Narayana number, Chebyshev polynomial and Motzkin path on RNA abstract shapes

Sang Kwan Choi, Chaiho Rim and Hwajin Um

Abstract We consider a certain abstract of RNA secondary structures, which is closely related to so-called RNA shapes. The generating function counting the number of the abstract structures is obtained in three different ways, namely, by means of Narayana numbers, Chebyshev polynomials and Motzkin paths. We show that a combinatorial interpretation on 2-Motzkin paths explains a relation between Motzkin paths and RNA shapes and also provides an identity related to Narayana numbers and Motzkin polynomial coefficients.

1 Introduction

Ribonucleic acid (RNA) is a single stranded molecule with a backbone of nucleotides, each of which has one of the four bases, adenine (A), cytosine (C), guanine (G) and uracil (U). Base pairs are formed intra-molecularly between A-U, G-C or G-U, leading the sequence of bases to form helical regions. The primary structure of a RNA is merely the sequence of bases and its three-dimensional conformation by base pairs is called the tertiary structure. As an intermediate structure between the primary and the tertiary, the secondary structure is a planar structure allowing only nested base pairs. This is easy to see in its diagrammatic representation, see figure 1. A sequence of n bases is that of labeled vertices $(1,2,\cdots,n)$ in a horizontal line and base pairs are drawn as arcs in the upper half-plane. The condition of nested base pairs means non-crossing arcs: for two arcs (i,j) and (k,l) where $i<j$,

Sang Kwan Choi
Center for Theoretical Physics, College of Physical Science and Technology
Sichuan University, Chengdu 6100064, China, e-mail: :hermit1231@sogang.ac.kr

Chaiho Rim
Department of Physics, Sogang University, Seoul 121-742, Korea e-mail: rimpine@sogang.ac.kr

Hwajin Um
Department of Physics, Sogang University, Seoul 121-742, Korea e-mail: um16@sogang.ac.kr

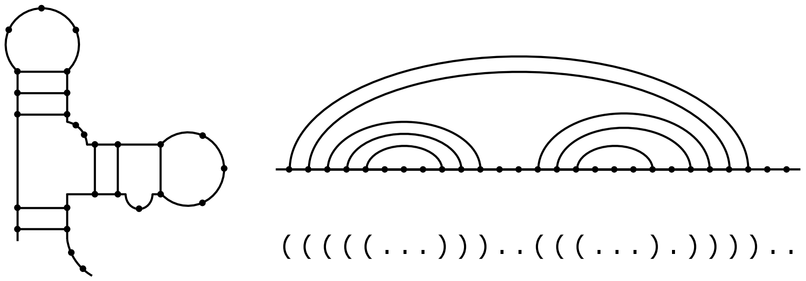


Fig. 1 Representations of secondary structures. The RNA structure on the left hand side is represented as the diagram (top right) and the dot-bracket string (bottom right).

$k < l$ and $i < k$, either $i < j < k < l$ or $i < k < l < i$. Since the functional role of a RNA depends mainly on its 3D conformation, prediction of RNA folding from the primary structure has long been an important problem in molecular biology. The most common approach for the prediction is free energy minimization and many algorithms to compute the structures with minimum free energy has been developed (see for instance, [13, 22, 21, 17]).

On the other hand, RNA structures are often considered as combinatorial objects in terms of representations such as strings over finite alphabets, linear trees or the diagrams. Combinatorial approaches enumerate the number of possible structures under various kinds of constraints and observe its statistics to compare with experimental findings [18, 9, 16, 1, 4]. They also provide classifications of structures to advance prediction algorithms [20, 8, 14, 15].

In this paper, we consider a certain abstract of secondary structures under a pure combinatorial point of view regardless of primary structures. The abstract structure is, in fact, closely related to so-called RNA shapes [8, 10, 12], see section 3. Although we will consider it apart from prediction algorithms, let us review briefly the background to RNA shapes in the context of prediction problem. In free energy minimization scheme, the lowest free energy structures are not necessarily native structures. One needs to search suboptimal foldings in a certain energy bandwidth and, in general, obtains a huge set of suboptimal foldings. RNA shapes classify the foldings according to their structural similarities and provide so-called shape representatives such that native structures can be found among those shape representatives. Consequently, it can greatly narrow down the huge set of suboptimal foldings to probe in order to find native structures.

