On the number of saturated and optimal extended 2-regular simple stacks in the Nussinov-Jacobson energy model

Qianghui Guo[1*], Yinglie Jin[1], Mengqin Li[1], Lisa Hui Sun[2], and Yanyan Xu[1]

[1]School of Mathematical Sciences, LPMC, Nankai University,

Tianjin 300071, P. R. China

[2]Center for Combinatorics, LPMC, Nankai University,

Tianjin 300071, P. R. China

[*]To whom correspondence should be addressed;

E-mail: guo@nankai.edu.cn.

March 19, 2022

Abstract: It is known that both RNA secondary structure and protein contact map can be presented using combinatorial diagrams, the combinatorial enumeration and related problems of which have been studied extensively. Motivated by previous enumeration works on saturated RNA secondary structures and extended stack structures of protein contact maps, we are interested in the enumeration

1

problems of saturated and optimal extended stacks in the Nussinov-Jacobson energy model, in which each base pair contributes energy -1. Then optimal structures are those with most arcs, and locally optimal structures are exactly the saturated structures, in which no more arc can be added without violating the structure definition. For saturated extended 2-regular simple stacks, whose degree configuration is related to the protein fold in 2D honeycomb lattice, we obtain generating function equation and asymptotic formula for its number. Moreover, an explicit formula for the number of optimal extended 2-regular simple stacks is also obtained.

1   Introduction

It is well-known that the function of RNA depends on its 3D structure and dynamics, and RNA 3D structure is largely determined by the secondary structure (Banerjee et al. , 1993). The case of protein is analogous, and the contact map of protein can be seen as a counterpart of the secondary structure of RNA. Contact plays a fundamental role in the well-known Hydrophobic-Hydrophilic (HP) model for protein folding (Dill , 1990), and recently, contact map also plays an extremely important role in the high accuracy prediction of protein structure through deep learning method (Xu , 2019; Yang et al. , 2020).

Due to the pioneer work of Waterman (Schmitt and Waterman , 1978), combinatorial theory and method were introduced into the field of computational molecular biology, and the combinatorial enumeration of various RNA secondary structures and protein contact maps have attracted extensive study since then. Schmitt and Waterman (1994) provided an explicit formula for the number of RNA secondary structures on n vertices and k arcs by establishing a bijection between RNA secondary structures and linear trees. Nebel (2002) de-

rived the generating function of m-regular RNA secondary structures by using the binary trees and the Horton-Strahler number. Clote (2006) obtained recurrence relations and asymptotic formulas for combinatorial problems related to the number of saturated secondary structures. Jin et al. (2008) derived the recursion formulas of 3-noncrossing RNA structures. A collection of new combinatorial and computational approaches in the study of RNA structures with pseudoknots was presented in a monograph (Reidys , 2011).

Recently, enumerative combinatorics has also made progress in the study of protein contact map. Based on the classic HP model (Dill , 1990), protein fold can be considered as a self-avoiding walk in some lattice model. Lattice models usually retain important features of protein structures, and enable us to focus on dominant aspects of the structure. Goldman et al. (1999) showed that for any protein fold in 2D square lattice self-avoiding walk HP model, the contact map can be decomposed into (at most) two stacks and one queue. Actually, this decomposition also holds for the case of 2D honeycomb lattice, which will be shown in Section 2. Agarwal et al. (2007) found a similar decomposition result for the contact maps of protein folds in the 3D cubic lattice self-avoiding walk HP model.

In combinatorial terms, a *stack* is a noncrossing diagram, a *queue* is a nonnesting diagram, they are the two elementary structures of protein contact map. When folding a protein in the lattice model, different lattice models will introduce different degree and arc length constraints to the corresponding contact map. For instance, on the 2D square lattice, each internal vertex in the contact map has maximum degree 2, while the the two terminal vertices can have maximum degree 3, and the arc length is at least 3. On the 2D honeycomb lattice, the degree of each internal vertex and terminal vertex is at most 1 and 2, respectively, while the length of each arc is at least 5. For a survey of various

lattice models used for protein folding, see (Pierri et al. , 2008). A stack with arc length at least m is called m-regular, and a stack with degree of each vertex bounded by one is called simple. Clearly, RNA secondary structures can be viewed as 2-regular simple stacks.

Istrail and Lam (2009) proposed the question concerning generalizations of the Schmitt-Waterman counting formulas for RNA secondary structures (Schmitt and Waterman , 1994) to enumerating protein stacks and queues, and they pointed out that the enumeration of stacks and of queues could provide insights into computing rigorous approximations of the partition function of protein folding in HP models.

To attack this question, Chen et al. (2014) proposed the primary decomposition method to study the combinatorial enumeration of m-regular (with arc length at least m) linear stack (with the degree of each vertex bounded by two), which is a generalization of the classic RNA secondary structure, and got enumeration results in the form of generating function equation, recurrence relation and asymptotic formula. Furthermore, combinatorial enumeration results were obtained for extended (with the degree of two terminal vertices bounded by three) m-regular linear stacks (Guo et al. , 2016), m-regular linear stacks with n vertices and k arcs (Guo and Sun , 2018), 2-regular and 3-regular simple (with the degree of each vertex bounded by one) queues (Guo et al. , 2017).

The RNA and protein folding problems can be formulated as energy minimization problems in some energy model, for example, in the classic Nussinov-Jacobson free energy model (Nussinov and Jacobson , 1980), the energy function is the negative of the number of base pairs (for RNA) or contacts (for protein) and the structures with minimum energy are called optimal. Another concept closely related to optimal structure is *saturated* structure, which means no base pairs (contacts) can be added without violating the definition of the structure

(Clote , 2006). Saturated structures are actually local minima in the energy landscape. The combinatorial problems related to the number of saturated RNA secondary structures have attracted much interests, e.g. Clote (2006); Clote et al. (2009, 2007); Fusy and Clote (2014).

Stimulated by the enumeration works on saturated and optimal RNA secondary structures and extended stacks of protein contact map, we are interested in the problem of enumerating saturated and optimal extended 2-regular simple stacks, in which the degree of two terminal vertices bounded by 2, and the degree of the internal vertices bounded by 1. This degree configuration is related to the protein fold in the 2D honeycomb lattice. The honeycomb lattice alleviates the " sharp turn " problem and models certain aspects of the protein secondary structure more realistically with reduced combinatorial complexity (Guo et al. , 2018; Jiang and Zhu , 2005). Denote the number and the generating function of all saturated extended 2-regular simple stacks with n vertices by $s(n)$ and $S(x)$, respectively. Based on the techniques of Clote (2006) and the idea of the structure decomposition as given in Chen et al. (2014), we obtain the following equation satisfied by $S(x)$.

