Review

General enumeration of RNA secondary structures based on new arc representation

Wenwen Wang*, Xiao Liu and Rui Liu

School of Sciences, China University of Mining and Technology, Xuzhou 221008, People's Republic of China.

We have presented a new arc representation, which differentiates the Watson-Crick base pairs between *AU* and *GC*. Based on the new representation, this paper also computes the number of various types of constrained secondary structures taking the minimum stack length 1 and minimum size m for each bonding loop as two parameters by virtue of the technique of generating function. Furthermore, asymptotes are derived from these recurrences relations, which are the generalizations of previous results.

Key words: RNA secondary structure, recurrence relation, generating function, asymptotic enumeration.

INTRODUCTION

In Biology, the nucleic acids play an important role in coding, transferring and retrieving genetic information, and in directing cell metabolism. The nucleic acid includes DNA and RNA molecule. RNA molecule is a single-stranded nucleic acid of four different kinds of nucleotides. The four nucleotides only differ by one part, called bases. Hence, one usually identifies nucleotides and bases. Generally, the bases are denoted by letters *A*, *U*, *G* and *C*. Each nucleotide is a polar molecule with two differing ends, usually denoted by end 5' and 3' end. The primary structure of a RNA molecule is the order in which these bases occur (Holley et al., 1965). Additionally, it has been known for some time that *A* base-*U* base and *G* base-*C* base may form hydrogen bonded base pairs and are said to be complementary, which is called Watson-Crick pairing rules. However, adjacent bases never form hydrogen bonds. The planar folding of a RNA molecule is then called its secondary structure. Understanding the secondary structure of single-stranded RNA molecule is crucial to advancing knowledge of its biological functions (Couzin, 2002; Doudna, 2000).

As it is known that the prediction of the shapes of biological molecules is a hot topic in Biology; suppose we wish to predict the secondary structure of a single RNA,

many plausible secondary structures can be drawn for a sequence. The number increases exponentially with sequence length. We are concerned with the enumeration problem of RNA secondary structures, which play an important role in guiding for the predictions of RNA secondary structures (Rivas and Eddy, 1999; Waterman and Smith, 1978; Waterman and Smith, 1986; Zuker and Sankoff, 1984). The research on the enumeration of RNA secondary structures is one of the hot topics in Computational Molecular Biology (Nebel, 2004; Nkwanta, 1997; Stein and Waterman, 1978; Wang et al., 2010; Waterman, 1978). In order to satisfy different needs of researches, RNA secondary structures are usually modeled by some discrete mathematic objects, such as linear tree (Schmitt and Waterman, 1994), lattice path (Doslic et al., 2004; Nkwanta,1997; Wang et al., 2010), and so on, which raise many mathematically interesting questions concerning their enumeration problems (Doslic et al., 2004; Hofacker et al., 1998; Nebel,2002; Stein and Waterman, 1978; Wang et al., 2010; Waterman and Smith, 1978).

Generally, researchers are concerned with the number of secondary structures with limited length in each hairpin loop, in which the influence of pseudoknots are often neglected. Such as, the initial steps are based on the work of Waterman and Smith (Howell et al., 1980; Schmitt and Waterman, 1994; Stein and Waterman, 1978; Waterman, 1978; Waterman and Smith, 1986;

_____

*Corresponding author. E-mail: wangwenwencumt@163.com.

Waterman, 1995). In 1978, M.S. Waterman firstly gave a mathematical definition of RNA secondary structures (Stein and Waterman, 1978), which is the key of the enumeration problem of RNA secondary structures. Twenty years later I.L. Hofacker with his coworkers gave a basic formal framework for the same topic (Hofacker et al., 1998). Of course, there are many other authors who have been devoting to the research of the enumeration of secondary structures including pseudoknots (Haslinger and Stadler, 1999; Jin et al., 2008a; Jin and Reidys, 2008b; Rodland, 2006). In this paper, we discussed the enumeration problem of secondary structures, which is mainly based on the following conditions:

1. We neglect the influence of pseudoknots;

2. We differentiate $\{A, U\}$ base pairs and $\{G, C\}$ base pairs (Hofacker et al., 1998; Liao and Wang, 2004; Waterman and Smith, 1978).