In the following preliminary, we introduce our combinatorial object, what we call island diagrams and present basic definitions needed to describe the diagrams. In section 2, we find the generating function counting the number of island diagrams in three different ways and through which, one may see the intertwining relations between Narayana numbers, Chebyshev polynomials and Motzkin paths. In particular, we find a combinatorial identity, see equation (15), which reproduces the following two identities that Coker provided [5] (see also [3] for a combinatorial interpretation):

$$\sum_{k=1}^{n}\frac{1}{n}\binom{n}{k}\binom{n}{k-1}x^{k-1}=\sum_{k=0}^{\lfloor\frac{n-1}{2}\rfloor}C_k\binom{n-1}{2k}x^k(1+x)^{n-2k-1} \tag{1}$$

$$\sum_{k=1}^{n}\frac{1}{n}\binom{n}{k}\binom{n}{k-1}x^{2(k-1)}(1+x)^{2(n-k)}=\sum_{k=1}^{n}C_k\binom{n-1}{k-1}x^{k-1}(1+x)^{k-1} \tag{2}$$

where C_k is the Catalan number defined by $C_k=\frac{1}{k+1}\binom{2k}{k}$ for $k\geq 0$. We also provide a combinatorial interpretation on 2-Motzkin paths to explain the identity (15). The interpretation implies the bijection between π-shapes and Motzkin paths which was shown in [7, 11].

Preliminary

A formal definition of secondary structures is given as follows:

Definition 1 (Waterman [20]). A secondary structure is a vertex-labeled graph on n vertices with an adjacency matrix $A=(a_{ij})$ (whose element $a_{ij}=1$ if i and j are adjacent, and $a_{ij}=0$ otherwise with $a_{ii}=0$) fulfilling the following three conditions:
1. $a_{i,i+1}=1$ for $1\leq i\leq n-1$.
2. For each fixed i, there is at most one $a_{ij}=1$ where $j\neq i\pm 1$
3. If $a_{ij}=a_{kl}=1$, where $i<k<j$, then $i\leq l\leq j$.

An edge (i,j) with $|i-j|\neq 1$ is said to be a base pair and a vertex i connected only to $i-1$ and $i+1$ is called unpaired. We will call an edge $(i,i+1)$, $1\leq i\leq n-1$, a backbone edge. Note that a base pair between adjacent two vertices is not allowed by definition and the second condition implies non-existence of base triples.

There are many other representations of secondary structures than the diagrammatic representation. In this paper, we often use the so-called dot-bracket representation, see figure 1. A secondary structure can be represented as a string $\mathbf{S}$ over the alphabet set $\{(,),.\}$ by the following rules [9]:
1. If vertex i is unpaired then $\mathbf{S}_i=$“.”.
2. If (i,j) is a base pair and $i<j$ then $\mathbf{S}_i=$ “(” and $\mathbf{S}_j=$“)”.

In the following, we present the basic definitions of structure elements needed for our investigations.

Definition 2. A secondary structure on $(1,2,\cdots,n)$ consists of the following structure elements (cf. Fig.2). By a base pair (i,j), we always assume $i<j$.
1. The sequence of unpaired vertices $(i+1,i+2,\cdots,j-1)$ is a *hairpin* if (i,j) is a base pair. The pair (i,j) is said to be the *foundation of the hairpin*.
2. The sequence of unpaired vertices $(i+1,i+2,\cdots,j-1)$ is a *bulge* if either (k,j), $(k+1,i)$ or $(i,k+1)$, (j,k) are base pairs.
3. The sequence of unpaired vertices $(i+1,i+2,\cdots,j-1)$ is a *join* if (k,i) and (j,l) are base pairs.
4. A *tail* is a sequence of unpaired vertices $(1,2,\cdots,i-1)$, resp. $(j+1,j+2,\cdots,n)$ such that i, resp. j is paired.
5. An *interior loop* is two sequences of unpaired vertices $(i+1,i+2,\cdots,j-1)$ and

$(k+1, k+2, \cdots, l-1)$ such that (i,l) and (j,k) are pairs, where $i < j < k < l$.
6. For any $k \geq 3$ and $0 \leq l, m \leq k$ with $l + m = k$, a *multi loop* is l sequences of unpaired vertices and m empty sequences $(i_1+1, \cdots, j_1-1), (i_2+1, \cdots, j_2-1), \cdots, (i_k+1, \cdots, j_k-1)$ such that $(i_1, j_k), (j_1, i_2), \cdots, (j_{k-1}, i_k)$ are base pairs. Here, a sequence $(i+1, \cdots, j-1)$ is an empty sequence if $i+1 = j$.
7. A *stack (or stem)* consists of uninterrupted base pairs $(i+1, j-1), (i+2, j-2), \cdots, (i+k, j-k)$ such that neither (i,j) nor $(i+k+1, j-k-1)$ is a base pair. Here the *length* of the stack is k.

