

# ON THE DECOMPOSITION OF $k$ -NONCROSSING RNA STRUCTURES

EMMA Y. JIN AND CHRISTIAN M. REIDYS\*

ABSTRACT. A  $k$ -noncrossing RNA structure can be identified with a  $k$ -noncrossing diagram over  $[n]$ , which in turn corresponds to a vacillating tableaux having at most  $(k-1)$  rows. In this paper we derive the limit distribution of irreducible substructures via studying their corresponding vacillating tableaux. Our main result proves, that the limit distribution of the numbers of irreducible substructures in  $k$ -noncrossing,  $\sigma$ -canonical RNA structures is determined by the density function of a  $\Gamma(-\ln \tau_k, 2)$ -distribution for some  $\tau_k < 1$ .

## 1. INTRODUCTION AND BACKGROUND

In this paper we analyze the number of irreducible substructures of  $k$ -noncrossing,  $\sigma$ -canonical RNA structures. We prove that the numbers of irreducible substructures of  $k$ -noncrossing,  $\sigma$ -canonical RNA structures are, in the limit of long sequence length, given via the density function of a  $\Gamma(-\ln \tau_k, 2)$ -distribution.

An RNA structure is the helical configuration of its primary sequence, i.e. the sequence of nucleotides **A**, **G**, **U** and **C**, together with Watson-Crick (**A-U**, **G-C**) and (**U-G**) base pairs. As RNA structure is oftentimes tantamount to its function, it is of key importance. The concept of irreducibility in RNA structures is of central importance since the computation of the minimum free energy (mfe) configuration of a given RNA molecule is determined by its largest, irreducible substructure.

---

*Date:* January, 2009.

*2000 Mathematics Subject Classification.* 05A16.

*Key words and phrases.* singularity analysis,  $k$ -noncrossing diagram,  $k$ -noncrossing RNA structure, irreducible substructure, nontrivial return.

Three decades ago, Waterman [20, 27, 28, 13, 29] pioneered the combinatorics of RNA secondary structures, an RNA structure class exhibiting only noncrossing bonds. Secondary structures can readily be identified with Motzkin-paths satisfying some minimum height and plateau-length, see Figure 1. The latter restrictions arise from biophysical constraints due to mfe and the limited

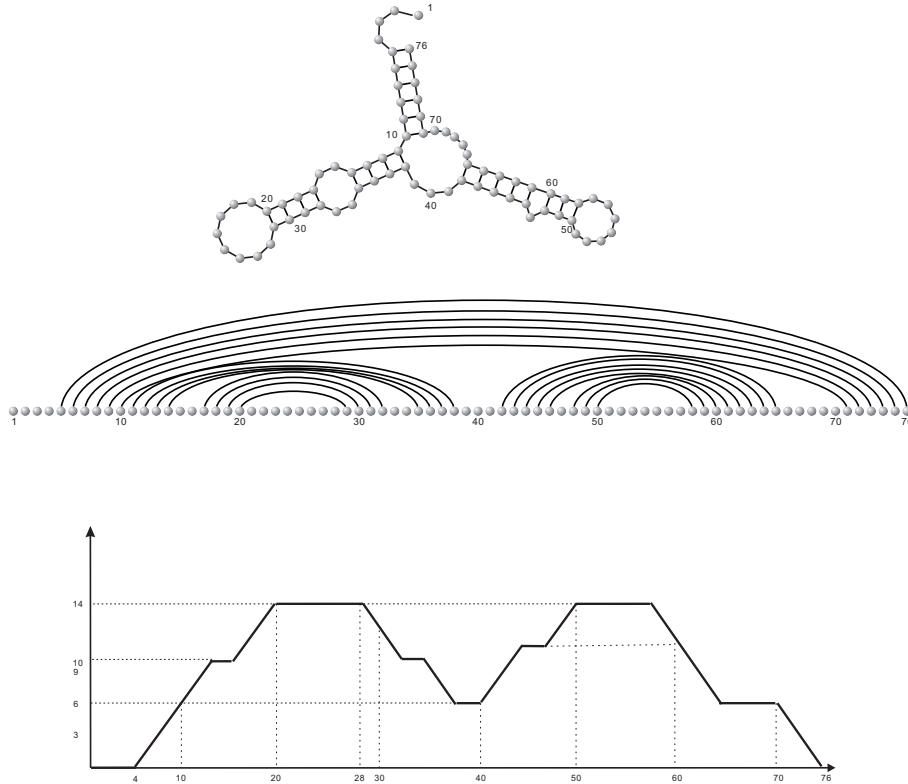


FIGURE 1. The phenylalanine tRNA secondary structure, as generated by the computer folding algorithm `cross` [14], represented as planar graph, diagram and Motzkin-path. The structure has arc-length  $\geq 8$  and stack-length  $\geq 3$  and uniquely corresponds to a Motzkin-path with minimum height 3 and minimum plateau-length 7.

flexibility of chemical bonds. It is clear from the particular bijection, that irreducible substructures in RNA secondary structures are closely related to the number of nontrivial returns, i.e. the number of non-endpoints, for which the Motzkin-path meets the  $x$ -axis.

For Dyck-paths this question has been studied by Shapiro [6], who showed that the expected number of nontrivial returns of Dyck-paths of length  $2n$  equals  $\frac{2n-2}{n+2}$ . Subsequently, Shapiro and Cameron [1] derived expectation and variance of the number of nontrivial returns for generalized Dyck-paths from  $(0,0)$  to  $((t+1)n,0)$

$$(1.1) \quad \mathbb{E}[\xi_t] = \frac{2n-2}{tn+2} \quad \text{and} \quad \mathbb{V}[\xi_t] = \frac{2tn(n-1)((t+1)n+1)}{(tn+2)^2(tn+3)}.$$

The bijection between a Dyck-path of length  $2n$  and a unique triangulation of the  $(n+2)$ -gon, due to Stanley [24], implies a combinatorial proof for  $\mathbb{E}[\xi_1]$ . An alternative approach is to employ the Riordan matrix [22], an infinite, lower triangular matrix  $L = (l_{n,k})_{n,k \geq 0} = (g, f)$ , where  $g(z) = \sum_{n \geq 0} g_n z^n$ ,  $f(z) = \sum_{n \geq 0} f_n z^n$  with  $f_0 = 0, f_1 \neq 0$ , such that  $\sum_{n \geq k} l_{n,k} z^n = g(z)f^k(z)$ . Clearly,

$$\mathbf{C}(z) = \sum_{n \geq 0} C_n z^n = \frac{1 - \sqrt{1-4z}}{2z} \quad \text{where} \quad C_n = \frac{1}{n+1} \binom{2n}{n}$$

is the generating function of Dyck-paths and let  $\zeta_{n,j}$  denote the number of Dyck-paths of length  $2n$  with  $j$  nontrivial returns. We consider the Riordan matrix  $L = (\zeta_{n,j})_{n,j \geq 0} = (z\mathbf{C}(z), z\mathbf{C}(z))$  and extract the coefficients  $\zeta_{n,j}$  from its generating function  $(z\mathbf{C}(z))^{j+1}$  by Lagrange inversion. Setting  $f(z) = zG(f(z))$  with  $f(z) = \mathbf{C}(z) - 1$  and  $G(z) = (1+z)^2$ , we obtain

$$\zeta_{n,j} = [z^{n-j-1}](f(z) + 1)^{j+1} = \frac{j+1}{2n-j-1} \binom{2n-j-1}{n},$$

where  $\sum_{j \geq 0} \zeta_{n,j} = C_n$ . From this we immediately compute  $\mathbb{E}[\xi_1] = \sum_{j \geq 1} j \cdot \frac{\zeta_{n,j}}{C_n}$  and  $\mathbb{V}[\xi_1] = \sum_{j \geq 1} j^2 \cdot \frac{\zeta_{n,j}}{C_n} - \left( \sum_{j \geq 1} j \cdot \frac{\zeta_{n,j}}{C_n} \right)^2$ , from which the expression of eq. (1.1), for  $t = 1$  follows.

In Section 3 we consider the bivariate generating function directly, which relates to the Riordan matrix in case of generalized Dyck-paths as follows

$$\sum_{n \geq 0} \sum_{j \geq 0} \zeta_{n,j} w^j z^n = \sum_{j \geq 0} z^{j+1} \mathbf{C}(z)^{j+1} w^j = \frac{z\mathbf{C}(z)}{1 - wz\mathbf{C}(z)}.$$

Our main idea is to derive the bivariate generating function from the Riordan matrix employing irreducible paths and to establish via singularity analysis a discrete limit law. This is done, however, for the far more general class of  $\mathfrak{C}$ -tableaux introduced in Section 2: in Theorem 7 we show that the limit distribution of nontrivial returns for these vacillating tableaux is given in terms of the density function of a  $\Gamma(\lambda, r)$ -distribution, which is, already for Motzkin-paths, a new result. For restricted Motzkin-paths satisfying specific height and plateau-lengths, the Riordan matrix Ansatz does not work “directly”, since the inductive decomposition of restricted Motzkin-paths is incompatible. Instead we introduce the notion of irreducible paths and express the Riordan

matrix in terms of the latter, see Lemma 2. This Ansatz allows us to compute the generating function of irreducible paths via setting one indeterminate of the bivariate generating function to one. The framework developed in Section 3 and Section 4, in fact works as long as the generating function of the particular path-class has a singular expansion and is explicitly known. We have, for instance, for nontrivial returns of Motzkin-paths with height  $\geq 3$  and plateau length  $\geq 3$ :  $\lim_{n \rightarrow \infty} \mathbb{E}[\eta_n] \approx 5.4526$  and  $\lim_{n \rightarrow \infty} \mathbb{V}[\eta_n] \approx 20.3179$ .

