Bull Math Biol (2017) 79:21–35
DOI 10.1007/s11538-016-0212-y

ORIGINAL ARTICLE

Regular Simple Queues of Protein Contact Maps

Qiang-Hui Guo[1] · **Lisa Hui Sun**[1,2] · **Jian Wang**[1]

Received: 24 March 2016 / Accepted: 21 September 2016 / Published online: 14 November 2016

Abstract A protein fold can be viewed as a self-avoiding walk in certain lattice model, and its contact map is a graph that represents the patterns of contacts in the fold. Goldman, Istrail, and Papadimitriou showed that a contact map in the 2D square lattice can be decomposed into at most two stacks and one queue. In the terminology of combinatorics, stacks and queues are noncrossing and nonnesting partitions, respectively. In this paper, we are concerned with 2-regular and 3-regular simple queues, for which the degree of each vertex is at most one and the arc lengths are at least 2 and 3, respectively. We show that 2-regular simple queues are in one-to-one correspondence with hill-free Motzkin paths, which have been enumerated by Barcucci, Pergola, Pinzani, and Rinaldi by using the Enumerating Combinatorial Objects method. We derive a recurrence relation for the generating function of Motzkin paths with k_i peaks at level i, which reduces to the generating function for hill-free Motzkin paths. Moreover, we show that 3-regular simple queues are in one-to-one correspondence with Motzkin paths avoiding certain patterns. Then we obtain a formula for the generating function of 3-regular simple queues. Asymptotic formulas for 2-regular and 3-regular simple queues are derived based on the generating functions.

✉ Lisa Hui Sun
sunhui@nankai.edu.cn

Qiang-Hui Guo
guo@nankai.edu.cn

Jian Wang
wj121313@126.com

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

2 Center for Applied Mathematics, Tianjin University, Tianjin 300072, People's Republic of China

Keywords Contact map · Stack · Queue · Motzkin path

AMS Classification 05A15 · 92B05

1 Introduction

The enumeration of contact maps plays a fundamental role in the study of protein folding prediction (Domany 2000; Vendruscolo et al. 1997), structure alignment (Goldman et al. 1999; Lancia et al. 2001; Agarwal et al. 2007), and protein structure data mining. Contact maps can be viewed as a combinatorial structure representing bonds between amino acid residues; see Goldman et al. (1999). In the 2D lattice model, two amino acid residues in a protein fold form a contact if they are adjacent in the lattice but not consecutive in the fold. The contact map of a protein fold can be obtained by representing the residues as vertices $1, 2, \ldots, n$ drawn on a horizontal line in increasing order, and connecting two adjacent vertices i and j by an arc in the upper-half plane. An arc is denoted by (i, j) with $i < j$, and the length of an arc (i, j) is defined to be $j - i$. Figure 1 illustrates a contact map of a protein fold in the 2D square lattice, where the arcs $(1, 16)$, $(2, 15)$, $(3, 6)$, $(5, 8)$, $(6, 15)$, $(7, 10)$, $(7, 14)$, and $(10, 13)$ correspond to the contacts between the residues in the protein fold.

Goldman et al. (1999) showed that the contact map for any protein fold in the 2D square lattice can be decomposed into at most two stacks and one queue. As an example, the contact map in Fig. 1 can be decomposed into one stack (consisting of dashed lines) and one queue (consisting of solid lines).

In the context of combinatorics, a *stack* is a noncrossing diagram, and a *queue* is a nonnesting diagram. Moreover, a contact map with vertex degree at most one corresponds to a partition in which each block contains at most two elements, which is also called a poor partition or a partial matching; see Chen et al. (2005, 2012) and Klazar (1998). A stack or a queue is said to be simple if the degree of each vertex is at most one. A stack or a queue is called m-regular if the length of each arc is at least m. Clearly, a simple stack is a noncrossing poor partition and a simple queue is a nonnesting poor partition.

Stacks and queues also arise in RNA structures. Recall that an RNA structure can be viewed as a diagram for which the degree of each vertex is at most one and the length of each arc is at least two. In combinatorial terms, an RNA structure can be viewed

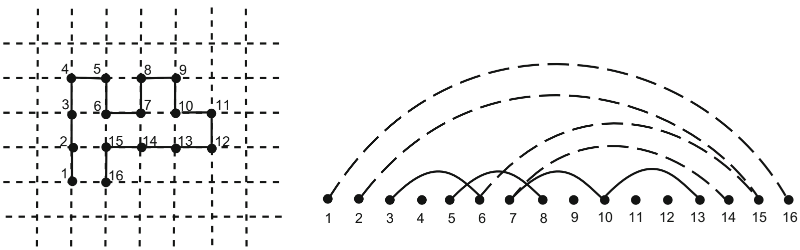

Fig. 1 A protein fold in the 2D square lattice and its contact map

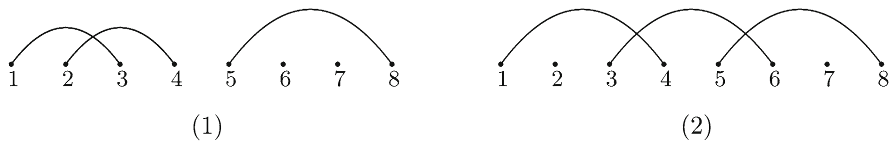

Fig. 2 A 2-regular simple queue and a 3-regular simple queue

as a 2-regular poor partition. An RNA secondary structure can be seen as a 2-regular simple stack. Note that a 2-regular simple queue can be seen as a nonnesting RNA structure and a 3-regular simple queue can be viewed as a nonnesting RNA structure in which the length of each arc is at least 3. As illustrated in Fig. 2, (1) is a 2-regular simple queue and (2) is a 3-regular simple queue.

