8) and (9, 12) w.r.t. the x-axis yields two secondary structures.

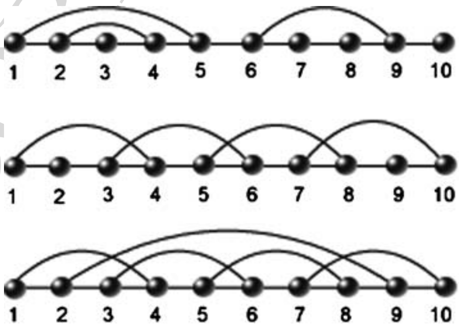

Fig. 4 k-noncrossing RNA structures. (a) Secondary structure (with isolated labels 3, 7, 8, 10), (b) Bi-secondary structure, 2, 9 being isolated, (c) 3-noncrossing structure, which is *not* a bi-secondary structure. In fact, this is *the* smallest 3-noncrossing RNA structure which is not a bi-secondary structure.

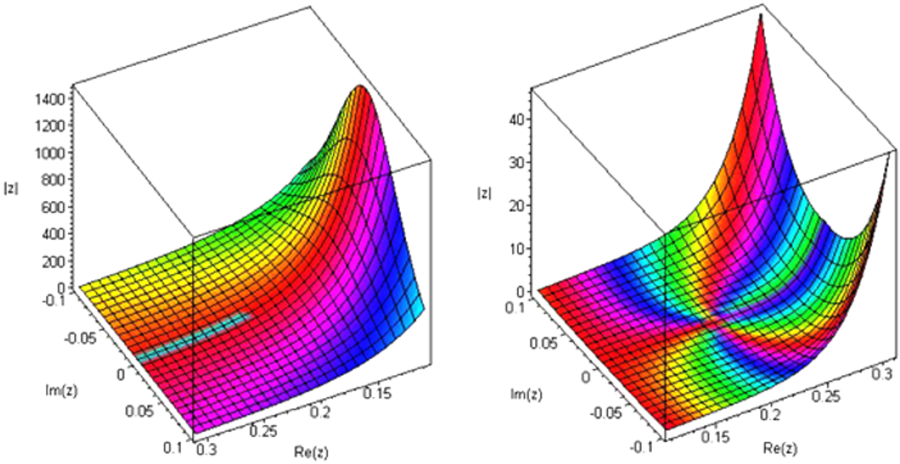

Fig. 5 Toroidal harmonics and its singular expansion. We display the analytic continuation of $\sum_{n\geq 0}\mathsf{S}_3(n)z^n$, the generating function of 3-noncrossing RNA structures (left) and its singular expansion (right) at the dominant singularity $\rho_3 = \frac{5-\sqrt{21}}{2}$.

For $k = 3$ we obtain an expression involving the Legendre polynomial $P^{-1}_{\frac{3}{2}}(z)$ indicating that the type of singularity is fundamentally different from the branch-point singularity of the square root. In Fig. 5, we display the analytic continuation of $\sum_{n\geq 0}\mathsf{S}_3(n)z^n$ at the dominant singularity, $\rho_3 = \frac{5-\sqrt{21}}{2}$ and its singular expansion. We proceed by proving that $f(z)$ the dominant singularity is indeed unique. The next step is to establish that there exists a singular expansion for $f(z)$, i.e., there exists a function h such that $f(z) = O(h(z))$ at the dominant singularity (see Section 4). Intuitively, this singular expansion approximates $f(z)$ well enough to retrieve precise asymptotic expansions of the coefficients via transfer theorems (Flajolet et al., 2005; Gao and Richmond, 1992; Odlyzko, 1995). The existence of the singular expansion can be deduced from the particular form of the generating function for k-noncrossing RNA structures. Due to Lemma 2, it suffices to analyze the coefficients $f_3(2n,0)$, which are known via the determinant formula of Bessel functions in Eq. (2.3). We then proceed using this particular form of $f_3(2n,0)$ to explicitly compute the singular expansion and show in the process how the logarithmic term arises naturally from elementary calculations. It should be remarked that we use the transfer theorems since our generating function is the composition of two analytic functions $f(\vartheta(z))$. We then show that the type of the singularity of $f(\vartheta(z))$ coincides with the type of singularity of the function $f(z)$. The phenomenon of the persistence of the singularity of the "outer" function $f(z)$ is known as the *supercritical case* (Flajolet et al., 2005). This will allow us to obtain the asymptotics of the coefficients of the function $f(\vartheta(z))$. One main result of the paper is the formula

$$\mathsf{S}_3(n) \sim \frac{10.47244!}{n(n-1)\dots(n-4)}\left(\frac{5+\sqrt{21}}{2}\right)^n. \qquad (1.3)$$

In order to assess the quality of our formula, let us list the subfactors for $k=2$ and $k=3$, obtained from Theorem 4 and Theorem 5:

$$\mathsf{s}_2(n) = 1.1002\left[\frac{1}{n^{\frac{3}{2}}} - \frac{7}{8n^{\frac{5}{2}}} - \frac{111}{128n^{\frac{7}{2}}} + \frac{893}{1024n^{\frac{9}{2}}} + O\big(n^{-\frac{11}{2}}\big)\right],$$

$$\mathsf{s}_3(n) = \frac{10.4724 \cdot 4!}{n(n-1)\dots(n-4)} \sim 251.3375\left[\frac{1}{n^5} - \frac{35}{4n^6} + \frac{1525}{32n^7} + O\big(n^{-8}\big)\right].$$

In the table below, we list the subexponential factors, i.e., we compare for $k=2,3$ the quantities $\mathsf{S}_k(n)/(\frac{3+\sqrt{5}}{2})^n$ and $\mathsf{s}_k(n)$, respectively. $\mathsf{S}_2(n)$ and $\mathsf{S}_3(n)$ are given by the generating function of k-noncrossing RNA structures.

