

# Combinatorics of RNA Structures with Pseudoknots

Emma Y. Jin, Jing Qin, Christian M. Reidys\*

*Center for Combinatorics, LMPC-TJKLC, Nankai University, Tianjin 300071, China*

Received: 19 April 2007 / Accepted: 31 May 2007 / Published online: 26 September 2007  
© Society for Mathematical Biology 2007

**Abstract** In this paper, we derive the generating function of RNA structures with pseudoknots. We enumerate all  $k$ -noncrossing RNA pseudoknot structures categorized by their maximal sets of mutually intersecting arcs. In addition, we enumerate pseudoknot structures over circular RNA. For 3-noncrossing RNA structures and RNA secondary structures we present a novel 4-term recursion formula and a 2-term recursion, respectively. Furthermore, we enumerate for arbitrary  $k$  all  $k$ -noncrossing, restricted RNA structures i.e.  $k$ -noncrossing RNA structures without 2-arcs i.e. arcs of the form  $(i, i + 2)$ , for  $1 \leq i \leq n - 2$ .

**Keywords** RNA secondary structure · Pseudoknot · Enumeration · Generating function · Reflection principle · Walks · Weyl-chamber

## 1. Introduction

In this paper, we study the combinatorics of helical structures of RNA sequences. RNA is described by its primary sequence of nucleotides **A**, **G**, **U** and **C** together with the Watson–Crick (**A–U**, **G–C**) and (**U–G**) base pairing rules specifying which pairs of nucleotides can potentially form bonds. Subject to these single stranded RNA form helical structures. The function of many RNA sequences depends on their structures. Therefore, it is important to understand RNA structure in the context of studying the function of biological RNA as well as in the design process of artificial RNA structures. Since RNA is capable of catalytic activity, for instance RNA ribozymes can cleave other RNA molecules, it is believed that RNA may have been instrumental for early evolution, before proteins emerged. A particularly well-studied sub-class of RNA structures, consisting of outer-planar graphs are the RNA secondary structures. Their combinatorics was pioneered by Waterman et al. in a series of seminal papers (Penner and Waterman, 1993; Waterman, 1978, 1979; Schmitt and Waterman, 1994; Howell et al., 1980). RNA secondary structures are coarse grained structures and systematic prediction of the full three dimensional structures, the tertiary structures seems at present time to be out of reach. It was shown in Waterman and Smith (1986) that the prediction of secondary structures

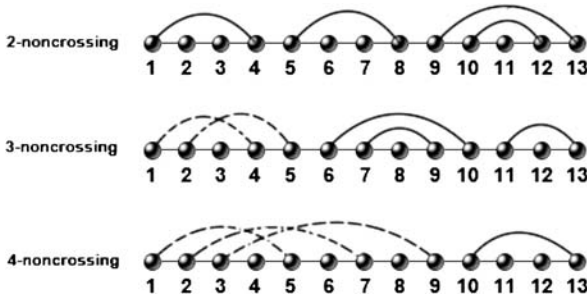
---

\*Corresponding author.

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

can be obtained in polynomial time and their combinatorics, specifically the existence of recursion relations is the key for all folding algorithms (Zuker and Sankoff, 1984; Hofacker et al., 1998). Over the last two decades, a variety of prediction algorithms, based on minimum free energy (Zuker and Sankoff, 1984; Waterman and Smith, 1986; Tacker et al., 1996), kinetic folding (Tacker et al., 1994) or the partition function (McCaskill, 1990) for RNA secondary structures have been derived.

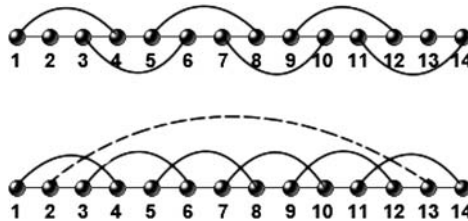
An increasing number of experimental findings, as well as results from comparative sequence analysis imply that there exist additional types of interactions between RNA nucleotides (Westhof and Jaeger, 1992). These bonds are called pseudoknots and occur in functional RNA like for instance RNaseP (Loria and Pan, 1996) as well as ribosomal RNA (Konings and Gutell, 1995). 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 (Tuerk et al., 1992) experiments have produced families of RNA structures with pseudoknot motifs, when binding HIV-1 reverse transcriptase. In addition, important mechanisms like ribosomal frame shifting (Parkin et al., 1991) also involve pseudoknot interactions. As a result, RNA pseudoknot structures have drawn over the last years a lot of attention (Mapping RNA, 2005). Several folding algorithms (Rivas and Eddy, 1999; Hasegawa et al., 1999; Akutsu, 2000; Lyngso and Pedersen, 1996) have been developed which include certain families of pseudoknots. The prediction problem in general is (although we have not seen formal proof) believed to be NP-hard. In difference to RNA secondary structures, a recursive enumeration for pseudoknot RNA is believed to be non-trivial but nevertheless of vital importance for prediction algorithms. Intuitively, if bonds can cross, it is much harder to enumerate since structural elements can now interact and as a result a structure cannot be straightforwardly decomposed into independent sub-structures. The important question is: how should pseudoknots be categorized? Clearly, secondary structures are outerplanar graphs and one approach has been to generalize to planar graphs, resulting in the concept of bi-secondary structures (Haslinger and Stadler, 1999). In the following, we will take a different point of view, effectively generalizing even further. We represent structures as diagrams (see Section 2 for details) and classify according to the number of mutually crossing arcs in this representation. To be precise, we consider  $k$ -noncrossing structures, i.e. structures that do not contain a  $k$ -set of mutually intersecting arcs. Intuitively, a higher number of pairwise crossing arcs represents higher structural complexity, see Fig. 1. In the simplest case, we have 2-noncrossing structures, which correspond to secondary structures. Relatively high numbers of pairwise crossing bonds do occur in natural RNA structures. For instance, the *gag-pro* ribosomal frame-shift signal of the simian retrovirus-1 (ten Dam et al., 1994), which is a 10-noncrossing RNA structure motif, displayed in Fig. 2. Little is known with respect to the combinatorics of pseudoknot RNA structures. Haslinger and Stadler (1999) suggested a classification of their knot-types based on a notion of inconsistency graphs and provided an upper bound for a certain class of pseudoknots. His bi-secondary structures are “superpositions” of two secondary structures, i.e. they can be drawn as a set of non intersecting arcs in the upper and lower half plane, respectively. Figure 3 illustrates that bi-secondary structures are exactly planar 3-noncrossing RNA structures. As for bi-secondary structures, the combinatorics developed in the following allows to derive precise asymptotics complementing the bound derived in Haslinger and Stadler (1999).



**Fig. 1** The first types of RNA pseudoknots. We consider a label of a nucleotide to be isolated if and only if there exists no arc different from the bonds of the backbone of the original sequence. (a) displays a secondary structure with isolated labels 2, 3, 6, 7, 11, (b) a 3-noncrossing structure with isolated labels 3, 8, 12. This structure exhibits intersecting arcs (dashed) but there are no three arcs which mutually intersect. (c) shows an example of a 4-noncrossing structure (isolated points: 4, 6, 8, 11, 12), where we have three pairwise intersecting arcs (dashed).

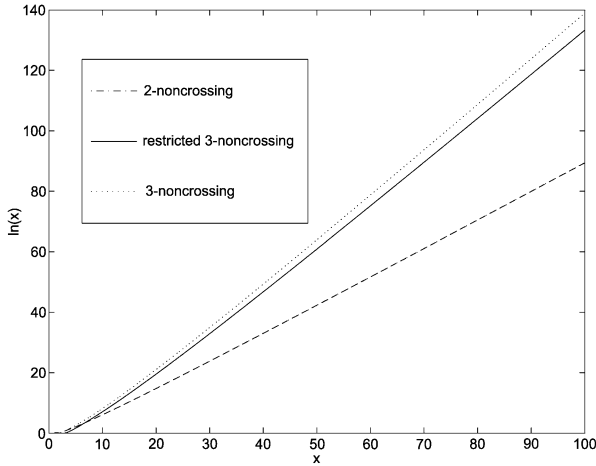


**Fig. 2** The proposed SRV-1 frame-shift (ten Dam et al., 1994) is a 10-noncrossing RNA structure motif.



**Fig. 3** Bi-secondary and 3-noncrossing RNA structures. We display a bi-secondary structure (top) and a 3-noncrossing RNA structure which is not a bi-secondary structure (bottom), obtained by adding the arc (2, 13) (dashed).

In this paper, we introduce a novel approach to the enumeration of RNA structures. Based on new concepts in enumerative combinatorics (Chen et al., 2007; Gessel and Zeilberger, 1992), we use an approach which has the potential to offer insight also into other lattice structure concepts. To be precise, Chen et al. have shown in 2007 that there is a bijection between certain types of matchings and walks inside Weyl-chambers. This bijection is obtained via Stanley’s construction of oscillating tableaux, i.e. families of Young diagrams in which any two consecutive shapes differ by exactly one square. The corresponding walks can then be enumerated via determinant formulas derived from a reflection principle due to Gessel and Zeilberger (1992) and Lindström (1973). The key idea behind the reflection principle is that walks which hit the wall of a Weyl-chamber can be reflected. The original (unreflected) and the reflected walk cancel themselves leaving



**Fig. 4** The numbers of RNA structures for large  $n$ . We display the numbers of 2-noncrossing RNA structures (secondary structures),  $S_2(n)$ , 3-noncrossing RNA structures  $S_3(n)$ , and restricted 3-noncrossing RNA structures  $S_3^{(r)}(n)$ . Numerically we obtained the following exponential growth rates:  $S_2(n) \sim 2.5913^n$  ( $n = 1000$ ),  $S_3(n) \sim 4.6542^n$  ( $n = 1000$ ), and  $S_3^{(r)}(n) \sim 4.2741^n$  ( $n = 400$ ).

just the walks that never hit a wall. Crucial for its applicability are restrictive symmetry assumptions since the reflected walk has to be of the same type and, more importantly, the reflection itself can occur at any step. These symmetries are non-existent in walks corresponding to RNA structures. However, our derivation of the generating function of RNA structures is based on these symmetric walks. The key idea is to introduce the asymmetries of RNA structures into the symmetric walks using a certain involution idea. We believe that our particular strategy can be applied to the enumeration of further structure classes. As a result, we have tried to keep this paper self-contained.