In a recent survey paper, Istrail and Lam (2009) raised the question of finding generalizations of the Schmitt–Waterman counting formulas to stacks (in full generality) and to queues. In this paper, we obtain generating functions of 2-regular simple queues and 3-regular simple queues.

The enumeration of special classes of stacks has been extensively studied. Schmitt and Waterman (1994) found an explicit formula for the number of RNA secondary structures of length n. Došlić et al. (2004) obtained the generating function of secondary structures of rank l, which are $(l+1)$-regular simple stacks. Höner zu Siederdissen et al. (2011) introduced the notion of extended RNA secondary structures, which are 2-regular stacks with maximum degree two. Müller and Nebel (2015) obtained a functional equation satisfied by the generating function of extended RNA secondary structures by using a context-free grammar approach. Applying the reduction operation on m-regular noncrossing partitions introduced by Chen et al. (2005), Chen et al. (2014) derived a functional equation for the generating function of m-regular linear stacks, which are m-regular stacks with maximum degree 2.

Whereas a stack has no intersections of arcs, a pseudoknot requires that each bond between amino acid residues has length at least two and every two bonds intersect with each other. So a pseudoknot is a queue. Pseudoknots play an important role in a variety of biological structures; see, for example, Anderson et al. (2013), Jin and Reidys (2008), Jin et al. (2008), and Parkin et al. (1991). More precisely, a pseudoknot is called of type $k-2$ if it does not contain any set of k mutually intersecting arcs. A set of k mutually interesting arcs is also called a k-crossing.

An RNA structure that does not contain any k-crossing is called an RNA pseudoknot of type $k-2$. In other words, an RNA pseudoknot of type $k-2$ can be viewed as a 2-regular k-noncrossing poor partition. Employing the bijection between k-noncrossing partitions and oscillating tableaux due to Chen et al. (2007), Jin et al. (2008) derived an expression of the number of RNA pseudoknots of type $k-2$. In particular, when $k=2$, their result reduces to Schmitt and Waterman's formula for RNA secondary structures. When $k=3$, Jin and Reidys (2008) obtained an asymptotic formula for the number of 3-noncrossing RNA pseudoknots. Anderson et al. (2013) derived the generating function of RNA pseudoknots of genus g.

This paper is organized as follows. In Sect. 2, we show that hill-free Motzkin paths are bijective with 2-regular simple queues. We give an expression of the generating function of Motzkin paths with k_i peaks at level i in terms of a recurrence relation. It reduces to the generating function of hill-free Motzkin paths, which has been derived by Barcucci et al. (2001) by using the Enumerating Combinatorial Objects (ECO) method. The ECO method is a recursive approach to constructing a class of combinatorial structures in order to establish a functional equation of the generating function. In Sect. 3, we show that 3-regular simple queues are in one-to-one correspondence with Motzkin paths avoiding certain patterns. By using the technique of object grammars, we obtain the generating function of these Motzkin paths. The object grammars, introduced by Dutour (1996), give another recursive method for enumerating decomposable combinatorial structures; see also Duchi et al. (2004). Finally, we present the asymptotic formulas of the numbers of 2-regular simple queues and 3-regular simple queues of length n, denoted by $q_2(n)$ and $q_3(n)$, respectively. More precisely, we obtain that

$$q_2(n) \sim 0.825 \cdot 3^n \cdot n^{-\frac{3}{2}}, \tag{1}$$

$$q_3(n) \sim 0.443 \cdot 3^n \cdot n^{-\frac{3}{2}}. \tag{2}$$

2 Hill-Free Motzkin Paths and 2-Regular Simple Queues

In this section, we show that 2-regular simple queues are in one-to-one correspondence with hill-free Motzkin paths. The hill-free Motzkin paths are enumerated by sequence A089372 in the OEIS (Sloane 1964). The generating function of hill-free Motzkin paths has been obtained by Barcucci et al. (2001) using the ECO method. We present an alternative approach to computing the generating function of hill-free Motzkin paths by considering the enumeration of Motzkin paths with k_i peaks at level i.

Recall that a Motzkin path of length n is a lattice path in $\mathbb{Z}^2$ from $(0, 0)$ to $(n, 0)$ consisting of up-steps $(1, 1)$, down-steps $(1, -1)$, and horizontal steps $(1, 0)$ which never goes below the x-axis. The level or the height of a vertex (x, y) is defined to be the y-coordinate. A peak of height k of a Motzkin path is defined to be the shape consisting of an up-step of height k immediately followed by a down-step. In particular, a peak of height one is called a hill. Moreover, a Motzkin path is called hill-free if it contains no hills.