Indeed, RNA structures are far more complex than secondary structures: they exhibit additional, cross-serial nucleotide interactions [21]. These interactions were observed in natural RNA structures, as well as via comparative sequence analysis [30]. They are called pseudoknots, see Figure 2, and widely occur in functional RNA, like for instance, eP RNA [17] as well as ribosomal RNA [16]. RNA pseudoknots are conserved also in the catalytic core of group I introns. In plant viral RNAs pseudoknots mimic tRNA structure and in vitro RNA evolution [25] experiments have produced families of RNA structures with pseudoknot motifs, when binding HIV-1 reverse transcriptase.

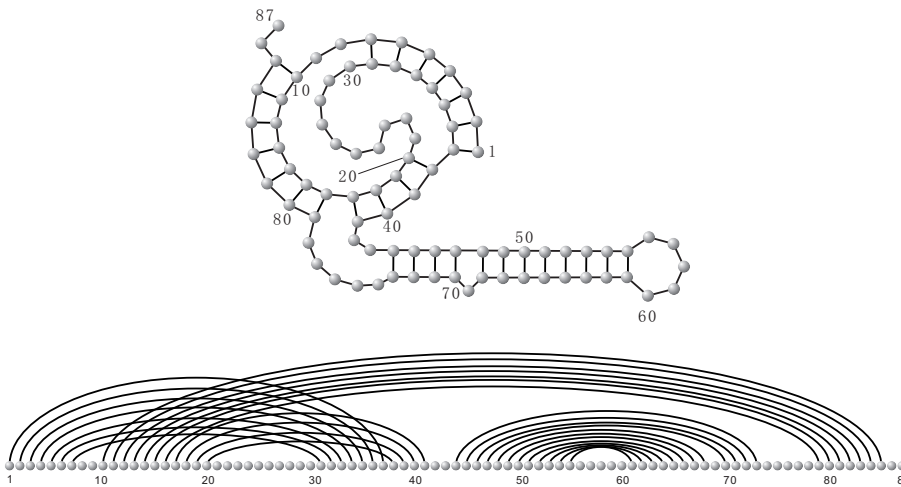


FIGURE 2. The hepatitis delta virus (HDV)-pseudoknot structure and its diagram representation. Top: the structure as folded by `cross` [14] for  $k = 3$  and minimum stack size 3 and the corresponding diagram representation (bottom).

Combinatorially, cross serial interactions are tantamount to crossing bonds. To this end, RNA pseudoknot structures have been modeled via  $k$ -noncrossing diagrams [9], i.e. labeled graphs over the vertex set  $[n] = \{1, \dots, n\}$  with degree  $\leq 1$ . Diagrams are represented by drawing their

vertices  $1, \dots, n$  in a horizontal line and its arcs  $(i, j)$ , where  $i < j$ , in the upper half plane. In the following, the degree of  $i$  refers to the number of non-horizontal arcs incident to  $i$ , i.e. the backbone of the primary sequence is not accounted for. The vertices and arcs correspond to nucleotides and Watson-Crick (**A-U**, **G-C**) and (**U-G**) base pairs, respectively, see Figure 2.

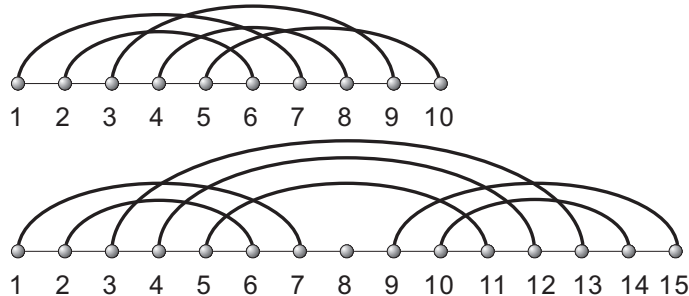


FIGURE 3.  $k$ -noncrossing diagrams: we display a 4-noncrossing, arc-length  $\lambda \geq 4$  and  $\sigma \geq 1$  diagram (top), where the edge set  $\{(1, 7), (3, 9), (5, 10)\}$  is a 3-crossing, the arc  $(2, 6)$  has length 4 and  $(5, 10)$  has stack-length 1. Below, we display a 3-noncrossing,  $\lambda \geq 4$  and  $\sigma \geq 2$  (lower) diagram, where  $(2, 6)$  has arc-length 4 and the stack  $((2, 6), (1, 7))$  has stack-length 2.

Natural RNA pseudoknot structures are typically 3-noncrossing [12]. However, relatively high numbers of pairwise crossing bonds occur in natural RNA structures. For instance, the *gag-pro* ribosomal frame-shift signal of the simian retrovirus-1 [5], which is a 10-noncrossing RNA structure motif, displayed in Figure 4.

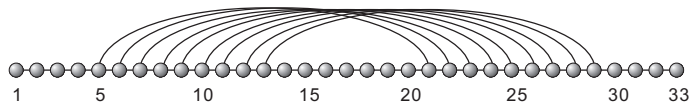


FIGURE 4. The proposed SRV-1 frame-shift [5] is a 10-noncrossing RNA structure motif.

Diagrams are characterized via their maximum number of mutually crossing arcs,  $k - 1$ , their minimum arc-length,  $\lambda$ , and their minimum stack-length,  $\sigma$ . A  $k$ -crossing is a set of  $k$  distinct arcs  $(i_1, j_1), (i_2, j_2), \dots, (i_k, j_k)$  with the property  $i_1 < i_2 < \dots < i_k < j_1 < j_2 < \dots < j_k$ . A diagram without any  $k$ -crossings is called a  $k$ -noncrossing diagram. The length of an arc  $(i, j)$  is  $j - i$  and

a stack of length  $\sigma$  is a sequence of “parallel” arcs of the form

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

A subdiagram of a  $k$ -noncrossing diagram is a subgraph over a subset  $M \subset [n]$  of consecutive vertices that starts with an origin and ends with a terminus of some arc. Let  $(i_1, \dots, i_m)$  be a sequence of isolated points, and  $(j_1, j_2)$  be an arc. We call  $(i_1, \dots, i_m)$  interior if and only if there exists some arc  $(j_1, j_2)$  such that  $j_1 < i_1 < i_m < j_2$  holds and exterior, otherwise. By abuse of language, a gap either contains no vertices, or is any exterior sequence of consecutive, isolated vertices. A diagram or subdiagram is called irreducible, if it cannot be decomposed into a sequence of gaps and subdiagrams, see Figure 5. Accordingly, any  $k$ -noncrossing diagram can be uniquely decomposed into an alternating sequence of gaps and irreducible subdiagrams. In fact irreducibility is quite common for natural RNA pseudoknot structures, see for instance, Figure 6.

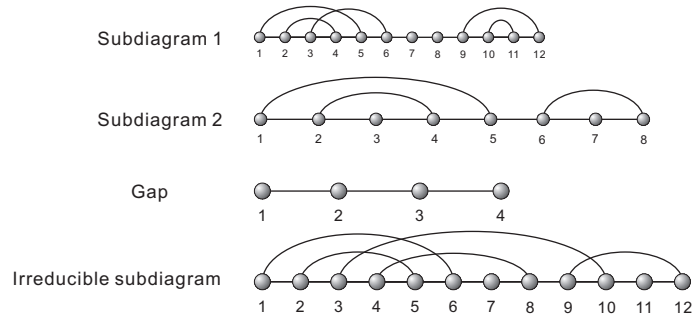


FIGURE 5. Subdiagrams, gaps and irreducibility: Subdiagram 1 is decomposed into the irreducible subdiagram over  $(1, 6)$ , the gap  $(7, 8)$  and the irreducible subdiagram over  $(9, 12)$ . Subdiagram 2 decomposes into the irreducible subdiagram over  $(1, 5)$ , the (empty) gap and the irreducible subdiagram over  $(6, 8)$ . Finally we display a gap and an irreducible diagram over  $(1, 12)$  (bottom).

We call a  $k$ -noncrossing diagram with arc-length  $\geq 4$  and stack-length  $\geq \sigma$ , a  $k$ -noncrossing,  $\sigma$ -canonical RNA structure, see Figure 3. We accordingly adopt the notions of gap, substructure and irreducibility for RNA structures.

Our main result is Theorem 6, which proves that the numbers of irreducible substructures are in the limit of long sequence length given via the density function of a  $\Gamma(-\ln \tau_k, 2)$ -distribution. Furthermore, we show that the probability generating function of the limit distribution is given

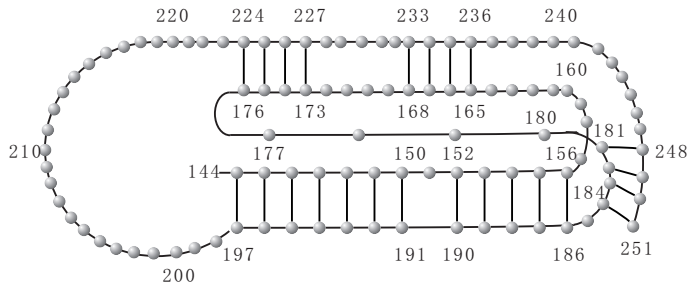


FIGURE 6. mRNA-Ec $_{\alpha}$ : the irreducible pseudoknot structure of the regulatory region of the  $\alpha$  ribosomal protein operon.

by  $q(u) = \frac{u(1-\tau_k)^2}{(1-\tau_k u)^2}$ , where  $\tau_k$  is expressed in terms of the generating function of  $k$ -noncrossing,  $\sigma$ -canonical RNA structures [18] and its dominant singularity  $\alpha_k$ . In Figure 7 we compare our analytic results with mfe secondary and 3-noncrossing structures generated by computer folding algorithms [26, 14], respectively. The data indicate that already for  $n = 75$ , the limit distribution of Theorem 6 provides a good fit for both structure classes.