The subexponential factor

n	$\mathsf{S}_2(n)/(\frac{3+\sqrt{5}}{2})^n$	$\mathsf{s}_2(n)$	$\mathsf{S}_3(n)/(\frac{5+\sqrt{21}}{2})^n$	$\mathsf{s}_3(n)$
10	2.796×10^{-2}	3.124×10^{-2}	5.229×10^{-4}	1.512×10^{-3}
20	1.100×10^{-2}	1.164×10^{-2}	3.358×10^{-5}	5.354×10^{-5}
30	6.215×10^{-3}	6.452×10^{-3}	5.776×10^{-6}	7.874×10^{-6}
40	4.114×10^{-3}	4.229×10^{-3}	1.576×10^{-6}	1.991×10^{-6}
50	2.980×10^{-3}	3.043×10^{-3}	5.627×10^{-7}	6.789×10^{-7}
60	2.284×10^{-3}	2.324×10^{-3}	2.397×10^{-7}	2.804×10^{-7}
70	1.822×10^{-3}	1.849×10^{-3}	1.156×10^{-7}	1.323×10^{-7}
80	1.500×10^{-3}	1.516×10^{-3}	6.123×10^{-8}	6.888×10^{-8}
90	1.259×10^{-3}	1.273×10^{-3}	3.483×10^{-8}	3.868×10^{-8}
100	1.078×10^{-3}	1.088×10^{-3}	2.098×10^{-8}	2.305×10^{-8}
1000	3.484×10^{-5}	3.475×10^{-5}	2.475×10^{-13}	2.492×10^{-13}
10000	1.104×10^{-6}	1.100×10^{-6}	2.517×10^{-18}	2.516×10^{-18}

2. *k*-noncrossing RNA structures

Suppose we are given the primary RNA sequence

AACCAUGUGGUACUUGAUGGCGAC.

We then identify an RNA structure with the set of all bonds different from the backbone-bonds of its primary sequence, i.e., the arcs $(i, i+1)$ for $1 \le i \le n-1$. Accordingly, an RNA structure is a combinatorial graph over the labels of the nucleotides of the primary sequence. These graphs can be represented in several ways. In Fig. 6, we represent a structure with loop-loop interactions in two ways. In the following, we will consider structures as diagram representations of digraphs. A digraph D_n is a pair of sets V_{D_n}, E_{D_n}, where $V_{D_n} = \{1, \dots, n\}$ and $E_{D_n} \subset \{(i,j) \mid 1 \le i < j \le n\}$. V_{D_n} and E_{D_n} are called vertex and arc set, respectively. A k-noncrossing digraph is a digraph in which all vertices have degree ≤ 1 and which does not contain a k-set of arcs that are mutually intersecting, or more formally

$$\nexists (i_{r_1}, j_{r_1}), (i_{r_2}, j_{r_2}), \dots, (i_{r_k}, j_{r_k}); \quad i_{r_1} < i_{r_2} < \dots < i_{r_k} < j_{r_1} < j_{r_2} < \dots < j_{r_k}. \tag{2.1}$$

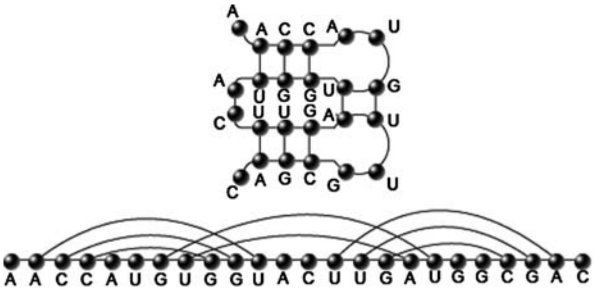

Fig. 6 A 3-noncrossing RNA structure, as a planar graph (top) and as a diagram (bottom).

We will represent digraphs as a diagrams (Fig. 6) by representing the vertices as integers on a line and connecting any two adjacent vertices by an arc in the upper-half plane. The direction of the arcs is implicit in the linear ordering of the vertices and accordingly omitted.

Definition 1. An k-noncrossing RNA structure is a digraph in which all vertices have degree ≤ 1, that does not contain a k-set of mutually intersecting arcs and 1-arcs, i.e., arcs of the form $(i, i+1)$, respectively. We denote the number of RNA structures by $\mathsf{S}_k(n)$ and the number of RNA structures with exactly ℓ isolated vertices and with exactly h arcs by $\mathsf{S}_k(n,\ell)$ and $\mathsf{S}'_k(n,h)$, respectively. Note that $\mathsf{S}'_k(n,h) = \mathsf{S}_k(n, n-2h)$.

Let $f_k(n,\ell)$ denote the number of k-noncrossing digraphs with ℓ isolated points. We have shown in (Jin et al., 2007) that

$$f_k(n,\ell) = \binom{n}{\ell} f_k(n-\ell, 0), \tag{2.2}$$

$$\det\big[I_{i-j}(2x) - I_{i+j}(2x)\big]\Big|_{i,j=1}^{k-1} = \sum_{n\ge 1} f_k(n,0)\,\frac{x^n}{n!}, \tag{2.3}$$

$$\begin{aligned} e^x \det\big[I_{i-j}(2x) - I_{i+j}(2x)\big]\Big|_{i,j=1}^{k-1} &= \left(\sum_{\ell\ge 0}\frac{x^\ell}{\ell!}\right)\left(\sum_{n\ge 1} f_k(n,0)\frac{x^n}{n!}\right) \\ &= \sum_{n\ge 1}\left\{\sum_{\ell=0}^{n} f_k(n,\ell)\right\}\frac{x^n}{n!}, \end{aligned} \tag{2.4}$$

where $I_r(2x) = \sum_{j\ge 0}\frac{x^{2r+j}}{j!(r+j)!}$ is the hyperbolic Bessel function of the first kind of order r. In particular, we obtain for $k=2$ and $k=3$

$$f_2(n,\ell) = \binom{n}{\ell} C_{(n-\ell)/2} \quad\text{and}\quad f_3(n,\ell) = \binom{n}{\ell}\left[C_{\frac{n-\ell}{2}+2}C_{\frac{n-\ell}{2}} - C^2_{\frac{n-\ell}{2}+1}\right], \tag{2.5}$$

where $C_m = \frac{1}{m+1}\binom{2m}{m}$ is the mth Catalan number. The derivation of the generating function of k-noncrossing RNA structures, given in Theorem 1 below uses advanced methods and novel constructions of enumerative combinatorics due to Chen et al. (2007), Gessel and Zeilberger (1992) and Stanley's mapping between matchings and oscillating tableaux i.e., families of Young diagrams in which any two consecutive shapes differ by exactly one square. The enumeration is obtained using the reflection principle due to Gessel and Zeilberger (1992) and Lindström (1973) combined with an inclusion-exclusion argument in order to eliminate the arcs of length 1. In Jin et al. (2007) generalizations to restricted (i.e., where arcs of the form $(i, i+2)$ are excluded) and circular RNA structures are given.