It is known that both nonnesting partial matchings and noncrossing partial matchings are in one-to-one correspondence with Motzkin paths; see Chen et al. (2005), Došlić et al. (2004) and Stanley (1999). To describe this bijection, recall that for a vertex v in a diagram, the left degree of v, denoted by $\mathrm{ldeg}(v)$, is the number of vertices u such that u is to the left of v and u is connected to v by an arc. The right degree $\mathrm{rdeg}(v)$ is defined similarly. Obviously, in a partial matching both $\mathrm{ldeg}(v)$ and $\mathrm{rdeg}(v)$ are either 0 or 1. For a given nonnesting partial matching M, its corresponding Motzkin path $\theta(M)$ is obtained by converting each vertex with right degree one into an up-step, each vertex with left degree one into a down-step, and each isolated vertex into a horizontal step. Conversely, from a Motzkin path P of length n, we can easily construct a nonnesting partial matching M by the inverse map of θ. Let $S_1, S_2, \ldots, S_n$ be the steps of P. First, draw n vertices $1, 2, \ldots, n$ from left to right on a horizontal line. Starting with the first step S_1, if S_1 is a horizontal step, then define 1 to be an isolated vertex. Otherwise S_1 is an up-step. Let S_i be the first down-step of P, and

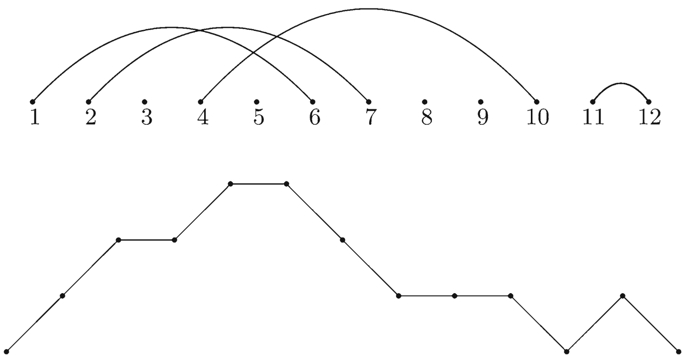

Fig. 3 A nonnesting partial matching and the corresponding Motzkin path

define $(1, i)$ to be an arc of M. Repeating the above process for the remaining steps and vertices, we are led to a nonnesting partial matching M. See Fig. 3 for an example.

Clearly, a 2-regular simple queue is a nonnesting partial matching without arcs of length one. In fact, the above bijection reduces to a one-to-one correspondence between 2-regular simple queues and hill-free Motzkin paths.

Theorem 2.1 *There is a one-to-one correspondence between* 2*-regular simple queues on* $[n]$ *and hill-free Motzkin paths of length* n.

Proof We claim that the above bijection θ between nonnesting partial matchings and Motzkin paths reduces to a bijection between the set $\mathcal{Q}_2(n)$ of 2-regular simple queues on $[n]$ and the set $\mathcal{M}_0(n)$ of hill-free Motzkin paths of length n.

Let Q be an arbitrary simple queue on $[n]$, and let $P = \theta(Q)$. We proceed to show that Q is 2-regular if and only if P is hill-free. Let $S_1, S_2, \ldots, S_n$ be the steps of P. First, assume that Q is 2-regular. Assume that $S_i S_{i+1}$ is an arbitrary peak of P; that is, S_i is an up-step and S_{i+1} is a down-step. By the correspondence θ, the vertex i has right degree one and the vertex $i + 1$ has left degree one in Q. Since Q is 2-regular, $(i, i + 1)$ is not an arc of Q. Hence there exist two vertices k and l such that $(k, i + 1)$ and (i, l) are arcs of Q, where $k < i < i + 1 < l$. Under the correspondence θ, we see that the peak $S_i S_{i+1}$ has height at least two. Thus, P is hill-free.

Conversely, assume that P is hill-free. We aim to show that Q is 2-regular. Suppose to the contrary that Q contains an arc of length one, say $(i, i+1)$. By the correspondence θ, we know that Q is a nonnesting partial matching. Clearly, $(i, i+1)$ cannot be nested by any arc of Q; that is, there does not exist an arc (u, v) of Q such that $u \in [1, i-1]$ and $v \in [i+2, n]$. Consider the partial matching on $[1, i-1]$. By the correspondence θ, the steps $S_1, S_2, \ldots, S_{i-1}$ form a Motzkin path on $[i - 1]$. Hence P returns to the x-axis at the $(i - 1)$-st step. This implies that $(i, i + 1)$ corresponds to a hill of P, which is a contradiction to the assumption that P is hill-free. Hence Q is 2-regular. This completes the proof. □

From the above theorem, we see that the enumeration of 2-regular simple queues can be achieved by the enumeration of hill-free Motzkin paths. In general, we obtain a recurrence relation for the generating function of the number of Motzkin paths of length n with k_i peaks at level i, which leads to an expression in terms of a continued fraction.

Let $\mathcal{M}$ be the set of all the Motzkin paths. For a given Motzkin path $P \in \mathcal{M}$, let $l(P)$ be the length of P, let $k(P)$ be the number of the peaks of P, and let $k_i = k_i(P)$ be the number of the peaks of P at level i. Define the type of P, denoted by $t(P)$, to be the vector $(k_1, k_2, \ldots)$.