The paper is organized as follows: in Section 2 we recall some basic combinatorial background. Of particular importance here is the bijection between  $k$ -noncrossing diagrams and vacillating tableaux of Theorem 1 with at most  $(k - 1)$  rows [4]. In Section 3, we present all the key ideas and derive the limit distribution of  $*$ -tableaux. In Section 4 we study the limit distribution of nontrivial returns using the framework developed in Section 3.

## 2. SOME BASIC FACTS

A *Ferrers diagram* (shape) is a collection of squares arranged in left-justified rows with weakly decreasing number of boxes in each row. A standard *Young tableau* (SYT) is a filling of the squares by numbers which is strictly decreasing in each row and in each column. We refer to standard Young tableaux as Young tableaux, see Figure 8. A *vacillating tableau*  $V_{\lambda}^{2n}$  of shape  $\lambda$  and length  $2n$  is a sequence of Ferrers diagrams  $(\lambda^0, \lambda^1, \dots, \lambda^{2n})$  of shapes such that (i)  $\lambda^0 = \emptyset$  and  $\lambda^{2n} = \lambda$ , and (ii)  $(\lambda^{2i-1}, \lambda^{2i})$  is derived from  $\lambda^{2i-2}$ , for  $1 \leq i \leq n$ , by one of the following operations.  $(\emptyset, \emptyset)$ : do nothing twice;  $(-\square, \emptyset)$ : first remove a square then do nothing;  $(\emptyset, +\square)$ : first do nothing then adding a square;  $(\pm\square, \pm\square)$ : add/remove a square at the odd and even steps, respectively. We denote the set of vacillating tableaux by  $\mathcal{V}_{\lambda}^{2n}$ . The *RSK-algorithm* is a process of

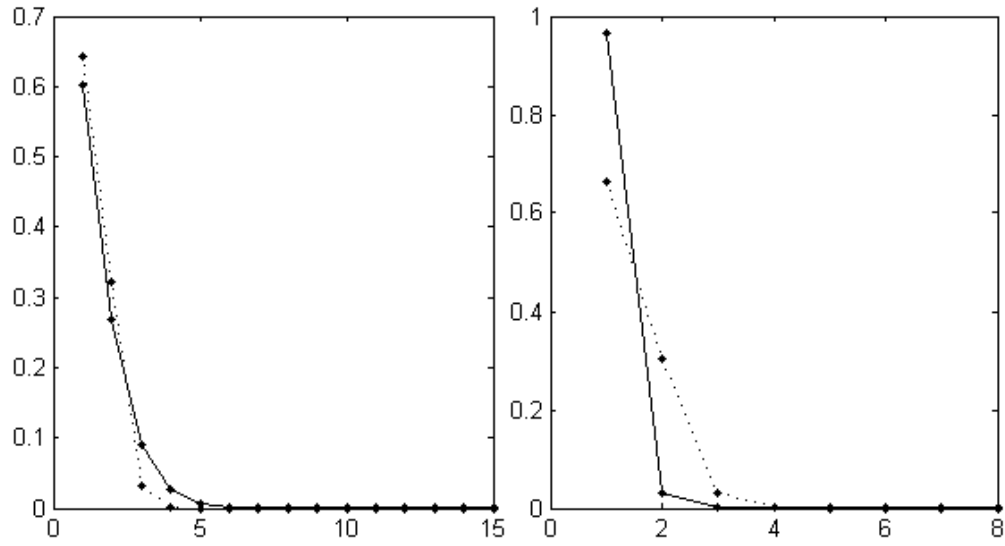
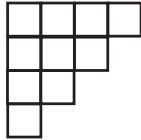


FIGURE 7. For  $n = 75$  the lhs displays the distribution of irreducible substructures obtained by folding  $10^4$  random sequences into their RNA secondary structures [26] (dashed), and the scaled density function of a  $\Gamma(-\ln(0.2241), 2)$ -distribution (solid) sampled at the positive integers. The rhs shows this distribution obtained by folding  $9 \times 10^3$  random sequences into 3-noncrossing, 3-canonical structures [14] (dashed) and the scaled density function of a  $\Gamma(-\ln(0.0167), 2)$ -distribution (solid) derived from Theorem 6.

Ferrers diagram



Young tableau

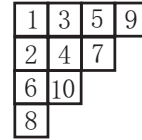


FIGURE 8. Ferrers diagram and Young tableau.

row-inserting elements into a Young tableau. Suppose we want to insert  $q$  into a standard Young tableau of shape  $\lambda$ . Let  $\lambda_{i,j}$  denote the element in the  $i$ -th row and  $j$ -th column of the Young tableau. Let  $j$  be the largest integer such that  $\lambda_{1,j-1} \leq q$ . (If  $\lambda_{1,1} > q$ , then  $j = 1$ .) If  $\lambda_{1,j}$  does not exist, then simply add  $q$  at the end of the first row. Otherwise, if  $\lambda_{1,j}$  exists, then replace  $\lambda_{1,j}$  by  $q$ . Next insert  $\lambda_{1,j}$  into the second row following the above procedure and continue until an element is inserted at the end of a row. As a result, we obtain a new standard Young tableau with



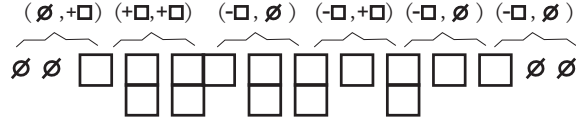


FIGURE 9. A vacillating tableau of shape  $\emptyset$  and length 12.

$q$  included. For instance, inserting the sequence 5, 2, 4, 1, 6, 3, starting with an empty shape yields the standard Young tableaux displayed in Figure 10.

The RSK-insertion algorithm has an inverse [4], see Lemma 1 below, which will be of central

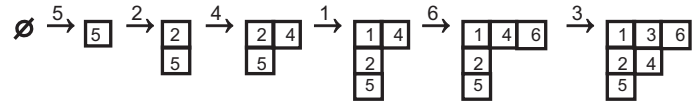


FIGURE 10. RSK-insertion of the elements 5, 2, 4, 1, 6, 3. The insertion of the above sequence successively constructs a standard Young tableau.

importance for constructing a vacillating tableaux from a tangled diagram.

**Lemma 1.** *Suppose we are given two shapes  $\lambda^i \subsetneq \lambda^{i-1}$ , which differ by exactly one square. Let  $T_{i-1}$  and  $T_i$  be SYT of shape  $\lambda^{i-1}$  and  $\lambda^i$ , respectively. Given  $\lambda^i$  and  $T_{i-1}$ , then there exists a unique  $j$  contained in  $T_{i-1}$  and a unique tableau  $T_i$  such that  $T_{i-1}$  is obtained from  $T_i$  by inserting  $j$  via the RSK-algorithm.*

In addition, Lemma 1 explicitly constructs this unique  $j$  such that  $T_{i-1}$  is obtained from  $T_i$  by inserting  $j$  via the RSK-algorithm, see Figure 11. Here  $r$  denotes the label of the row out of which the entry  $x_r$  has been extracted (which in turn is inserted into row  $r-1$ , leading to the extraction of  $x_{r-1}$ ). As for Figure 11, we begin by extracting 4 out of the third row. Next we remove the square in this row and replace the entry 2 of the second row (being maximal subject to the condition  $x_2 < 4$ ) by 4. We proceed accordingly, extracting 1 from the first row.

**2.1. From diagrams to vacillating tableaux and back.** RNA tertiary interactions, in particular the interactions between helical and non-helical regions give rise to consider tangled diagrams [4]. The key feature of tangled diagrams (tangles) is to allow for *two* interactions: one being Watson-Crick or **G-U** and the other being a hydrogen bond for each nucleotide. A *tangled diagram*,  $G_n$ , over  $[n]$  is obtained by drawing its arcs in the upper halfplane having vertices of degree

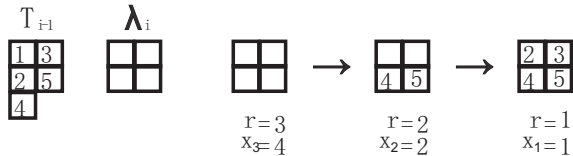


FIGURE 11. How Lemma 1 works: given the Young tableau,  $T_{i-1}$  and the shape  $\lambda_i$ , we show how to find the unique  $j$  (note here we have  $j = x_1 = 1$ ) such that  $T_{i-1}$  is obtained from  $T_i$  by inserting 1 via the RSK-algorithm.

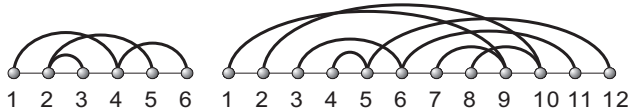


FIGURE 12. Tangled diagrams: the first tangled diagram represents the key bonds of the hammerhead ribosome and the second tangle represents key bonds of the catalytic core region of the Group I self-splicing intron [2].

at most two and a specific notion of crossings and nestings [4]. The *inflation*, of a tangle is a diagram, obtained by “splitting” each vertex of degree two,  $j$ , into two vertices  $j$  and  $j'$  having degree one, see Figure 13. Accordingly, a tangled diagram with  $\ell$  vertices of degree two is expanded into a diagram over  $n + \ell$  vertices. Obviously, the inflation has its unique inverse, obtained by simply identifying the vertices  $j, j'$ . By construction, the inflation preserves the maximal number

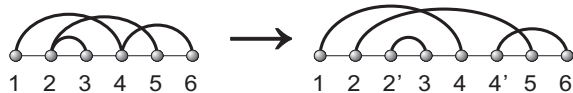


FIGURE 13. The inflation of the first tangled diagram displayed in Figure 12.

of mutually crossing and nesting arcs [4]. Given a  $k$ -noncrossing tangle, we can construct a vacillating tableaux, using the following algorithm: starting from right to left, we take three types of actions: we either RSK-insert, extract (via Lemma 1) or do nothing, depending on whether we are given an terminus, origin or isolated point of the inflated tangle. In fact, each arc in the inflated structure is considered twice: for its terminus and origin, respectively, see Figure 14.