Compared with previous investigations, in this paper we chose the minimum stack length 1 and minimum size m, for bonding loops as two parameters, which generalize the results obtained in Wang et al. (2008ab), and it shows that there are a large number of various types of constrained RNA secondary structures to a certain extent.

THE BASIC DEFINITIONS AND A NEW ARC REPRESENTATION

An RNA molecule is a string of n characters $R = r_1 r_2 \cdots r_n,$ such that $r_i \in A(U) \ or \ G(C)$. The secondary structure is a vertex-labeled graph on n vertices with an adjacency matrix $A = (r_{ij})$ fulfilling:

(1) $r_{i,i+1} = 1, \ 1 \leq i \leq n-1;$

(2) If $r_{i,k} = 1, \ k \neq i-1, i+1, \ r_i \ pairs \ with \ r_k.$

Note that if, r_i pairs with r_k, that is, $r_{i,k} = 1$ then, r_i and r_k all belong to $A(U) \ or \ G(C);$

(3) For each i there is at most a single $k \neq i-1, i+1,$ such that $r_{i,k} = 1;$

(4) If $r_{i,j} = r_{k,l} = 1 \ and \ i < k < j, \ then \ i < l < j,$ which prohibits certain knot structures.

We call an edge $(i, j), \ |i - j| \neq 1$ a bond or a base pair. A vertex i connected only to i-1 and i+1 is called

unpaired. A vertex i is said to be interior to the base pair (k, l) if, k<i<l. If, in addition, there is no base pair (p, q) such that k <p <i <q <l, we will say that i is immediately interior to the base pair (k, l).

A *stack* consists of subsequent base pairs (p-k, q+k), (p-k+1, q+k-1)… (P, q) such that neither (p-k-1, q+k+1) nor (p+1, q-1) is a base pair. k+1 are the length of the ladder. (p-k, q+k) is the terminal base pair of the stack. A stack [(p, q),…(p+k, q-k)] is called *terminal* if p-1=0 or q+1=n+1 or if the two vertices p-1 and q+1 are not interior to any base pair. The sub-structure enclosed by the terminal base pair (p, q) of a terminal stack is called a *component* of secondary structure. We say that a structure on n vertices has a terminal base pair if (1, n) is a base pair.

'A bonding loop' consists of a terminal base pair and unpaired vertices. The number of unpaired vertices is the length of the bonding loop. 'An external vertex' is an unpaired vertex which does not belong to a loop. A collection of adjacent external vertices is called an 'external element'. If it contains the vertex 1 or n it is 'a free end', otherwise it is called joint. 'An internal vertex' is an unpaired vertex which is interior to a base pair. Previous researches on the representation of RNA secondary structures were presented in Hofacker et al. (1998), Schmitt and Waterman, (1994) and Waterman (1978). All bases (e.g. A, U, G and C) are regarded as the same. In previous paper (Wang et al., 2008a), we have presented a new arc representation. That is to let circles ○ represent the bases A or U and dots ● represent the bases G or C, which is more meaningful than classical representation. Here an example of a secondary structure of size 91 is shown in Figure 1. Note that the minimum stack length l = 2 and each bonding loop has at least three bases, that is, m = 3.

Recurrence relations of RNA secondary structures

In order to compute the various types of constrained RNA secondary structures, let us introduce the Lemma given in Wang et al. (2008b) which is the base of our main results obtained in this paper.