Let m_n denote the number of Motzkin paths of length n, which is known as Motzkin number. Let

$$M(x) = \sum_{n\geq 0} m_n x^n.$$

Donaghey and Shapiro (1977, Eq. 7) derived that

$$M(x) = \frac{1 - x - \sqrt{1 - 2x - 3x^2}}{2x^2}. \tag{3}$$

Let $c(n; k_1, k_2, \ldots)$ denote the number of Motzkin paths of length n and type $(k_1, k_2, \cdots)$, and let

$$M(x; t_1, t_2, \ldots) = \sum_{n,k_1,k_2,\ldots\geq 0} c(n; k_1, k_2, \ldots)x^n \prod_{i\geq 1} t_i^{k_i}. \tag{4}$$

The following theorem gives a relation satisfied by $M(x; t_1, t_2, \ldots)$, which recursively leads to an expression in terms of a continued fraction.

Theorem 2.2 *We have*

$$M(x; t_1, t_2, \ldots) = \frac{1}{1 - x - (t_1 - 1)x^2 - x^2 M(x; t_2, t_3, \ldots)}. \tag{5}$$

Proof Recall that a Motzkin path is prime if it starts from the x-axis, ends at the x-axis, and never goes back to the x-axis in the middle steps. Clearly, a horizontal step at level 0 and a hill are prime Motzkin paths.

Suppose that P is a Motzkin path of length n and of type $(k_1, k_2, \ldots)$. It is obvious that P can be decomposed into a sequence of prime Motzkin paths. Assume that P can be decomposed into k_1 hills, r nonhill prime Motzkin paths of height at least one, and s horizontal steps at level 0.

We next give a recursive procedure to construct a Motzkin path of type $(k_1, k_2, \ldots)$ from Motzkin paths of lower heights. For a Motzkin path P, we define the map η as follows. First, remove all the steps of P under the line $y = 1$. Then join the Motzkin paths on or above the line $y = 1$ to build a shorter Motzkin path of lower height. Figure 4 gives an illustration of η.

Let Q be the path $\eta(P)$. If $r = 0$, then all the steps of P lie below the line $y = 1$. It can be easily seen that in this case Q is empty and $n = 2k_1 + s$. Moreover, the number of such Motzkin paths is equal to the number of the permutations of k_1 hills and s horizontal steps on the x-axis. In other words, the number of such Motzkin paths is

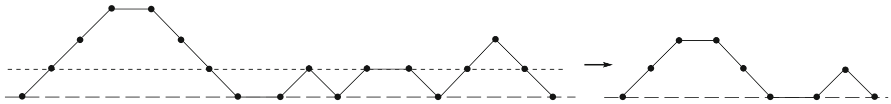

Fig. 4 Map η

$$\binom{k_1 + s}{s}. \tag{6}$$

If $r \geq 1$, then Q consists of a sequence $P_1, P_2, \ldots, P_r$ of Motzkin paths which are constructed by the steps of P on and above the line $y = 1$. Below the line $y = 1$, P contains k_1 hills, r pairs of one up-step and one down-step, and s horizontal steps at level 0. The number of permutations of these steps equals

$$\binom{k_1 + r + s}{k_1, r, s}.$$

For $1 \leq i \leq r$, let $l(P_i) = n_i$ and let $k_{i,j}$ denote the number of peaks of P_i at level j. Clearly, $n_i \geq 1$ for any i and $n_1 + n_2 + \cdots + n_r = n - s - 2(r + k_1)$. Moreover, $k_{1,j} + k_{2,j} + \cdots + k_{r,j} = k_{j+1}$. Note that P_i is of type $(k_{i,1}, k_{i,2}, \ldots)$. The number of such paths equals $c(n_i; k_{i,1}, k_{i,2}, \ldots)$. Hence the number of such sequences $(P_1, P_2, \ldots, P_r)$ is

$$\prod c(n_i; k_{i,1}, k_{i,2}, \ldots),$$

where the product ranges over $n_i \geq 1$, $k_{i,j} \geq 0$ such that

$$\sum_{i=1}^{r} n_i = n - s - 2(r + k_1)$$

and

$$\sum_{i=1}^{r} k_{i,j} = k_{j+1}.$$

It follows that

$$c(n; k_1, k_2, \ldots) = \sum_{r \geq 0, s \geq 0} \binom{k_1 + r + s}{k_1, r, s} \prod_{\substack{n_i \geq 1 \\ \sum_{i=1}^{r} n_i = n - s - 2(r + k_1) \\ \sum_{i=1}^{r} k_{i,1} = k_2 \\ \sum_{i=1}^{r} k_{i,2} = k_3 \\ \cdots}} c(n_i; k_{i,1}, k_{i,2}, \ldots). \tag{7}$$

Note that when $r = 0$, (7) reduces to (6).