To be explicit: we first read the vertices of the inflated tangle from right-to-left. For an inflated tangle having  $n$  vertices, we will construct the sequence of Young tableaux  $\{T_m\}_{m=0}^{2n}$ . To this

end we set  $T_{2n} = \emptyset$ . Starting from vertex  $i = n, n-1, \dots, 1$  we derive inductively the sequence  $(T_{2n}, T_{2n-1}, \dots, T_0)$  as follows:

- (1) If the  $i$ -th vertex,  $x$ , is a terminus of an arc  $(j, x)$ : set  $T_{2i-1} = T_{2i}$ . Next we derive  $T_{2i-2}$  from  $T_{2i-1}$  by adding  $j$  via the RSK-algorithm to  $T_{2i-1}$ .
- (2) If the  $i$ -th vertex,  $x$ , is an isolated vertex, set  $T_{2i} = T_{2i-1} = T_{2i-2}$ .
- (3) If the  $i$ -th vertex,  $x$ , is the origin of an arc  $(x, k)$ , then let  $T_{2i-1}$  be the standard Young tableau obtained by removing the square containing  $x$  from  $T_{2i}$ . Furthermore set  $T_{2i-1} = T_{2i-2}$ .

In fact, the above algorithm has a unique inverse: from a vacillating tableaux, we can derive a unique tangle, see Figure 15. For  $+\square$  steps one simply inserts into the tableaux, does nothing for  $\emptyset$  steps and RSK-extracts (Lemma 1) for  $-\square$  steps. As a result (see Figure 14 and Figure 15) we derive the following theorem [4].

**Theorem 1.** *There exists a bijection between  $k$ -noncrossing tangled diagrams and vacillating tableaux of type  $\mathcal{V}_{\emptyset}^{2n}$  having shapes  $\lambda^i$  with less than  $k$  rows.*

Theorem 1 implies bijections between various subclasses of vacillating tableaux and subclasses of tangles. Most notably the bijection [3] between  $k$ -noncrossing diagrams and vacillating tableaux (of empty shape) such that (i)  $\lambda^0 = \emptyset$  and  $\lambda^{2n} = \emptyset$ , and (ii)  $(\lambda^{2i-1}, \lambda^{2i})$  is derived from  $\lambda^{2i-2}$ , for  $1 \leq i \leq n$ , by one of the following operations.  $(\emptyset, \emptyset)$ : do nothing twice;  $(-\square, \emptyset)$ : first remove a square then do nothing;  $(\emptyset, +\square)$ : first do nothing then adding a square. We refer to the latter as  $\dagger$ -tableaux. Obviously, the latter are completely determined by the sequence of shapes  $(\lambda^2, \lambda^4, \dots, \lambda^{2n-2})$ .

**2.2.  $k$ -noncrossing RNA structures.** The combinatorics of  $k$ -noncrossing RNA pseudoknot structures has been derived in [9, 10]. The set (number) of  $k$ -noncrossing,  $\sigma$ -canonical RNA structures is denoted by  $T_{k,\sigma}(n)$  ( $\mathbb{T}_{k,\sigma}(n)$ ) and let  $f_k(n, \ell)$  denote the number of  $k$ -noncrossing diagrams with arbitrary arc-length and  $\ell$  isolated vertices over  $[n]$ . It follows from Theorem 1, that the number of  $k$ -noncrossing matchings on  $[2n]$  equals the number of walks from  $(k-1, k-2, \dots, 1)$  to itself that stay inside the Weyl Chamber  $x_1 > x_2 > \dots > x_{k-1} > 0$  with steps  $\pm e_i$ ,  $1 \leq i \leq k-1$ . The latter is given by Grabiner *et al.* [8]. It is exactly the situation  $\eta = \lambda = (k-1, k-2, \dots, 1)$

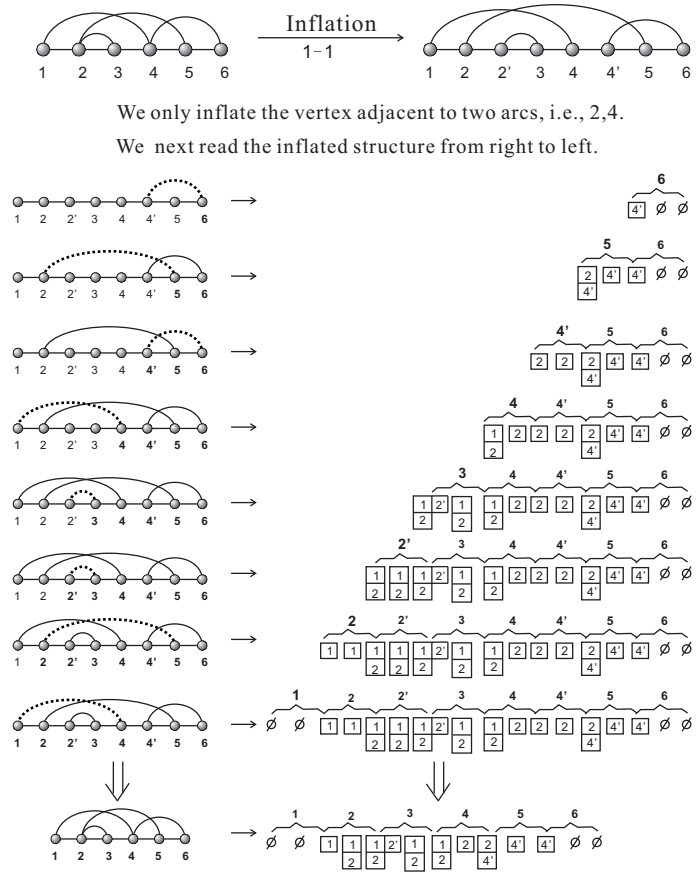


FIGURE 14. From tangled diagrams to vacillating tableaux via inflation: for the first tangled diagram displayed in Figure 12 we first inflate (top) and second derive the bijection into the corresponding tangled diagram  $(T_{16}, T_{15}, \dots, T_0)$  starting from right to left, by RSK-insertion. Finally, identifying  $i$  and  $i'$ , the vertices of the inflated tangle (lhs, bottom) and the labels of the steps in the vacillating tableaux (rhs, bottom), respectively, we obtain the desired mapping.

of equation (38) in [8]. As shown in detail in [9], Lemma 2

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

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

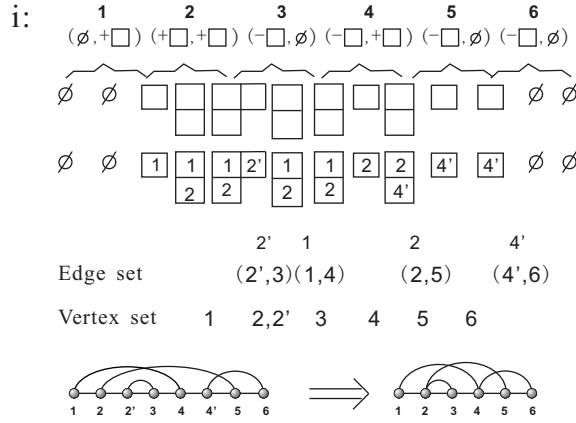


FIGURE 15. From vacillating tableaux to tangled diagrams: For  $+\square$  steps one simply inserts into the tableaux, does nothing for  $\emptyset$  and RSK-extracts (Lemma 1) for  $-\square$  steps. The labels  $2', 1, 2, 4'$  are obtained by RSK-extraction and correspond to the “ $-\square$ ” steps for  $i = 3, 4, 5, 6$ .

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  $r$ . In particular for  $k = 2$  and  $k = 3$  we have the formulas

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

In view of  $f_k(n, \ell) = \binom{n}{\ell} f_k(n - \ell, 0)$  everything can be reduced to matchings, where we have the following situation: there exists an asymptotic approximation of the determinant of hyperbolic Bessel function for general order  $k$  due to [15] and employing the subtraction of singularities-principle [19] one can prove [15]

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

Let  $\mathbf{F}_k(z) = \sum_{n \geq 0} f_k(2n, 0) z^{2n}$  denote the generating function of  $k$ -noncrossing matchings. Setting

$$w_0(x) = \frac{x^{2\sigma-2}}{1-x^2+x^{2\sigma}} \quad \text{and} \quad v_0(x) = 1-x+w_0(x)x^2+w_0(x)x^3+w_0(x)x^4$$

we can now state the following result [18].

**Theorem 2.** *Let  $k, \sigma \in \mathbb{N}$ , where  $k \geq 2, \sigma \geq 3$ , let  $x$  be an indeterminat and  $\rho_k = \frac{1}{2(k-1)}$  the dominant, positive real singularity of  $\mathbf{F}_k(z)$ . Then  $\mathbf{T}_{k,\sigma}(x)$ , the generating function of  $k$ -noncrossing,  $\sigma$ -canonical structures, is given by*

$$(2.5) \quad \mathbf{T}_{k,\sigma}(x) = \frac{1}{v_0(x)} \mathbf{F}_k \left( \frac{\sqrt{w_0(x)}x}{v_0(x)} \right).$$

Furthermore,

$$(2.6) \quad \mathbf{T}_{k,\sigma}(n) \sim c_k n^{-(k-1)^2-(k-1)/2} \left( \frac{1}{\gamma_{k,\sigma}} \right)^n, \quad \text{for } k = 2, 3, 4, \dots, 9,$$

holds, where  $\gamma_{k,\sigma}$  is the minimal positive real solution of the equation  $\frac{\sqrt{w_0(x)}x}{v_0(x)} = \rho_k = \frac{1}{2(k-1)}$ .