Theorem 1 (Jin et al., 2007). *Let $k \in \mathbb{N}$, $k \geq 2$, then the number of RNA structures with ℓ isolated vertices, $\mathsf{S}_k(n, \ell)$, is given by*

$$\mathsf{S}_k(n, \ell) = \sum_{b=0}^{(n-\ell)/2} (-1)^b \binom{n-b}{b} f_k(n-2b, \ell), \tag{2.6}$$

where $f_k(n-2b, \ell)$ is given by the generating function in Eq. (2.3). Furthermore the number of k-noncrossing RNA structures, $\mathsf{S}_k(n)$ is

$$\mathsf{S}_k(n) = \sum_{b=0}^{\lfloor n/2 \rfloor} (-1)^b \binom{n-b}{b} \left\{ \sum_{\ell=0}^{n-2b} f_k(n-2b, \ell) \right\}, \tag{2.7}$$

where $\{\sum_{\ell=0}^{n-2b} f_k(n-2b, \ell)\}$ is given by the generating function in Eq. (2.4).

3. The exponential factor

In this section, we obtain the exponential growth factor of the coefficients $\mathsf{S}_k(n)$. Let us begin by considering the generating function $\sum_{n\geq 0} \mathsf{S}_k(n) z^n$ as a power series over $\mathbb{R}$. Since $\sum_{n\geq 0} \mathsf{S}_k(n) z^n$ has monotonously increasing coefficients $\lim_{n\to\infty} \mathsf{S}_k(n)^{\frac{1}{n}}$ exists and determines via Hadamard's formula its radius of convergence. As we already mentioned, due to the inclusion-exclusion form of the terms $\mathsf{S}_k(n)$, it is not obvious, however, how to compute this radius of convergence. Our strategy consists in first showing that $\mathsf{S}_k(n)$ is closely related to $f_k(2n, 0)$ via a functional relation of generating functions.

Lemma 1. *Let z be an indeterminant over $\mathbb{R}$ and $w \in \mathbb{R}$ a parameter. Let furthermore $\rho_k(w)$ denote the radius of convergence of the power series $\sum_{n\geq 0}[\sum_{h\leq n/2} \mathsf{S}'_k(n, h) w^{2h}] z^n$. Then, for $|z| < \rho_k(w)$ holds*

$$\sum_{n\geq 0} \sum_{h\leq n/2} \mathsf{S}'_k(n, h) w^{2h} z^n = \frac{1}{w^2z^2 - z + 1} \sum_{n\geq 0} f_k(2n, 0) \left(\frac{wz}{w^2z^2 - z + 1} \right)^{2n}. \tag{3.1}$$

In particular, we have for $w = 1$,

$$\sum_{n\geq 0} \mathsf{S}_k(n) z^n = \frac{1}{z^2 - z + 1} \sum_{n\geq 0} f_k(2n, 0) \left(\frac{z}{z^2 - z + 1} \right)^{2n}. \tag{3.2}$$

« BMAB 11538 layout: SPEC (bmab)reference style: basic file: bmab9265.tex (Ingrida) aid: 9265 doctopic: OriginalPaper class: spr-spec-bmab v.2007/09/18 Prn:18/09/2007; 12:55 p. 10»

The proof of Lemma 1 is a bit technical and consists in a series of changes of orders of summations and Laplace transforms. We give the proof in Section 5. In Section 4, we will employ basic complex analysis and extend Eq. (3.1) to complex z. Lemma 1 is the key to prove Theorem 2 below, where we obtain the exponential factor for any $k > 1$. In its proof, we recruit the theorem of Pringsheim (Titchmarsh, 1939) which asserts that a power series $\sum_{n\geq 0} a_n z^n$ with $a_n \geq 0$ has its radius of convergence as dominant (but not necessarily unique) singularity.

Theorem 2. *Let k be a positive integer, $k > 1$ and let r_k be the radius of convergence of the power series $\sum_{n\geq 0} f_k(2n,0)z^{2n}$. Then, the power series $\sum_{n\geq 0} \mathsf{S}_k(n)z^n$ has the real valued, dominant singularity at $\rho_k = \frac{1+\frac{1}{r_k}}{2} - \sqrt{\left(\frac{1+\frac{1}{r_k}}{2}\right)^2 - 1}$ and for the number of k-noncrossing RNA structures holds*

$$\mathsf{S}_k(n) \sim \left(\frac{1}{\rho_k}\right)^n. \tag{3.3}$$

We will prove later in Theorem 4 and Theorem 5 that for $k = 2$ and $k = 3$, the dominant singularities ρ_2 and ρ_3 are unique, respectively.

Proof: Suppose we are given r_k, then $r_k \leq \frac{1}{2}$ (this follows immediately from $C_n \sim 2^{2n}$ via Stirling's formula) and obviously, $(z-\frac{1}{2})^2 + \frac{3}{4}$ has no roots over $\mathbb{R}$. The functional identity of Lemma 1 allows us to derive the radius of convergence of $\sum_{n\geq 0} S_k(n)z^n$. Setting $w = 1$ Lemma 1 yields

$$\sum_{n\geq 0} \mathsf{S}_k(n)z^n = \frac{1}{\left(z-\frac{1}{2}\right)^2 + \frac{3}{4}} \sum_{n\geq 0} f_k(2n,0)\left(\frac{z}{\left(z-\frac{1}{2}\right)^2 + \frac{3}{4}}\right)^{2n}. \tag{3.4}$$

$f_k(2n,0)$ is monotone, whence the limit $\lim_{n\to\infty} f_k(2n,0)^{\frac{1}{2n}}$ exists and applying Hadamard's formula: $\lim_{n\to\infty} f_k(2n,0)^{\frac{1}{2n}} = \frac{1}{r_k}$. For $z \in \mathbb{R}$, we proceed by computing the roots of $|\frac{z}{z^2-z+1}| = r_k$ which for $r_k \leq \frac{1}{2}$ has the minimal root $\rho_k = \frac{1+\frac{1}{r_k}}{2} - ((\frac{1+\frac{1}{r_k}}{2})^2 - 1)^{1/2}$. We next show that ρ_k is indeed the radius of convergence of $\sum_{n\geq 0} \mathsf{S}_k(n)z^n$. For this purpose, we observe that the map

$$\vartheta : \left[0, \frac{1}{2}\right] \longrightarrow \left[0, \frac{2}{3}\right], \quad z \mapsto \frac{z}{\left(z-\frac{1}{2}\right)^2 + \frac{3}{4}}, \quad \text{where } \vartheta(\rho_k) = r_k \tag{3.5}$$

is bijective, continuous and strictly increasing. Continuity and strict monotonicity of ϑ guarantee in view of Eq. (3.4) that ρ_k, is indeed the radius of convergence of the power series $\sum_{n\geq 0} \mathsf{S}_k(n)z^n$. In order to show that ρ_k is a dominant singularity, we consider $\sum_{n\geq 0} \mathsf{S}_k(n)z^n$ as a power series over $\mathbb{C}$. Since $\mathsf{S}_k(n) \geq 0$, the theorem of Pringsheim (Titchmarsh, 1939) guarantees that ρ_k itself is a singularity. By construction, ρ_k has minimal absolute value and is accordingly dominant. Since $\mathsf{S}_k(n)$ is monotone $\lim_{n\to\infty} \mathsf{S}_k(n)^{\frac{1}{n}}$ exists and we obtain using Hadamard's formula