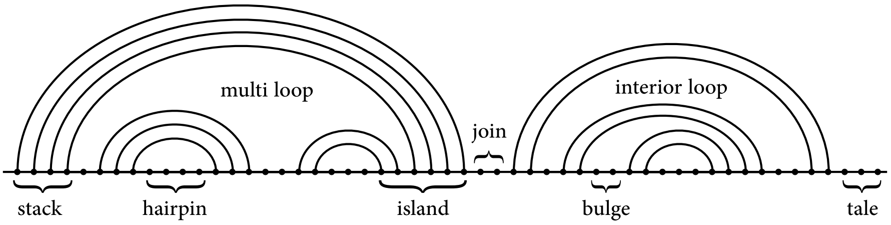


Fig. 2 Structure elements of secondary structures.

Note that, while other structure elements consist of at least one vertex, a multiloop does not necessarily have a vertex. In the diagrammatic representation, a multiloop is a structure bounded by three or more base pairs and backbone edges.

Definition 3. An *island* is a sequence of paired vertices $(i, i+1, \cdots, j)$ such that
1. $i-1$ and $j+1$ are both unpaired, where $1 < i \leq j < n$.
2. $j+1$ is unpaired, where $i = 1$ and $1 < j < n$.
3. $i-1$ is unpaired, where $1 < i < n$ and $j = n$.

Now we introduce the abstract structures to consider throughout this paper. From here on, we will call the structures *island diagrams* for convenience. An island diagram (cf. Fig.3) is obtained from secondary structures by
1. Removing tails.
2. Representing a sequence of consecutive unpaired vertices between two islands by a single blank.
Accordingly, we retain unpaired regions except for tails but do not account for the number of unpaired vertices. In terms of the dot-bracket representation, we shall use the underscore "_" for the blank: for example, the island diagram "((_)_)" abstracts the secondary structure "((...)....)". Since the abstraction preserves all the structure elements (except for tails) in the definition 2, we will use them to describe island diagrams in such a way that, for instance, the blank is a hairpin if its left and right vertices are paired to each other.

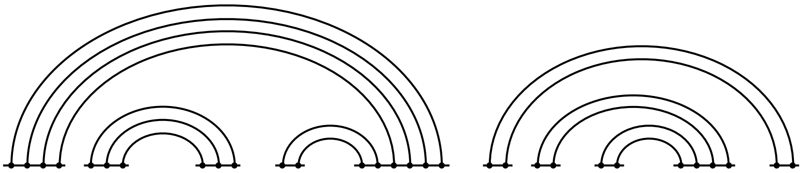


Fig. 3 An example of island diagrams. This island diagram is the abstract structure of the secondary structure given in the figure 2.

2 Generating function

We enumerate the number of island diagrams $g(h,I,\ell)$, filtered by the number of hairpins(h), islands(I) and basepairs(ℓ). Let $G(x,y,z) = \sum_{h,I,\ell} g(h,I,\ell)x^h y^I z^\ell$ denotes the corresponding generating function. We obtain the generating function in three different ways, by means of Narayana numbers, Chebyshev polynomials and Motzkin paths. In particular, we provide a bijection map between 2-Motzkin paths and sequences of matching brackets.

2.1 Narayana number

The easiest way to obtain the generating function $G(x,y,z)$ is to use a combinatorial interpretation of the Narayana numbers, which are defined by

$$N(n,k) = \frac{1}{n}\binom{n}{k}\binom{n}{k-1}, \quad 1 \leq k \leq n. \tag{3}$$

The Narayana number $N(n,k)$ counts the number of ways arranging n pairs of brackets to be correctly matched and contain k pairs as “()”. For instance, the bracket representations for $N(4,2) = 6$ are given as follows:

(()(()))   ((()()))   ((())())   ()((()))   (())(())   ((()))()

It is easy to recover island diagrams from this representation.

Proposition 1 *The generating function has the form*

$$G(x,y,z) = \sum_{\ell,h} N(\ell,h)\, x^h\, y^{h+1}\, (1+y)^{2\ell-1-h}\, z^\ell\,. \tag{4}$$

Its closed form is

$$G(x,y,z) = \left(\frac{y}{1+y}\right)\frac{1-A(1+B)-\sqrt{1-2A(1+B)+A^2(1-B)^2}}{2A} \tag{5}$$

where $A = z(1+y)^2$ and $B = xy/(1+y)$.