Via Theorem 1 each  $k$ -noncrossing,  $\sigma$ -canonical structure corresponds to a unique  $\dagger$ -tableau. We refer to the set of these tableaux as  $\mathfrak{C}$ -tableaux.

**2.3. Singularity analysis.** In view of Theorem 2 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 [7]. We use the notation

$$(2.7) \quad (f(z) = O(g(z)) \text{ as } z \rightarrow \rho) \iff (f(z)/g(z) \text{ is bounded as } z \rightarrow \rho)$$

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

**Theorem 3.** [7] *Let  $f(z), g(z)$  be  $D$ -finite functions with unique dominant singularity  $\rho$  and suppose  $f(z) = O(g(z))$  for  $z \rightarrow \rho$ . Then we have*

$$(2.8) \quad [z^n]f(z) = K \left( 1 - O\left(\frac{1}{n}\right) \right) [z^n]g(z),$$

where  $K$  is some constant.

Theorem 3 and eq. (2.4) imply

$$(2.9) \quad \mathbf{F}_k(z) = \begin{cases} O\left(\left(1 - \frac{z}{\rho_k}\right)^{(k-1)^2+(k-1)/2-1} \ln\left(1 - \frac{z}{\rho_k}\right)\right) & \text{for } k \text{ odd, } z \rightarrow \rho_k \\ O\left(\left(1 - \frac{z}{\rho_k}\right)^{(k-1)^2+(k-1)/2-1}\right) & \text{for } k \text{ even, } z \rightarrow \rho_k, \end{cases}$$

in accordance with basic structure theorems for singular expansions of  $D$ -finite functions [7]. Furthermore, Theorem 3, eq. (2.4) and the so called subcritical case of singularity analysis [7], VI.9., p. 411, imply the following result tailored for our functional equations [11]. Let  $\rho_k$  denote the dominant positive real singularity of  $\mathbf{F}_k(z)$ .

**Theorem 4.** *Suppose  $\vartheta_\sigma(z)$  is algebraic over  $K(z)$ , analytic for  $|z| < \delta$  and satisfies  $\vartheta_\sigma(0) = 0$ . Suppose further  $\gamma_{k,\sigma}$  is the real unique solution with minimal modulus  $< \delta$  of the two equations  $\vartheta_\sigma(z) = \rho_k$  and  $\vartheta_\sigma(z) = -\rho_k$ . Then*

$$(2.10) \quad [z^n] \mathbf{F}_k(\vartheta_\sigma(z)) \sim c_k n^{-((k-1)^2+(k-1)/2)} \left(\gamma_{k,\sigma}^{-1}\right)^n.$$

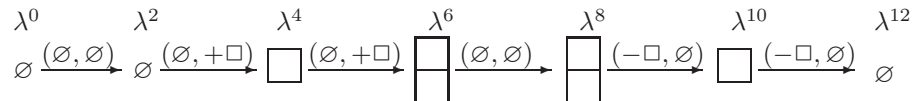
The below continuity theorem of discrete limit laws will be used in the proofs of Theorem 6 and Theorem 7. It ensures that under certain conditions the point-wise convergence of probability generating functions implicates the convergence of its coefficients.

**Theorem 5.** *Let  $u$  be an indeterminat and  $\Omega$  be a set contained in the unit disc, having at least one accumulation point in the interior of the disc. Assume  $P_n(u) = \sum_{k \geq 0} p_{n,k} u^k$  and  $q(u) = \sum_{k \geq 0} q_k u^k$  such that  $\lim_{n \rightarrow \infty} P_n(u) = q(u)$  for each  $u \in \Omega$  holds. Then we have for any finite  $k$ ,*

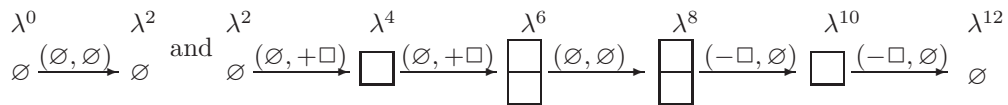
$$(2.11) \quad \lim_{n \rightarrow \infty} p_{n,k} = q_k \quad \text{and} \quad \lim_{n \rightarrow \infty} \sum_{j \leq k} p_{n,j} = \sum_{j \leq k} q_j.$$

### 3. IRREDUCIBLE SUBSTRUCTURES

In the following we shall identify a  $\mathfrak{C}$ -tableaux with the subsequence of even-indexed shapes, i.e. the sequence  $(\lambda^2, \dots, \lambda^{2n-2})$ . Subsequences of two or more consecutive  $\emptyset$ -shapes result from the elementary move  $(\emptyset, \emptyset)$ . For instance, consider the  $\mathfrak{C}$ -tableaux

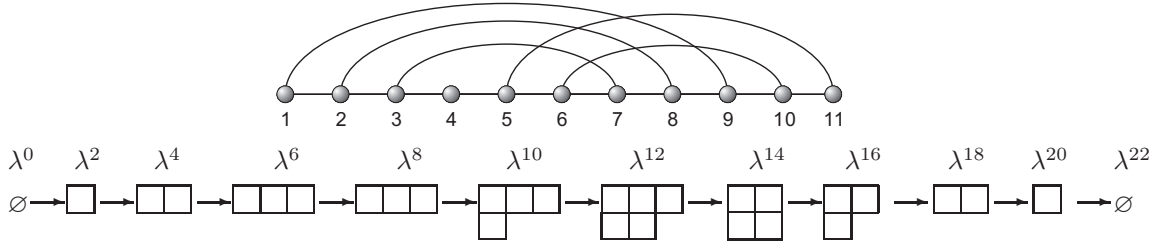


The above tableaux splits at  $\lambda^2 = \emptyset$  into two  $\mathfrak{C}$ -subtableaux, i.e.

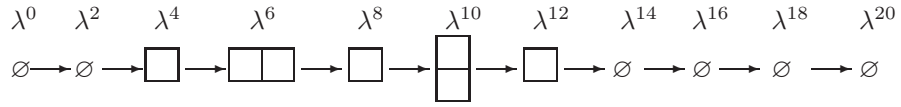


We call a sequence of consecutive  $\emptyset$ -shapes of length  $(r + 1)$ ,  $(\emptyset, \dots, \emptyset)$  a gap of length  $r$ . In particular, the empty gap is the  $\emptyset$ -shape of length 0. Theorem 1 implies that these  $\emptyset$ -gaps correspond uniquely to the gaps of diagrams, introduced in Section 2. A  $*$ -tableaux is a  $\mathfrak{C}$ -tableaux,

with the property  $\lambda^i \neq \emptyset$  for  $2 \leq i \leq 2n - 2$ . It is evident that a  $*$ -tableaux corresponds via the bijection of Theorem 1 to an irreducible  $k$ -noncrossing,  $\sigma$ -canonical RNA structure. For instance,



Obviously, any  $\mathfrak{C}$ -tableaux can be uniquely decomposed into a sequences of gaps and  $*$ -tableaux. For instance,



splits into the gap  $(\lambda^0, \lambda^2)$ , the  $*$ -tableaux over  $(\lambda^2, \dots, \lambda^{14})$  and the gap  $(\lambda^{14}, \dots, \lambda^{20})$ . Let  $\delta_{n,j}^{(k)}$  denote the number of  $\mathfrak{C}$ -tableaux of length  $2n$  with less than  $k$  rows, containing exactly  $j$   $*$ -tableaux. Furthermore, let

$$(3.1) \quad \mathbf{U}_k(z, u) = \sum_{n \geq 0} \sum_{j \geq 0} \delta_{n,j}^{(k)} u^j z^n,$$

and  $\delta_n^{(k)} = \sum_{j \geq 0} \delta_{n,j}^{(k)}$ . We set  $\mathbf{T}_k(z) = \mathbf{T}_{k,\sigma}(z) = \sum_{n \geq 0} \delta_n^{(k)} z^n$  and denote the generating function of  $*$ -tableaux by  $\mathbf{R}_k(z)$ .