$$\lim_{n\to\infty} \mathsf{S}_k(n)^{\frac{1}{n}} = \frac{1}{\rho_k}, \quad \text{or equivalently} \quad \mathsf{S}_k(n) \sim \left(\frac{1}{\rho_k}\right)^n, \tag{3.6}$$

from which Eq. (3.3) follows and the proof of the theorem is complete. □

4. Asymptotics of 3-noncrossing RNA structures

In this section, we provide the asymptotics for RNA secondary and 3-noncrossing RNA structures. For $k = 2$ and $k = 3$, i.e., for RNA secondary and 3-noncrossing RNA structures, respectively, we will explicitly obtain analytic continuations of the power series $\sum_{n\geq 0} \mathsf{S}_2(n)z^n$ and $\sum_{n\geq 0} \mathsf{S}_3(n)z^n$, respectively. As a result, the dominant singularity relevant for the asymptotics is known and Theorem 2 becomes obsolete. However, it is not entirely trivial to derive the analytic continuations for arbitrary crossing numbers k. In the context of complexity of prediction algorithms for RNA pseudoknot structures, it suffices to obtain the exponential factor which is given via Theorem 2.

We begin by revealing the "true" power of Lemma 1 in the context of analytic functions.

Lemma 2. *Let $k \geq 1$ be an integer, then we have for arbitrary $z \in \mathbb{C}$ with the property $|z| < \rho_k$ the equality*

$$\sum_{n\geq 0} \mathsf{S}_k(n)z^n = \frac{z}{z^2 - z + 1} \sum_{n\geq 0} f_k(2n, 0)\left(\frac{z}{z^2 - z + 1}\right)^{2n}. \tag{4.1}$$

Proof: The power series $\sum_{n\geq 0} \mathsf{S}_k(n)z^n$ and $\sum_{n\geq 0} f_k(2n, 0)(\frac{z}{z^2-z+1})^{2n}$ are analytic in a disc of radius $0 < \epsilon < \rho_k$ and according to Lemma 1 coincide on the interval $]-\epsilon, \epsilon[$. Therefore, both functions are equal on the sequence $(\frac{1}{n})_{n\in\mathbb{N}}$ which converges to 0 and standard results of complex analysis (zeros of nontrivial analytic functions are isolated) imply that Eq. (4.1) holds for any $z \in \mathbb{C}$ with $|z| < \rho_k$, hence the lemma. □

The derivation of the subexponential factors is based on singular expansions in combination with a transfer theorem, which recruits Hankel contours; see Fig. 7. Let us begin by specifying a suitable domain for our Hankel contours tailored for Theorem 3.

Definition 2. 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

$$\Delta_\rho(\phi, R) = \{z \mid |z| < R, z \neq \rho, \left|\operatorname{Arg}(z - \rho)\right| > \phi\}. \tag{4.2}$$

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.

Since the Taylor coefficients have the property

$$\forall \gamma \in \mathbb{C} \setminus 0; \quad [z^n]f(z) = \gamma^n [z^n] f\left(\frac{z}{\gamma}\right), \tag{4.3}$$

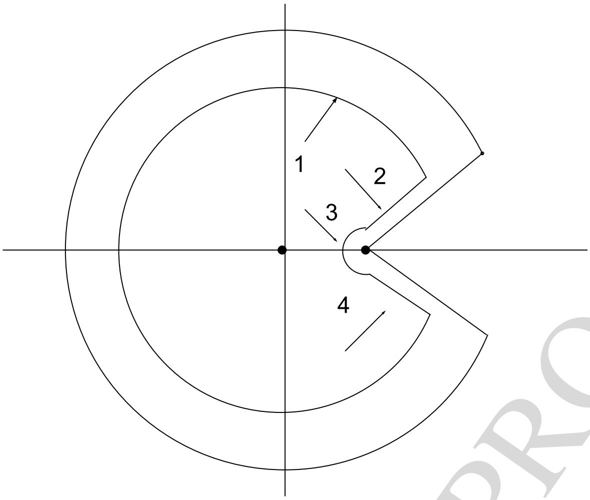

Fig. 7 Δ_1-domain enclosing a Hankel contour. We assume $z = 1$ to be the unique dominant singularity. The coefficients are obtained via Cauchy's integral formula and the integral path is decomposed in 4 segments. Segment 1 becomes asymptotically irrelevant since by construction the function involved is bounded on this segment. Relevant are the rectilinear segments 2 and 4 and the inner circle 3. The only contributions to the contour integral are being made here, which shows why the singular expansion allows to approximate the coefficients so well.

we can, w.l.o.g. reduce our analysis to the case where 1 is the dominant singularity. We use $U(a,r) = \{z \in \mathbb{C} \mid |z-a| < r\}$ to denote the open neighborhood of a in $\mathbb{C}$.

We use the notation

$$\big(f(z) = O\big(g(z)\big) \text{ as } z \to \rho\big) \quad \Longleftrightarrow \quad \big(f(z)/g(z) \text{ is bounded as } z \to \rho\big) \tag{4.4}$$

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 3 (Flajolet et al., 2005). *Let α be an arbitrary complex number in $\mathbb{C} \setminus \mathbb{Z}_{\le 0}$, $f(z)$ be a Δ_1-analytic function. Suppose $f(z) = O((1-z)^{-\alpha})$ in the intersection of a neighborhood of 1 and the Δ_1-domain, then*

$$[z^n]\,f(z) \sim K n^{\alpha-1}\bigg[1 + \frac{\alpha(\alpha-1)}{2n} + \frac{\alpha(\alpha-1)(\alpha-2)(3\alpha-1)}{24n^2} + \frac{\alpha^2(\alpha-1)^2(\alpha-2)(\alpha-3)}{48n^3} + \cdots\bigg] \quad \text{for some } K > 0.$$

Suppose $r \in \mathbb{Z}_{\ge 0}$, and $f(z) = O((1-z)^r \ln(\frac{1}{1-z}))$ in the intersection of a neighborhood of 1 and the Δ_1-domain, then we have

$$[z^n]\,f(z) \sim K\,(-1)^r \frac{r!}{n(n-1)\dots(n-r)} \quad \text{for some } K > 0. \tag{4.5}$$

Let us first analyze the case $k=2$, which illustrates the general strategy without the technicality of establishing the existence of a suitable singular expansion. Here, the generating function itself can be used directly (i.e., is its own singular expansion). Our particular proof, given in Section 5, exercises the base strategy used in the proof of Theorem 5. In particular, Theorem 4 improves on the quality of approximation providing a subexponential factor of higher order compared to (Hofacker et al., 1998).