Proof. One may immediately associate bracket representations of the Narayana numbers with island diagrams. Without regard to underscores, the pair of brackets is associated with the basepair and the sub-pattern “()” corresponds to the foundation of the hairpin. It clearly explains the factor $N(\ell,h)x^h z^\ell$. Now we consider the insertions of underscores to recover the string representation of island diagrams. Recall that, in secondary structures, a hairpin consists of at least one unpaired vertices. Therefore, the foundation of the hairpin “()” must contain a underscore “(_)”. The number h of underscores are so inserted that we have the factor y^{h+1}. After the insertion of hairpin underscores, there are $(2\ell-1-h)$ places left to possibly insert underscores. The numbers of all possible insertions are summarized by the factor $(1+y)^{2\ell-1-h}$. The generating function of the Narayana numbers is well-known (see for instance [2]) so that one writes the closed form.

2.2 Chebyshev polynomial

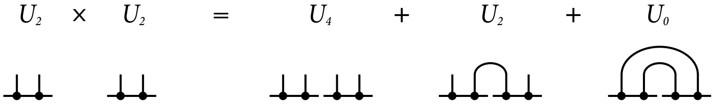
One can also count the number of island diagrams by using the Chebyshev polynomials of the second kind, which are defined by the recurrence relation:

$$U_0(\xi)=1\,,\quad U_1(\xi)=2\xi\,,\quad U_{n+1}(\xi)=2\xi U_n(\xi)-U_{n-1}(\xi)\,. \tag{6}$$

The product of the polynomials expands as

$$U_m(\xi)U_n(\xi)=\sum_{k=0}^{n}U_{m-n+2k}(\xi)\quad \text{for}\quad n\leq m\,. \tag{7}$$

The relation between island diagrams and Chebyshev polynomials are based on the Feynman diagram of the Hermitian matrix model, refer to [4]. One may have an insight from the simplest example:



The polynomial U_k corresponds to the island with k vertices. The product U_2U_2 expands to U_4(no basepair), U_2(one basepair) and U_0(all vertices are paired). The island diagram is the one associated with U_0 in the expansion of the product. In general, we have the following theorem. See [4] for its proof.

Theorem 1. *Suppose that there exist the number I of islands such that each of which has $k_a\geq 1$ vertices for $a\in\{1,\cdots,I\}$. The number of island diagrams one finds by making base pairs is given by*

$$\left\langle \prod_{a=1}^{I} U_{k_a}, U_0 \right\rangle := \frac{2}{\pi} \int_{-1}^{1} \prod_{a=1}^{I} U_{k_a}(\xi) U_0(\xi) \sqrt{1-\xi^2} d\xi \,. \tag{8}$$

where $U_k(\xi)$ is the second kind Chebyshev polynomial of degree k.

The Chebyshev polynomials of the second kind are orthogonal with respect to the weight $\sqrt{1-\xi^2}$: $\langle U_m, U_n \rangle = \delta_{m,n}$. Thus, the theorem 1 means that the number of island diagrams is the coefficient of $U_0 = 1$ when the product $\prod_{a=1}^{I} U_{k_a}(\xi)$ expands to the linear combination of Chebyshev polynomials.

In order to reproduce the generating function given in (4), we need to take the number of hairpins into account as well. Let us first consider the case of island diagrams in which every blank(underscore) is a hairpin. A hairpin is accompanied with the foundation of the hairpin, that is, h basepairs are assigned as the foundations. Since those basepairs are the most nested ones, the number of the island diagrams is simply given by $\left\langle U_{k_1-1} \prod_{j=2}^{h} U_{k_j-2} U_{k_{h+1}-1}, U_0 \right\rangle$. The foundations of the hairpin take one vertex from the outermost islands and take two vertices from the others. In fact, the island diagrams having only hairpins are no different from strings of matching brackets which represents Narayana numbers as shown in the previous subsection. By just putting (_) → (), we recover the bracket representations. Thus, we have the following corollary:

Corollary 1.1 *For any $\ell \in \mathbb{N}$ and $1 \le h \le \ell$,*

$$N(\ell, h) = \sum_{k_1 + \cdots + k_{h+1} = 2(\ell - h)} \left\langle \prod_{a=1}^{h+1} U_{k_a}, U_0 \right\rangle \tag{9}$$

where k_a for $a \in \{1, \cdots, h+1\}$ are non-negative integers.