Substituting (7) into (4), and using the following binomial identity

$$\sum_{k\geq 0}\binom{m+k}{k}x^k = \frac{1}{(1-x)^{m+1}},$$

we are led to

$$\begin{aligned}
M(x,t_1,t_2,\ldots) &= \sum_{n,k_1,k_2,\ldots\geq 0} c(n;k_1,k_2,\ldots)x^n \prod_{i\geq 1} t_i^{k_i} \\
&= \sum_{n,k_1,k_2,\ldots\geq 0}\sum_{r,s\geq 0}\binom{k_1+r+s}{k_1,r,s} \prod_{\substack{n_i\geq 1\\ \sum_{i=1}^r n_i = n-s-2(r+k_1)\\ \sum_{i=1}^r k_{i,j}=k_{j+1}}} c(n_i;k_{i,1},k_{i,2},\ldots)x^n \prod_{i\geq 1} t_i^{k_i} \\
&= \sum_{r,s,k_1\geq 0}\binom{k_1+r+s}{k_1,r,s} t_1^{k_1}x^s(x^2)^{r+k_1}\prod_{i=1}^{r}\sum_{n_i\geq 1,k_{i,1},k_{i,2},\ldots\geq 0} c(n_i;k_{i,1},k_{i,2},\ldots)x^{n_i} \\
&\quad \prod_{j\geq 1} t_{j+1}^{k_{i,j}} \\
&= \sum_{r,s,k_1\geq 0}\binom{k_1+r+s}{k_1,r,s} t_1^{k_1}x^s(x^2)^{r+k_1}\Big(M(x,t_2,t_3,\cdots)-1\Big)^r \\
&= \frac{1}{1-x-(t_1-1)x^2-x^2M(x,t_2,t_3,\ldots)}.
\end{aligned}$$

These complete the proof. □

Taking special values of $t_1, t_2, \ldots$, Eq. (5) reduces to some known results. Let $c(n,k)$ denote the number of Motzkin paths of length n with k peaks and let

$$M(x;t) = \sum_{n,k\geq 0} c(n,k)x^n t^k.$$

Setting $t_1 = t_2 = t_3 = \cdots = t$ in (5), we obtain that

$$M(x;t) = \frac{1}{1-x-(t-1)x^2-x^2M(x;t)}. \tag{8}$$

Furthermore, taking $t = 1$ in (8) and solving the resulting equation, we are led to the generating function (3) of Motzkin paths.

Setting $t = 0$ in (8), we get the generating function of Motzkin paths without peaks

$$F(x) = \frac{1-x+x^2-\sqrt{(1+x+x^2)(1-3x+x^2)}}{2x^2},$$

which can also be derived from Eq. (3) in Došlić et al. (2004).

Fixing $t_1 = 0$ and $t_2 = t_3 = \cdots = 1$ in (5), we obtain an equation for the generating function of hill-free Motzkin paths, which implies the following equation of the generating function of 2-regular simple queues

$$Q_2(x) = \frac{1}{1 - x + x^2 - x^2 M(x)}.$$

Substituting (3) into the above equation, we obtain

$$Q_2(x) = \frac{1 - x + 2x^2 - \sqrt{1 - 2x - 3x^2}}{2x^2(2 - x + x^2)}. \tag{9}$$

This generating function has been obtained by Barcucci et al. (2001) by using the ECO method.

Next, we give the generating functions of several classes of Motzkin paths based on (5). As the first example, let $f_m(n)$ denote the number of Motzkin paths of length n without peaks of height less than $m + 1$ and denote the generating function by

$$F_m(x) = \sum_{n \geq 0} f_m(n) x^n.$$

Setting $t_1 = t_2 = \cdots = t_m = 0$ and $t_{m+1} = t_{m+2} = \cdots = 1$ in (5), we obtain that

$$F_m(x) = \frac{1}{1 - x + x^2 - x^2 F_{m-1}(x)},$$

and $F_0(x) = M(x)$.

Let $g_m(n)$ denote the number of Motzkin paths of length n without peaks at level m and denote

$$G_m(x) = \sum_{n \geq 0} g_m(n) x^n.$$

Substituting $t_m = 0$ and $t_1 = t_2 = \cdots = t_{m-1} = t_{m+1} = t_{m+2} = \cdots = 1$ in (5), we see that $G_m(x)$ satisfies the following relation

$$G_m(x) = \frac{1}{1 - x - x^2 G_{m-1}(x)},$$

and $G_0(x) = M(x)$.

Let $h_m(n)$ be the number of Motzkin paths of length n with peaks only at level m and denote

$$H_m(x) = \sum_{n \geq 0} h_m(n) x^n.$$

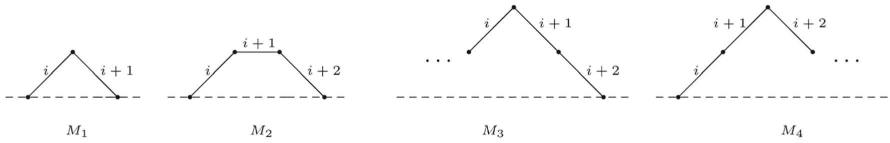

Fig. 5 Four patterns in Motzkin paths

Setting $t_m = 1$ and $t_1 = t_2 = \cdots = t_{m-1} = t_{m+1} = t_{m+2} = \cdots = 0$ in (5), we obtain that

$$H_m(x) = \frac{1}{1 - x + x^2 - x^2 H_{m-1}(x)},$$

and $H_0(x) = F(x)$.

Finally, we consider Motzkin paths of length n of type $t = (k_1, k_2, \ldots, k_\ell)$ which have height at most ℓ. Denote the generating function of such Motzkin paths by $M_\ell(x; t_1, t_2, \ldots, t_\ell)$. From (5), we obtain that

$$M_\ell(x; t_1, t_2, \ldots, t_\ell) = \frac{1}{1 - x - (t_1 - 1)x^2 - x^2 M_{\ell-1}(x; t_2, t_3, \ldots, t_\ell)},$$

where $M_0(x)$ is the generating function of Motzkin paths of height zero, namely

$$M_0(x) = \frac{1}{1 - x}.$$

3 Motzkin Paths Avoiding Patterns and 3-Regular Simple Queues

In this section, we show that the 3-regular simple queues are in one-to-one correspondence with Motzkin paths avoiding patterns M_1, M_2, M_3, M_4 illustrated in Fig. 5. By applying the technique of object grammars, we obtain the generating function of Motzkin paths avoiding patterns M_1, M_2, M_3, M_4, from which we obtain a recurrence relation for the number of 3-regular simple queues related to Motzkin numbers.