Theorem 4. *The number of RNA secondary, i.e.,* 2*-noncrossing RNA structures is asymptotically given by*

$$\mathsf{S}_2(n) \sim \frac{1.1002}{\sqrt{n}}\left(\frac{1}{n+1} - \frac{1}{8n(n+1)} + \frac{1}{128n^3} + \frac{5}{1024n^4} + O\left(n^{-5}\right)\right) \times \left(\frac{3+\sqrt{5}}{2}\right)^n. \tag{4.6}$$

We next analyze the 3-noncrossing RNA structures. Here the situation changes dramatically since it has to be shown that a suitable singular expansion exists. We will prove this using the determinant formula arising in the context of the exponential generating function of $f_k(2n,0)$ given in Eq. (2.3).

Theorem 5. *The number of* 3*-noncrossing RNA structures is asymptotically given by*

$$\mathsf{S}_3(n) \sim \frac{10.4724 \cdot 4!}{n(n-1)\dots(n-4)}\left(\frac{5+\sqrt{21}}{2}\right)^n.$$

Proof: *Claim* 1. The dominant singularity ρ_3 of the power series $\sum_{n\geq 0}\mathsf{S}_3(n)z^n$ is unique.

In order to prove Claim 1, we use Lemma 2 according to which the analytic function $\Xi_3(z)$ is the analytic continuation of the power series $\sum_{n\geq 0}\mathsf{S}_3(n)z^n$. We proceed by showing that $\Xi_3(z)$ has exactly 6 singularities in $\mathbb{C}$, 4 of which have distinct moduli. The first two singularities are the roots of the quadratic polynomial $P(z)=(z-\frac{1}{2})^2+\frac{3}{4}$, given by $\alpha_1=\frac{1}{2}+i\frac{\sqrt{3}}{2}$ and $\alpha_2=\frac{1}{2}-i\frac{\sqrt{3}}{2}$. Both of them have moduli 1. Next, we observe that the power series $\sum_{n\geq 0} f_3(2n,0)y^n$ has the analytic continuation $\Psi(y)$ (obtained by MAPLE sumtools) given by

$$\Psi(y) = \frac{-(1-16y)^{\frac{3}{2}}P_{3/2}^{-1}(-\frac{16y+1}{16y-1})}{16y^{\frac{5}{2}}}, \tag{4.7}$$

where $P_\nu^m(x)$ denotes the Legendre polynomial of the first kind with the parameters $\nu=\frac{3}{2}$ and $m=-1$. $\Psi(y)$ has one dominant singularity at $y=\frac{1}{16}$, which in view of $\vartheta(z)=(\frac{z}{z^2-z+1})^2$ induces exactly 4 singularities of $\Xi_3(z)=\frac{1}{z^2-z+1}\Psi((\frac{z}{z^2-z+1})^2)$. Indeed, $\Psi(y^2)$ has the two singularities $\mathbb{C}$: $\beta_1=\frac{1}{4}$ and $\beta_2=-\frac{1}{4}$ which produce for $\Xi_3(z)$ the four singularities $\rho_3=\frac{5-\sqrt{21}}{2}$, $\zeta_2=\frac{5+\sqrt{21}}{2}$, $\zeta_3=\frac{-3-\sqrt{5}}{2}$ and $\zeta_4=\frac{-3+\sqrt{5}}{2}$. Therefore, $\rho_3,\zeta_2,\zeta_3,\zeta_4$ have distinct moduli and Claim 1 follows.

Claim 2: the singular expansion. $\Psi(z)$ is $\Delta_{\frac{1}{16}}(\phi, R)$-analytic and has the singular expansion $(1-16z)^4 \ln(\frac{1}{1-16z})$.

$$\forall z \in \Delta_{\frac{1}{16}}(\phi, R) \cap U\left(\frac{1}{16}, \epsilon\right); \quad \Psi(z) = O\left((1-16z)^4 \ln\left(\frac{1}{1-16z}\right)\right). \tag{4.8}$$

First, $\Delta_{\frac{1}{16}}(\phi, R)$-analyticity of the function $\Psi(z)$ is obvious. We proceed by proving that $(1-16z)^4 \ln(\frac{1}{1-16z})$ is its singular expansion in the intersection of a neighborhood of $\frac{1}{16}$ and the Δ-domain $\Delta_{\frac{1}{16}}(\phi, R)$. Using the notation of falling factorials $(n-1)_4 = (n-1)(n-2)(n-3)(n-4)$, we observe

$$f_3(2n, 0) = C_{n+2}C_n - C_{n+1}^2 = \frac{1}{(n-1)_4} \frac{12(n-1)_4(2n+1)}{(n+3)(n+1)^2(n+2)^2} \binom{2n}{n}^2.$$

With this expression for $f_3(2n, 0)$, we arrive at the formal identity

$$\sum_{n\geq 5} 16^{-n} f_3(2n, 0) z^n = O\left(\sum_{n\geq 5}\left[16^{-n} \frac{1}{(n-1)_4} \frac{12(n-1)_4(2n+1)}{(n+3)(n+1)^2(n+2)^2} \binom{2n}{n}^2 - \frac{4!}{(n-1)_4} \frac{1}{\pi} \frac{1}{n}\right] z^n + \sum_{n\geq 5} \frac{4!}{(n-1)_4} \frac{1}{\pi} \frac{1}{n} z^n\right),$$

where $f(z) = O(g(z))$ denotes that the limit $f(z)/g(z)$ is bounded for $z \to 1$, Eq. (4.4). It is clear that

$$\lim_{z\to 1}\left(\sum_{n\geq 5}\left[16^{-n} \frac{1}{(n-1)_4} \frac{12(n-1)_4(2n+1)}{(n+3)(n+1)^2(n+2)^2} \binom{2n}{n}^2 - \frac{4!}{(n-1)_4} \frac{1}{\pi} \frac{1}{n}\right] z^n\right)$$
$$= \sum_{n\geq 5}\left[16^{-n} \frac{1}{(n-1)_4} \frac{12(n-1)_4(2n+1)}{(n+3)(n+1)^2(n+2)^2} \binom{2n}{n}^2 - \frac{4!}{(n-1)_4} \frac{1}{\pi} \frac{1}{n}\right] < \kappa$$

for some $\kappa < 0.0784$. Therefore, we can conclude

$$\sum_{n\geq 5} 16^{-n} f_3(2n, 0) z^n = O\left(\sum_{n\geq 5} \frac{4!}{(n-1)_4} \frac{1}{\pi} \frac{1}{n} z^n\right). \tag{4.9}$$