Our main result is the enumeration of all RNA structures. We classify RNA structures by their specific crossing types under the assumption that all base pairs can occur. For arbitrary but fixed  $k$  we enumerate all RNA structures with no  $k$ -set of mutually intersecting bonds, see Fig. 4. In case of  $k = 2$ , our results reduce to noncrossing structures, i.e. RNA secondary structures (Schmitt and Waterman, 1994) and Waterman's formula for the number of RNA secondary structures with exactly  $k$  bonds is derived directly. We obtain from the generating function a novel 4-term recursion formula for RNA structures of length  $n$  with no 3-set of mutually intersecting bonds and having  $\ell$  isolated vertices. We believe that this recursion can be instrumental for developing new prediction algorithms for RNA structures. Also, we derive the generating function for circular RNA structures, i.e. for sequences where the bond between 1 and  $n$  is considered part of the primary sequence. Finally, we enumerate restricted RNA structures, i.e. structures in which two interacting nucleotides have at least distance 3.

The paper is structured as follows. We will begin by introducing several important combinatorial concepts needed for our derivations. Young tableaux, oscillating Young diagrams, RSK algorithm, Weyl-chambers and the reflection principle. We discuss these concepts, illustrate all key ideas and give the corresponding proofs in Appendix A. Our derivation is obtained in three steps. First (Theorem 1), we show that each structure (rep-

resented as a  $k$ -noncrossing digraph) corresponds uniquely to a walk starting and ending at  $(k - 1, k - 2, \dots, 1)$  in  $\mathbb{Z}^{k-1}$  and which never touches a wall of the Weyl-chamber  $C_0$ . Secondly, we apply the reflection principle (Theorem 2) in order to count the symmetric walks that remain in the interior of  $C_0$ . Thirdly, (Theorem 3) we incorporate the specific properties of RNA into these symmetric walks and compute the generating function of  $k$ -noncrossing RNA structures. We show how our results relate to known formulas of RNA secondary structures for which we present a new two term recursion formula. For 3-noncrossing RNA structures, we give a novel 4-term recursion formula. We finally generalize our strategy (Theorem 5) and enumerate restricted RNA structures.

**2. From structures to walks and back**

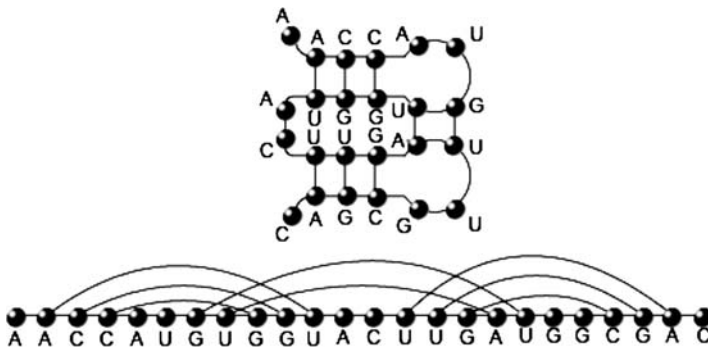
Let us begin by illustrating the concept of RNA structures. Suppose we are given the primary sequence

**AACCAUGUGGUACUUGAUGGCGAC.**

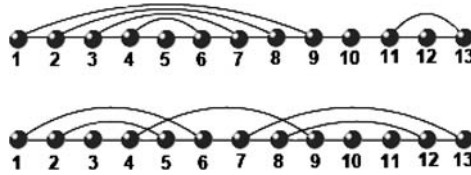
Despite the fact that the nucleotides of the primary sequence are connected via its backbone, we consider in the diagram representation of structures the labels (i.e. the vertices of the diagram graph) of the nucleotides of the primary sequence to be isolated if they are not incident to any arc in the upper halfplane. In the following, structures are combinatorial graphs over the labels of the nucleotides of the primary sequence. These graphs can be represented in several ways. In Fig. 5, we represent a structure with loop-loop interactions: first, we display the structure as an planar graph and secondly as a diagram, where the bonds are drawn as arcs in the positive half-plane.

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  $E_{D_n} \subset \{(i, j) \mid 1 \leq i < j \leq n\}$ .  $V_{D_n}$  and  $E_{D_n}$  are called vertex and arc set, respectively. A  $k$ -noncrossing digraph,  $G_{k,n}$ , is a digraph in which all vertices have degree  $\leq 1$  and which does not contain a  $k$ -set of arcs that are mutually intersecting, i.e.

$$\begin{aligned} &\bar{A}(i_{r_1}, j_{r_1}), (i_{r_2}, j_{r_2}), \dots, (i_{r_k}, j_{r_k}); \\ &i_{r_1} < i_{r_2} < \dots < i_{r_k} < j_{r_1} < j_{r_2} < \dots < j_{r_k}. \end{aligned} \tag{1}$$



**Fig. 5** Two representations of RNA structures, planar graphs (top) and diagrams (bottom).



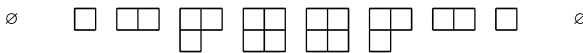
**Fig. 6** RNA structures represented as diagrams, i.e. arcs over  $\{1, \dots, n\}$  in the upper half-plane.  $k$ -noncrossing digraphs are precisely those which have no  $k$ -set of mutually intersecting arcs. 2-noncrossing diagrams without 1-arcs (top) correspond to secondary structures. 3-noncrossing diagrams without 1-arcs (bottom) represent RNA structures with pseudoknots.

The set of all  $k$ -noncrossing digraphs  $G_{k,n}$  is denoted by  $\mathcal{G}_{n,k}$  and we set  $\mathbf{G}_{n,k} = |\mathcal{G}_{n,k}|$ . The (formal) direction of the edges will have procedural convenience when we map a  $k$ -noncrossing digraph into an oscillating tableaux (Theorem 1). We will represent digraphs as diagrams (Fig. 5) 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 RNA structure (of pseudoknot type  $k - 2$ ),  $S_{k,n}$ , is a digraph in which all vertices have degree  $\leq 1$ , 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  $S_k(n)$  and the number of RNA structures with exactly  $\ell$  isolated vertices by  $S_k(n, \ell)$ , respectively. We call an RNA structure restricted if and only if it does not contain any 2-arcs, i.e. an arc of the form  $(i, i + 2)$ .

We derive the enumeration of RNA structures in three steps. First, we establish a bijection from  $k$ -noncrossing digraphs into a certain class of walks. Secondly, we will use the reflection principle in order to count these walks. Thirdly, we enumerate all walks subject to specific conditions recruiting a certain involution idea. Let us first discuss two basic concepts needed for our arguments.

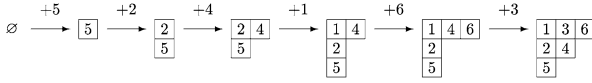
**Young tableaux and the RSK algorithm.** A Young diagram (shape) is a collection of squares arranged in left-justified rows with weakly decreasing number of boxes in each row. A Young tableau is a filling of the squares by numbers which is weakly decreasing in each row and strictly decreasing in each column. A tableau is called standard if each entry occurs exactly once. An oscillating tableau is a sequence  $\emptyset = \mu^0, \mu^1, \dots, \mu^n = \emptyset$  of standard Young diagrams, such that for  $1 \leq i \leq n$ ,  $\mu^i$  is obtained from  $\mu^{i-1}$  by either adding one square or removing one square or doing nothing. For instance the sequence



is an oscillating tableaux.

The RSK algorithm is a process of row-inserting elements into a tableau. Suppose we want to insert  $k$  into a standard Young tableau  $\lambda$ . Let  $\lambda_{i,j}$  denote the element in the  $i$ -th row and  $j$ -th column of the Young tableau. Let  $i$  be the largest integer such that

$\lambda_{1,i-1} \leq k$ . (If  $\lambda_{1,1} > k$ , then  $i = 1$ .) If  $\lambda_{1,i}$  does not exist, then simply add  $k$  at the end of the first row. Otherwise, if  $\lambda_{1,i}$  exists, then replace  $\lambda_{1,i}$  by  $k$ . Next insert  $\lambda_{1,i}$  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  $k$  included. For instance inserting the number sequence 5, 2, 4, 1, 6, 3 starting with an empty shape yields the following sequence of standard Young tableaux:



**Symmetry groups and Weyl-chambers.** We consider the lattice  $\mathbb{Z}^{k-1}$  and walks in  $\mathbb{Z}^{k-1}$  having the steps  $s$  contained in  $\{\pm e_i, 0 \mid 1 \leq i \leq k-1\}$ , where  $e_i$  denotes the  $i$ -th unit vector. That is for  $a, b \in \mathbb{Z}^{k-1}$  a walk from  $a$  to  $b$ ,  $\gamma_{a,b}$ , of length  $n$  is an  $n$  tuple  $(s_1, \dots, s_n)$  where  $s_i \in \{\pm e_i, 0 \mid 1 \leq i \leq k-1\}$  such that  $b = a + \sum_{h=1}^n s_h$ . We set  $\gamma_{a,b}(s_r) = a + \sum_{h=1}^r s_h \in \mathbb{Z}^{k-1}$  i.e. the element at which the walk  $(s_1, \dots, s_r)$  resides at step  $r$ . The symmetric group  $S_{k-1}$  and the group  $E_{k-1} = \langle \epsilon_i \mid 1 \leq i \leq k-1 \rangle$  act on  $\mathbb{Z}^{k-1}$  via  $\sigma(x_i)_{1 \leq i \leq k-1} = (x_{\sigma^{-1}(i)})_{1 \leq i \leq k-1}$  and  $\epsilon_i(x_1, \dots, x_i, \dots, x_{k-1}) = (x_1, \dots, -x_i, \dots, x_{k-1})$ , respectively. It is straightforward to verify that  $\{\epsilon_i \sigma \mid \sigma \in S_{k-1}, \epsilon_i \in E_{k-1}\}$  carries a natural group structure via

$$(\epsilon_i \sigma) \cdot (\epsilon_j \sigma') = \epsilon_i \sigma \epsilon_j \sigma^{-1} \sigma \sigma' = \epsilon_i \epsilon_{\sigma^{-1}(j)} \sigma \sigma'. \tag{2}$$

This group, denoted by  $B_{k-1}$ , is the semidirect product  $E_{k-1} \rtimes S_{k-1}$  and is generated by  $M_{k-1} = \{\epsilon_{k-1}, \rho_j \mid 2 \leq j \leq k-1\}$ , where  $\rho_j = (j, j-1)$ , i.e.  $\rho$  transposes the coordinates  $x_{j-1}$  and  $x_j$ . By definition  $B_{k-1}$  acts on the set

$$\Delta_{k-1} = \{\pm e_i \mid 1 \leq i \leq k-1\} \cup \{e_i \pm e_j \mid 1 \leq i, j \leq k-1\}. \tag{3}$$