Theorem 3.1 *There is a one-to-one correspondence between* 3*-regular simple queues and Motzkin paths avoiding patterns* M_1, M_2, M_3, M_4 *in Fig.* 5.

Proof Let $\mathcal{Q}_3(n)$ be the set of 3-regular simple queues and $\mathcal{M}_a(n)$ be the set of Motzkin paths avoiding patterns M_1, M_2, M_3, M_4 in Fig. 5. We claim that the correspondence θ given in Sect. 2 also reduces to a bijection between $\mathcal{Q}_3(n)$ and $\mathcal{M}_a(n)$.

It is obvious that any 3-regular simple queue in $\mathcal{Q}_3(n)$ contains no arc of length one or two. If a simple queue contains an arc of length one, then it corresponds to pattern Q_1 in Fig. 6. If a simple queue contains an arc of length two, say $(i, i+2)$, then the vertex $i+1$ is either an isolated vertex, or a vertex having left degree one or right degree one. These three cases correspond to the patterns Q_2, Q_3, and Q_4 in Fig. 6, respectively. In summary, a simple queue is 3-regular if and only if it avoids patterns Q_1, Q_2, Q_3, and Q_4.

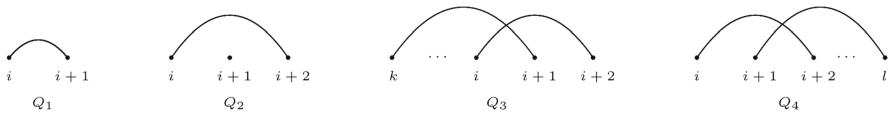

Fig. 6 Four patterns in simple queues

Let $Q \in \mathcal{Q}_3(n)$ and $P = \theta(Q)$. Let $S_1, S_2, \ldots, S_n$ be the steps of P. Thus, to prove the theorem, it is sufficient to show that for each $1 \leq i \leq 4$, Q avoids the pattern Q_i if and only if P avoids pattern M_i.

First, from the proof of Theorem 2.1, we see that Q contains no arc of length one if and only if P is hill-free, or equivalently, Q avoids the pattern Q_1 if and only if P avoids the pattern M_1.

Next, to prove the equivalence of the pattern avoidance with respect to Q_2 and M_2, we first suppose that P contains a subsequence $S_i S_{i+1} S_{i+2}$ that is of pattern M_2. Then $S_i S_{i+1} S_{i+2}$ starts from and ends at the x-axis. Applying the map θ, we see that $(i, i+2)$ is an arc of Q and $i+1$ is an isolated vertex of Q. Hence the subqueue on the vertices i, $i+1$ and $i+2$ is of the pattern Q_2. Conversely, suppose that Q contains the pattern Q_2. Obviously, the arc $(i, i+2)$ cannot be nested by any arc of Q; that is, there does not exist an arc (u, v) of Q such that $u \in [1, i-1]$ and $v \in [i+3, n]$. Consider the partial matching on $[1, i-1]$. By the correspondence θ, the steps $S_1, S_2, \ldots, S_{i-1}$ form a Motzkin path on $[i-1]$. Hence P returns to the x-axis at the $(i-1)$-st step. This implies that $S_i S_{i+1} S_{i+2}$ is of pattern M_2. This proves the equivalence of the pattern avoidance with respect to Q_2 and M_2.

Now we claim that Q avoids pattern Q_3 if and only if P avoids pattern M_3. First, suppose that P contains a subsequence $S_i S_{i+1} S_{i+2}$ that is of pattern M_3. Noting that P returns to the x-axis at the step S_{i+2}, then under θ we see that S_{i+2} should be paired with S_i; thus, $(i, i+2)$ is an arc of Q. Furthermore, since S_{i+1} is a down-step, it must be paired with an up-step lying on the left of S_i. Thus, the vertex $i+1$ has left degree one in Q. Therefore, the subqueue on the vertices $i, i+1, i+2$ in Q is of pattern Q_3.

On the other hand, suppose that Q contains a subqueue on the vertices $i, i+1$ and $i+2$ of pattern Q_3. Then under θ, the steps S_i, S_{i+1}, and S_{i+2} form an up-step followed by two down-steps. To show that $S_i S_{i+1} S_{i+2}$ is of pattern M_3, it is necessary to verify that P returns to the x-axis at the step S_{i+2}. Otherwise, if the down-step S_{i+2} is of height h with $h \geq 2$, then among the steps $S_1, S_2, \ldots, S_{i+2}$, there are more up-steps than down-steps. It implies that there must exist an up-step S_j on the left of S_{i+2} paired with a down-step S_k on the right of S_{i+2}, where $j \leq i < i+2 < k$. Under θ, the steps S_j and S_k are mapped to an arc (j, k) of Q. If $j = i$, then (i, k) and $(i, i+2)$ are arcs of Q, which is impossible since Q is simple. If $j < i$, then the arc $(i, i+2)$ is nested by the arc (j, k), contradicting the assumption that Q is nonnesting. Hence the height of S_{i+2} must be $h = 1$. This shows that $S_i S_{i+1} S_{i+2}$ is of pattern M_3, which completes the proof of the equivalence of the pattern avoidance with respect to Q_3 and M_3.