We proceed by interpreting the power series on the rhs, observing

$$\forall n \geq 5; \quad [z^n]\left((1-z)^4 \ln \frac{1}{1-z}\right) = \frac{4!}{(n-1)\dots(n-4)} \frac{1}{n}, \tag{4.10}$$

when $((1-z)^4 \ln \frac{1}{1-z})$ is the unique analytic continuation of $\sum_{n\geq 5} \frac{4!}{(n-1)_4} \frac{1}{\pi} \frac{1}{n} z^n$. Using the scaling property of Taylor coefficients $[z^n]f(z) = \gamma^n [z^n] f(\frac{z}{\gamma})$, we obtain

$$\forall z \in \Delta_{\frac{1}{16}}(\phi, R) \cap U\left(\frac{1}{16}, \epsilon\right); \quad \Psi(z) = O\left((1-16z)^4 \ln\left(\frac{1}{1-16z}\right)\right). \tag{4.11}$$

Therefore, we have proved that $(1-16z)^4\ln(\frac{1}{1-16z})$ is the singular expansion of $\Psi(z)$ at $z=\frac{1}{16}$, hence Claim 2. Our last step consists in verifying that the type of the singularity does not change when passing from $\Psi(z)$ to $\Xi_3(z)=\frac{1}{z^2-z+1}\Psi((\frac{z}{z^2-z+1})^2)$. That is, we show that the singular expansion is not affected by substituting $\vartheta(z)=(\frac{z}{z^2-z+1})^2$.

Claim 3: the singularity persists. For $z\in\Delta_{\rho_3}(\phi,R)\cap U(\rho_3,\epsilon)$ we have $\Xi_3(z)=O((1-\frac{z}{\rho_3})^4\ln(\frac{1}{1-\frac{z}{\rho_3}}))$. To prove the claim, we first observe that Claim 2 and Lemma 2 imply

$$\Xi_3(z)=O\left(\frac{1}{z^2-z+1}\left[\left(1-16\left(\frac{z}{z^2-z+1}\right)^2\right)^4\ln\frac{1}{(1-16(\frac{z}{z^2-z+1})^2)}\right]\right).$$

The Taylor expansion of $q(z)=1-16(\frac{z}{z^2-z+1})^2$ at ρ_3 is given by $q(z)=\frac{\sqrt{21}}{5-\sqrt{21}}(\rho_3-z)+O(z-\rho_3)^2$ and setting $\alpha=\frac{\sqrt{21}}{5-\sqrt{21}}$ we compute

$$\begin{aligned}\frac{1}{z^2-z+1}\left[q(z)^4\ln\frac{1}{q(z)}\right]&=\frac{(\alpha(\rho_3-z)+O(z-\rho_3)^2)^4\ln\frac{1}{\alpha(\rho_3-z)+O(z-\rho_3)^2}}{(z-\rho_3)^2+(2\rho_3-1)(z-\rho_3)+\rho_3^2-\rho_3+1}\\&=\frac{\left([\alpha+O(z-\rho_3)](\rho_3-z)^4\ln\frac{1}{[\alpha+O(z-\rho_3)](\rho_3-z)}\right)}{O(z-\rho_3)+\rho_3^2-\rho_3+1}\\&=O\left((\rho_3-z)^4\ln\frac{1}{\rho_3-z}\right),\end{aligned}$$

whence Claim 3. Now we are in the position to employ Theorem 3, and obtain for $\mathsf{S}_3(n)$

$$\mathsf{S}_3(n)\sim K'[z^n]\left((\rho_3-z)^4\ln\frac{1}{\rho_3-z}\right)\sim K'\frac{4!}{n(n-1)\ldots(n-4)}\left(\frac{1}{\rho_3}\right)^n.$$

Of course K' can be computed from Theorem 1, explicitly $K'=10.4724$ and the proof of the theorem is complete. □

5. Proofs

Proof of Lemma 1: First, we observe that for $z,w\in[-1,1]$ the term w^2z^2-z+1 is strictly positive. We set

$$F_k(z,w)=\sum_{n\geq 0}\sum_{h\leq n/2}\mathsf{S}'_k(n,h)w^{2h}z^n \tag{5.1}$$

and compute

$$\begin{aligned}F_k(z,w) &= \sum_{n\ge 0}\sum_{h\le n/2}\sum_{j=0}^{h}(-1)^j\binom{n-j}{j}\binom{n-2j}{2(h-j)}f_k\bigl(2(h-j),0\bigr)w^{2h}z^n\\&= \sum_{n\ge 0}\sum_{j\le n/2}\sum_{h=j}^{n/2}(-1)^j\binom{n-j}{j}\binom{n-2j}{2(h-j)}f_k\bigl(2(h-j),0\bigr)w^{2h}z^n\\&= \sum_{j\ge 0}\sum_{n\ge 2j}\sum_{h=j}^{n/2}(-1)^j\binom{n-j}{j}\binom{n-2j}{2(h-j)}f_k\bigl(2(h-j),0\bigr)w^{2h}z^n\\&= \sum_{j\ge 0}(-1)^j\frac{(wz)^{2j}}{j!}\sum_{n\ge 2j}(n-j)!\sum_{h=j}^{n/2}\binom{n-2j}{2(h-j)}\\&\quad\times f_k\bigl(2(h-j),0\bigr)\frac{w^{2(h-j)}}{(n-2j)!}z^{n-2j}.\end{aligned}$$

We shift summation indices $n'=n-2j$ and $h'=h-j$ and derive for the rhs the following expression

$$\begin{aligned}&= \sum_{j\ge 0}(-1)^j\frac{(wz)^{2j}}{j!}\sum_{n'\ge 0}(n'+j)!\sum_{h=j}^{n/2}\binom{n'}{2(h-j)}f_k\bigl(2(h-j),0\bigr)\frac{w^{2(h-j)}}{n'!}z^{n-2j}\\&= \sum_{j\ge 0}(-1)^j\frac{(wz)^{2j}}{j!}\sum_{n'\ge 0}(n'+j)!\left\{\sum_{h'=0}^{n/2-j=n'/2}\binom{n'}{2h'}f_k(2h',0)w^{2h'}\right\}\frac{z^{n'}}{n'!}.\end{aligned}$$