Lemma 1

Let $\phi_n(l)$ be the number of RNA secondary structures on n vertices, where a bonding loop must contain at least m unpaired bases and the minimal stack length must be l, and $\phi_n^*(l)$ be the number of structures on n vertices which have only stacks of length at least l if an additional terminal base pair is attached. Furthermore, let $\phi_n^{**}(l)$ be the number of structures on n vertices which have all stacks of length at least l for which (1, n) is not a base
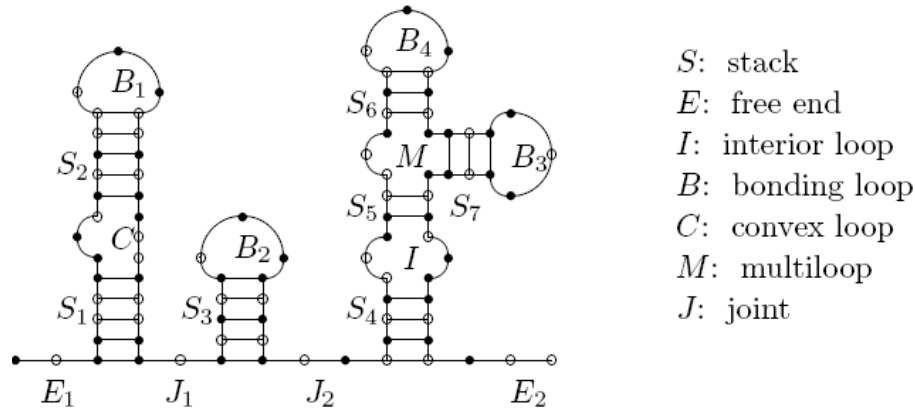
Figure 1. A new arc representation of secondary structures.

pair. Then, $\phi_n(l), \phi_n^*(l), \phi_n^{**}(l)$ fulfill the following recursions following recursions.

$$\phi_{n+1}(l) = 2\phi_n(l) + \sum_{k=m+2l-2}^{n-1} \phi_k^*(l)\phi_{n-k-1}(l), n \geq m + 2l - 1;$$

$$\phi_n^*(l) = 2\sum_{p=l-1}^{[\frac{n-m}{2}]} 2^p \phi_{n-2p}^{**}(l), n \geq m + 2l - 2;$$

$$\phi_n(l) = \phi_n^{**}(l) + \phi_{n-2}^*(l), n \geq m + 2l - 1;$$

$$2\phi_n(l) = \phi_{n+1}^{**}(l) = 2^{n+1}, n < m + 2l - 1;$$

$$\phi_n^*(l) = 0, n \leq m + 2l - 3.$$

The proof of above Lemma is given in detail (Wang et al., 2008b). Due to the fact that the stack structures and bonding loop structures are two basic elements of any RNA secondary structure, in this paper, we took the lengths of two elements of structures as parameters and discussed the enumeration problem of various types' secondary structures. Note that for the special case $l = 1$, the number 87 of structures with minimum stack length 1,

that is, $\phi_n(1)$ is equal to S_n given in Wang et al. (2008a) because the length of any stack is at least 1, and in Wang et al. (2008b), it is pointed that

$$\phi_n^*(l) = 2\phi_n(l).$$

Of course, it is obvious that $\phi_n^*(l) = 2S_n$ for $l = 1$. We now calculate the number of different elements of RNA secondary structures.

Theorem 1

Let Jn(b,l) denote the number of structures on n vertices with exactly b components, then:

$$J_{n+1}(b,l) = 2J_n(b,l) + \sum_{k=1}^{n-m-2l+2} \phi_{n-k}^*(l)J_{k-1}(b-1,l), n \geq m + 2l - 1;$$

$$J_n(b,l) = 0, \text{for } b > 0, n \leq m + 2l - 1; J_n(0,l) = 2^n, n \geq 0.$$

Proof

The number $J_{n+1}(b,l)$ of structures on $n+1$ vertex with b components can be computed as follows:

i. Adding an unpaired base belonging to $\{A, U\}$ to a structure on n digits doesn't change the number of components, we obtain $J_n(b,l)$ structures with b components. In the same way, if the unpaired base belonging to $\{G, C\}$ we get the same number. So if we add an unpaired base, we can get the total number $2J_n(b,l)$;

ii. Inserting an additional pair makes the bracketed part of length k a single component and doesn't affect the remainder of the sequence. There are b-1 components in the remainder; we have $\phi_{n-k}^*(l)$ times all the structures with b-1 components in the remainder of the sequence. Summing over k, we can get the desired result.