Now we find the generating function $G(x,y,z)$. Note that a basepair must be made across at least one hairpin. Conversely, no basepair can be made amongst consecutive islands that do not have a hairpin inbetween. We regard a group of maximally consecutive islands with no hairpin inbetween as one effective island. Then, a backbone of island diagram can be seen as an alternate arrangement of effective island and hairpin. This is nothing but the case that every blank is a hairpin. One additional thing to consider is the number of ways to make an effective island having k_a vertices out of I_a islands, which is given by $\binom{k_a-1}{I_a-1}$. Therefore, we find

$$g(h, I, \ell) = \sum_{\{k_a, I_a\}} \prod_{a=1}^{h+1} \binom{k_a - 1}{I_a - 1} \left\langle U_{k_1-1} \prod_{j=2}^{h} U_{k_j-2} U_{k_{h+1}-1}, U_0 \right\rangle \tag{10}$$

where the summation runs over $k_1 + \cdots + k_{h+1} = 2\ell$ and $I_1 + \cdots + I_{h+1} = I$. By means of the corollary 1.1, one can obtain the generating function (4).

We mention that one may also find the generating function by direct calculation of the integral in (10). Using the generating function of the Chebyshev polynomial,

$$\sum_{k\geq 0}\sum_{i=0}^{k}\binom{k}{i} z^{k/2} y^i U_k(\xi) = \frac{1}{1-2\sqrt{z}(1+y)\xi + z(1+y)^2}\,, \tag{11}$$

the integral is calculated to give

$$G(x,y,z) = \sum_h x^h z^h y^{h+1}\,(1+y)^{h-1}{}_2F_1(h+1,h;2;z(1+y)^2) \tag{12}$$

where ${}_2F_1(a,b;c;z)$ is the hypergeometric function. One may easily show that ${}_2F_1(h+1,h;2;z) = \sum_{k\geq 0} N(h+k,h) z^k$ and therefore obtains the generating function (4).

2.3 Motzkin path

The generating function $G(x,y,z)$ can also be written in terms of Motzkin polynomial coefficients. The Motzkin numbers M_n and the Motzkin polynomial coefficients $M(n,k)$ are defined as

$$M_n = \sum_{k=0}^{\lfloor n/2\rfloor} M(n,k) \quad \text{where} \quad M(n,k) = \binom{n}{2k} C_k\,. \tag{13}$$

Let us consider the combinatorial identity in the following theorem. It is easy to prove using the generating function of the Motzkin polynomials:

$$\sum_{\ell\geq 1}\sum_{p=0}^{\lfloor(\ell-1)/2\rfloor} M(\ell-1,p) A^{\ell-1} B^p = \frac{1-A-\sqrt{(1-A)^2-4A^2B}}{2A^2B}\,. \tag{14}$$

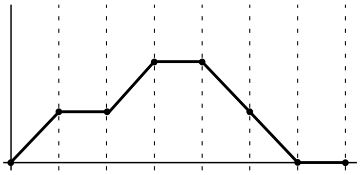
Theorem 2. *For any integer $\ell \geq 1$, there holds*

$$\begin{aligned} &\frac{y}{1+y}\sum_{h=1}^{\ell} N(\ell,h)\,(xy)^h\,(1+y)^{2\ell-h} \\ &= xy^2 \sum_{p=0}^{\lfloor\frac{\ell-1}{2}\rfloor} M(\ell-1,p)\,\big(xy\,(1+y)^3\big)^p\,\big((1+y)(1+y+xy)\big)^{\ell-2p-1}\,. \end{aligned} \tag{15}$$

Proof. The left hand side is $[z^\ell]G(x,y,z)$ given in (4). Multiplying z^ℓ and taking the summation over ℓ at each side, one can check that the right hand side is indeed the generating function $G(x,y,z)$.

Note that the identity (15) reproduces the Coker's two identities. When we substitute x/y for x and then put $y = 0$, we get the identity (1). Furthermore, the substitution $x \to y/(1+y)$ leads to the identity (2).[1]

We will investigate how the right hand side in (15) represents island diagrams. In order to do that, we need a combinatorial interpretation of 2-Motkzin paths. Let us first introduce the Motzkin paths, that can also be called 1-Motkzin paths. A Motzkin path of size n is a lattice path starting at $(0,0)$ and ending at $(n,0)$ in the integer plane $\mathbb{Z}\times\mathbb{Z}$, which satisfies two conditions: (i) It never passes below the x-axis. (ii) Its allowed steps are the up step $(1,1)$, the down step $(1,-1)$ and the horizontal step $(1,0)$. We denote by U, D and H an up step, a down step and a horizontal step, respectively. The Motzkin polynomial coefficient $M(n,k)$ is the number of Motzkin paths of size n with k up steps. Since the Motkzin number M_n is given by the sum of $M(n,k)$ over the number of up steps, M_n is the number of Motzkin paths of size n. See for instance, the following figure depicting a Motzkin path of $M(7,2)$:



On the other hand, 2-Motzkin paths allow two kinds of horizontal steps, which often distinguish one from another by a color, let us say, R and B denoting a red and a blue step, respectively. We provide a bijection map between 2-Motzkin paths and strings of matching brackets.[2] Suppose we have a 2-Motzkin path of size n given by a string $q_1 q_2 \cdots q_n$ over the set $\{U,D,R,B\}$. The corresponding string of brackets S_n can be obtained by the following rules:
(i) We begin with "()" : Let $S_0 = ()$.
(ii) For any $1 \leq k \leq n$, suppose there exist a string of brackets S' and a string of matching brackets S'' which are possibly empty such that S_{k-1} has the form $S'(S'')$. Then S_k is given by

$$\begin{array}{rlrl} S'((S'')() & \text{if } q_k = U, & S'(S'')) & \text{if } q_k = D, \\ S'(S'')() & \text{if } q_k = R, & S'((S'')) & \text{if } q_k = B. \end{array}$$

For example, the string of matching brackets corresponding to the 2-Motzkin path $UBURDD$ is obtained as follows:

[1] In order to deduce the identity, one may need the Touchard's identity [19]: $C_n = \sum_k C_k \binom{n-1}{2k} 2^{n-2k-1}$, which can be also derived from (1) when $x = 1$.

[2] Sequences of matching brackets are only Dyck paths. A bijection map between Dyck paths and 2-Motzkin paths was introduced by Delest and Viennot [6]. But here we present a different way of mapping than the well-known one.

$$
\begin{aligned}
() \xrightarrow{U} (()() \xrightarrow{B} (()(()) \xrightarrow{U} (()((())()& \\
\xrightarrow{R} (()((())()() \xrightarrow{D} (()((())()()) \xrightarrow{D} (()((())()()))&
\end{aligned}
$$

We remark here that only blue steps can make a stack. In other words, directly nested structures such as $(())$ never occur without blue steps. Therefore, a 1-Motzkin path can be translated into a string of matching brackets without directly nested brackets. This is one of the 14 interpretations of Motzkin numbers provided by Donaghey and Shapiro in [7]. Later, in [11], it was also shown using context-free grammars in the context of RNA shapes. We also remark that the Motzkin polynomial coefficient $M(\ell-1,u)$ is the number of ways arranging ℓ pairs of brackets to be correctly matched and contain $\ell-u$ pairs as "()" with no occurrence of directly nested bracket.

Now we go back to the generating function on the right hand side in (15) and rewrite it as

$$
\begin{aligned}
G(x,y,z) = \sum_{\ell,u} & M(\ell-1,u)\,(xy^2z)\,\big((1+y)\sqrt{z}\big)^u \big(xy(1+y)z\big)^u \\
& \times \big((1+y)\sqrt{z}\big)^d \big((1+y)^2z + xy(1+y)z\big)^s
\end{aligned}
\tag{16}
$$

where u, d and s stand for the number of up, down and horizontal steps, respectively ($u=d$, $u+d+s=\ell-1$). Let us explain each factor in detail by means of the above rules. The term xy^2z is merely the starting hairpin "(_)" (recall that the exponent of x, y and z are the number of hairpins, islands and basepairs, resp.). At each up step, one has a left bracket and a hairpin to add. For a given non-empty string S of island diagrams, suppose that we add a left bracket then there are the two possibilities, (S and (_S corresponding to $\sqrt{z}$ and $y\sqrt{z}$, respectively. Thus, we get the factor $(1+y)\sqrt{z}$ at every up step and, in the same manner, at every down step. Likewise, adding a hairpin introduces the factor $xy(1+y)z$ since S(_) and S_(_) corresponds to xyz and xy^2z, respectively. On the other hand, a horizontal step can be either R or B. A red step is to add a hairpin and corresponds to $xy(1+y)z$. A blue step is to add one basepair nesting the string "(S)" and there are three possibilities: the stack $((S))$ for z, the two bulges (_(S)) and ((S)_) for yz and the interior loop (_(S)_) for y^2z. Therefore, we get $((1+y)^2z+xy(1+y)z)$ at each horizontal step.

Note that the number of up steps is the number of multiloops since every up step opens a new multiloop. Thus the generating function written in terms of Motzkin polynomials can be said to classify island diagrams by the number of basepairs and multiloops while the one written in terms of Narayana numbers classify island diagrams by the number of basepairs and hairpins.