**Lemma 2.** *The bivariate generating function of the number of  $\mathfrak{C}$ -tableaux of length  $2n$  with less than  $k$  rows, which contain exactly  $i$   $*$ -tableaux, is given by*

$$\mathbf{U}_k(z, u) = \frac{\frac{1}{1-z}}{1 - u \left( 1 - \frac{1}{(1-z)\mathbf{T}_k(z)} \right)}.$$

*Proof.* Since each  $\mathfrak{C}$ -tableau can be uniquely decomposed into a sequence of gaps and  $*$ -tableaux we obtain for fixed  $j$

$$(3.2) \quad \sum_{n \geq j} \delta_{n,j} z^n = \mathbf{R}_k(z)^j \left( \frac{1}{1-z} \right)^{j+1}.$$



As a result the bivariate generating function of  $\delta_{n,j}$  is given by

$$(3.3) \quad \mathbf{U}_k(z, u) = \sum_{j \geq 0} \sum_{n \geq j} \delta_{n,j} z^n u^j = \sum_{j \geq 0} \mathbf{R}_k(z)^j \left( \frac{1}{1-z} \right)^{j+1} u^j = \frac{1}{1-z-u\mathbf{R}_k(z)}.$$

Setting  $u = 1$  we derive

$$(3.4) \quad \mathbf{T}_k(z) = \mathbf{U}_k(z, 1) = \frac{1}{1-z-\mathbf{R}_k(z)}$$

which allows us to express the generating function of  $*$ -tableaux via  $\mathbf{T}_k(z)$

$$(3.5) \quad \mathbf{R}_k(z) = 1-z-\frac{1}{\mathbf{T}_k(z)}.$$

Consequently,  $\mathbf{U}_k(z, u)$  is given by

$$(3.6) \quad \mathbf{U}_k(z, u) = \frac{1}{1-z-u\mathbf{R}_k(z)} = \frac{\frac{1}{1-z}}{1-u\left(1-\frac{1}{(1-z)\mathbf{T}_k(z)}\right)}$$

and the lemma follows.  $\square$

Setting  $g(z) = \frac{1}{1-z}$  and  $h(z) = 1 - \frac{1}{(1-z)\mathbf{T}_k(z)}$ , Lemma 2 implies

$$(3.7) \quad \mathbf{U}_k(z, u) = g(z) \cdot \frac{1}{1-uh(z)} = g(z) \cdot g(uh(z)).$$

Let  $\xi_n^{(k)}$  be a r.v. such that  $\mathbb{P}(\xi_n^{(k)} = i) = \frac{\delta_{n,i}^{(k)}}{\delta_n^{(k)}}$  and let  $\rho_p$  and  $\rho_w$  denote the radius of convergence of the power series  $p(z)$  and  $w(z)$ , respectively. We denote  $\tau_w = \lim_{z \rightarrow \rho_w^-} w(z)$  and call a function  $F(z, u) = p(u \cdot w(z))$  subcritical if and only if  $\tau_w < \rho_p$ .

**Theorem 6.** *Let  $\alpha_k$  be the real positive dominant singularity of  $\mathbf{T}_k(z)$  and  $\tau_k = 1 - \frac{1}{(1-\alpha_k)\mathbf{T}_k(\alpha_k)}$ . Then the r.v.  $\xi_n^{(k)}$  satisfies the discrete limit law*

$$(3.8) \quad \lim_{n \rightarrow \infty} \mathbb{P}(\xi_n^{(k)} = i) = q_i \quad \text{where} \quad q_i = \frac{(1-\tau_k)^2}{\tau_k} i \tau_k^i.$$

*That is,  $\xi_n^{(k)}$  is determined by the density function of a  $\Gamma(-\ln \tau_k, 2)$ -distribution. Furthermore, the probability generating function of the limit distribution  $q(u) = \sum_{n \geq 1} q_n u^n$  satisfies  $q(u) = \frac{u(1-\tau_k)^2}{(1-\tau_k u)^2}$ .*

*Proof.* Since  $g(z) = \frac{1}{1-z}$  and  $h(z) = 1 - \frac{1}{(1-z)\mathbf{T}_k(z)}$  have non negative coefficients and  $h(0) = 0$ , the composition  $g(h(z))$  is well defined as a formal power series. According to eq. (3.7) we may express  $\mathbf{U}_k(z, u)$  as  $\mathbf{U}_k(z, u) = g(z)g(uh(z))$ . For  $z = \alpha_k$  we have  $\tau_k = 1 - \frac{1}{(1-\alpha_k)\mathbf{T}_k(\alpha_k)} < 1 = \rho_g$ , i.e. we are given the subcritical case.

*Claim 1.*  $h(z)$  has a singular expansion at its dominant singularity  $z = \alpha_k$  and there exists some constant  $c_k > 0$  such that

$$(3.9) \quad h(z) = \begin{cases} \tau_k - c_k \left(1 - \frac{z}{\alpha_k}\right)^\mu \ln\left(1 - \frac{z}{\alpha_k}\right) (1 + o(1)) & \text{for } k \equiv 1 \pmod{2} \\ \tau_k - c_k \left(1 - \frac{z}{\alpha_k}\right)^\mu (1 + o(1)) & \text{for } k \equiv 0 \pmod{2} \end{cases}$$

for  $z \rightarrow \alpha_k$  and  $\mu = (k-1)^2 + \frac{k-1}{2} - 1$ .

Since  $\mathbf{F}_k(z)$  is  $D$ -finite, the composition  $\mathbf{F}_k(\vartheta(z))$  where  $\vartheta(z) = \frac{\sqrt{w_0(z)}z}{v_0(z)}$  and  $\vartheta(0) = 0$ , is also  $D$ -finite [23]. As a result,  $\mathbf{T}_k(z)$  is, being a product of the two  $D$ -finite functions  $\frac{1}{v_0(z)}$  and  $\mathbf{F}_k(\vartheta(z))$ ,  $D$ -finite. Its  $D$ -finiteness guarantees that  $\mathbf{T}_k(z)$  has an analytic continuation  $\mathbf{T}_k^*(z)$  for which  $\mathbf{T}_k(z) = \mathbf{T}_k^*(z)$  holds for some simply connected  $\Delta_{\alpha_k}$ -domain [23]. In light of the fact that  $h(z) = 1 - \frac{1}{(1-z)\mathbf{T}_k(z)}$  has the same dominant singularity as  $\mathbf{T}_k(z)$ ,  $1 - \frac{1}{(1-z)\mathbf{T}_k(z)} = 1 - \frac{1}{(1-z)\mathbf{T}_k^*(z)}$  holds for some simply connected  $\Delta_{\alpha_k}$ -domain and  $1 - \frac{1}{(1-z)\mathbf{T}_k^*(z)}$  is the analytic continuation of  $h(z)$ . Consequently, the singular expansion of  $h(z)$  at  $z = \alpha_k$  does exist and

$$\begin{aligned} h(z) &= \tau_k + h'(\alpha_k)(z - \alpha_k) + \frac{h''(\alpha_k)}{2!}(z - \alpha_k)^2 + \dots \\ &= \tau_k + \frac{\mathbf{T}'_k(\alpha_k)}{\mathbf{T}_k^2(\alpha_k)}(z - \alpha_k) + \left[ \frac{\mathbf{T}''_k(\alpha_k)}{2\mathbf{T}_k^2(\alpha_k)} - \frac{(\mathbf{T}'_k(\alpha_k))^2}{\mathbf{T}_k^3(\alpha_k)} \right] (z - \alpha_k)^2 + \dots \end{aligned}$$

We next observe that Theorem 3, the singular expansion of  $\mathbf{F}_k(z)$  at  $\rho_k$  and Theorem 4 imply

$$(3.10) \quad \mathbf{T}_k(z) = \begin{cases} O\left(\left(1 - \frac{z}{\alpha_k}\right)^{(k-1)^2 + (k-1)/2 - 1} \ln\left(1 - \frac{z}{\alpha_k}\right)\right) & \text{for } k \text{ odd, } z \rightarrow \alpha_k \\ O\left(\left(1 - \frac{z}{\alpha_k}\right)^{(k-1)^2 + (k-1)/2 - 1}\right) & \text{for } k \text{ even, } z \rightarrow \alpha_k. \end{cases}$$

Suppose first that  $k \equiv 1 \pmod{2}$  and set  $\mu = (k-1)^2 + \frac{k-1}{2} - 1$ . For  $z \rightarrow \alpha_k$ , eq. (3.10) guarantees

$$(3.11) \quad \mathbf{T}_k(z) = \ell(\alpha_k) \left(1 - \frac{z}{\alpha_k}\right)^\mu \ln\left(1 - \frac{z}{\alpha_k}\right) + r(\alpha_k),$$

where  $\ell(\alpha_k) < 0$ . Since  $\ell(\alpha_k) < 0$ ,  $\mathbf{T}_k(z)$  is a power series with positive coefficients and in view of  $\mu \geq \frac{1}{2}$ , for any  $k \geq 2$ ,  $\mathbf{T}_k(\alpha_k) < \infty$ . Accordingly, we obtain for  $z \rightarrow \alpha_k$

$$(3.12) \quad \mathbf{T}'_k(z) = -\frac{\mu}{\alpha_k} \cdot \ell(\alpha_k) \left(1 - \frac{z}{\alpha_k}\right)^{\mu-1} \ln\left(1 - \frac{z}{\alpha_k}\right) - \frac{\ell(\alpha_k)}{\alpha_k} \left(1 - \frac{z}{\alpha_k}\right)^{\mu-1}$$

$$(3.13) \quad \mathbf{T}''_k(z) = \frac{\mu(\mu-1)}{\alpha_k^2} \cdot \ell(\alpha_k) \left(1 - \frac{z}{\alpha_k}\right)^{\mu-2} \ln\left(1 - \frac{z}{\alpha_k}\right) + \frac{(2\mu-1)\ell(\alpha_k)}{\alpha_k^2} \left(1 - \frac{z}{\alpha_k}\right)^{\mu-2}.$$

Eq. (3.12) and eq. (3.13) imply

$$\begin{aligned} h'(\alpha_k)(z - \alpha_k) &= \frac{\mu \ell(\alpha_k)}{\mathbf{T}_k^2(\alpha_k)} \left(1 - \frac{z}{\alpha_k}\right)^\mu \ln \left(1 - \frac{z}{\alpha_k}\right) (1 + o(1)) \\ \frac{h''(\alpha_k)}{2}(z - \alpha_k)^2 &= \frac{\mu(\mu - 1)\ell(\alpha_k)}{2\mathbf{T}_k^2(\alpha_k)} \left(1 - \frac{z}{\alpha_k}\right)^\mu \ln \left(1 - \frac{z}{\alpha_k}\right) (1 + o(1)) \\ \frac{h'''(\alpha_k)}{3!}(z - \alpha_k)^3 &= \frac{\mu(\mu - 1)(\mu - 2)\ell(\alpha_k)}{3!\mathbf{T}_k^2(\alpha_k)} \left(1 - \frac{z}{\alpha_k}\right)^\mu \ln \left(1 - \frac{z}{\alpha_k}\right) (1 + o(1)). \end{aligned}$$