Similarly, one can prove the equivalence of the pattern avoidance involved with Q_4 and M_4. This completes the proof of the theorem. □

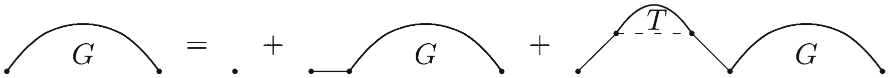

Fig. 7 Classification of Motzkin paths avoiding patterns M_1, M_2, M_3, and M_4

By applying the technique of object grammars to the Motzkin path avoiding patterns M_1, M_2, M_3, and M_4, we obtain the generating function of the 3-regular simple queues.

Theorem 3.2 *The generating function of* 3*-regular simple queues equals*

$$Q_3(x) = \frac{x^5 - 3x^4 + 4x^2 - x + 1 - (x-1)^2(x+1)^2\sqrt{1-2x-3x^2}}{2x^2(x^8 - x^7 - x^6 - x^4 + 5x^3 - 2x^2 - 4x + 4)}. \tag{10}$$

Proof By Theorem 3.1, we see that $Q_3(x)$ can be determined by the generating function of Motzkin paths avoiding patterns M_1, M_2, M_3, and M_4. Let g_n be the number of Motzkin paths in $\mathcal{M}_a(n)$ and let $G(x)$ be the generating function of g_n. Denote by $\mathcal{G}$ the set of all the Motzkin paths avoiding patterns M_1, M_2, M_3, and M_4. In view of the first steps of Motzkin paths, $\mathcal{G}$ consists of three disjoint subsets: the set of a singleton and the following two subsets

$$\mathcal{G}_1 := \{w \in \mathcal{G}\colon \text{the first step is horizontal}\},$$
$$\mathcal{G}_2 := \{w \in \mathcal{G}\colon \text{the first step is an up-step}\}.$$

This decomposition is demonstrated in Fig. 7, where G stands for an arbitrary Motzkin path avoiding patterns M_1, M_2, M_3, M_4 and T stands for the Motzkin paths on and above the line $y = 1$ in the first prime component of $G \in \mathcal{G}_2$.

It is easy to see that T cannot be any of the following structures:

(1) a singleton;
(2) a horizontal step;
(3) a hill;
(4) a Motzkin path preceded or followed by a hill.

Denote the generating functions of the Motzkin paths in $\mathcal{G}_1$ and $\mathcal{G}_2$ by $G_1(x)$ and $G_2(x)$, respectively. Let $T(x)$ denote the generating function of Motzkin paths in T. Clearly,

$$G(x) = 1 + G_1(x) + G_2(x), \tag{11}$$

where

$$G_1(x) = xG(x)$$

and

$$G_2(x) = x^2T(x)G(x).$$

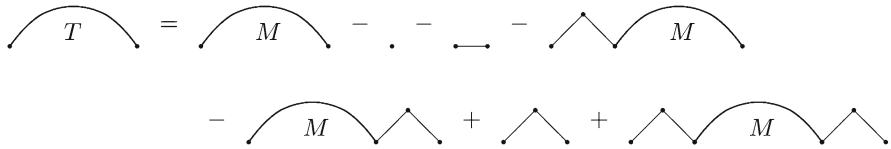

Fig. 8 Structure of a Motzkin path of type T

Employing the principle of inclusion and exclusion, T can be enumerated as shown in Fig. 8, where M stands for an arbitrary Motzkin path.

Thus, we have

$$T(x) = M(x) - 1 - x - 2x^2M(x) + x^2 + x^4M(x)$$
$$= (1 - 2x^2 + x^4)M(x) - 1 - x + x^2. \quad (12)$$

Substituting (3) and (12) into (11), we are led to

$$G(x) = \frac{x^5 - 3x^4 + 4x^2 - x + 1 - (x-1)^2(x+1)^2\sqrt{1 - 2x - 3x^2}}{2x^2(x^8 - x^7 - x^6 - x^4 + 5x^3 - 2x^2 - 4x + 4)},$$

which completes the proof since $Q_3(x) = G(x)$. □

The first few values of $q_3(n)$ are given as follows.

n	0	1	2	3	4	5	6	7	8
$q_3(n)$	1	1	1	1	2	5	14	37	96
n	9	10	11	12	13	14	15	16	17
$q_3(n)$	249	653	1732	4640	12,532	34,080	93,231	256,395	708,445

Note that the sequence $q_3(n)$ does not appear in OEIS (Sloane 1964).

Furthermore, employing (3) to eliminate the square root in (10), we obtain a relation between $Q_3(x)$ and $M(x)$, which leads to a recurrence relation of $q_3(n)$

$$4q_3(n) - 4q_3(n-1) - 2q_3(n-2) + 5q_3(n-3) - q_3(n-4)$$
$$- q_3(n-6) - q_3(n-7) + q_3(n-8) = m_n - 2m_{n-2} + m_{n-4},$$

where m_n is the Montzkin number.