Theorem 1. *We have*

$$p_3(x)S^3(x) + p_2(x)S^2(x) + p_1(x)S(x) + p_0(x) = 0. \tag{1}$$

where

$$p_0(x) = -x^3(x^{17} + 4x^{16} - 9x^{15} - 15x^{14} - 17x^{13} + 13x^{12} + 34x^{11} + 65x^{10} + 121x^9$$
$$+ 99x^8 + 117x^7 + 110x^6 + 84x^5 + 70x^4 + 30x^3 + 16x^2 + 3x + 1),$$

$$p_1(x) = x^{17} + 7x^{16} + x^{15} - x^{14} - 25x^{13} - 62x^{12} - 93x^{11} - 105x^{10} - 95x^9 - 92x^8$$
$$- 53x^7 - 52x^6 - 40x^5 - 11x^4 - 3x^3 + 7x^2 + x + 1,$$

$$p_2(x) = -x^4(2x^9 + 7x^8 + 18x^7 + 25x^6 + 29x^5 + 33x^4 + 22x^3 + 7x^2 + 2x + 2),$$

$$p_3(x) = x^8(x + 1).$$

Based on equation (1), we furthermore obtain the following asymptotic formulas for $s(n)$ through the singularity analysis method.

$$s(n) \sim 2.818464011 \times 2.354673605^n \cdot n^{-\frac{3}{2}}.$$

The explicit expression for $so(n)$, the number of optimal extended 2-regular simple stacks with n vertices, is also obtained.

Theorem 2. *We have for $n \geq 5$,*

$$so(n) = \begin{cases} \dfrac{1}{12}(n^3 - 3n^2 - 7n + 69), & if \ n \ is \ odd, \\ n - 3, & if \ n \ is \ even. \end{cases} \tag{2}$$

This paper is organized as follows. In Section 2, we give basic definitions and notations, and some previous results which are the foundation of this paper. In Section 3, we study the combinatorial enumeration problem of the saturated extended 2-regular simple stacks. At last, Section 4 focuses on the number of optimal extended 2-regular simple stacks.

2   Basic Definitions and Notations

First, we recall the definition of RNA secondary structure and protein contact map in a combinatorial way. A secondary structure on RNA sequence $r = r_1 r_2 \cdots r_n$ is a set of ordered base pairs (i, j) with three properties: (i) Each vertex can be bonded to at most one other vertex; (ii) If r_i and r_j are bonded, then $j - i \geq m$, where m means the minimum length of two bonded vertices;

(iii) If r_i and r_j are bonded, then any bonding of r_k $(i < k < j)$ must be with a vertex between r_i and r_j.

For the case of protein, when two nonconsecutive amino acids residues in a protein fold come very close to each other, say, closer than a predetermined threshold, they presumably form some kind of bond, which is called a contact. A protein contact map on protein sequence $p = p_1 p_2 \cdots p_n$ is a set of ordered contact pairs (i, j), where p_i and p_j form a contact in the 3D structure of p.

In combinatorics, both RNA secondary structure and protein contact map can be presented by a diagram, i.e., drawing vertices $1, 2, \ldots, n$ on a horizontal line in increasing order and connecting two vertices by an arc if they are bonded (for RNA) or in contact (for protein). In the diagram presentation, for any two arcs (i, j) and (k, l), if $i < k < l < j$, we say that they form a nesting; if $i < k < j < l$, we say that they form a crossing. A noncrossing diagram is called a stack, and a nonnesting diagram is called a queue.

Stack and queue are the two elementary structures of protein contact map. Following Chen et al. (2014); Guo et al. (2017), a structure (stack or queue) with arc length at least m is called m-regular, a structure with degree of each vertex bounded by one is called simple, a structure with degree of each vertex bounded by two is called linear. Furthermore, an m-regular simple stack with the degrees of the two terminal vertices bounded by two is called an extended m-regular simple stack. Note that this degree configuration is related to the protein contact map in the 2D honeycomb lattice.

Recall that the contact map of any protein fold in 2D square lattice can be decomposed into (at most) two stacks and one queue (Goldman et al. , 1999). We claim that the same decomposition also holds in the 2D honeycomb lattice.

Theorem 3. *Any self-avoiding walk in the two-dimensional honeycomb lattice can be decomposed into (at most) 2 stacks and 1 queue.*

Proof. For each vertex in the walk, we assign a label O or U to its adjacent edges in the lattice which are not edges in the walk. Edges in the lattice will then have multisets of labels consisting of 0, 1, or 2 members. Labels are assigned inductively as follows:

- Label non-walk edges adjacent to vertex 1 as follows: assign O to one of the edges and assign U to the other.

- Label non-walk edges adjacent to vertex i, where $2 \leq i \leq n$, as follows: if the walk edge adjacent to vertex i lie in the same honeycomb lattice as an edge labelled by vertex $i - 1$ with label L, assign label L to it, otherwise, assign label $\{O, U\} \backslash \{L\}$ to it.

After the labelling procedure, the edges in the contact map will be assigned exactly two labels. Following the similar arguments of Goldman et al. (1999), the graph consisting of edges labelled by $\{O, O\}$ (or $\{U, U\}$) is a stack, and the graph consisting of edges labelled by $\{O, U\}$ is a queue. $\square$

See Figure 1 for an illustration.

Next, let us recall some previous results by Clote (2006) on enumeration of saturated and optimal RNA secondary structures. Call a vertex $i \in [n]$ visible if it is not covered by any arc. Let $a(n)$ denote the number of all saturated RNA secondary structures (2-regular simple stacks) of length n and $b(n)$ denote the number of such stacks with no visible vertex. For $n < 0$, we define $a(n) = b(n) = 0$. For $n \geq 0$, $a(n)$ and $b(n)$ satisfy the following recurrence relations.

Lemma 1. (Clote , 2006, Proposition 3, Proposition 4)

$$a(n) = \begin{cases} 1, & 0 \leq n \leq 2, \\ \sum\limits_{k=1}^{n-2} a(k-1) \cdot a(n-k-1) + \sum\limits_{k=1}^{2} b(n-k), & \textit{otherwise,} \end{cases}$$
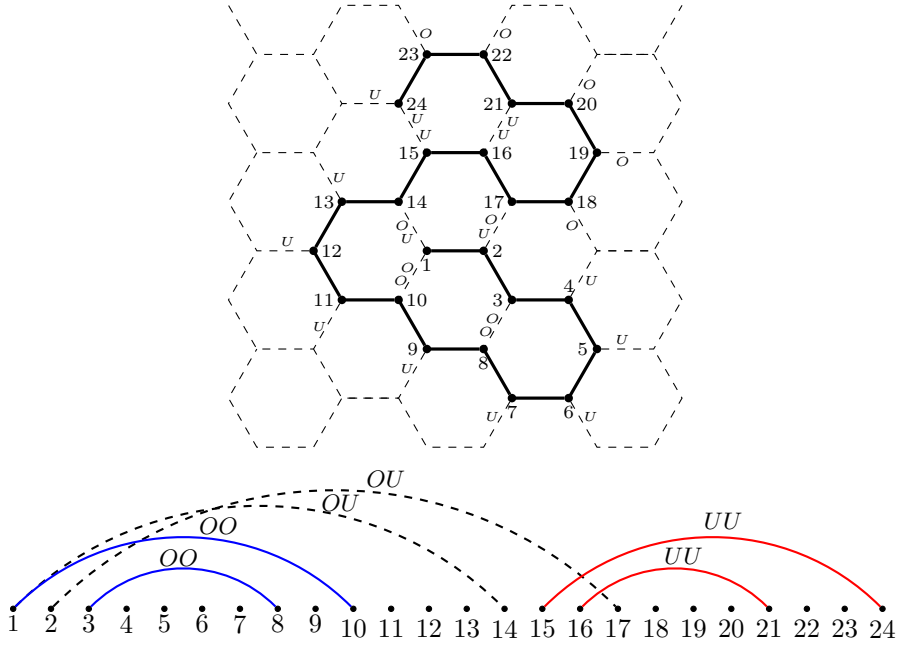
Figure 1: The contact map of a protein fold in two-dimensional honeycomb lattice. The graph consisting of arcs (solid) labelled by $\{O, O\}$ (or $\{U, U\}$) is a stack, and the graph consisting of arcs (dotted) labelled by $\{O, U\}$ is a queue.

$$b(n) = \begin{cases} 1, & n = 0, \\ 0, & 1 \leq n \leq 2, \\ \sum\limits_{k=1}^{n-2} b(k-1) \cdot a(n-k-1), & otherwise. \end{cases}$$

Denote the generating functions of $a(n)$, $b(n)$ by

$$y(x) = \sum_{n=0}^{\infty} a_n x^n, \qquad z(x) = \sum_{n=0}^{\infty} b_n x^n, \tag{3}$$

respectively, they satisfy the following relations.