We observe that there exists a bijection between  $\Delta'_{k-1} = \{e_{k-1}, e_{j-1} - e_j \mid 2 \leq j \leq k-1\}$  and the set of generators  $M_{k-1}$  which maps each  $\alpha \in \Delta'_{k-1}$  into a corresponding reflection (in particular:  $B_{k-1}$  is generated by reflections)

$$\{e_{k-1}, e_{j-1} - e_j \mid 2 \leq j \leq k-1\} \longrightarrow \{\epsilon_{k-1}, \rho_j \mid 2 \leq j \leq k-1\}, \tag{4}$$

$$\alpha \mapsto \left( \beta_\alpha : x \mapsto x - 2 \frac{\langle \alpha, x \rangle}{\langle \alpha, \alpha \rangle} \alpha \right), \tag{5}$$

where  $\langle x, x' \rangle$  denotes the standard scalar product in  $\mathbb{R}^{k-1}$ . It is clear that  $\Delta'_{k-1}$  is a basis of  $\mathbb{R}^{k-1}$ . We refer to the sub spaces  $\langle e_i \rangle$  for  $1 \leq i \leq k-1$  and  $\langle e_{j-1} - e_j \rangle$  for  $2 \leq j \leq k-1$  as walls. A  $B_{k-1}$ -chamber is defined as the set of  $x \in \mathbb{Z}^{k-1}$  with the property that  $\langle \alpha, x \rangle \geq 0$  for all  $\alpha \in \Delta'_{k-1}$ . We denote the Weyl chamber

$$C_0 = \{x \in \mathbb{Z}^{k-1} \mid 0 \leq x_{k-1} \leq x_{k-2} \leq \dots \leq x_1\}. \tag{6}$$

For RNA secondary structures we have  $k-1 = 1$ , and  $B_1 = E_1 = \{\epsilon_1, 1\}$  and  $\Delta' = \{e_1\}$ . For 3-noncrossing RNA, we have  $k-1 = 2$  and  $B_2 = E_2 \rtimes_\varphi S_2 \cong D_4$  (where  $\varphi : S_2 \rightarrow \text{Aut}(E_2)$ ) is the dihedral group of order 8.



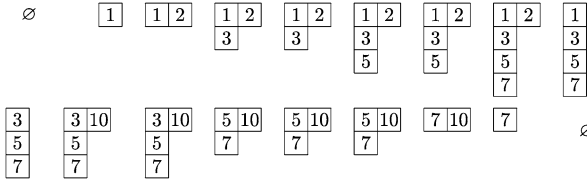
**Fig. 7** A 5-noncrossing digraph. In the text we show how to derive from this digraph the oscillating tableau below and subsequently the corresponding walk  $\gamma_{a,a}$  in  $\mathbb{Z}^4$ .

The following theorem is the first step for the enumeration of RNA structures. It will allow us to interpret a certain class of digraphs as walks in  $\mathbb{Z}^{k-1}$  which remain in the interior of the Weyl chamber  $C_0$ . The result is due to Chen et al. (2007), where it is formulated for matchings. The original bijection between oscillating tableaux and matchings is due to Stanley and was generalized by Sundaram (1990). We give a proof of Theorem 1 in Appendix A.

**Theorem 1** (Chen et al., 2007). *There exists a bijection between  $k$ -noncrossing digraphs and walks of length  $n$  in  $\mathbb{Z}^{k-1}$  which start and end at  $a = (k - 1, k - 2, \dots, 1)$  having steps  $0, \pm e_i, 1 \leq i \leq k - 1$  such that  $0 < x_{k-1} < \dots < x_1$  at any step. In other words we have a bijection*

$$\mathcal{G}_{n,k} \longrightarrow \{\gamma_{a,a} \mid \gamma_{a,a}, \text{ remains inside the Weyl-chamber } C_0\}, \tag{7}$$

where  $\mathcal{G}_{n,k}$  denotes the set of  $k$ -noncrossing digraphs of length  $n$ .



The 5-noncrossing digraph in Fig. 7 corresponds to an oscillating tableau as follows: from right to left start at vertex 17, which is a terminus. The corresponding origin is 7 which is inserted via the RSK algorithm into the empty shape. Next, insert the origin corresponding to 16 and 15, respectively. At 14 and 13, nothing happens since they are isolated vertices. At 12, origin 3 is inserted and 11 is isolated. 10 is an origin of an arc and accordingly removed from the tableaux. It is now clear how to proceed. The walk  $\gamma_{a,a}$  is obtained from the above oscillating tableau (starting with the rightmost  $\emptyset$  and reading right to left) as follows: its  $x_i$ -coordinate is the number of squares in the  $i$ -th row, i.e.  $\gamma_{a,a}$  is given by

- (4, 3, 2, 1), (5, 3, 2, 1), (6, 3, 2, 1), (6, 4, 2, 1), (6, 4, 2, 1), (6, 4, 2, 1),
- (6, 4, 3, 1), (6, 4, 3, 1), (5, 4, 3, 1), (5, 4, 3, 2), (6, 4, 3, 2), (6, 4, 3, 1),
- (6, 4, 3, 1), (6, 4, 2, 1), (6, 4, 2, 1), (6, 3, 2, 1), (5, 3, 2, 1), (4, 3, 2, 1)

We show in Appendix A in detail why this is a bijection.

Let us next discuss the reflection principle. The key idea is to count walks that remain in the interior of a Weyl chamber by counting all walks. Then one utilizes the fact that

all walks that touch a wall at some step can be paired and eventually cancel themselves in the enumeration. The particular way to obtain this pairing is by reflecting the walk at the corresponding wall. The following observation is essential for the reflection principle, formulated in Theorem 2 below.

**Lemma 1.** *Let  $\Delta'_{k-1} = \{e_{k-1}, e_{j-1} - e_j \mid 2 \leq j \leq k-1\}$ . Then every walk starting at some lattice point in the interior of  $C$  having steps  $\pm e_i, 0$  that crosses from inside  $C$  into outside  $C$  touches a subspace  $\langle e_i - e_{i-1} \mid 2 \leq i \leq k-1 \rangle$  or  $\langle e_i \mid 1 \leq i \leq k-1 \rangle$ .*

Let  $\Gamma_n(a, b)$  be the number of walks  $\gamma_{a,b}$ . For  $a, b \in C_0$  (Eq. (6)) let  $\Gamma_n^+(a, b)$  denote the number of walks  $\gamma_{a,b}$  that never touch a wall, i.e. remain in the interior of  $C_0$ . Finally for  $a, b \in \mathbb{Z}^{k-1}$ , let  $\Gamma_n^-(a, b)$  denote the number of walks  $\gamma_{a,b} = (s_1, \dots, s_n)$  that hit a wall at some step  $s_r$ .  $\ell(\beta)$  denotes the length of  $\beta \in \mathbb{B}_{k-1}$ . For  $a = b = (k-1, \dots, 1)$  we have according to Theorem 1

$$\Gamma_n^+(a, a) = \mathbf{G}_{n,k}, \quad (8)$$

where  $\mathbf{G}_{n,k} = |\mathcal{G}_{n,k}|$ .

**Theorem 2** (Reflection principle; Gessel and Zeilberger, 1992). *Suppose  $a, b \in C_0$ , then we have*

$$\Gamma_n^+(a, b) = \sum_{\beta \in \mathbb{B}_{k-1}} (-1)^{\ell(\beta)} \Gamma_n(\beta(a), b). \quad (9)$$

Theorem 2 allows us to compute the exponential generating function for  $\Gamma_n^+(a, b)$ , which is the number of walks from  $a$  to  $b$ , that remain in the interior of  $C_0$  (Grabiner and Magyar, 1993).

**Lemma 2** (Grabiner and Magyar, 1993). *Let  $I_r(2x) = \sum_{j \geq 0} \frac{x^{2r+j}}{j!(r+j)!}$  be the hyperbolic Bessel function of the first kind of order  $r$ . Then the generating functions for the numbers of  $k$ -noncrossing digraphs of length  $n$  and for  $k$ -noncrossing digraphs of length  $n$  without isolated points,  $\Gamma_n^+(a, b)$  and  $\Gamma_n'^+(a, b)$  are given by*

$$\sum_{n \geq 0} \Gamma_n^+(a, b) \frac{x^n}{n!} = e^x \det[I_{a_i-b_j}(2x) - I_{a_i+b_j}(2x)] \Big|_{i,j=1}^{k-1}, \quad (10)$$

$$\sum_{n \geq 0} \Gamma_n'^+(a, b) \frac{x^n}{n!} = \det[I_{a_i-b_j}(2x) - I_{a_i+b_j}(2x)] \Big|_{i,j=1}^{k-1}. \quad (11)$$

Now we can combine our results in order to enumerate  $k$ -noncrossing digraphs using the bijection between digraphs and walks. Theorem 1 implies that the number of  $k$ -noncrossing digraphs is equal to  $\Gamma_k^+(a, a)$ , the number of lattice walks in  $\mathbb{Z}^{k-1}$  of length  $n$  that remain in the interior of  $C_0$  from  $(k-1, \dots, 1)$  to itself with steps  $0, \pm e_i, 1 \leq i \leq k-1$ . According to Lemma 2, the generating functions for walks with steps  $\pm e_i, 0$  and  $\pm e_i$ , where  $1 \leq i \leq k-1$  are given by

$$e^x \det[I_{i-j}(2x) - I_{i+j}(2x)] \Big|_{i,j=1}^{k-1} \quad \text{and} \quad \det[I_{i-j}(2x) - I_{i+j}(2x)] \Big|_{i,j=1}^{k-1}, \quad (12)$$

respectively. Let  $f_k(n, \ell)$  denote the number of  $k$ -noncrossing digraphs with  $\ell$  isolated points. Then

$$f_k(n, \ell) = \binom{n}{\ell} f_k(n - \ell, 0), \quad (13)$$

$$\det[I_{i-j}(2x) - I_{i+j}(2x)]_{i,j=1}^{k-1} = \sum_{n \geq 1} f_k(n, 0) \cdot \frac{x^n}{n!}, \quad (14)$$

$$\begin{aligned} e^x \det[I_{i-j}(2x) - I_{i+j}(2x)]_{i,j=1}^{k-1} \\ = \left( \sum_{\ell \geq 0} \frac{x^\ell}{\ell!} \right) \left( \sum_{n \geq 1} f_k(n, 0) \frac{x^n}{n!} \right) = \sum_{n \geq 1} \left\{ \sum_{\ell=0}^n f_k(n, \ell) \right\} \cdot \frac{x^n}{n!}. \end{aligned} \quad (15)$$

In particular, we obtain for  $k = 2$  and  $k = 3$

$$f_2(n, \ell) = \binom{n}{\ell} C_{(n-\ell)/2} \quad \text{and} \quad f_3(n, \ell) = \binom{n}{\ell} [C_{\frac{n-\ell}{2}+2} C_{\frac{n-\ell}{2}} - C_{\frac{n-\ell}{2}+1}^2], \quad (16)$$

where  $C_m$  denotes the  $m$ th Catalan number.