The idea is now to interpret the term $\sum_{h'=0}^{n'/2}\binom{n'}{2h'}f_k(2h',0)w^{2h'}\frac{z^n}{n!}$ as a product of the two power series e^z and $\sum_{n\ge 0}f_k(2n,0)\frac{(wz)^{2n}}{(2n)!}$:

$$\begin{aligned}\sum_{\ell\ge 0}\frac{z^\ell}{\ell!}\sum_{n\ge 0}f_k(2n,0)\frac{(wz)^{2n}}{(2n)!} &= \sum_{n'\ge 0}\sum_{2n+\ell=n'}\left\{\frac{1}{\ell!}\frac{1}{(2n)!}f_k(2n,0)w^{2n}\right\}z^{n'}\\&= \sum_{n'\ge 0}\left\{\sum_{n=0}^{n'/2}\binom{n'}{2n}f_k(2n,0)w^{2n}\right\}\frac{z^{n'}}{n'!}.\end{aligned}$$

We set $\eta_{n'}=\{\sum_{h'=0}^{n'/2}\binom{n'}{2h'}f_k(2h',0)w^{2h'}\}$. By assumption, we have $|z|<\rho_k(w)$ and we next derive, using the Laplace transformation and interchanging integration and summation

$$\sum_{n'\ge 0}(n'+j)!\eta_n\frac{z^{n'}}{n'!}=\int_0^\infty\sum_{n'\ge 0}\eta_{n'}\frac{(zt)^{n'}}{n'!}t^je^{-t}dt. \tag{5.2}$$

Since $|z|<\rho_k(w)$, the above transformation is valid and using

$$\sum_{n'\ge 0}\left\{\sum_{n=0}^{n'/2}\binom{n'}{2n}f_k(2n,0)w^{2n}\right\}\frac{z^{n'}}{n'!}=\sum_{\ell\ge 0}\frac{z^\ell}{\ell!}\sum_{n\ge 0}f_k(2n,0)\frac{(wz)^{2n}}{(2n)!} \tag{5.3}$$

we accordingly obtain

$$\sum_{n'\geq 0}\eta_{n'}\frac{(zt)^{n'}}{n'!}t^j e^{-t}dt=\int_0^\infty e^{tz}\sum_{n\geq 0}f_k(2n,0)\frac{(wzt)^{2n}}{(2n)!}t^j e^{-t}dt. \tag{5.4}$$

The next step is to substitute the term $\sum_{n'\geq 0}(n'+j)!\eta_n\frac{z^{n'}}{n'!}$ in Eq. (5.2), hence consequently

$$\begin{aligned}F_k(z,w)&=\sum_{j\geq 0}(-1)^j\frac{(wz)^{2j}}{j!}\int_0^\infty e^{tz}\sum_{n\geq 0}f_k(2n,0)\frac{(wzt)^{2n}}{(2n)!}t^j e^{-t}dt\\&=\int_0^\infty\sum_{j\geq 0}(-1)^j\frac{(wz)^{2j}}{j!}e^{tz}\sum_{n\geq 0}f_k(2n,0)\frac{(wzt)^{2n}}{(2n)!}t^j e^{-t}dt.\end{aligned}$$

The summation over the index j is just an exponential function and we derive

$$\begin{aligned}&=\int_0^\infty e^{-(w^2z^2-z+1)t}\sum_{n\geq 0}f_k(2n,0)\frac{(wzt)^{2n}}{(2n)!}dt\\&=\int_0^\infty e^{-(w^2z^2-z+1)t}\sum_{n\geq 0}f_k(2n,0)\frac{1}{(2n)!}\left(\frac{wz}{w^2z^2-z+1}\right)^{2n}\left((w^2z^2-z+1)t\right)^{2n}dt.\end{aligned}$$

We proceed by transforming the integral introducing $u=(w^2z^2-z+1)t$, i.e. $dt=(w^2z^2-z+1)^{-1}du$ and accordingly arrive at

$$\begin{aligned}F_k(z,w)&=\sum_{n\geq 0}f_k(2n,0)\frac{1}{(2n)!}\left(\frac{wz}{w^2z^2-z+1}\right)^{2n}\\&\quad\times\int_0^\infty e^{-(w^2z^2-z+1)t}\left((w^2z^2-z+1)t\right)^{2n}dt\\&=\sum_{n\geq 0}f_k(2n,0)\frac{1}{(2n)!}\left(\frac{wz}{w^2z^2-z+1}\right)^{2n}\frac{1}{w^2z^2-z+1}(2n)!\\&=\frac{1}{w^2z^2-z+1}\sum_{n\geq 0}f_k(2n,0)\left(\frac{wz}{w^2z^2-z+1}\right)^{2n},\end{aligned}$$

hence the lemma. □

Proof of Theorem 4: We shall begin by deriving the asymptotics of $f_2(2n,0)=C_n$. Since $\sum_{n\geq 0}\binom{2n}{n}z^n=(1-4z)^{-\frac{1}{2}}$, we observe

$$C_n=\frac{1}{n+1}[z^n](1-4z)^{-\frac{1}{2}}$$

and according to Theorem 3, we can express C_n asymptotically as

$$C_n\sim\frac{4^n}{\sqrt{\pi n}}\left(\frac{1}{n+1}-\frac{1}{8n(n+1)}+\frac{1}{128n^3}+\frac{5}{1024n^4}+O\left(n^{-5}\right)\right). \tag{5.5}$$

The generating function of the Catalan numbers is given by

$$\Psi(y)=\sum_{n\geq 0} C_n y^n=\frac{1-\sqrt{1-4y}}{2y} \tag{5.6}$$

having a branch-point singularity at $\frac{1}{4}$. Lemma 2 allows us to express the analytic continuation of $\sum_{n\geq 0}\mathsf{S}_2(n)z^n$ via Ψ:

$$\Xi_2(z)=\frac{1}{z^2-z+1}\Psi\left(\left(\frac{z}{z^2-z+1}\right)^2\right) \tag{5.7}$$

$$=\frac{1}{z^2-z+1}\,\frac{1-\sqrt{1-4(\frac{z}{z^2-z+1})^2}}{2(\frac{z}{z^2-z+1})^2}=\frac{1-\sqrt{1-4(\frac{z}{z^2-z+1})^2}}{2\frac{z^2}{z^2-z+1}}. \tag{5.8}$$