Lemma 2. (Clote , 2006, Proposition 5, Proposition 6)

$$x^2 y z = z - 1 + x^2 z, \tag{4}$$

$$x^2 y^2 = y(x^2 + 1) - z(x^2 + x) - 1. \tag{5}$$

Moreover, the number of optimal 2-regular simple stacks of length n is given by the following formula.

Lemma 3. (Clote , 2006, Corollary 13)

$$LO_0(n) = \begin{cases} 1, & n = 2m - 1, \\ \dfrac{m(m+1)}{2}, & n = 2m, \end{cases} \tag{6}$$

where $m \geq 1$ *and* $LO_0(0) = 1$.

3   Saturated extended 2-regular simple stacks

In this section, we will use the method of combinatorial structure decomposition method to study the generating function for the number of saturated extended 2-regular simple stacks.

Following the structure decomposition idea proposed by Chen et al. (2014), we call the component containing both vertices 1 and n the primary component. For the primary components of the saturated extended 2-regular simple stacks with n vertices, we distinguish six classes according to the degrees of the two terminal vertices and whether they form an arc or not, see Figure 2.

Let $\langle i, j \rangle$ denote the interval between i and j, that is, $\{i+1, i+2, …, j-1\}$. We use this notation $\langle i, j \rangle$ to distinguish with the notation (i, j) for an arc. Note that the interval is allowed to be empty.

As shown in Figure 2, the primary component splits $[n]$ into disjoint intervals, on which substructures can be constructed. To meet the restrictions of saturated extended 2-regular simple stack, the substructures can be classified into the following six types:

	$(\deg(1),\ \deg(n))$	the primary component decomposition
1	$(0,2),\ (2,0)$	$\underset{(1)}{1\,\boxed{T_1}\,\boxed{T_2}\,\boxed{T_3}n}$ $\qquad$ $\underset{(1')}{1\,\boxed{T_3}\,\boxed{T_2}\,\boxed{T_1'}n}$
2	$(1,2),\ (2,1)$ $1,\ n$ form an arc	$\underset{(2)}{1\,\boxed{T_1}\,\boxed{T_3}n}$ $\qquad$ $\underset{(2')}{1\,\boxed{T_3}\,\boxed{T_1'}n}$
3	$(1,2),\ (2,1)$ $1,\ n$ do not form an arc	$\underset{(3)}{1\,\boxed{T_4}\,\boxed{T_5}\,\boxed{T_2}\,\boxed{T_3}n}$ $\qquad$ $\underset{(3')}{1\,\boxed{T_3}\,\boxed{T_2}\,\boxed{T_5}\,\boxed{T_4'}n}$
4	$(1,1)$	$\underset{(4)}{1\quad\boxed{T_6}\quad n}$
5	$(2,2)$ $1,\ n$ form an arc	$\underset{(5)}{1\,\boxed{T_3}\,\boxed{T_2}\,\boxed{T_3}n}$
6	$(2,2)$ $1,\ n$ do not form an arc	$\underset{(6)}{1\,\boxed{T_3}\,\boxed{T_2}\,\boxed{T_2}\,\boxed{T_2}\,\boxed{T_3}n}$

Figure 2: The six cases of the primary components of the saturated extended 2-regular simple stacks.

- T_1: an isolated vertex followed by an arbitrary saturated structure with no visible vertex, or just an arbitrary saturated structure with no visible vertex; For a structure of type T_1, its reverse order structure is denoted of type T_1';

- T_2: an arbitrary saturated structure;

- T_3: an arbitrary nonempty saturated structure;

- T_4: an isolated vertex followed by an arbitrary saturated structure with no visible vertex, or just an arbitrary nonempty saturated structure with no visible vertex; T_4' denotes the type of the reverse of a structure of type T_4;

- T_5: an arbitrary saturated structure with no visible vertex;

- T_6: an isolated vertex, or an arbitrary nonempty saturated structure with no visible vertex.

Note that, the substructures of types T_1, T_2, T_5 may be empty.

Let $s(n)$ denote the number of all saturated extended 2-regular simple stacks of length n and $\overline{s}(n)$ denote the number of such stacks with no visible vertex. Denote the generating functions of $s(n), \overline{s}(n)$ by

$$S(x) = \sum_{n=0}^{\infty} s(n)x^n, \qquad \overline{S}(x) = \sum_{n=0}^{\infty} \overline{s}(n)x^n,$$

respectively. Denote the number of the structures with n vertices of Case (i) in Figure 2 by $s_i(n)$, and the number of such structures with no visible vertex by $\overline{s}_i(n)$. Let

$$S_i(x) = \sum_{n=0}^{\infty} s_i(n)x^n, \qquad \overline{S}_i(x) = \sum_{n=0}^{\infty} \overline{s}_i(n)x^n \qquad (1 \leq i \leq 6).$$

Employing the method of primary component decomposition, we obtain that $S(x), \overline{S}(x)$ can be expressed in terms of the generating functions y and z (3).

Theorem 4. *We have*

$$\begin{aligned}
S(x) =& 2(x+1)x^5y^2z^2 + 2x^4y^2z - 2(x+1)x^5yz^2 + 2x^3yz + x^6y^5 - 2x^6y^4 \\
&+ (x^2+1)x^4y^3 - 2x^4y^2 + x^4y - (2x^2+2x-1)x^2z + (x-1)x^2. \quad (7)
\end{aligned}$$

Proof. For a saturated extended 2-regular simple stack with n vertices, suppose that its primary component splits $[n]$ into disjoint intervals $I_1, I_2, \ldots$ from left to right, each having k_i vertices.

Obviously, Case (1) and Case (1′) are symmetric, we thus consider only Case (1). As shown in Figure 2, there are three intervals I_1, I_2, I_3 of type T_1, T_2 and T_3, respectively. Note that $n \geq 5$ since there are four vertices in the primary component and I_3 can not be empty. The total number of vertices in I_1, I_2, I_3 is $k_1 + k_2 + k_3 = n - 4$, where $k_1, k_2 \geq 0$ since I_1 and I_2 can be empty, and

$k_3 \geq 1$ since I_3 is not allowed to be empty.