### 3. RNA structures

In this section, we derive the generating function for  $k$ -noncrossing RNA structures. The successful strategy consists in counting the “wrong” objects “multiple” times. To be precise, we will enumerate all  $k$ -noncrossing digraphs with  $j$  1-arcs by relating them to a *family* whose cardinality we can easily compute. We denote the number of RNA structures with exactly  $\ell$  isolated vertices by  $\mathfrak{S}_k(n, \ell)$ . Suppose  $k \geq 2$  and let  $\mathcal{G}_{n,k}(\ell, j)$  be the set of all  $k$ -noncrossing digraphs having exactly  $\ell$  isolated points and exactly  $j$  1-arcs, where a 1-arc is an arc of the form  $(i, i + 1)$ ,  $1 \leq i \leq n - 1$ . Setting  $\mathfrak{G}_k(n, \ell, j) = |\mathcal{G}_{n,k}(\ell, j)|$ , we have in particular  $\mathfrak{S}_k(n, \ell) = \mathfrak{G}_k(n, \ell, 0)$ .

**Theorem 3.** *Let  $k \in \mathbb{N}$ ,  $k \geq 2$ , let  $C_m$  denote the  $m$ th Catalan number and  $f_k(n, \ell)$  be the number of  $k$ -noncrossing digraphs over  $n$  vertices with exactly  $\ell$  isolated vertices. Then the number of RNA structures with  $\ell$  isolated vertices,  $\mathfrak{S}_k(n, \ell)$ , is given by*

$$\mathfrak{S}_k(n, \ell) = \sum_{b=0}^{(n-\ell)/2} (-1)^b \binom{n-b}{b} f_k(n-2b, \ell), \quad (17)$$

where  $f_k(n-2b, \ell)$  is given by the generating function in Eq. (14). Furthermore, the number of  $k$ -noncrossing RNA structures,  $\mathfrak{S}_k(n)$  is

$$\mathfrak{S}_k(n) = \sum_{b=0}^{\lfloor n/2 \rfloor} (-1)^b \binom{n-b}{b} \left\{ \sum_{\ell=0}^{n-2b} f_k(n-2b, \ell) \right\}, \quad (18)$$

where  $\{\sum_{\ell=0}^{n-2b} f_k(n-2b, \ell)\}$  is given by the generating function in Eq. (15).

*Proof:* We first prove

$$\sum_{j \geq b} \binom{j}{b} \mathbf{G}_k(n, \ell, j) = \binom{n-b}{b} f_k(n-2b, \ell). \quad (19)$$

For this purpose we construct a family  $\mathcal{F}$  of  $\mathcal{G}_{n,k}$ -digraphs, having exactly  $\ell$  isolated points and having at least  $b$  1-arcs as follows: select (a)  $b$  1-arcs, and (b) an arbitrary  $k$ -noncrossing digraph with exactly  $\ell$  isolated points over the remaining  $n-2b$  vertices. Let  $\mathcal{F}$  be the resulting family of digraphs.

*Claim 1.* Each element  $\theta \in \mathcal{F}$  is contained in  $\mathcal{G}_{n,k}(\ell, j)$  for some  $j \geq b$ .

To prove this, we observe that a 1-arc cannot cross any other arc, i.e. cannot be contained in a set of mutually crossing arcs. As a result for  $k \geq 2$  our construction generates digraphs that are  $k$ -noncrossing. Clearly,  $\theta$  has exactly  $\ell$  isolated vertices and in step (b) we potentially derive additional 1-arcs, whence  $j \geq b$ .

*Claim 2.*

$$|\mathcal{F}| = \binom{n-b}{b} f_k(n-2b, \ell). \quad (20)$$

Let  $\lambda(n, b)$  denote the number of ways to select  $b$  1-arcs over  $\{1, \dots, n\}$ . We observe that  $\lambda(n, b) = \binom{n-b}{b}$ . Identifying the two incident vertices of an 1-arc we conclude that we can choose the  $b$  1-arcs in  $\binom{n-b}{b}$  ways. Obviously,  $\ell$  isolated vertices can be obtained in  $\binom{n-2b}{\ell}$  different ways and it remains to select an arbitrary  $k$ -noncrossing digraph with exactly  $\ell$  isolated points over  $n-2b$  vertices. The number of those is given by  $f_k(n-2b, \ell)$  which we can compute via Lemma 2, whence Eq. (20) and Claim 2 is proved.

In view of the fact that any of the  $k$ -noncrossing digraphs can introduce additional 1-arcs we set

$$\mathcal{F}(j) = \{\theta \in \mathcal{F} \mid \theta \text{ has exactly } j \text{ 1-arcs}\}.$$

Obviously,  $\mathcal{F} = \dot{\bigcup}_{j \geq b} \mathcal{F}(j)$ . Suppose  $\theta \in \mathcal{F}(j)$ . According to Claim 1,  $\theta \in \mathcal{G}_{n,k}(\ell, j)$  and furthermore  $\theta$  occurs with multiplicity  $\binom{j}{b}$  in  $\mathcal{F}$  since by construction any  $b$ -element subset of the  $j$  1-arcs is counted respectively in  $\mathcal{F}$ . Therefore, we have

$$|\mathcal{F}(j)| = \binom{j}{b} \mathbf{G}_k(n, \ell, j) \quad (21)$$

and

$$\sum_{j \geq b} \binom{j}{b} \mathbf{G}_k(n, \ell, j) = \sum_{j \geq b} |\mathcal{F}(j)| = \binom{n-b}{b} f_k(n-2b, \ell),$$

whence Eq. (19). We next set  $F_k(x) = \sum_{j \geq 0} \mathbf{G}_k(n, \ell, j) x^j$ . Taking the  $b$ th derivative and let  $x = 1$  we obtain

$$\frac{1}{b!} F_k^{(b)}(1) = \sum_{j \geq b} \binom{j}{b} \mathbf{G}_k(n, \ell, j) 1^{j-b}. \quad (22)$$

Claim 2 provides an interpretation of the r.h.s. of Eq. (22)

$$\sum_{j \geq b} \binom{j}{b} \mathbf{G}_k(n, \ell, j) 1^{j-b} = \binom{n-b}{b} f_k(n-2b, \ell). \quad (23)$$

In order to connect  $F_k(x)$  and  $\frac{1}{b!} F^{(b)}(1)$  we consider the Taylor expansion of  $F_k(x)$  at  $x = 1$  and compute

$$F_k(x) = \sum_{b \geq 0} \frac{1}{b!} F^{(b)}(1) (x-1)^b = \sum_{b=0}^{(n-\ell)/2} \binom{n-b}{b} f_k(n-2b, \ell) (x-1)^b.$$

In view of  $\mathbf{S}_k(n, \ell) = \mathbf{G}_k(n, \ell, 0)$  is the constant term of  $F_k(x)$ , i.e.  $F_k(0)$ , whence

$$\mathbf{S}_k(n, \ell) = \sum_{b=0}^{(n-\ell)/2} (-1)^b \binom{n-b}{b} f_k(n-2b, \ell). \quad (24)$$

It remains to prove Eq. (18). Summing over all possible values of isolated vertices, we get

$$\begin{aligned} S_k(n) &= \sum_{\ell=0}^n \sum_{b=0}^{(n-\ell)/2} (-1)^b \binom{n-b}{b} f_k(n-2b, \ell) \\ &= \sum_{b=0}^{\lfloor n/2 \rfloor} (-1)^b \binom{n-b}{b} \left\{ \sum_{\ell=0}^{n-2b} f_k(n-2b, \ell) \right\}, \end{aligned}$$

where  $\sum_{\ell=0}^{n-2b} f_k(n-2b, \ell)$  is given by Eq. (15) and the proof of the theorem is complete.  $\square$

A first implication of Theorem 3 is a new proof for Waterman's formula (Schmitt and Waterman, 1994) for the number of RNA secondary structures with exactly  $k$  base pairs or equivalently having  $\ell = n - 2k$  isolated vertices. In Table 1 we list the first 15 numbers of 3-noncrossing RNA structures.

**Corollary 1.** *The number of RNA secondary structures having exactly  $\ell$  isolated vertices,  $\mathbf{S}_2(n, \ell)$ , is given by*

$$\mathbf{S}_2(n, \ell) = \frac{2}{n-\ell} \binom{\frac{n+\ell}{2}}{\frac{n-\ell}{2}+1} \binom{\frac{n+\ell}{2}-1}{\frac{n-\ell}{2}-1}. \quad (25)$$

Furthermore  $\mathbf{S}_2(n, \ell)$  satisfies the recursion

$$(n-\ell)(n-\ell+2) \cdot \mathbf{S}_2(n, \ell) - (n+\ell)(n+\ell-2) \cdot \mathbf{S}_2(n-2, \ell) = 0. \quad (26)$$

**Table 1** The first 15 numbers of 3-noncrossing RNA structures

$n$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\mathbf{S}_3(n)$	1	1	2	5	13	36	105	321	1018	3334	11216	38635	135835	486337	1769500

*Proof:* We actually give two independent proofs of Eq. (25): the first being a direct computation based on Eq. (17) and the second using the recursion in Eq. (26) derived by Zeilberger's algorithm (Wilf et al., 1996). Let  $\frac{n-\ell}{2} = k$  we compute

$$\begin{aligned}
\mathbf{S}_2(n, \ell) &= \sum_{b=0}^{(n-\ell)/2} (-1)^b \binom{n-b}{b} \binom{n-2b}{\ell} C_{\frac{n-\ell-2b}{2}} \\
&= \sum_{b=0}^k (-1)^b \frac{(n-b)!}{b!(n-2k)!} \cdot \frac{1}{(k-b+1)!(k-b)!} \\
&= \frac{(n-k-1)!}{(n-2k)! \cdot k!} \sum_{b=0}^k (-1)^b \frac{(n-b)!}{(k-b+1)!(n-k-1)!} \frac{k!}{b!(k-b)!} \\
&= \frac{1}{n-k} \binom{n-k}{k} \sum_{b=0}^k (-1)^b \binom{k}{b} \binom{n-b}{k-b+1} \\
&= (-1)^{k+1} \frac{1}{n-k} \binom{n-k}{k} \sum_{b=0}^k \binom{k}{b} \binom{k-n}{k-b+1} \\
&= \frac{1}{n-k} \binom{n-k}{k+1} \binom{n-k}{k} = \frac{1}{k} \binom{n-k}{k+1} \binom{n-k-1}{k-1}.
\end{aligned}$$