We proceed by computing

$$\begin{aligned} h(z) &= \tau_k + \frac{\mathbf{T}'_k(\alpha_k)}{\mathbf{T}_k^2(\alpha_k)}(z - \alpha_k) + \left[ \frac{\mathbf{T}''_k(\alpha_k)}{2\mathbf{T}_k^2(\alpha_k)} - \frac{(\mathbf{T}'_k(\alpha_k))^2}{\mathbf{T}_k^3(\alpha_k)} \right] (z - \alpha_k)^2 + \dots \\ &= \tau_k + \frac{\ell(\alpha_k)}{\mathbf{T}_k^2(\alpha_k)} \left[ \mu + \frac{\mu(\mu - 1)}{2} + \frac{\mu(\mu - 1)(\mu - 2)}{3!} + \dots \right] \left(1 - \frac{z}{\alpha_k}\right)^\mu \ln \left(1 - \frac{z}{\alpha_k}\right) (1 + o(1)) \\ &= \tau_k - c_k \left(1 - \frac{z}{\alpha_k}\right)^\mu \ln \left(1 - \frac{z}{\alpha_k}\right) (1 + o(1)), \quad \text{where } c_k > 0. \end{aligned}$$

The case  $k \equiv 0 \pmod{2}$  is proved analogously and Claim 1 follows.  $\mathbf{U}_k(z, 1)$ , as the product of  $g(z)$  and  $g(h(z))$  has an analytic continuation since  $h(z)$  has an analytic continuation and  $g(h(z)) = \frac{1}{1-h(z)}$  is accordingly analytic continuable at  $z = \alpha_k < 1$ . Consequently  $\mathbf{U}_k(z, 1)$  has a singular expansion at  $z = \alpha_k$ . Without loss of generality, we may restrict ourselves in the following to the case  $k \equiv 1 \pmod{2}$  and proceed by computing

$$\begin{aligned} \mathbf{U}_k(z, 1) &= g(\alpha_k)g(\tau_k) + (g \cdot g(h))'(\alpha_k)(z - \alpha_k) + \frac{(g \cdot g(h))''(\alpha_k)}{2!}(z - \alpha_k)^2 + \dots \\ &= g(\alpha_k)g(\tau_k) - c_k g(\alpha_k)g'(\tau_k) \left(1 - \frac{z}{\alpha_k}\right)^\mu \ln \left(1 - \frac{z}{\alpha_k}\right) (1 + o(1)). \end{aligned}$$

Therefore we derive

$$(3.14) \quad [z^n]\mathbf{U}_k(z, 1) = -c_k g(\alpha_k)g'(\tau_k)\alpha_k^{-n}n^{-\mu-1}(1 + o(1)).$$

For any fixed  $u \in (0, 1)$  the singular expansion of  $\mathbf{U}_k(z, u)$  at  $z = \alpha_k$  is given by

$$\begin{aligned} \mathbf{U}_k(z, u) &= g(\alpha_k)g(u\tau_k) + (g \cdot g(uh))'(\alpha_k)(z - \alpha_k) + \frac{(g \cdot g(uh))''(\alpha_k)}{2!}(z - \alpha_k)^2 + \dots \\ &= g(\alpha_k)g(u\tau_k) - c_k u g(\alpha_k)g'(u\tau_k) \left(1 - \frac{z}{\alpha_k}\right)^\mu \ln \left(1 - \frac{z}{\alpha_k}\right) (1 + o(1)) \end{aligned}$$

and we consequently obtain, setting  $\tau_k = 1 - \frac{1}{(1-\alpha_k)\mathbf{T}_k(\alpha_k)}$

$$(3.15) \quad \lim_{n \rightarrow \infty} \frac{[z^n]\mathbf{U}_k(z, u)}{[z^n]\mathbf{U}_k(z, 1)} = \frac{u g'(u\tau_k)}{g'(\tau_k)} = \frac{u(1 - \tau_k)^2}{(1 - \tau_k u)^2} = q(u).$$

In view of eq. (3.15) and  $[u^i]q(u) = \frac{(1-\tau_k)^2}{\tau_k} i\tau_k^i = q_i$ , Theorem 5 implies the discrete limit law

$$(3.16) \quad \lim_{n \rightarrow \infty} \mathbb{P}(\xi_n^{(k)} = i) = \lim_{n \rightarrow \infty} \frac{\delta_{n,i}^{(k)}}{\delta_n^{(k)}} = q_i \quad \text{where} \quad q_i = \frac{(1-\tau_k)^2}{\tau_k} i\tau_k^i.$$

Since the density function of a  $\Gamma(\lambda, r)$ -distribution is given by

$$(3.17) \quad f_{\lambda,r}(x) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, & x > 0 \\ 0 & x \leq 0, \end{cases}$$

where  $\lambda > 0$  and  $r > 0$ , we obtain, setting  $r = 2$  and  $\lambda = -\ln \tau_k > 0$

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(\xi_n^{(k)} = i) &= \frac{(1-\tau_k)^2}{\tau_k} (i \cdot \tau_k^i) = \frac{(1-\tau_k)^2}{\tau_k} \frac{1}{(\ln \tau_k)^2} (\ln \tau_k)^2 i \cdot \tau_k^i \\ &= \frac{(1-\tau_k)^2}{\tau_k} \frac{1}{(\ln \tau_k)^2} f_{-\ln \tau_k, 2}(i) \end{aligned}$$

and the proof of the theorem is complete.  $\square$

#### 4. THE LIMIT DISTRIBUTION OF NONTRIVIAL RETURNS

Let  $\beta_n^{(k)}$  denote the number of  $\mathfrak{C}$ -tableaux of length  $2n$ , which are in correspondence to  $k$ -noncrossing,  $\sigma$ -canonical RNA structures. Let  $\beta_{n,i}^{(k)}$  denote the number of  $\mathfrak{C}$ -tableaux of length  $2n$ , having exactly  $i$   $\emptyset$ -shapes contained in the sequence  $(\lambda^2, \dots, \lambda^{2n})$ . Let  $\mathbf{W}_k(z, u)$  denote the bivariate generating function of  $\beta_{n,i}^{(k)}$ . Then  $\beta_{n,j}^{(k)} = [z^n u^j] \mathbf{W}_k(z, u)$  and  $\mathbf{W}_k(z, u) = \sum_{j \geq 0} \sum_{n \geq j} \beta_{n,j}^{(k)} z^n u^j$ . Furthermore we set  $\beta_n^{(k)} = [z^n] \mathbf{W}_k(z, 1)$ .

**Lemma 3.** *The bivariate generating function of the number of  $\mathfrak{C}$ -tableaux of length  $2n$ , with less than  $k$  rows, containing exactly  $i$   $\emptyset$ -shapes, is given by*

$$(4.1) \quad \mathbf{W}_k(z, u) = \frac{1}{1 - u \left(1 - \frac{1}{\mathbf{T}_k(z)}\right)}.$$

*Proof.* Suppose the  $\mathfrak{C}$ -tableaux  $(\lambda^2, \dots, \lambda^{2n})$  contains exactly  $i$   $\emptyset$ -shapes. These  $\emptyset$ -shapes split  $(\lambda^2, \dots, \lambda^{2n})$  uniquely into exactly  $i$   $\mathfrak{C}$ -subtableaux, each of which either being a gap of length 2 or an irreducible  $*$ -tableaux. We conclude from this, that for fixed  $j$

$$(4.2) \quad \sum_{n \geq j} \beta_{n,j}^{(k)} z^n = (z + \mathbf{R}_k(z))^j$$

holds. Therefore the bivariate generating function  $\mathbf{W}_k(z, u)$  satisfies

$$\begin{aligned} \mathbf{W}_k(z, u) &= \sum_{j \geq 0} \sum_{n \geq j} \beta_{n,j} z^n u^j = \sum_{j \geq 0} (z + \mathbf{R}_k(z))^j u^j \\ &= \frac{1}{1 - u(z + \mathbf{R}_k(z))} \\ &= \frac{1}{1 - u(1 - \frac{1}{\mathbf{T}_k(z)})}, \end{aligned}$$

where the last equality follows from eq. (3.5), proving the lemma.  $\square$

We set  $g(z) = \frac{1}{1-z}$ ,  $h(z) = 1 - \frac{1}{\mathbf{T}_k(z)}$  and let  $\eta_n^{(k)}$  denote the random variable having probability distribution  $\mathbb{P}(\eta_n^{(k)} = i) = \frac{\beta_{n,i}^{(k)}}{\beta_n^{(k)}}$ . In case of  $\mathbf{W}_k(z, u) = g(uh(z))$  we have  $\rho_g = 1$  while  $\tau_h < 1$ , i.e. we are given the subcritical case. In our next theorem, we prove that the limit distribution of  $\eta_n^{(k)}$  is determined by the density function of a  $\Gamma(\lambda, r)$ -distribution.

**Theorem 7.** *Let  $\alpha_k$  denote the real, positive, dominant singularity of  $\mathbf{T}_k(z)$  and let  $\tau_k = 1 - \frac{1}{\mathbf{T}_k(\alpha_k)}$ . Then the r.v.  $\eta_n^{(k)}$  satisfies the discrete limit law*

$$(4.3) \quad \lim_{n \rightarrow \infty} \mathbb{P}(\eta_n^{(k)} = i) = q_i, \quad \text{where} \quad q_i = \frac{(1 - \tau_k)^2}{\tau_k} i \tau_k^i.$$

That is,  $\eta_n^{(k)}$  is determined by the density function of a  $\Gamma(-\ln \tau_k, 2)$ -distribution and the limit distribution has the probability generating function  $q(u) = \sum_{n \geq 1} q_n u^n = \frac{u(1-\tau_k)^2}{(1-\tau_k u)^2}$ .