We have two cases for the substructures on I_1. For the case of an isolated vertex followed by an arbitrary saturated 2-regular simple stack with no visible vertex, there will be $b(k_1 - 1)$ substructures can be built. For the case of an arbitrary saturated substructures with no visible vertex or empty, there will be $b(k_1)$ substructures can be built. Clearly, the numbers of substructures can be built on I_2 and I_3 are $a(k_2)$ and $a(k_3)$, respectively. It deduces that the number of structures of Case (1) is

$$s_1(n) = \sum_{\substack{k_1+k_2+k_3=n-4 \\ k_1 \geq 0,\ k_2 \geq 0,\ k_3 \geq 1}} \big(b(k_1 - 1) + b(k_1)\big) a(k_2) a(k_3),$$

from which we can obtain the generating function as follows

$$S_1(x) = \sum_{n \geq 5} s_1(n) x^n$$

$$= \sum_{n \geq 5} \sum_{\substack{k_1+k_2+k_3=n-4 \\ k_1 \geq 0,\ k_2 \geq 0,\ k_3 \geq 1}} \big(b(k_1 - 1) + b(k_1)\big) a(k_2) a(k_3) x^n$$

$$= x^5 \sum_{k_1 \geq 0} b(k_1 - 1) x^{k_1 - 1} \sum_{k_2 \geq 0} a(k_2) x^{k_2} \sum_{k_3 \geq 1} a(k_3) x^{k_3}$$

$$+ x^4 \sum_{k_1 \geq 0} b(k_1) x^{k_1} \sum_{k_2 \geq 0} a(k_2) x^{k_2} \sum_{k_3 \geq 1} a(k_3) x^{k_3}$$

$$= x^5 zy(y - 1) + x^4 zy(y - 1)$$

$$= (x + 1) x^4 y^2 z - (x + 1) x^4 yz. \tag{8}$$

Case (2) and Case (2′) are also symmetric, so we only consider Case (2). Following the similar discussions in Case (1), we have $n \geq 4$. Moreover, there are two intervals I_1, I_2 of type T_1 and T_3 with k_1 and k_2 vertices, respectively, such that $k_1 + k_2 = n - 3$ and $k_1 \geq 0, k_2 \geq 1$. Consequently, we have that the

number of structures of Case (2) is

$$s_2(n) = \sum_{\substack{k_1+k_2=n-3 \\ k_1 \geq 0, \ k_2 \geq 1}} (b(k_1 - 1) + b(k_1))a(k_2),$$

whose generating function is

$$S_2(x) = \sum_{n \geq 4} s_2(n)x^n$$

$$= \sum_{n \geq 4} \sum_{\substack{k_1+k_2=n-3 \\ k_1 \geq 0, \ k_2 \geq 1}} \big(b(k_1 - 1) + b(k_1)\big)a(k_2)x^n$$

$$= x^4 \sum_{k_1 \geq 0} b(k_1 - 1)x^{k_1-1} \sum_{k_2 \geq 1} a(k_2)x^{k_2} + x^3 \sum_{k_1 \geq 0} b(k_1)x^{k_1} \sum_{k_2 \geq 1} a(k_2)x^{k_2}$$

$$= x^4 z(y - 1) + x^3 z(y - 1)$$

$$= (x + 1)x^3 z(y - 1)$$

$$= (x + 1)x^3 yz - (x + 1)x^3 z. \tag{9}$$

Case (3) and Case (3′) are symmetric, so we only consider Case (3). There are four intervals I_1, I_2, I_3, I_4 of types T_4, T_5, T_2 and T_3, respectively. Note that $n \geq 7$ and the total number of vertices in I_1, I_2, I_3, I_4 is $k_1 + k_2 + k_3 + k_4 = n - 5$, where $k_1, k_4 \geq 1$ since I_1 and I_4 are not allowed to be empty, and $k_2, k_3 \geq 0$. Similar to Case (1), there are $b(k_1 - 1) + b(k_1)$ substructures can be built on I_1. Clearly, the numbers of substructures can be built on I_2, I_3 and I_4 are $b(k_2), a(k_3)$ and $a(k_4)$, respectively. So we obtain the number of structures of Case (3) is

$$s_3(n) = \sum_{\substack{k_1+k_2+k_3+k_4=n-5 \\ k_1 \geq 1, \ k_4 \geq 1, \ k_2 \geq 0, \ k_3 \geq 0}} \big(b(k_1 - 1) + b(k_1)\big)b(k_2)a(k_3)a(k_4),$$

whose generating function is

$$S_3(x) = \sum_{n \geq 7} s_3(n) x^n$$

$$= \sum_{n \geq 7} \sum_{\substack{k_1+k_2+k_3+k_4=n-5 \\ k_1 \geq 1,\ k_4 \geq 1,\ k_2 \geq 0,\ k_3 \geq 0}} \big(b(k_1-1) + b(k_1)\big) b(k_2) a(k_3) a(k_4) x^n$$

$$= x^6 \sum_{k_1 \geq 1} b(k_1-1) x^{k_1-1} \sum_{k_2 \geq 0} b(k_2) x^{k_2} \sum_{k_3 \geq 0} a(k_3) x^{k_3} \sum_{k_4 \geq 1} a(k_4) x^{k_4}$$

$$+ x^5 \sum_{k_1 \geq 1} b(k_1) x^{k_1} \sum_{k_2 \geq 0} b(k_2) x^{k_2} \sum_{k_3 \geq 0} a(k_3) x^{k_3} \sum_{k_4 \geq 1} a(k_4) x^{k_4}$$

$$= x^6 y z^2 (y-1) + x^5 y z (y-1)(z-1)$$

$$= (x+1) x^5 y^2 z^2 - x^5 y^2 z - (x+1) x^5 y z^2 + x^5 y z. \tag{10}$$

For Case (4), there is only one interval of type T_6. Obviously, $n \geq 3$ and the number of vertices in the interval is $n - 2$. If $n = 3$, the substructure can be built on the interval is just an isolated vertex. If $n \geq 4$, there will be $b(n-2)$ substructures can be built. Therefore, the numbers of structures of Case (4) is

$$s_4(n) = \begin{cases} 1, & n = 3, \\ b(n-2), & n \geq 4 \end{cases}$$

with generating function

$$S_4(x) = x^3 + x^2(z - 1)$$

$$= x^2 z + (x - 1) x^2. \tag{11}$$

For Case (5), there are three intervals I_1, I_2 and I_3 of type T_3, T_2 and T_3, respectively. Note that $n \geq 6$ and the total number of vertices in I_1, I_2, I_3 is $k_1 + k_2 + k_3 = n - 4$, where $k_1, k_3 \geq 1$ and $k_2 \geq 0$. Thus, the number of

structures of Case (5) is

$$s_5(n) = \sum_{\substack{k_1+k_2+k_3=n-4 \\ k_1\geq 1,\ k_3\geq 1,\ k_2\geq 0}} a(k_1)a(k_2)a(k_3),$$

from which we can get the generating function as follows

$$S_5(x) = x^4 y(y-1)^2$$
$$= x^4 y^3 - 2x^4 y^2 + x^4 y. \tag{12}$$

For Case (6), there are five intervals I_i $(1 \leq i \leq 5)$, in which I_1, I_5 are of type T_3, and I_2, I_3, I_4 are of type T_2. Obviously $n \geq 8$, the total number of vertices in the intervals is $\sum_{i=1}^{5} k_i = n - 6$, where $k_1, k_5 \geq 1$ and $k_2, k_3, k_4 \geq 0$. Thus the number of structures of Case (6) is

$$s_6(n) = \sum_{\substack{k_1+k_2+k_3+k_4+k_5=n-6 \\ k_1\geq 1,\ k_5\geq 1,\ k_2\geq 0,\ k_3\geq 0,\ k_4\geq 0}} a(k_1)a(k_2)a(k_3)a(k_4)a(k_5),$$

whose generating function is

$$S_6(x) = x^6 (y-1)^2 y^3$$
$$= x^6 y^5 - 2x^6 y^4 + x^6 y^3. \tag{13}$$

Above all, substituting (8)—(13) into

$$S(x) = 2\big(S_1(x) + S_2(x) + S_3(x)\big) + S_4(x) + S_5(x) + S_6(x),$$

we obtain (7), which completes the proof. $\qquad\square$

Now, we are ready to prove Theorem 1 by eliminating the variables y and z from (4), (5) and (7).