As for the second proof we use  $\mathbf{S}_2(n, \ell) = \sum_{b=0}^{(n-\ell)/2} (-1)^b \binom{n-b}{b} f_2(n-2b, \ell)$  as the input for Zeilberger's algorithm (Wilf et al., 1996) and obtain that  $\mathbf{S}_2(n, \ell)$  satisfies the recursion formula

$$(n-\ell)(n-\ell+2) \cdot \mathbf{S}_2(n, \ell) - (n+\ell)(n+\ell-2) \cdot \mathbf{S}_2(n-2, \ell) = 0. \quad (27)$$

Using a bijection between RNA secondary structures and linear trees, Waterman computed in Schmitt and Waterman (1994) the number of RNA secondary structures with exactly  $h$  arcs,  $s(n, h)$

$$s(n, h) = \frac{1}{h} \binom{n-h}{h+1} \binom{n-h-1}{h-1}. \quad (28)$$

It follows by direct computation that  $\mathbf{S}_2(n, n-2h) = s(n, h) = \frac{1}{h} \binom{n-h}{h+1} \binom{n-h-1}{h-1}$  satisfies the recursion in Eq. (27), from which we can conclude  $\mathbf{S}_2(n, \ell) = s(n, (n-\ell)/2)$ .  $\square$

**Corollary 2.** *The number of 3-noncrossing RNA structures having exactly  $\ell$  isolated vertices,  $\mathbf{S}_3(n, \ell)$ , is given by*

$$\mathbf{S}_3(n, \ell) = \sum_{b=0}^{(n-\ell)/2} (-1)^b \binom{n-b}{b} \binom{n-2b}{\ell} \left[ C_{\frac{n-\ell-2b}{2}} C_{\frac{n-\ell-2b}{2}+2} - C_{\frac{n-\ell-2b}{2}+1}^2 \right]. \quad (29)$$

Using the expression of Corollary 2 for  $\mathbf{S}_3(n, \ell)$  as an input for Zeilberger's algorithm (Wilf et al., 1996) we derive

**Corollary 3.** *The number of 3-noncrossing RNA structures having exactly  $\ell$  isolated vertices,  $S_3(n, \ell)$ , satisfies the 4-term recursion*

$$\begin{aligned} & p_1(n, \ell)S_3(n-6, \ell) - p_2(n, \ell)S_3(n-4, \ell) \\ & - p_3(n, \ell)S_3(n-2, \ell) + p_4(n, \ell)S_3(n, \ell) = 0, \end{aligned} \quad (30)$$

where the coefficients  $p_1(n, \ell)$ ,  $p_2(n, \ell)$ ,  $p_3(n, \ell)$  and  $p_4(n, \ell)$  are given by

$$\begin{aligned} p_1(n, \ell) &= \frac{1}{2}n(n-1)(n-10+\ell)(n-4+\ell)(n-8+\ell), \\ p_2(n, \ell) &= \frac{1}{2}n(n-3)(13n^3 - 126n^2 + 13n^2\ell - 88n\ell + 392n \\ &\quad + 3n\ell^2 + 216\ell - 384 - 42\ell^2 + 3\ell^3), \\ p_3(n, \ell) &= (n-1)\left(\frac{1}{2}n-2\right)(13n^3 - 30n^2 - 13n^2\ell + 8n + 16n\ell \\ &\quad + 3n\ell^2 + 30\ell^2 - 72\ell - 3\ell^3), \\ p_4(n, \ell) &= (n-3)\left(\frac{1}{2}n-2\right)(n-\ell)(n-\ell+6)(n-\ell+4). \end{aligned}$$

Theorem 3 immediately allows us to derive the generating function for circular  $k$ -noncrossing RNA structures. Circular RNA structures are  $k$ -noncrossing digraphs without arcs of the form  $(1, n)$ , representing molecular structures over circular sequences. In circular sequences, the arc  $(n, 1)$  is considered a bond of the primary sequence and consequently does not occur as an arc in the corresponding digraph representation. Suppose  $k \geq 2$  and let  $\mathcal{G}_{n,k}^{(c)}(\ell, j)$  be the set of all  $k$ -noncrossing digraphs having exactly  $\ell$  isolated points and exactly  $j$  1-arcs, where a 1-arc is an arc of the form  $(i, i+1)$ , where  $i$  is considered modulo  $n$ . We set  $G_k^{(c)}(n, \ell, j) = |\mathcal{G}_{n,k}^{(c)}(\ell, j)|$ .

**Theorem 4.** *Let  $k \in \mathbb{N}$ ,  $k \geq 2$ , then the number of circular  $k$ -noncrossing RNA structures, with exactly  $\ell$  isolated vertices  $S_k^{(c)}(n, \ell)$ , is given by*

$$S_k^{(c)}(n, \ell) = \sum_{b=0}^{(n-\ell)/2} (-1)^b \left[ \binom{(n-2)-(b-1)}{b-1} + \binom{n-b}{b} \right] f_k(n-2b, \ell), \quad (31)$$

where  $\lambda^{(c)}(n, b)(1, 0) = 0$ ,  $\lambda^{(c)}(n, b)(1, 1) = 1$ ,  $\lambda^{(c)}(2, 0) = 0$  and  $\lambda^{(c)}(2, 2) = 1$ .

$$S_k^{(c)}(n) = \sum_{b=0}^{\lfloor n/2 \rfloor} (-1)^b \left[ \binom{(n-2)-(b-1)}{b-1} + \binom{n-b}{b} \right] \left\{ \sum_{\ell=0}^{n-2b} f_k(n-2b, \ell) \right\}, \quad (32)$$

where  $\sum_{\ell=0}^{n-2b} f_k(n-2b, \ell)$  is given by Eq. (15).

*Proof:* For circular RNA structures, the 1-arcs are considered modulo  $n$ . Again we derive a family  $\mathcal{F}$  of  $\mathcal{G}_{n,k}$ -digraphs, having exactly  $\ell$  isolated points and at least  $b$  1-arcs.

We select (a)  $b$  1-arcs, and (b) an arbitrary  $k$ -noncrossing digraph with exactly  $\ell$  isolated points over the remaining  $n - 2b$  vertices. In complete analogy we derive that each element  $\theta \in \mathcal{F}$  is contained in  $\mathcal{G}_{n,k}^{(c)}(\ell, j)$  for some  $j \geq b$ . Let  $\lambda^{(c)}(n, b)$ , denote the number of ways to select  $b$  1-arcs over  $\{1, \dots, n\}$  including the arc  $(n, 1)$ . Then  $\lambda^{(c)}(n, b)$  is given by

$$\lambda^{(c)}(n, b) = \binom{(n-2) - (b-1)}{b-1} + \binom{n-b}{b}, \quad (33)$$

where  $\lambda^{(c)}(n, b)(1, 0) = 0$ ,  $\lambda^{(c)}(n, b)(1, 1) = 1$ ,  $\lambda^{(c)}(2, 0) = 0$  and  $\lambda^{(c)}(2, 2) = 1$ . Indeed, either the arc  $(n, 1)$  is selected in which case we are left with exactly  $\binom{(n-2)-(b-1)}{b-1}$  ways to select the remaining 1-arcs or  $(n, 1)$  is not selected, in which case according to Theorem 3 there are exactly  $\binom{n-b}{b}$  ways to select the 1-arcs. Therefore, we obtain

$$\sum_{j \geq b} \binom{j}{b} \mathbf{G}_k^{(c)}(n, \ell, j) = \left[ \binom{(n-2) - (b-1)}{b-1} + \binom{n-b}{b} \right] f_k(n - 2b, \ell). \quad (34)$$

In complete analogy to the argument in Theorem 3, we can conclude

$$\mathbf{S}_k^{(c)}(n, \ell) = \sum_{b=0}^{(n-\ell)/2} (-1)^b \left[ \binom{(n-2) - (b-1)}{b-1} + \binom{n-b}{b} \right] f_k(n - 2b, \ell). \quad (35)$$

Equation (32) follows analogously and the proof of the theorem is complete.  $\square$

#### 4. Restricted RNA structures

Realistic RNA structures are typically subject to the constraint of having no arcs of the form  $(i, i + 1)$ ,  $(i, i + 2)$  and  $(i, i + 3)$ . We now generalize the ideas in Section 3 to the enumeration of restricted RNA structures, i.e. RNA structures without any arcs of the form  $(i, i + 2)$  (2-arcs). We restrict ourselves to the case of 2-arcs for two reasons: first the combinatorics of structures without 3-arcs becomes significantly more technical. In particular, certain coefficients are only known via a recursion or a rational generating function and secondly excluding 3-arcs implies eventually  $k > 3$ , i.e. the structures exhibit more complicated crossings. However, our framework does allow in principle to analyze these structures. For restricted RNA structures, we need the condition  $k > 2$  instead of  $k \geq 2$ , since our construction of the family  $\mathcal{F}$  of  $\mathcal{G}_{n,k}$ -digraphs in Theorem 5 can produce pairs of mutually crossing arcs. Let  $\mathcal{G}_{n,k}(\ell, j_1, j_2)$  be the set of all  $k$ -noncrossing digraphs having exactly  $\ell$  isolated points and exactly  $j_1$  and  $j_2$  1- and 2-arcs. We set  $\mathbf{G}_k(n, \ell, j_1, j_2) = |\mathcal{G}_{n,k}(\ell, j_1, j_2)|$ . In particular, we have  $\mathbf{G}_k(n, \ell, 0, 0) = \mathbf{S}_k^{(r)}(n, \ell)$ . In Table 2 we list the first 15 numbers of restricted RNA structures.

**Theorem 5.** *Let  $k \in \mathbb{N}$ ,  $k > 2$ . Then the numbers of restricted RNA structures  $\mathbf{S}_k^{(r)}(n, \ell)$  and  $\mathbf{S}_k^{(r)}(n)$  are given by*

$$S_k^{(r)}(n, \ell) = \sum_{b_1 \geq 0, b_2 \geq 0} (-1)^{b_1+b_2} \lambda(n, b_1, b_2) f_k(n - 2(b_1 + b_2), \ell), \quad (36)$$

$$S_k^{(r)}(n) = \sum_{b_1 \geq 0, b_2 \geq 0}^{[n/2]} (-1)^{b_1+b_2} \lambda(n, b_1, b_2) \left\{ \sum_{\ell=0}^{n-2(b_1+b_2)} f_k(n - 2(b_1 + b_2), \ell) \right\}. \quad (37)$$

Here  $\lambda(n, b_1, b_2)$  satisfies the recursion

$$\begin{aligned} \lambda(n, b_1, b_2) &= \lambda(n - 2, b_1 - 1, b_2) + \lambda(n - 1, b_1, b_2) + \lambda(n - 4, b_1, b_2 - 2) \\ &\quad + \lambda(n - 3, b_1, b_2 - 1) \end{aligned} \quad (38)$$

and the initial conditions for Eq. (38) are  $\lambda(n, 0, 0) = 1$ ,  $\lambda(n, b_1, 0) = \binom{n-b_1}{b_1}$ ,  $\lambda(n, 0, b_2) = \gamma(n, b_2)$  and  $\gamma(n, 1) = 0$  for  $n = 1$ ,  $\gamma(n, 1) = n - 2$  for  $n \geq 2$  and  $\gamma(n, 2) = 0$  for  $n = 2, 3$ .