*Proof.* Since  $g(z) = \frac{1}{1-z}$  and  $h(z) = 1 - \frac{1}{\mathbf{T}_k(z)}$  have non negative coefficients and  $h(0) = 0$ , the composition  $g(h(z))$  is again a power series.  $\mathbf{W}_k(z, u) = g(uh(z))$  has a singularity at  $z = \alpha_k$  and  $\tau_k = 1 - \frac{1}{\mathbf{T}_k(\alpha_k)} < 1 = \rho_g$ , whence we are given the subcritical case. Furthermore we observe, that regardless of the singularity arising from  $\mathbf{T}_k(z) = 0$ , the dominant singularity of  $h(z) = 1 - \frac{1}{\mathbf{T}_k(z)}$  equals the dominant singularity of  $\mathbf{T}_k(z)$ , i.e.,  $z = \alpha_k$ .

*Claim 1.*  $h(z)$  has a singular expansion at  $z = \alpha_k$  and there exists some constant  $c_k > 0$  such that

$$(4.4) \quad h(z) = \begin{cases} \tau_k - c_k \left(1 - \frac{z}{\alpha_k}\right)^\mu \ln \left(1 - \frac{z}{\alpha_k}\right) (1 + o(1)) & \text{for } k \equiv 1 \pmod{2} \\ \tau_k - c_k \left(1 - \frac{z}{\alpha_k}\right)^\mu (1 + o(1)) & \text{for } k \equiv 0 \pmod{2} \end{cases}$$

for  $z \rightarrow \alpha_k$  and  $\mu = (k-1)^2 + \frac{k-1}{2} - 1$ .

The proof of Claim 1 is analogous to that of Theorem 6. Again, we restrict ourselves to the case

$k \equiv 1 \pmod{2}$ .  $\mathbf{W}_k(z, 1) = g(h(z))$  has an analytic continuation and its Taylor expansion of at  $z = \alpha_k$  is given by

$$\begin{aligned} \mathbf{W}_k(z, 1) &= g(\tau_k) + (gh)'(\alpha_k)(z - \alpha_k) + \frac{(gh)''(\alpha_k)}{2!}(z - \alpha_k)^2 + \dots \\ &= g(\tau_k) + \frac{g'(\tau_k)\ell(\alpha_k)}{\mathbf{T}_k^2(\alpha_k)} \left[ \mu + \frac{\mu(\mu-1)}{2} + \dots \right] \left(1 - \frac{z}{\alpha_k}\right)^\mu \ln \left(1 - \frac{z}{\alpha_k}\right) (1 + o(1)) \\ &= g(\tau_k) - c_k g'(\tau_k) \left(1 - \frac{z}{\alpha_k}\right)^\mu \ln \left(1 - \frac{z}{\alpha_k}\right) (1 + o(1)). \end{aligned}$$

Therefore we arrive at

$$(4.5) \quad [z^n] \mathbf{W}_k(z, 1) = c_k g'(\tau_k) \alpha_k^{-n} n^{-\mu-1} (1 + o(1)).$$

For any fixed  $u \in (0, 1)$  the singular expansion of  $\mathbf{W}_k(z, u) = g(uh(z))$  at  $z = \alpha_k$  is given by

$$\begin{aligned} \mathbf{W}_k(z, u) &= g(u\tau_k) + (g(uh))'(\alpha_k)(z - \alpha_k) + \frac{(g(uh))''(\alpha_k)}{2!}(z - \alpha_k)^2 + \dots \\ &= g(u\tau_k) - c_k u g'(u\tau_k) \left(1 - \frac{z}{\alpha_k}\right)^\mu \ln \left(1 - \frac{z}{\alpha_k}\right) + (1 + o(1)) \end{aligned}$$

from which we conclude

$$(4.6) \quad \lim_{n \rightarrow \infty} \frac{[z^n] \mathbf{W}_k(z, u)}{[z^n] \mathbf{W}_k(z, 1)} = \frac{u g'(u\tau_k)}{g'(\tau_k)} = \frac{u(1 - \tau_k)^2}{(1 - \tau_k u)^2} \quad \text{where} \quad \tau_k = 1 - \frac{1}{\mathbf{T}_k(\alpha_k)}.$$

In view of  $[u^i]q(u) = \frac{(1-\tau_k)^2}{\tau_k} i \tau_k^i = q_i$ , Theorem 5 implies the discrete limit law

$$(4.7) \quad \lim_{n \rightarrow \infty} \mathbb{P}(\eta_n^{(k)} = i) = \lim_{n \rightarrow \infty} \frac{\beta_{n,i}^{(k)}}{\beta_n^{(k)}} = q_i.$$

In view of eq. (3.17), setting  $r = 2$  and  $\lambda = -\ln \tau_k > 0$ , we analogously obtain

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(\eta_n^{(k)} = i) &= \frac{(1 - \tau_k)^2}{\tau_k} (i \cdot \tau_k^i) = \frac{(1 - \tau_k)^2}{\tau_k} \frac{1}{(\ln \tau_k)^2} (\ln \tau_k)^2 i \cdot \tau_k^i \\ &= \frac{(1 - \tau_k)^2}{\tau_k} \frac{1}{(\ln \tau_k)^2} f_{-\ln \tau_k, 2}(i) \end{aligned}$$

and Theorem 7 is proved. □

**Acknowledgments.** We are grateful to W.Y.C. Chen for stimulating discussions. 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] N.T. Cameron and L. Shapiro, Random walks, trees and extensions of Riordan group techniques, talk, Annual Joint Mathematics Meetings, Baltimore, MD, U.S. (2003).
- [2] M. Chastain and I.Jr. Tinoco, A base-triple structural domain in RNA, *Biochemistry* 31 (1992) 12733–12741.
- [3] 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 (4) (2007) 1555–1575.
- [4] W.Y.C. Chen, J. Qin and C.M. Reidys, Crossings and Nestings of tangled-diagrams, *Electron. J. Combin.* 15 (2008) #R86 (electronic).
- [5] E. Dam, I. Brierly, S. Inglis and C. Pleij, Identification and analysis of the pseudoknot containing *gag-pro* ribosomal frameshift signal of simian retrovirus-1, *Nucl. Acids Res.* 22 (1994) 2304–2310.
- [6] E. Deutsch and L. Shapiro, A survey of the Fine numbers, *Discr. Math.* 241 (2001) 241–265.
- [7] P. Flajolet and R. Sedgewick, *Analytic combinatorics*, (2008) (electronic).
- [8] D.J. Grabiner and P. Magyar, Random walks in Weyl chambers and the decomposition of tensor powers, *J. Alg. Comb.* 2 (1993) 239–260.
- [9] E.Y. Jin, J. Qin and C.M. Reidys, Combinatorics of RNA structures with pseudoknots, *Bull. Math. Biol.* 70 (1) (2008) 45–67.
- [10] E.Y. Jin and C.M. Reidys, Asymptotic enumeration of RNA structures with pseudoknots, *Bull. Math. Biol.* 70 (4) (2008) 951–970.
- [11] E.Y. Jin and C.M. Reidys, Combinatorial Design of Pseudoknot RNA, *Adv. Appl. Math.* 42 (2009) 135–151.
- [12] C.Haslinger and P.F. Stadler, RNA structures with pseudo-knots, *Bull. Math. Biol.* 61 (1999) 437–467.
- [13] 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.
- [14] F.W.D. Huang, W.W.J. Peng and C.M. Reidys, Folding 3-Noncrossing RNA Pseudoknot Structures, (2008) submitted.
- [15] E.Y. Jin, C.M. Reidys and R. Wang, Asymptotic analysis k-noncrossing matchings, (2008) submitted.
- [16] 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.
- [17] A. Loria and T. Pan, Domain structure of the ribozyme from eubacterial ribonuclease P, *RNA.* 2 (1996) 551–563.
- [18] G. Ma and C.M. Reidys, Canonical RNA Pseudoknot Structures, *J. Comp. Biol.* 15 (10) (2008) 1257–1273.
- [19] A.M. Odlyzko, Asymptotic enumeration methods, *Handbook of combinatorics*, 2 (1995) 1021–1231 MIT press.
- [20] R.C. Penner and M.S. Waterman, Spaces of RNA secondary structures, *Adv. Math.* 101 (1993) 31–49.
- [21] D.B. Searls, Formal language theory and biological macromolecules, *Series in Discr. Math. and Theor. Comp. Sci.* 47 (1999) 117–140.
- [22] L. Shapiro, S. Getu, W. Woan and L. Woodson, The Riordan Group, *Discr. Appl. Math.* 34 (1991) 229–239.
- [23] R.P. Stanley, Differentiably finite power series, *Europ. J. Combinatorics.* 1 (1980) 175–188.
- [24] R.P. Stanley, *Enumerative Combinatorics*, 2 (1999) Cambridge University Press.
- [25] C. Tuerk, S. MacDougall and L. Gold, RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl. Acad. Sci. USA* 89 (1992) 6988–6992.
- [26] Vienna RNA Package—RNA Secondary Structure Prediction and Comparison. [www.tbi.univie.ac.at/RNA/](http://www.tbi.univie.ac.at/RNA/).
- [27] M.S. Waterman, Combinatorics of RNA hairpins and cloverleaves, *Stud. Appl. Math.* 60 (1979) 91–96.
- [28] M.S. Waterman, Secondary structure of single - stranded nucleic acids, *Adv. Math.I (suppl.)* 1 (1978) 167–212.

- [29] M.S. Waterman and W.R. Schmitt, Linear trees and RNA secondary structure, *Discr. Appl. Math.* 51 (1994) 317–323.
- [30] E. Westhof and L. Jaeger, RNA pseudoknots, *Curr. Opin. 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