Proof of Theorem 1. To derive the functional equation satisfied by $S(x)$ in Theorem 1, by (4) it is direct to see that

$$y = \frac{x^2 z + z - 1}{x^2 z}. \tag{14}$$

Substituting (14) into (7), we obtain

$$
\begin{aligned}
x^4 z^5 S(x) &- 2(x^3 + x^2 + x + 1)x^5 z^7 + (2x^4 + 2x^3 - x^2 + 2x - 2)x^4 z^6 \\
&+ (-x^7 + 2x^6 - 4x^2 - 1)z^5 + (2x^6 + 9x^4 + 15x^2 + 5)z^4 - (x^6 + 10x^4 \\
&+ 21x^2 + 10)z^3 + (3x^4 + 13x^2 + 10)z^2 - (3x^2 + 5)z + 1 = 0.
\end{aligned}
\tag{15}
$$

By computing the resultant of (4) and (5) with respect to y, we have

$$(x + 1)x^3 z^3 + x^2 z^2 - (x^2 + 1)z + 1 = 0. \tag{16}$$

Then eliminating z by computing the resultant of (15) and (16), we obtain (1). $\qquad\square$

From the functional equation (1) of $S(x)$, we can deduce the following recurrence relation of $s(n)$ by using the `Maple` commands `algfuntodiffeq` and `diffeqtorec` in the `gfun` package (Salvy and Zimmerman , 1994).

Corollary 1. *The number of saturated extended 2-regular simple stacks on n vertices $s(n)$ satisfies the following recurrence relation*

$$\sum_{i=0}^{26} p_i(n)s(n+i) = 0, \tag{17}$$

in which

$$p_0(n) = -80n^2 + 560n - 960, \qquad\qquad p_1(n) = -424n^2 + 1976n - 2112,$$

$p_2(n) = 224n^2 - 940n + 2120,$ $\qquad$ $p_3(n) = 5920n^2 - 2844n + 3600,$

$p_4(n) = 14236n^2 + 28268n + 6320,$ $\qquad$ $p_5(n) = 1042n^2 + 7068n - 52316,$

$p_6(n) = -66150n^2 - 453864n - 903872,$ $\qquad$ $p_7(n) = -166716n^2 - 1543338n - 3600520,$

$p_8(n) = -192882n^2 - 2175708n - 5799422,$ $\quad$ $p_9(n) = -74278n^2 - 855668n - 1405342,$

$p_{10}(n) = 94943n^2 + 1787565n + 9945244,$ $\qquad$ $p_{11}(n) = 187411n^2 + 3757775n + 20379846,$

$p_{12}(n) = 187813n^2 + 4238697n + 23752086,$ $\quad$ $p_{13}(n) = 157739n^2 + 4068271n + 23817290,$

$p_{14}(n) = 177075n^2 + 5271923n + 35846974,$ $\quad$ $p_{15}(n) = 251299n^2 + 8276931n + 65487426,$

$p_{16}(n) = 289002n^2 + 10257420n + 89937374,$ $\quad$ $p_{17}(n) = 228044n^2 + 8645464n + 82079842,$

$p_{18}(n) = 104272n^2 + 4199356n + 42802142,$ $\quad$ $p_{19}(n) = 1624n^2 + 71614n + 1164296,$

$p_{20}(n) = -35127n^2 - 1578391n - 17593210,$ $\quad$ $p_{21}(n) = -26613n^2 - 1262021n - 14949286,$

$p_{22}(n) = -9296n^2 - 465608n - 5843760,$ $\qquad$ $p_{23}(n) = -254n^2 - 16130n - 255584,$

$p_{24}(n) = 1234n^2 + 66010n + 880404,$ $\qquad$ $p_{25}(n) = 590n^2 + 33210n + 466900,$

$p_{26}(n) = 120n^2 + 7020n + 102600.$

The first 29 initial values of $s(n)$ are given in Table 1. The saturated extended 2-regular simple stacks of length five and six are listed in Figure 3 and Figure 4, respectively.

Table 1: The first 29 values of $s(n)$.

n	0	1	2	3	4	5	6	7	8	9
$s(n)$	1	1	1	1	2	7	12	26	57	116

n	10	11	12	13	14	15	16	17	18	
$s(n)$	251	545	1159	2517	5503	11962	26204	57711	127054	

n	19	20	21	22	23
$s(n)$	280704	622425	1381923	3074897	6858928

n	24	25	26	27	28
$s(n)$	15323958	34293674	76885723	172630454	388146408

Next, we use the singularity analysis method to derive the asymptotic formula of $s(n)$. Here, we give a sketch of the computation process, and refer to the classic book Flajolet and Sedgewick (2009) for more details.

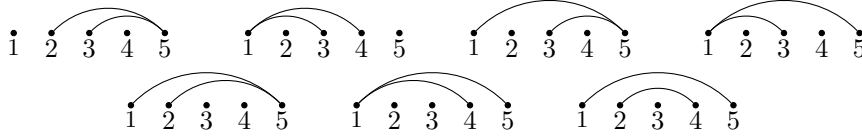First, by equation (1), we find the dominant singularity of $S(x)$ to be $\zeta =$

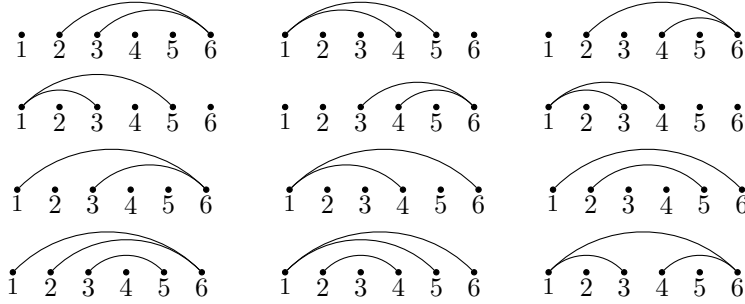Figure 3: The saturated extended 2-regular simple stacks of length five.

Figure 4: The saturated extended 2-regular simple stacks of length six.

$0.4246873104\ldots$, then applying Newton-Puiseux Expansion Theorem, we obtain the following expansion of $S(x)$ at $x = \zeta$,

$$S(x) = \sigma - \alpha\sqrt{\frac{x-\zeta}{-\alpha}} + O(x-\zeta),$$

where $\sigma = 1.969506702\ldots$, and $\alpha = 235.0528747\ldots$.