*Proof:* Suppose  $\lambda(n, b_1, b_2)$  is the number of ways to select exactly  $b_1$  1-arcs and  $b_2$  2-arcs over  $\{1, \dots, n\}$  vertices.

*Claim.*  $\lambda(n, b_1, b_2)$  satisfies the recursion of Eq. (38) with the respective initial conditions, and we have

$$\sum_{j_1 \geq b_1, j_2 \geq b_2} \binom{j_1}{b_1} \binom{j_2}{b_2} \mathbf{G}_k(n, \ell, j_1, j_2) = \lambda(n, b_1, b_2) f_k(n - 2(b_1 + b_2), \ell). \quad (39)$$

In analogy to the proof of Theorem 3, we derive a family  $\mathcal{F}$  of  $\mathcal{G}_{n,k}$ -digraphs, having exactly  $\ell$  isolated points and at least  $b_1$  and  $b_2$  1-arcs and 2-arcs, respectively. We first prove that this construction generates elements of  $\mathcal{G}_{n,k}(\ell, j_1, j_2)$  and then express  $|\mathcal{F}|$  via the numbers  $\mathbf{G}_k(n, \ell, j_1, j_2)$ . We select (a)  $b_1$  1-arcs and  $b_2$  2-arcs and (b) an arbitrary  $k$ -noncrossing digraph over the remaining  $n - 2(b_1 + b_2)$  vertices with exactly  $\ell$  isolated points. Let  $\mathcal{F}$  be the family of digraphs obtained this way.

*Claim 1.* Each element  $\theta \in \mathcal{F}$  is contained in  $\mathcal{G}_{n,k}(\ell, j_1, j_2)$  for some  $j_1 \geq b_1$  and  $j_2 \geq b_2$ .

To prove this, we observe that any 1-arc or 2-arc can only cross at most one other arc. Therefore, 1-arcs and 2-arcs cannot be contained in a set of more than 2-mutually crossing arcs. As a result, for  $k > 2$  we generate digraphs that are  $k$ -noncrossing. Clearly,  $\theta$  has exactly  $\ell$  isolated vertices and in step (b) we potentially derive additional 1-arcs and 2-arcs, whence  $j_1 \geq b_1$  and  $j_2 \geq b_2$ , respectively.

*Claim 2.*

$$|\mathcal{F}| = \lambda(n, b_1, b_2) f_k(n - 2(b_1 + b_2), \ell). \quad (40)$$

We prove that the number of ways to select 1 and 2-arcs satisfies the recursion in Eq. (38) by induction on  $n$ . For the induction step we distinguish the following cases:

Case 1. The arc  $(1, 2)$  is selected. Then we have  $\lambda(n - 2, b_1 - 1, b_2)$  ways to select  $(b_1 - 1)$  1-arcs and  $b_2$  2-arcs over the vertices  $\{3, \dots, n\}$ .

Case 2. The arc  $(1, 2)$  is not selected. Then we distinguish the scenarios:  $(1, 3)$  is selected and  $(1, 3)$  is not selected. In the latter case we have  $\lambda(n - 1, b_1, b_2)$  ways to choose  $b_1$

1-arcs and  $b_2$  2-arcs over the vertices  $\{2, \dots, n\}$ . Suppose (1, 3) is selected. Then we have either that (2, 4) is selected, in which case we can select the remaining  $b_1$  1-arcs and  $b_2 - 2$  2-arcs over  $\{5, \dots, n\}$  in exactly  $\lambda(n - 4, b_1, b_2 - 2)$  different ways. In case (2, 4) is not selected, we can freely choose  $b_1$  1-arcs and  $(b_2 - 1)$  2-arcs over  $\{4, \dots, n\}$ , i.e. there are  $\lambda(n - 3, b_1, b_2 - 1)$  ways. Therefore, we derive the recursion

$$\begin{aligned} \lambda(n, b_1, b_2) &= \lambda(n - 2, b_1 - 1, b_2) + \lambda(n - 1, b_1, b_2) + \lambda(n - 4, b_1, b_2 - 2) \\ &\quad + \lambda(n - 3, b_1, b_2 - 1). \end{aligned}$$

As for the initial conditions, we have  $\lambda(n, 0, 0) = 1$ ,  $\lambda(n, b_1, 0) = \binom{n-b_1}{b_1}$ ,  $\lambda(n, 0, b_2) = \gamma(n, b_2)$  and  $\gamma(n, 1) = 0$  for  $n = 1$ ,  $\gamma(n, 1) = n - 2$  for  $n \geq 2$  and  $\gamma(n, 2) = 0$  for  $n = 2, 3$ . It remains to select an arbitrary  $k$ -noncrossing digraph with  $\ell$  isolated vertices over  $n - 2(b_1 + b_2)$  vertices. According to Lemma 2, the latter number is given by  $f_k(n - 2(b_1 + b_2), \ell)$ , whence Eq. (40) and Claim 2 is proved. In view of the fact that any of the  $k$ -noncrossing digraphs over  $n - 2(b_1 + b_2)$  vertices can introduce additional 1-arcs or 2-arcs, we set

$$\mathcal{F}(j_1, j_2) = \{\theta \in \mathcal{F} \mid \theta \text{ has exactly } j_1 \text{ 1-arcs and } j_2 \text{ 2-arcs}\}.$$

Obviously, we have the partition  $\mathcal{F} = \bigcup_{j_1 \geq b_1, j_2 \geq b_2} \mathcal{F}(j_1, j_2)$ . Suppose  $\theta \in \mathcal{F}(j_1, j_2)$ . According to Claim 1,  $\theta \in \mathcal{G}_{n,k}(\ell, j_1, j_2)$  and furthermore  $\theta$  occurs with multiplicity  $\binom{j_1}{b_1} \binom{j_2}{b_2}$  in  $\mathcal{F}$  since by construction any  $b_1$ -element subset of the  $j_1$  1-arcs and  $b_2$ -element subset of the  $j_2$  2-arcs are counted respectively in  $\mathcal{F}$ . Therefore, we have

$$|\mathcal{F}(j_1, j_2)| = \binom{j_1}{b_1} \binom{j_2}{b_2} \mathbf{G}_k(n, \ell, j_1, j_2) \quad (41)$$

and

$$\begin{aligned} \sum_{j_1 \geq b_1, j_2 \geq b_2} \binom{j_1}{b_1} \binom{j_2}{b_2} \mathbf{G}_k(n, \ell, j_1, j_2) &= \sum_{j_1 \geq b_1, j_2 \geq b_2} |\mathcal{F}(j_1, j_2)| \\ &= \lambda(n, b_1, b_2) f_k(n - 2(b_1 + b_2), \ell). \end{aligned}$$

We next set  $F_k(x, y) = \sum_{j_1 \geq 0} \sum_{j_2 \geq 0} \mathbf{G}_k(n, \ell, j_1, j_2) x^{j_1} y^{j_2}$ . Taking the  $b_1$ -th and  $b_2$ -th derivatives w.r.t.  $x$  and  $y$ , we obtain

$$\frac{1}{b_1!} \frac{1}{b_2!} F_k^{(b_1, b_2)}(1) = \sum_{j_1 \geq b_1, j_2 \geq b_2} \binom{j_1}{b_1} \binom{j_2}{b_2} \mathbf{G}_k(n, \ell, j_1, j_2) 1^{j_1-b_1} 1^{j_2-b_2}. \quad (42)$$

Then we have

$$\begin{aligned} &\sum_{j_1, j_2 \geq 0} \mathbf{G}_k(n, \ell, j_1, j_2) x^{j_1} y^{j_2} \\ &= \sum_{b_1 \geq 0, b_2 \geq 0} \left[ \sum_{j_1 \geq b_1, j_2 \geq b_2} \binom{j_1}{b_1} \binom{j_2}{b_2} \mathbf{G}_k(n, \ell, j_1, j_2) \right] (x-1)^{b_1} (y-1)^{b_2} \\ &= \sum_{b_1 \geq 0, b_2 \geq 0} \lambda(n, b_1, b_2) f_k(n - 2(b_1 + b_2), \ell) (x-1)^{b_1} (y-1)^{b_2}. \end{aligned}$$

**Table 2** The first 15 numbers of 3-noncrossing restricted RNA structures

$n$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$S_3^r(n)$	1	1	1	2	5	14	40	119	364	1145	3688	12139	40734	139071	482214

By construction  $\mathbb{G}(n, \ell, 0, 0)$  is the constant term of the  $F_k(x, y)$ . That is, the number of  $k$ -noncrossing RNA structures with  $\ell$  isolated vertices and no 2-arcs is given by

$$\mathbb{G}(n, \ell, 0, 0) = \sum_{b_1 \geq 0, b_2 \geq 0} (-1)^{b_1+b_2} \lambda(n, b_1, b_2) f_k(n - 2(b_1 + b_2), \ell) \quad (43)$$

and taking the sum over all  $\ell$  Eq. (37) follows

$$S_k^{(r)}(n) = \sum_{b_1 \geq 0, b_2 \geq 0}^{[n/2]} (-1)^{b_1+b_2} \lambda(n, b_1, b_2) \left\{ \sum_{\ell=0}^{n-2(b_1+b_2)} f_k(n - 2(b_1 + b_2), \ell) \right\},$$

where  $\{\sum_{\ell=0}^{n-2(b_1+b_2)} f_k(n - 2(b_1 + b_2), \ell)\}$  is given by Eq. (15) and the proof of the theorem is complete.  $\square$

## Acknowledgements

We are grateful to Prof. W.Y.C. Chen and Prof. Xin for helpful comments. Many thanks to J.Z.M. Gao and F.W.D. Huang for their help and L.C. Zuo for her suggestions. Special thanks to one reviewer who's suggestions led to Fig. 3. 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.