3 Single-stack diagrams and RNA shapes

An island diagram is called a single-stack diagram if the length of each stack in the diagram is 1 so that each basepair is a stem by itself. Let $s(h,I,k)$ denotes the number of single-stack diagrams classified by the number of hairpins(h), islands(I) and stems(k) and let $S(x,y,z)=\sum_{h,I,k}s(h,I,k)x^hy^Iz^k$ denotes its generating function. The island diagrams with k stems and ℓ basepairs build on the single-stack diagrams with k stems. The number of ways stacking $\ell-k$ basepairs on k stems is $\binom{\ell-1}{k-1}$ and we have

$$g(h,I,\ell)=\sum_{k=1}^{\ell}\binom{\ell-1}{k-1}s(h,I,k)\,. \tag{17}$$

Multiplying $x^hy^Iz^\ell$ at each side and summing over h, I, ℓ, one finds the relation

$$G(x,y,z)=S\Big(x,y,\frac{z}{1-z}\Big) \tag{18}$$

and equivalently, $S(x,y,z)=G(x,y,z/(1+z))$. In terms of Motzkin polynomials, the generating function $S(x,y,z)$ expands to

$$\begin{aligned} S(x,y,z)=&\sum_{k,u}M(\ell-1,u)\,(xy^2z)\,\big((1+y)\sqrt{z}\big)^u\,\big(xy(1+y)z\big)^u\\ &\times\big((1+y)\sqrt{z}\big)^d\,\big((2y+y^2)z+xy(1+y)z\big)^s \end{aligned} \tag{19}$$

where u, d and s stand for the number of up, down and horizontal steps, respectively ($u=d$, $u+d+s=k-1$). This is the same as (16) except for one thing. Recall that only blue steps make a directly nested bracket and from which we get three possibilities by putting underscores, *i.e.,* a stack for z, two bulges for yz and an interior loop for y^2z. One obtains single-stack diagrams by getting rid of the possibility of stacking and hence the one different thing is the factor z such that one has $(2y+y^2)z$ instead of $(1+y)^2z$.

We mention that the single-stack diagram is closely related to the π'-shape (or type 1), which is one of the five RNA abstract shapes provided in [8] classifying secondary structures according to their structural similarities. π'-shape is an abstraction of secondary structures preserving their loop configurations and unpaired regions. A stem is represented as one basepair and a sequence of maximally consecutive unpaired vertices is considered as an unpaired region regardless of the number of unpaired vertices in it. In terms of the dot-bracket representation, a length k stem $(^k\cdots)^k$ is represented by a pair of squared brackets $[\cdots]$ and an unpaired region is depicted by an underscore. For instance, the π'-shape “ _[[[_]_[_]]_]” can abstract from the secondary structure “ ...((((...)..((...))))..)”. The only difference between single-stack diagrams and π'-shapes is whether or not to retain tales.

On the other hand, π-shape (or type 5) ignores unpaired regions such that, for example, the π'-shape “_[[[_]_[_]]_]” results in the π-shape “[[][]]”. Consequently, π-shapes retain only hairpin and multiloop configurations. One may immediately notice that the string representations of π-shapes are nothing but the sequences of matching brackets without directly nested brackets. Therefore, as was shown in the previous section, there is a bijection map between π-shapes and 1-Motzkin paths. Accordingly, one finds the theorem 3.1 in [11] that the number of π-shapes with ℓ pairs of squared brackets is the Motzkin number $M_{\ell-1}$. Furthermore, the Motzkin polynomial coefficient $M(\ell-1,u)$ is the number of π-shapes with u multiloops and $\ell-u$ hairpins.

Acknowledgements The authors acknowledge the support of this work by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (NRF-2017R1A2A2A05001164). S.K. Choi is partially supported by the National Natural Science Foundation of China under grant 11575119.