At last, leveraging the transfer theorem, we transfer the approximation of $S(x)$ near $x = \zeta$ into the following approximation of its coefficients,

$$s(n) = [x^n]S(x) = \frac{\gamma}{2\sqrt{\pi}}\omega^n n^{-\frac{3}{2}}\left(1 + O\left(\frac{1}{n}\right)\right),$$

where $\omega = \dfrac{1}{\zeta} = 2.354673605\ldots$, $\gamma = \sqrt{\alpha \cdot \zeta} = 9.99119478\ldots$.

Finally, we arrive at

$$s(n) \sim 2.818464011 \times 2.354673605^n \cdot n^{-\frac{3}{2}}.$$

The Maple source codes of the above procedures can be found at

`https://github.com/aimerbam/asymptotic_formulas_RNA/tree/master`.

For the saturated extended 2-regular simple stacks of length n with no visible vertex, we obtain the following expression for the generating function $\overline{S}(x)$.

Corollary 2. *We have*

$$\overline{S}(x) = x^6 y^4 z - 2x^6 y^3 z + 2(x+1)x^5 y^2 z^2 + (x-2)x^5 y^2 z - 2(x+1)x^5 y z^2$$
$$+ 2(x^2 + x + 1)x^3 yz + x^4 y^3 - 2x^4 y^2 + x^4 y - (2x^2 + 2x - 1)x^2 z + (x-1)x^2.$$

$$(18)$$

Proof. Similar to the proof of Theorem 4, we continue to use the method of primary component decomposition and discuss the six cases as shown in Figure 2.

Obviously, it is impossible for Case (1) to build a stack with no visible vertex. According to the proof of Theorem 7, we can see that the structures of cases (2), (2′), (3), (3′), (4) and (5) are all saturated 2-regular simple stacks with no visible vertex. Thereby for $i = 2, 3, 4, 5$,

$$\overline{S}_i(x) = S_i(x).$$

For Case (6), the interval I_3 of type T_2 must have no visible vertex. So the number of saturated extended 2-regular simple stacks of this case is

$$\overline{s}_6(n) = \sum_{\substack{k_1 + k_2 + k_3 + k_4 + k_5 = n - 6 \\ k_1 \geq 1,\ k_5 \geq 1,\ k_2 \geq 0,\ k_3 \geq 0,\ k_4 \geq 0}} a(k_1)a(k_2)b(k_3)a(k_4)a(k_5),$$

whose generating function is

$$\overline{S}_6(x) = x^6 (y-1)^2 y^2 z$$

$$= x^6 y^4 z - 2x^6 y^3 z + x^6 y^2 z. \tag{19}$$

Substituting (9)–(12) and (19) into

$$\overline{S}(x) = 2\big(S_2(x) + S_3(x)\big) + S_4(x) + S_5(x) + \overline{S}_6(x),$$

we obtain equation (18), which completes the proof. □

Similarly, by eliminating y and z from (4), (5) and (18), we obtain the following functional equation satisfied by $\overline{S}(x)$.

Theorem 5. *We have*

$$q_3(x)\overline{S}^3(x) + q_2(x)\overline{S}^2(x) + q_1(x)\overline{S}(x) + q_0(x) = 0, \tag{20}$$

where

$$q_0(x) = -x^3(8x^{14} - 3x^{13} + 22x^{12} + 7x^{11} - 9x^{10} - 24x^9 - 27x^8 - 6x^7 + 5x^6$$
$$+ 24x^5 + 12x^4 + 5x^3 - 4x^2 - 2x - 1),$$

$$q_1(x) = 14x^{14} + 7x^{13} + 24x^{12} + 16x^{11} + 8x^{10} - 12x^9 - 24x^8 - 24x^7 - 21x^6$$
$$+ 5x^4 + 9x^3 + x^2 - 1,$$

$$q_2(x) = -x^4(7x^7 + 8x^6 + 13x^5 + 11x^4 + 3x^3 - 3x^2 - 3x - 1),$$

$$q_3(x) = x^7(x + 1).$$

Consequently, the asymptotic formula of $\overline{s}(n)$ is

$$\overline{s}(n) \sim 1.673111891 \times 2.354673605^n \cdot n^{-\frac{3}{2}}.$$

4   The optimal extended 2-regular simple stacks

In this section, we consider the enumeration of optimal extended 2-regular simple stacks, which correspond to the structures with the maximum number of arcs in the Nussinov-Jacobson energy model.

Let us first consider the number of arcs in the optimal structures.

Lemma 4. *The number of arcs of the optimal extended 2-regular simple stacks with n vertices is $\left\lfloor \dfrac{n}{2} \right\rfloor$, where $n \geq 3$.*

Proof. We follow the six cases as shown in Figure 2 to discuss the number of arcs in the optimal structures.

We first consider the case when n is odd. Suppose that $n = 2m - 1$ with $m \geq 1$. From the restrictions for the structures of each case as discussed in the proof of Theorem 4, we claim that the maximum number of arcs for the saturated extended 2-regular simple stacks is $m - 1$. We explain only Case (3), the other cases are similar and therefore omitted.

In Case (3), the primary component splits $[n]$ into four intervals I_1, I_2, I_3, I_4 of types T_4, T_5, T_2 and T_3, respectively. The total number of vertices in these intervals is $2m - 6$. Since I_1, I_4 can not be empty, the remaining $2m - 6$ vertices can form at most $\dfrac{2m - 6 - 2}{2} = m - 4$ arcs. Adding the three arcs in the primary component, the optimal structures of Case (3) have exactly $m - 1$ arcs.

When n is even, suppose that $n = 2m$ with $m \geq 2$. We claim that the maximum number of arcs for the saturated extended 2-regular simple stacks is m. Similar to the above discussions, it is easy to see that for the cases (1), (3) and (4), we can construct saturated structures with at most $m - 1$ arcs, and for the cases (2), (5), (6), we can construct saturated structures with at most m arcs. Therefore, the maximum number of arcs for the saturated extended 2-regular simple stacks is m, namely, only cases (2), (5) and (6) can be optimal.

Above all, the number of arcs of the optimal extended 2-regular simple stacks with n vertices is $\left\lfloor \dfrac{n}{2} \right\rfloor$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Now, we are ready to prove Theorem 2.

Proof of Theorem 2. Denote the number of optimal extended 2-regular simple stacks of length n by $so(n)$. When n is odd, suppose that $n = 2m - 1$. By Lemma 4, we see that the optimal structures have $m - 1$ arcs. So we only need to discuss the structures with $m - 1$ arcs in each case as shown in Figure 2. Denote the number of optimal structures in Case (i) by $so_i(n)$ for $1 \leq i \leq 6$.