The explicit form of $\Xi_2(z)$ allows us to conclude that $\rho_2=\frac{3-\sqrt{5}}{2}$ is the unique dominant singularity. We denote the map $z\mapsto(\frac{z}{z^2-z+1})^2$ by ϑ and compute the first terms of the Taylor series at $z=\rho_2$, i.e. where $\vartheta(\rho_2)=\frac{1}{16}$:

$$\vartheta(z)=\frac{1}{4}+\frac{5+3\sqrt{5}}{8}(z-\rho_2)+(z-\rho_2)^2T(z), \tag{5.9}$$

where $T(z)=\sum_{i\geq 0}c_i(z-\rho_2)^i$, $c_i\in\mathbb{R}$. Analyzing $\Xi_2(z)$ in an intersection of an ϵ-disc around ρ_2 with Δ_{ρ_2} produces

$$\Xi_2(z)=\frac{1-\sqrt{(\frac{5+3\sqrt{5}}{2})(\rho_2-z)-(z-\rho_2)^2T(z)}}{2\frac{z^2}{z^2-z+1}} \tag{5.10}$$

from which we immediately conclude

$$\Xi_2(\rho_2 z)=O\big(\Psi(4z)\big). \tag{5.11}$$

Theorem 3 and the scaling property of Taylor coefficients $[z^n]f(z)=\gamma^n[z^n]f(\frac{z}{\gamma})$ imply

$$K\big[z^n\big]\Xi_2(\rho_2 z)\sim\big[z^n\big]\Psi(4z),\quad \text{for some } K>0 \tag{5.12}$$

and we accordingly arrive substituting $\alpha=-\frac{1}{2}$ at

$$\big[z^n\big]\Xi_2(z)=\frac{K}{\sqrt{n}}\left(\frac{1}{n+1}-\frac{1}{8n(n+1)}+\frac{1}{128n^3}+\frac{5}{1024n^4}+O\big(n^{-5}\big)\right)\times\left(\frac{3+\sqrt{5}}{2}\right)^n,$$

for some $K>0$. Via Theorem 1, the coefficients $\mathsf{S}_2(n)$ are explicitly known and we compute $K=1.9572$ from which the theorem follows. □

Acknowledgements

We are grateful to Professor Jason Z. Gao for helpful comments. Many thanks to J.Z.M. Gao for helping to draw the figures. 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

Akutsu, T., 2000. Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots. Discret. Appl. Math. 104, 45–62.

Chamorro, M., Parkin, N., Varmus, H.E., 1991. An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. Proc. Natl. Acad. Sci. USA 89, 713–717.

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

Flajolet, P., 1999. Singularity analysis and asymptotics of Bernoulli sums. Theor. Comput. Sci. 215(1–2), 371–381.

Flajolet, P., Fill, J.A., Kapur, N., 2005. Singularity analysis, Hadamard products, and tree recurrences. J. Comput. Appl. Math. 174, 271–313.

Flajolet, P., Grabiner, P., Kirschenhofer, P., Prodinger, H., Tichy, R.F., 1994. Mellin transforms and asymptotics: digital sums. Theor. Comput. Sci. 123, 291–314.

Gao, Z., Richmond, L.B., 1992. Central and local limit theorems applied to asymptotic enumeration. J. Appl. Comput. Anal. 41, 177–186.

Gessel, I.M., Zeilberger, D., 1992. Random walk in a Weyl chamber. Proc. Am. Math. Soc. 115, 27–31.

Haslinger, C., Stadler, P.F., 1999. RNA Structures with pseudo-knots. Bull. Math. Biol. 61, 437–467.

Hofacker, I.L., Schuster, P., Stadler, P.F., 1998. Combinatorics of RNA secondary structures. Discret. Appl. Math. 88, 207–237.

Howell, J.A., Smith, T.F., Waterman, M.S., 1980. Computation of generating functions for biological molecules. SIAM J. Appl. Math. 39, 119–133.

Jin, E.Y., Qin, J., Reidys, C.M., 2007. Combinatorics of RNA structures with pseudoknots. Bull. Math. Biol., in press.

Konings, D.A.M., Gutell, R.R., 1995. A comparison of thermodynamic foldings with comparatively derived structures of 16s and 16s-like rRNAs. RNA 1, 559–574.

Lindstroem, B., 1973. On the vector representation of induced matroids. Bull. Lond. Math. Soc. 5, 85–90.

Loria, A., Pan, T., 1996. Domain structure of the ribozyme from eubacterial ribonuclease p. RNA 2, 551–563.

Lyngso, R., Pedersen, C., 1996. Pseudoknots in RNA secondary structures. In: H. Flyvbjerg, J. Hertz, M.H. Jensen, O.G. Mouritsen, K. Sneppen (Eds.), Physics of Biological Systems: From Molecules to Species. Springer, Berlin.

Mapping RNA Form and Function, 2005. Science 2, September 2005.

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

Odlyzko, A.M., 1992. Explicit tauberian estimates for functions with positive coefficients. J. Comput. Appl. Math. 41, 187–197.

Odlyzko, A.M., 1995. Handbook of Combinatorics. Elsevier, Amsterdam. Chapter 22.

Penner, R.C., Waterman, M.S., 1993. Spaces of RNA secondary structures. Adv. Math. 101, 31–49.

Popken, A., 1953. Asymptotic expansions from an algebraic standpoint. Indag. Math. 15, 131–143.

Rivas, E., Eddy, S., 1999. A dynamic programming algorithm for RNA structure prediction inclusing pseudoknots. J. Mol. Biol. 285, 2053–2068.

Schmitt, W.R., Waterman, M.S., 1994. Linear trees and RNA secondary structure. Discret. Appl. Math. 51, 317–323.

Tacker, M., Fontana, W., Stadler, P.F., Schuster, P., 1994. Statistics of RNA melting kinetics. Eur. Biophys. J. 23, 29–38.

Tacker, M., Stadler, P.F., Bauer, E.G., Hofacker, I.L., Schuster, P., 1996. Algorithm independent properties of RNA secondary structure predictions. Eur. Biophys. J. 25, 115–130.
Titchmarsh, E.C., 1939. The Theory of Functions. Oxford University Press, London.
Tuerk, C., MacDougal, S., Gold, L., 1992. RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. Proc. Natl. Acad. Sci. USA 89, 6988–6992.
Uemura, Y., Hasegawa, A., Kobayashi, S., Yokomori, T., 1999. Tree adjoining grammars for RNA structure prediction. Theor. Comput. Sci. 210, 277–303.
Waterman, M.S., 1978. Secondary structure of single-stranded nucleic acids. Adv. Math. I (suppl.) 1, 167–212.
Waterman, M.S., 1979. Combinatorics of RNA hairpins and cloverleafs. Stud. Appl. Math. 60, 91–96.
Waterman, M.S., Smith, T.F., 1986. Rapid dynamic programming algorithms for RNA secondary structure. Adv. Appl. Math. 7, 455–464.
Westhof, E., Jaeger, L., 1992. RNA pseudoknots. Curr. Opin. Struct. Biol. 2, 327–333.
Wong, R., Wyman, M., 1974. The method of Darboux. J. Approx. Theory 10, 159–171.
Zuker, M., Sankoff, D., 1984. RNA secondary structures and their prediction. Bull. Math. Biol. 46(4), 591–621.