We conclude this paper with asymptotic formulas of $q_2(n)$ and $q_3(n)$, which can be derived by using the method of singularity analysis (Flajolet and Odlyzko 1990) to the generating functions $Q_2(x)$ and $Q_3(x)$.

Note that $x = \frac{1}{3}$ is the singularity closest to the origin for both $Q_2(x)$ and $Q_3(x)$. Using the Taylor theorem to approximate $Q_2(x)$ and $Q_3(x)$ at $x = \frac{1}{3}$, we have

$$Q_2(x) \sim \frac{9}{4} - \frac{27\sqrt{3}}{16}\sqrt{1-3x} + o(1-3x)^{1/2},$$
$$Q_3(x) \sim \frac{243}{131} - \frac{3064\sqrt{3}}{3381}\sqrt{1-3x} + o(1-3x)^{1/2}.$$

By using the asymptotic expansion given by Flajolet and Odlyzko (1990, Eq. 2.3)

$$[x^n](1-x)^{-\alpha} \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)},$$

where α cannot be zero or negative integers, we obtain that

$$q_2(n) \sim \frac{27}{32\sqrt{\pi/3}} \cdot 3^n \cdot n^{-\frac{3}{2}},$$
$$q_3(n) \sim \frac{1532}{3381\sqrt{\pi/3}} \cdot 3^n \cdot n^{-\frac{3}{2}}.$$

So, approximately we get (1) and (2).

Acknowledgements This work was supported by the 863 Project of the Ministry of Science and Technology, the 973 Project and the PCSIRT Project of the Ministry of Education, the National Science Foundation of China, and the Natural Science Foundation of Tianjin, China.

References

Agarwal PK, Mustafa NH, Wang Y (2007) Fast molecular shape matching using contact maps. J Comput Biol 14(2):131–143

Anderson JE, Penner RC, Reidys CM, Waterman MS (2013) Topological classification and enumeration of RNA structures by genus. J Math Biol 67:1261–1278

Barcucci E, Pergola E, Pinzani R, Rinaldi S (2001) ECO method and hill-free generalized Motzkin paths. Sem Loth Comb 46:B46b

Chen WYC, Deng EYP, Du RRX (2005) Reduction of m-regular noncrossing partitions. Eur J Comb 26:237–243

Chen WYC, Deng EYP, Du RRX, Stanley RP, Yan CH (2007) Crossings and nestings of matchings and partitions. Trans Am Math Soc 359:1555–1575

Chen WYC, Guo Q-H, Sun LH, Wang J (2014) Zigzag stacks and m-regular linear stacks. J Comput Biol 21(12):915–935

Chen WYC, Fan NJY, Zhao AFY (2012) Partitions and partial matchings avoiding neighbor patterns. Eur J Comb 33:491–504

Domany E (2000) Protein folding in contact map space. Phys A 288:1–9

Donaghey R, Shapiro LW (1977) Motzkin numbers. J Comb Theory Ser A 23:291–301

Došlić T, Svrtan D, Veljan D (2004) Enumerative aspects of secondary structures. Discrete Math 285:67–82

Duchi E, Fédou J-M, Rinaldi S (2004) From object grammars to ECO systems. Theor Comput Sci 314:57–95

Dutour I (1996) Grammaires d'objets: énumération, bijections et génération aléatoire. PhD thesis, Université Bordeaux I, France

Flajolet P, Odlyzko AM (1990) Singularity analysis of generating functions. SIAM J Discrete Math 3(2):216–240

Goldman D, Istrail S, Papadimitriou CH (1999) Algorithmic aspects of protein structure similarity. In: Proceedings of 40th IEEE symposium on foundations of computer science, pp 512–522

Höner zu Siederdissen C, Bernhart SH, Stadler PF, Hofacker IL (2011) A folding algorithm for extended RNA secondary structures. Bioinformatics 27(13):129–136

Istrail S, Lam F (2009) Combinatorial algorithms for protein folding in lattice models: a survey of mathematical results. Commun Inf Syst 9(4):303–346

Jin EY, Qin J, Reidys CM (2008) Combinatorics of RNA structures with pseudoknots. Bull Math Biol 70:45–67

Jin EY, Reidys CM (2008) Asymptotic enumeration of RNA structures with pseudoknots. Bull Math Biol 70:951–970

Klazar M (1998) On trees and noncrossing partitions. Discrete Appl Math 82:263–269

Lancia G, Carr R, Walenz B, Istrail S (2001) 101 optimal PDB structure alignments: a branch-and-cut algorithm for the maximum contact map overlap problem. In: Proceedings of the 5th annual international conference on computational biology, pp 193–202

Müller R, Nebel ME (2015) Combinatorics of RNA secondary structures with base triples. J Comput Biol 22(7):619–648

Parkin N, Chamorro M, Varmus HE (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

Schmitt WR, Waterman M (1994) Linear trees and RNA secondary structure. Discrete Appl Math 51:317–323

Sloane NJA (1964) On-line encyclopedia of integer sequences. http://oeis.org/

Stanley RP (1999) Enumerative combinatorics, vol 2. Cambridge University Press, Cambridge

Vendruscolo M, Kussell E, Domany E (1997) Recovery of protein structure from contact maps. Fold Des 2(5):295–306