For Case (1), we see that $m \geq 3$. The primary component splits $[n]$ into three intervals I_1, I_2, I_3 of types T_1, T_2 and T_3, respectively. To be optimal, the substructures on I_1 and I_2 must be empty, and the substructures on I_3 must be optimal on $2m - 5$ vertices. By using (6), we have

$$so_1(2m - 1) = LO_0(2m - 5) = 1.$$

For Case (2), $m \geq 3$, there are two intervals I_1, I_2 of type T_1 and T_3 with k_1, k_2 vertices, respectively. To be optimal, we have two possibilities: one is that the substructure on I_1 is empty, the substructures on I_2 are optimal with $2m - 4$ vertices, the other one is that both of the substructures on I_1 and I_2 are optimal and k_1, k_2 are odd with $k_1 + k_2 = 2m - 4$. By using (6), we have

$$
\begin{aligned}
so_2(2m - 1) &= LO_0(2m - 4) + \sum_{\substack{k_1 + k_2 = 2m - 4 \\ k_1, k_2 \text{ odd} \\ k_1 \geq 1,\ k_2 \geq 1}} LO_0(k_1) LO_0(k_2) \\
&= \frac{(m - 1)(m - 2)}{2} + \sum_{\substack{k_1 + k_2 = 2m - 4 \\ k_1, k_2 \text{ odd} \\ k_1 \geq 1,\ k_2 \geq 1}} 1 \\
&= \frac{(m - 1)(m - 2)}{2} + m - 2 \\
&= \frac{1}{2}(m^2 - m - 2).
\end{aligned}
\tag{21}
$$

For Case (3), $m \geq 4$, there are four intervals I_1, I_2, I_3, I_4 of type T_4, T_5, T_2 and T_3, respectively. To be optimal, the substructures on I_2 and I_3 must be empty, and the substructures on I_1 and I_4 must be optimal and k_1, k_4 are odd with $k_1 + k_4 = 2m - 6$. By using (6), we have

$$so_3(2m - 1) = \sum_{\substack{k_1+k_4=2m-6 \\ k_1,k_4 \text{ odd} \\ k_1 \geq 1, \ k_4 \geq 1}} LO_0(k_1)LO_0(k_4)$$

$$= m - 3. \tag{22}$$

For Case (4), $m \geq 2$, there is only one interval I_1 of type T_6, and the substructures on I_1 must be optimal with $2m - 3$ vertices. By using (6), we have

$$so_4(2m - 1) = LO_0(2m - 3) = 1. \tag{23}$$

For Case (5), $m \geq 4$, there are three intervals I_1, I_2, I_3 of type T_3, T_2, and T_3, respectively. To be optimal, we have two possibilities: one is that the substructures on I_1, I_2 and I_3 are all optimal and k_1, k_2, k_3 are odd numbers with $k_1 + k_2 + k_3 = 2m - 5$, the other one is that the substructure on I_2 is empty, and the substructures on I_1 and I_3 are both optimal with $k_1 + k_3 = 2m - 5$. By using (6), we have

$$so_5(2m - 1) \tag{24}$$

$$= \sum_{\substack{k_1+k_2+k_3=2m-5 \\ k_1,k_2,k_3 \text{ odd} \\ k_1 \geq 1, \ k_2 \geq 1, \ k_3 \geq 1}} LO_0(k_1)LO_0(k_2)LO_0(k_3) + \sum_{\substack{k_1+k_3=2m-5 \\ k_1 \geq 1, \ k_3 \geq 1}} LO_0(k_1)LO_0(k_3)$$

$$= \sum_{\substack{k_1+k_2+k_3=2m-5 \\ k_1,k_2,k_3 \text{ odd} \\ k_1 \geq 1, \ k_2 \geq 1, \ k_3 \geq 1}} 1 + \sum_{\substack{k_1+k_3=2m-5 \\ k_1 \text{ even}, \ k_3 \text{ odd} \\ k_1 \geq 1, \ k_3 \geq 1}} LO_0(k_1)LO_0(k_3) + \sum_{\substack{k_1+k_3=2m-5 \\ k_1 \text{ odd}, \ k_3 \text{ even} \\ k_1 \geq 1, \ k_3 \geq 1}} LO_0(k_1)LO_0(k_3)$$

$$= \sum_{\substack{k_1+k_2+k_3=2m-5 \\ k_1,k_2,k_3 \text{ odd} \\ k_1\geq 1,\ k_2\geq 1,\ k_3\geq 1}} 1 + 2 \sum_{\substack{1\leq k_1\leq 2m-5 \\ k_1 \text{ even}}} LO_0(k_1)$$

$$= \frac{(m-2)(m-3)}{2} + 2 \sum_{1\leq j\leq m-3} LO_0(2j) \qquad (2j=k_1)$$

$$= \frac{(m-2)(m-3)}{2} + 2 \sum_{1\leq j\leq m-3} \frac{j(j+1)}{2}$$

$$= \frac{(m-2)(m-3)}{2} + \frac{(m-3)(m-2)(m-1)}{3}$$

$$= \frac{1}{6}(2m^3 - 9m^2 + 7m + 6). \tag{25}$$

For Case (6), $m \geq 5$, there are five intervals I_1, I_2, I_3, I_4, I_5 of type T_3, T_2, T_2, T_2 and T_3, respectively. To be optimal, the possibilities structures may be: the substructures on I_1, I_i and I_5 are all optimal, and k_1, k_i, k_5 are odd such that $k_1 + k_i + k_5 = 2m - 7$, where $2 \leq i \leq 4$; or the substructures on I_2, I_3 and I_4 are empty, and the substructures on I_1 and I_5 are both optimal with $k_1 + k_5 = 2m - 7$. By using (6), we have

$$so_6(2m-1) = 3 \sum_{\substack{k_1+k_2+k_5=2m-7 \\ k_1,k_2,k_5 \text{ odd} \\ k_1\geq 1,\ k_2\geq 1,\ k_5\geq 1}} LO_0(k_1)LO_0(k_2)LO_0(k_5) + \sum_{\substack{k_1+k_5=2m-7 \\ k_1\geq 1,\ k_5\geq 1}} LO_0(k_1)LO_0(k_5)$$

$$= 3 \times \frac{(m-3)(m-4)}{2} + \frac{(m-4)(m-3)(m-2)}{3}$$

$$= \frac{1}{6}(2m^3 - 9m^2 - 11m + 60). \tag{26}$$

Substituting (4)–(26) into

$$so(2m-1) = 2 \sum_{i=1,2,3} so_i(2m-1) + \sum_{i=4,5,6} so_i(2m-1),$$

we obtain

$$so(2m-1) = \frac{1}{3}(2m^3 - 6m^2 + m + 18),$$

thus for n being odd, we have that

$$so(n) = \frac{1}{12}(n^3 - 3n^2 - 7n + 69).$$

When n is even, suppose that $n = 2m$. By Lemma 4, we see that the optimal structures have m arcs and we only need to discuss the cases (2), $(2')$, (5) and (6) as shown in Figure 2.

For Case (2), $m \geq 2$, to be optimal, the substructure on interval I_1 must be empty, the substructures on interval I_2 must be optimal with $2m - 3$ vertices. Thus, we have

$$so_2(2m) = LO_0(2m - 3) = 1. \tag{27}$$

For Case (5), $m \geq 3$, to be optimal, the substructures on I_2 must be empty, the substructures on I_1 and I_3 must be optimal and $k_1 + k_3 = 2m - 4$ with k_1, k_2 are odd numbers. Thus, we have

$$so_5(2m) = \sum_{\substack{k_1+k_3=2m-4 \\ k_1,k_3 \text{ odd} \\ k_1 \geq 1, \ k_3 \geq 1}} LO_0(k_1)LO_0(k_3)$$

$$= m - 2. \tag{28}$$

For Case (6), $m \geq 4$, to be optimal, the substructures on intervals I_2, I_3, I_4 must be empty, and the substructures on intervals I_1 and I_5 must be optimal and k_1, k_5 are odd numbers with $k_1 + k_5 = 2m - 6$. Thus, we have

$$so_6(2m) = \sum_{\substack{k_1+k_5=2m-6 \\ k_1,k_5 \text{ odd} \\ k_1 \geq 1, \ k_5 \geq 1}} LO_0(k_1)LO_0(k_5)$$

$$= m - 3. \tag{29}$$

Substituting (27)–(29) into

$$so(2m) = 2so_2(2m) + so_5(2m) + so_6(2m),$$

we obtain

$$so(2m) = 2m - 3.$$

Thus for n being even,

$$so(n) = n - 3.$$

We complete the proof of Theorem 2. □

The first 22 initial values of $so(n)$ are given in Table 2.