## Appendix A:

*Proof of Theorem 1:* Suppose we have two shapes  $\mu^i \subsetneq \mu^{i-1}$  and  $T_{i-1}$  is a standard Young tableau of shape  $\mu^{i-1}$ . We first observe that there exists a unique  $j$  and a unique  $T_i$  such that  $T_{i-1}$  is obtained from  $T_i$  by row-inserting  $j$  with the RSK algorithm.

Suppose  $\mu^{i-1}$  differs from  $\mu^i$  in the first row. Then  $j$  is the element at the end of the first row in  $T_{i-1}$ . Otherwise suppose  $\ell$  is the row of the square being removed from  $T_{i-1}$ . Remove the square and insert its element  $x$  into the  $(\ell - 1)$ th row at precisely the position, where the removed element  $y$  would push it down via the RSK-algorithm. That is,  $y$  is maximal subject to  $y < x$ . Since each column is strictly increasing  $y$  always exists. Iterating this process results in exactly one element  $j$  being removed from  $T_i$  and a new filling of  $\mu_{i-1}$ , i.e. an unique tableau  $T_{i-1}$ . By construction, inserting  $j$  with the RSK algorithm produces  $T_{i-1}$ .

*Claim 1.* There exists a bijection between the set of oscillating tableaux of length  $n$  and digraphs with vertices of degree  $\leq 1$ .

Given an oscillating tableau  $(\mu^i)_{i=0}^n$  ( $\mu^i$  differs from  $\mu^{i-1}$  by at most one square), we define a sequence  $(G_0, T_0), (G_1, T_1), \dots, (G_n, T_n)$ , recursively, where  $G_i$  is a digraph and  $T_i$  is a standard Young tableau. We define  $G_0$  to be the digraph with empty edge-set and  $T_0$  to be the empty standard Young tableau. The tableau  $T_i$  is obtained from  $T_{i-1}$  and the digraph  $G_i$  is obtained from  $G_{i-1}$  by the following procedure:

- (1) (Insert origins) For  $\mu^i \supseteq \mu^{i-1}$ , then  $T_i$  is obtained from  $T_{i-1}$  by adding the entry  $i$  in the square  $\mu^i \setminus \mu^{i-1}$ .
- (2) (Isolated vertices) For  $\mu^i = \mu^{i-1}$  then set  $T_i = T_{i-1}$
- (3) (Remove origins) For  $\mu^i \subsetneq \mu^{i-1}$ , then let  $T_i$  be the unique standard Young tableau of shape  $\mu^i$

and  $j$  be the unique number such that  $T_{i-1}$  is obtained from  $T_i$  by row-inserting  $j$  with the RSK algorithm. Then set  $E_{G_i} = E_{G_{i-1}} \cup \{(j, i)\}$ .

Obviously,  $G_n$  is a digraph, and the set of  $i$  where  $\mu^i = \mu^{i-1}$  equals the set of isolated vertices of  $G_n$ . By construction each entry  $j$  is removed exactly once whence no edges of the form  $(j, i)$  and  $(j, i')$  can be obtained. Therefore  $G_n$  has degree  $\leq 1$  and we have a well defined mapping

$$\beta: \{(\mu_i)_{i=0}^n \mid (\mu_i)_{i=0}^n \text{ is an oscillating tableau}\} \\ \longrightarrow \{G_n \mid G_n \text{ is a digraph with degree } \leq 1\}.$$

It is clear from the procedure that  $G_n$  is a labeled graph and  $\beta$  is injective. To prove surjectivity, we observe that each digraph  $G_n$  induces an oscillating tableau as follows. We set  $\mu_{G_n}^0 = \emptyset$  and  $T_n = \emptyset$ . Starting from vertex  $i = n, n-1, \dots, 1, 0$  we derive a sequence of Young tableaux  $(T_n, T_{n-1}, \dots, T_0)$  as follows:

I. If  $i$  is a terminus of an  $G_n$ -arc  $(j, i)$  add  $j$  via the RSK-algorithm to  $T_i$  set  $\mu_{G_n}^{i-1} \supseteq \mu_{G_n}^i$  to be the shape of  $T_{i-1}$  (corresponds to (3))

II. If  $i$  is an isolated  $G_n$ -vertex set  $\mu_{G_n}^{i-1} = \mu_{G_n}^i$  (corresponds to (2))

III. If  $i$  is the origin of an  $G_n$ -arc  $(i, k)$  let  $\mu_{G_n}^{i-1} \subsetneq \mu_{G_n}^i$  be the shape of  $T_{i-1}$ , the standard Young tableau obtained by removing the square containing  $i$  (corresponds to (1)).

Then we have  $\beta((\mu_{G_n})_0^n) = G_n$ , whence  $\beta$  is surjective.

*Claim 2.*  $G_n$  is  $k$ -noncrossing if and only if all shapes  $\mu^i$  in the oscillating tableau have less than  $k$  rows.

From Claim 1 we know  $\beta^{-1}(G_n) = (\emptyset = \mu^0, \mu^1, \dots, \mu^n = \emptyset)$ , so it suffices to prove that the maximal number of rows in the shape set  $\beta^{-1}(G_n)$  is less than  $k$ . First we observe that the arcs  $(i_1, j_1), \dots, (i_\ell, j_\ell)$  form a  $\ell$ -crossing of  $G_n$  if and only if there exists a tableau  $T_i$  such that elements  $i_1, i_2, \dots, i_\ell$  are in the  $\ell$  squares of  $T_i$  and being deleted in increasing order  $i_1 < i_2 < \dots < i_\ell$  afterwards. Next, we will obtain a permutation  $\pi_i$  from the entries in each tableau  $T_i$  recursively as follows:

1. If  $T_{i-1}$  is obtained from  $T_i$  by row-inserting  $j$  with the RSK algorithm, then  $\pi_{i-1} = \pi_i j$ .
2. If  $T_i = T_{i-1}$ , then  $\pi_i = \pi_{i-1}$ .
3. If  $T_{i-1}$  is obtained from  $T_i$  by deleting the entry  $i$ , then  $\pi_{i-1}$  is obtained from  $\pi_i$  by deleting  $i$ .

If  $\pi = r_1 r_2 \dots r_t$ , then the entries being deleted afterwards are in the order  $r_t, \dots, r_2, r_1$ .

Using the RSK algorithm w.r.t. the permutation  $\pi_i$ , the resulting row-inserting Young tableau is exactly  $T_i$ . We prove this by induction in reverse order of the oscillating tableau. It is trivial for the case  $i = n$ . Suppose it holds for  $j$ ,  $1 \leq j \leq n$ . Consider the above three cases: inserting an element, doing nothing and deleting an element. In the first case, the assertion is implied the RSK algorithm in the construction of the oscillating tableau. In the second case, it holds by the induction hypothesis on step  $j$ .

It remains to consider the third case, that is, removing the entry from  $T_j$  to get  $T_{j-1}$ . We show that also in this case the insertion Young tableau of  $\pi_i$  equals the labeled oscillating tableau  $T_i$ . Write  $\pi_j = x_1 x_2 \cdots x_p j y_1 y_2 \cdots y_q$  and  $\pi_{j-1} = x_1 x_2 \cdots x_p y_1 y_2 \cdots y_q$ . In view of step 3  $j$  is larger than elements  $x_1, x_2, \dots, x_p, y_1, \dots, y_q$ . We need to prove that the insertion tableau  $S_{j-1}$  of  $\pi_{j-1}$  by the RSK algorithm is exactly the same as deleting the entry  $j$  in  $T_j$ . We proceed by induction on  $q$ . In the case  $q = 0$ ,  $T_j$  is obtained from  $T_{j-1}$  by adding  $j$  at the end of the first row. Suppose the assertion holds for  $q - 1$ , that is  $S_{j-1}(x_1 x_2 \cdots x_p y_1 y_2 \cdots y_{q-1}) = S_j(x_1 x_2 \cdots x_p j y_1 y_2 \cdots y_{q-1}) \setminus \boxed{j}$ . Consider inserting  $y_q$  into  $S_{j-1}$ , via the RSK algorithm. If the insertion track path never touches the position of  $j$ , then  $S_{j-1}(x_1 x_2 \cdots x_p y_1 y_2 \cdots y_{q-1} y_q) = S_j(x_1 x_2 \cdots x_p j y_1 y_2 \cdots y_{q-1} y_q) \setminus \boxed{j}$ . Otherwise, if the insertion path touched  $j$  and pushed  $j$  into the next row, then since  $j$  is greater than any other entry,  $j$  must be moved to the end of next row and the push process stops. Accordingly, the insertion path in  $S_{j-1}(x_1 x_2 \cdots x_p y_1 y_2 \cdots y_{q-1})$  is the same path as in  $S_j(x_1 x_2 \cdots x_p j y_1 y_2 \cdots y_{q-1})$  except the last step moving  $j$  to a new position  $j$ , so deleting  $j$  will get  $S_{j-1}(x_1 x_2 \cdots x_p y_1 y_2 \cdots y_{q-1} y_q) = S_j(x_1 x_2 \cdots x_p j y_1 y_2 \cdots y_{q-1} y_q) \setminus \boxed{j}$ . According to Schensted's Theorem, for any permutation  $\pi$ , assume  $A$  is the corresponding insertion Young tableau by using the RSK algorithm on  $\pi$ . Then the length of the longest decreasing subsequences of  $\pi$  is the number of rows in  $A$ , whence the assertion.

Now we can prove Claim 2. A digraph is a  $\ell$ -crossing if and only if there exists a  $\pi_i$  which has decreasing subsequence of length  $\ell$ . And the insertion Young tableau of  $\pi_i$  equals the labeled oscillating tableau  $T_i$ . According to Schensted's theorem,  $\pi$  has a decreasing sequence of length  $\ell$  if and only if rows of  $T_i$  is  $\ell$ .

*Claim 3.* There is a bijection between oscillating tableaux with at most  $k - 1$  rows of length  $n$  and walks with steps  $\pm e_i, 0$  which stay in the interior of  $C_0$  starting and ending at  $(k - 1, k - 2, \dots, 1)$ .