References

1. Barrett, C.L., Li, T.J., Reidys, C.M.: RNA Secondary Structures Having a Compatible Sequence of Certain Nucleotide Ratios. Journal of Computational Biology **23**, 857–873 (2016)
2. Barry, P., Hennessy, A.: A note on narayana triangles and related polynomials, Riordan arrays, and MIMO capacity calculations. Journal of Integer Sequences **14**(3), 1–26 (2011)
3. Chen, W.Y., Yan, S.H., Yang, L.L.: Identities from weighted motzkin paths. Advances in Applied Mathematics **41**(3), 329 – 334 (2008). DOI https://doi.org/10.1016/j.aam.2004.11.007. URL http://www.sciencedirect.com/science/article/pii/S0196885808000158
4. Choi, S.K., Rim, C., Um, H.: RNA substructure as a random matrix ensemble (2016)
5. Coker, C.: Enumerating a class of lattice paths. Discrete Mathematics **271**(1), 13 – 28 (2003). DOI https://doi.org/10.1016/S0012-365X(03)00037-2. URL http://www.sciencedirect.com/science/article/pii/S0012365X03000372
6. Delest, M.P., Viennot, G.: Algebraic languages and polyominoes enumeration. Theoretical Computer Science **34**(1), 169 – 206 (1984). DOI https://doi.org/10.1016/0304-3975(84)90116-6. URL http://www.sciencedirect.com/science/article/pii/0304397584901166
7. Donaghey, R., Shapiro, L.W.: Motzkin numbers. Journal of Combinatorial Theory, Series A **23**(3), 291 – 301 (1977). DOI https://doi.org/10.1016/0097-3165(77)90020-6. URL http://www.sciencedirect.com/science/article/pii/0097316577900206
8. Giegerich, R., Voss, B., Rehmsmeier, M.: Abstract shapes of rna. Nucleic Acids Research **32**(16), 4843–4851 (2004). DOI 10.1093/nar/gkh779. URL + http://dx.doi.org/10.1093/nar/gkh779
9. Hofacker, I.L., Schuster, P., Stadler, P.F.: Combinatorics of RNA secondary structures. Discrete Applied Mathematics **88**(1-3), 207–237 (1998)
10. Janssen, S., Reeder, J., Giegerich, R.: Shape based indexing for faster search of rna family databases. BMC Bioinformatics **9**(1), 131 (2008). DOI 10.1186/1471-2105-9-131. URL https://doi.org/10.1186/1471-2105-9-131
11. Lorenz, W.A., Ponty, Y., Clote, P.: Asymptotics of rna shapes. Journal of Computational Biology **15**(1), 31–63 (2008)
12. Nebel, M.E., Scheid, A.: On quantitative effects of RNA shape abstraction. Theory in Biosciences **128**(4), 211–225 (2009). DOI 10.1007/s12064-009-0074-z
13. Nussinov, R., Jacobson, A.B.: Fast algorithm for predicting the secondary structure of single-stranded RNA. Proceedings of the National Academy of Sciences of the United States of America **77**(11), 6309–13 (1980). DOI 10.1073/pnas.77.11.6309
14. Orland, H., Zee, A.: Rna folding and large n matrix theory. Nucl. Phys. **B620**, 456–476 (2002)
15. Reidys, C.M., Huang, F.W.D., Andersen, J.E., Penner, R.C., Stadler, P.F., Nebel, M.E.: Topology and prediction of RNA pseudoknots. Bioinformatics **27**(8), 1076–1085 (2011). DOI 10.1093/bioinformatics/btr090. URL https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btr090

16. Schmitt, W.R., Waterman, M.S.: Linear trees and rna secondary structure. Discrete Applied Mathematics **51**(3), 317 – 323 (1994). DOI https://doi.org/10.1016/0166-218X(92)00038-N. URL http://www.sciencedirect.com/science/article/pii/0166218X9200038N
17. Schuster, P., Stadler, P.F., Renner, A.: Rna structures and folding: from conventional to new issues in structure predictions. Current Opinion in Structural Biology **7**(2), 229 – 235 (1997). DOI https://doi.org/10.1016/S0959-440X(97)80030-9. URL http://www.sciencedirect.com/science/article/pii/S0959440X97800309
18. Stein, P., Waterman, M.: On some new sequences generalizing the catalan and motzkin numbers. Discrete Mathematics **26**(3), 261 – 272 (1979). DOI https://doi.org/10.1016/0012-365X(79)90033-5. URL http://www.sciencedirect.com/science/article/pii/0012365X79900335
19. Touchard, J.: Sur certaines *é*quations fonctionnelles. Proceedings of the International Congress on Mathematics, Toronto (1924) **1**, 465 – 472 (1928)
20. Waterman, M.S.: Secondary structure of single-stranded nucleic acids. Adv. math. suppl. studies **1**, 167–212 (1978)
21. Zuker, M., Sankoff, D.: Rna secondary structures and their prediction. Bulletin of Mathematical Biology **46**(4), 591 – 621 (1984). DOI https://doi.org/10.1016/S0092-8240(84)80062-2. URL http://www.sciencedirect.com/science/article/pii/S0092824084800622
22. Zuker, M., Stiegler, P.: Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Research **9**(1), 133–148 (1981). DOI 10.1093/nar/9.1.133