Table 2: The fisrt 22 values of $so(n)$.

n	0	1	2	3	4	5	6	7	8	9	10
$so(n)$	1	1	1	1	2	7	3	18	5	41	7

n	11	12	13	14	15	16	17	18	19	20	21
$so(n)$	80	9	139	11	222	13	333	15	476	17	655

For $n = 5$, the seven saturated extended 2-regular simple stacks as in Figure 3 are all optimal. For $n = 6$, the last three structures with three arcs as in Figure 4 are optimal among the twelve saturated structures.

At last, the growth curves of the numbers of saturated and optimal extended 2-regular simple stacks are drawn together in Figure 5.
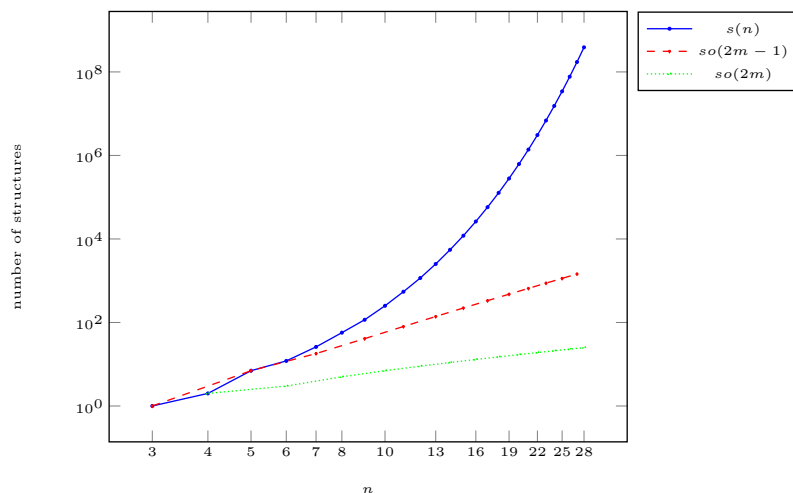
Acknowledgments

Figure 5: The growth curves of the numbers of saturated and optimal extended 2-regular simple stacks.

Author Disclosure Statement

The authors declare they have no competing financial interests.

Funding Information

References

Agarwal, P.K., Mustafa, N.H., Wang, Y. 2007. Fast molecular shape matching using contact maps. *J. Comput. Biol.* 14(2), 131–143.

Banerjee, A.R., Jaeger, J.A., Turner, D.H. 1993. Thermal unfolding of a group

I ribozyme: the low-temperature transition is primarily disruption of tertiary structure. *Biochemistry.* 32 (1), 153–163.

Chen, W.Y.C., Guo, Q.-H., Sun, L.H., et al. 2014. Zigzag stacks and m-regular linear stacks. *J. Comput. Biol.* 21 (12), 915–935.

Clote, P. 2006. Combinatorics of saturated secondary structures of RNA. *J. Comput. Biol.* 13(9), 1640–1657.

Clote, P., Kranakis, E., Krizanc, D., et al. 2009. Asymptotics of canonical and saturated RNA secondary structures. *J. Bioinform. Comput. Biol.* 7, 869–893.

Clote, P., Kranakis, E., Krizanc, D., et al. 2007. Asymptotic expected number of base pairs in optimal secondary structure for random RNA using the Nussinov-Jacobson energy model. *Discrete Appl. Math.* 155, 759–787.

Dill, K.A. 1990. Dominant forces in protein folding. *Biochemistry.* 29 (31), 7133–7155.

Flajolet, P., and Sedgewick, R. 2009. Analytic Combinatorics. Cambridge University Press, Cambridge.

Fusy, E., and Clote, P. 2014. Combinatorics of locally optimal RNA secondary structures.*J. Math. Biol.* 68(1-2), 341–375.

Goldman, D., Istrail, S., Papadimitriou, C.H. 1999. Algorithmic aspects of protein structure similarity (extended abstract). 40*th Annual Symposium on Foundations of Computer Science (New York, 1999)*, 512-522, *IEEE Computer Soc., Los Alamitos, CA*, 1999.

Guo, Q.-H., Wang, J., Xu Z. 2018. Approximation Algorithms for Protein Folding in the Hydrophobic-Polar Model on 3D Hexagonal Prism Lattice. *J. Comput. Biol.* 25(5), 487–498.

Guo, Q.-H., and Sun, L.H. 2018. Combinatorics of Contacts in Protein Contact Maps. *Bull. Math. Biol.* 80 (2), 385–403.

Guo, Q.-H., Sun, L.H., Wang, J. 2016. Enumeration of Extended m-Regular Linear Stacks. *J. Comput. Biol.* 23, 943–956.

Guo, Q.-H., Sun, L.H., Wang, J. 2017. Regular Simple Queues of Protein Contact Maps. *Bull. Math. Biol.* 79(1), 11–35.

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

Jiang, M., and Zhu, B. 2005. Protein folding on the hexagonal lattice in the hp model. *J. Bioinform. Comput. Biol.* 03 (01), 19–34.

Jin, E.Y., Qin, J., Reidys, C.M. 2008. Combinatorics of RNA structures with pseudoknots. *Bull. Math. Biol.* 70, 45–67.

Nebel, M.E. 2002. Combinatorial properties of RNA secondary structures. *J. Comput. Biol.* 9 (3), 541–573.

Nussinov, R., and Jacobson, A.B. 1980. Fast algorithm for predicting the secondary structure of single stranded RNA. *Proc. Natl. Acad. Sci. USA.* 77 (11), 6309–6313.

Pierri, C.L., De Grassi, A., and Turi, A. 2008. Lattices for ab initio protein structure prediction. *Proteins* 73, 351¨C-361.

Reidys, C. 2011. Combinatorial Computational Biology of RNA Pseudoknots and Neutral Networks. Springer-Verlag, New York.

Salvy, B., and Zimmerman, P. 1994. GFUN: a Maple package for the manipulation of generating and holonomic functions in one variable. *ACM T. Math. Software* 20, 163–177.

Schmitt, W.R., and Waterman, M. 1994. Linear trees and RNA secondary structure. *Discrete Appl. Math.* 51 (3), 317–323.

Waterman, M., and Smith, T. 1978. RNA secondary structure: a complete mathematical analysis. *Math. Biosci.* 42 (3-4), 257–266.

Xu, J. 2019. Distance-based protein folding powered by deep learning. *Proc. Natl. Acad. Sci. USA.* 116, 16856–16865.

Yang, J., Anishchenko, I., Park, H., et al. 2020. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. USA.* 117, 1496–1503.