This bijection is obtained by setting for  $1 \leq \ell \leq k - 1$ ,  $x_\ell$  to be the length of the  $\ell$ -th row. By definition of standard Young tableaux, we have  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$  i.e. the length of each row is weakly decreasing. This property also characterizes walks that stay within the Weyl-chamber  $C_0$ , i.e. where we have  $x_1 > x_2 > \cdots > x_{k-1} > 0$  since a walk from  $(k - 1, \dots, 2, 1)$  to itself in the interior of  $C_0$  corresponds to a walk from the origin to itself in the region  $x_1 \geq x_2 \geq \cdots \geq x_{k-1} \geq 0$ . In an oscillating tableau  $\mu^i$  differs from  $\mu^{i-1}$  by at most one square and adding or deleting a square in the  $\ell$ -th row or doing nothing corresponds to steps  $\pm e_\ell$  and  $0$ , respectively. Since the oscillating tableau is an empty shape, we have walks from the origin to itself, whence Claim 3 follows and the proof of the Theorem is complete.  $\square$

*Proof of Lemma 1:* To prove the lemma we can without loss of generality assume  $C = C_0 = \{(x_1, \dots, x_{k-1}) \mid x_1 \geq x_2 \geq \cdots \geq x_{k-1} \geq 0\}$ . Then the assertion is that every walk having steps  $\pm e_i, 0$  starting at  $a = (k - 1, k - 2, \dots, 1)$  that crosses from inside  $C_0$  into outside  $C_0$  intersects one of the sub-spaces  $\langle e_{k-1} \rangle$  or  $\langle e_{j-1} - e_j \rangle$  for  $2 \leq j \leq k - 1$ . This

is correct since to leave  $C$  implies that there exists some  $i$  such that  $x_i \leq x_{i+1}$ . Let  $s_j$  be minimal w.r.t.  $a + \sum_h^{j+1} s_h \notin C_0$ . Since we have steps  $\pm e_i, 0$  we conclude  $x_{k-1} = 0$  or  $x_j = x_{j-1}$  for some  $2 \leq j \leq k - 1$ , whence the lemma.  $\square$

*Proof of Theorem 2:* Totally order the roots of  $\Delta$ . Let  $\Gamma_n^-(a, b)$  be the number of walks from  $a$  to  $b, a, b \in \mathbb{Z}^{k-1}$  of length  $n$  using the steps  $s, s \in \{\pm e_i, 0\}$  such that  $\langle \gamma(s_r), \alpha \rangle = 0$  for some  $\alpha \in \Delta$  (i.e. the walk intersects with the subspace  $\langle \alpha \rangle$ ). According to Lemma 1, every walk that crosses from inside  $C$  into outside  $C$  touches a wall from which we can draw two conclusions:

$$\Gamma_n(a, b) = \Gamma_n^+(a, b) + \Gamma_n^-(a, b), \tag{A.1}$$

$$\beta \neq \text{id} \implies \Gamma_n(\beta(a), b) = \Gamma_n^-(\beta(a), b). \tag{A.2}$$

*Claim.*  $\sum_{\beta \in \mathbf{B}_{k-1}} (-1)^{\ell(\beta)} \Gamma_n^-(\beta(a), b) = 0$ .

Let  $(s_1, \dots, s_n)$  be a walk from  $\beta(a)$  to  $b$ . By assumption there exists some step  $s_r$  at which we have  $\langle \gamma_{\beta(a), b}(s_r), \alpha \rangle = 0$ , for  $\alpha \in \Delta$ . Let  $\alpha^*$  be the largest root for which we have  $\langle \gamma_{\beta(a), b}(s_r), \alpha^* \rangle = 0$  and  $\beta_{\alpha^*}(x) = x - \frac{2\langle \alpha^*, x \rangle}{\langle \alpha^*, \alpha^* \rangle} \alpha^*$  its associated reflection (Eq. (4)). We consider the walk

$$(\beta_{\alpha^*}(s_1), \dots, \beta_{\alpha^*}(s_r), s_{r+1}, \dots, s_n). \tag{A.3}$$

Now by definition  $(\beta_{\alpha^*}(s_1), \dots, \beta_{\alpha^*}(s_r), s_{r+1}, \dots, s_n)$  starts at  $(\beta_{\alpha^*} \circ \beta)(a)$  and has sign  $(-1)^{\ell(\beta)+1}$  since  $\ell(\beta) + 1 = \ell(\beta_{\alpha^*} \circ \beta)$ . Therefore to each element  $\gamma_{\beta(a), b}$  of  $\Gamma_n^-(\beta(a), b)$  having sign  $(-1)^{\ell(\beta)}$  there exists a  $\gamma_{\beta_{\alpha^*}\beta(a), b} \in \Gamma_n^-(\beta_{\alpha^*}\beta(a), b)$  with sign  $(-1)^{\ell(\beta)+1}$  and the claim follows. We immediately derive

$$\begin{aligned} & \sum_{\beta \in \mathbf{B}_{k-1}} (-1)^{\ell(\beta)} \Gamma_n(\beta(a), b) \\ &= \Gamma_n(a, b) + \sum_{\beta \in \mathbf{B}_{k-1}, \beta \neq \text{id}} (-1)^{\ell(\beta)} \underbrace{\Gamma_n(\beta(a), b)}_{= \Gamma_n^-(\beta(a), b)} \\ &= \Gamma_n^+(a, b) + \Gamma_n^-(a, b) + \underbrace{\sum_{\beta \in \mathbf{B}_{k-1}, \beta \neq \text{id}} (-1)^{\ell(\beta)} \Gamma_n^-(\beta(a), b)}_{\sum_{\beta \in \mathbf{B}_{k-1}} (-1)^{\ell(\beta)} \Gamma_n^-(\beta(a), b) = 0} \end{aligned}$$

whence the theorem.  $\square$

*Proof of Lemma 2:* Let  $u_i, 1 \leq i \leq k - 1$ , be indeterminants and  $u = (u_i)_1^{k-1}$ . We define  $u^{b-a} = \prod_{i=1}^{k-1} u_i^{b_i - a_i}$ . Let  $F(x, u)$  be a generating function, then  $F(x, u)|_{u^{b-a}}$  equals the family of coefficients  $a_i(u)$  at  $u^{b-a}$  of  $\sum_{i \geq 0} a_i(u)x^i$ . We first observe

$$\Gamma_n(a, b) = \left[ 1 + \sum_{i=1}^{k-1} (u_i + u_i^{-1}) \right] \Big|_{u^{b-a}}^n$$

The exponential generating function for  $\Gamma_n(a, b)$  is

$$\begin{aligned} \sum_{n \geq 0} \Gamma_n(a, b) \frac{x^n}{n!} &= \sum_{n \geq 0} \left[ 1 + \sum_{i=1}^{k-1} (u_i + u_i^{-1}) \right] \Big|_{u^{b-a}} \frac{x^n}{n!} \\ &= \sum_{n \geq 0} \frac{[1 + \sum_{i=1}^{k-1} (u_i + u_i^{-1})]^n}{n!} x^n \Big|_{u^{b-a}} \\ &= e^x \cdot \exp \left[ x \sum_{i=1}^{k-1} (u_i + u_i^{-1}) \right] \Big|_{u^{b-a}} \\ &= e^x \cdot \prod_{i=1}^{k-1} (\exp(x(u_i + u_i^{-1})) \Big|_{u_i^{b_i - a_i}}). \end{aligned}$$

We furthermore derive

$$\sum_{n \geq 0} \Gamma_n^+(a, b) \frac{x^n}{n!} = e^x \sum_{\beta \in B_{k-1}} (-1)^{\ell(\beta)} \prod_{i=1}^{k-1} \exp(x(u_i + u_i^{-1})) \Big|_{u^{b-\beta(a)}}$$

and writing  $\beta = \epsilon_h^r \sigma$ , where  $r = 0, 1$  we obtain

$$\begin{aligned} \sum_{n \geq 0} \Gamma_n^+(a, b) \frac{x^n}{n!} &= e^x \sum_{\sigma \in S_{k-1}} \sum_{\substack{h=1 \\ r=0,1}}^{k-1} (-1)^{\ell(\epsilon_h^r)} \operatorname{sgn}(\sigma) \prod_{i=1}^{k-1} (\exp(x(u_i + u_i^{-1})) \Big|_{u_i^{b_i - \epsilon_h^r a_{\sigma_i}}}) \\ &= e^x \sum_{\sigma \in S_{k-1}} \operatorname{sgn}(\sigma) \sum_{\substack{h=1 \\ r=0,1}}^{k-1} (-1)^{\ell(\epsilon_h^r)} \prod_{i=1}^{k-1} (\exp(x(u_i + u_i^{-1})) \Big|_{u_i^{b_i - \epsilon_h^r a_{\sigma_i}}}) \\ &= e^x \sum_{\sigma \in S_{k-1}} \operatorname{sgn}(\sigma) \left\{ \prod_{i=1}^{k-1} (\exp(x(u_i + u_i^{-1})) \Big|_{u_i^{b_i - a_{\sigma_i}}}) \right. \\ &\quad \left. - \prod_{i=1}^{k-1} (\exp(x(u_i + u_i^{-1})) \Big|_{u_i^{b_i + a_{\sigma_i}}}) \right\} \\ &= e^x \det_{(k-1) \times (k-1)} [I_{a_i - b_j}(2x) - I_{a_i + b_j}(2x)] \Big|_{i,j=1}^{k-1} \end{aligned}$$

and the lemma follows.  $\square$

## References

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

- Gessel, I.M., Zeilberger, D., 1992. Random walk in a Weyl chamber. Proc. Am. Math. Soc. 115, 27–31.
- Grabiner, D.J., Magyar, P., 1993. Random walks in Weyl chambers and the decomposition of tensor powers. J. Algebr. Comb. 2, 239–260.
- Hasegawa, A., Uemura, Y., Kobayashi, S., Yokomori, T., 1999. Tree adjoining grammars for RNA structure prediction. Theor. Comput. Sci. 210, 277–303.
- 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.
- 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.
- Lindström, 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.
- Lynsgo, 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.
- McCaskill, J.S., 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. Biopolymers 29, 1105–1119.
- Parkin, N., Chamorro, M., 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. J. Proc. Natl. Acad. Sci. USA 89, 713–717.
- Penner, R.C., Waterman, M.S., 1993. Spaces of RNA secondary structures. Adv. Math. 101, 31–49.
- Rivas, E., Eddy, S., 1999. A Dynamic Programming Algorithm for RNA structure prediction including 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.
- Sundaram, S., 1990. The Cauchy identity for  $Sp(2n)$ . J. Comb. Theory (A) 53, 209–238.
- 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.
- ten Dam, E., Brierly, I., Inglis, S., Pleij, C., 1994. Identification and analysis of the pseudoknot containing *gag-pro* ribosomal frameshift signal of simian retrovirus-1. Nucl. Acids Res. 22, 2304–2310.
- 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.
- 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 cloverleaves. 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.
- Wilf, H.S., Petkovsek, M., Zeilberger, D., 1996.  $A = B$ . A.K. Peters Ltd., Wellesly.
- Zuker, M., Sankoff, D., 1984. RNA secondary structures and their prediction. Bull. Math. Biol. 46(4), 591–621.