For the special case $l = 1$, we consider above recurrence relation. Submitting $l = 1$, we can get

$$J_{n+1}(b,1) = 2J_n(b,1) + \sum_{k=1}^{n-m} \phi_{n-k}^*(1)J_{k-1}(b-1,1), n \geq m+1.$$

Application of the identity $\phi_n^*(1) = 2S_n$, we can immediately get the following relation.

$$J_{n+1}(b,1) = 2J_n(b,1) + 2\sum_{k=1}^{n-m} S_{n-k}J_{k-1}(b-1,1), n \geq m+1.$$

Here, $J_n(b,1)$ is the same as $J_n(b)$ used in Wang et al. (2008a). It is obvious that the result coincides with the Theorem 5.1.2 in Wang et al. (2008a).

Theorem 2

Let $I_{n+1}(l)$ denote the total number of components in the set of all secondary structures on n vertices, then;

$$I_{n+1}(l) = 2I_n(l) + \sum_{k=1}^{n-m-2l+2} \phi_{n-k}^*(l)[I_{k-1}(l) + \phi_{k-1}(l)], n \geq m + 2l - 1;$$

$$I_n(l) = 0, n \leq m + 2l - 1.$$

Proof

The number $I_{n+1}(l)$ of secondary structures on $n+1$ vertices can be computed as follows:

i. Adding an unpaired base belonging to $\{A,U\}$ to a structure on n digits doesn't change the number of components, we obtain $I_n(l)$ structures. In the same way, if the unpaired base belonging to $\{G,C\}$, we get the same number. So if we add an unpaired base, we get the total number $2I_n(l)$;

ii. Inserting an additional pair, we have $\phi_{n-k}^*(l)$ times all the components in the reminder of the sequence plus the number of structures that can be formed the remainder of the structure. Summing over k, we can get the desired result.

For the special case $l = 1$, we consider above recurrence relation. Submitting $l = 1$, we can get;

$$I_{n+1}(1) = 2I_n(1) + \sum_{k=1}^{n-m} \phi_{n-k}^*(1)[I_{k-1}(1) + \phi_{k-1}(1)], n \geq m + 1;$$

Application of the identity $\phi_n^*(1) = 2S_n$, we can immediately get the following relation.

$$I_{n+1}(1) = 2I_n(1) + 2\sum_{k=1}^{n-m} S_{n-k}[I_{k-1}(1) + \phi_{k-1}(1)], n \geq m + 1.$$

Based on the fact that $\phi_n(1)$ is equal to S_n given in Wang et al. (2008a) $I_{k-1}(1)$ is the same as I_{k-1}. It is obvious that the result coincides with the Theorem 5.2.4 in Wang et al. (2008a).

Theorem 3

Let $E_n(b,l)$ denote the number of structures on n vertices with exactly b external digits then;

$$E_{n+1}(b,l) = 2E_n(b-1,l) + \sum_{k=1}^{n-m-2l+2} \phi_{n-k}^*(l)E_{k-1}(b,l), n \geq m + 2l - 1.$$

$$E_n(b,l) = 0, \; for \; b \neq n, \; n \leq m + 2l - 1; \; E_n(n,l)$$

For sake of completeness, let b = 0, then we have the following relations:

$$E_{n+1}(0,l) = \sum_{k=1}^{n-m-2l+2} \phi_{n-k}^* E_{k-1}(0,l), n \geq m + 2l - 1;$$

Proof

The number $E_{n+1}(b,l)$ of structures on $n + 1$ vertices with b external digits can be computed as follows:

i. Adding an unpaired base belonging to $\{A,U\}$ to a structure on n digits, because adding an external digit, we obtain $E_n(b-1,l)$ structures with $b-1$ external digits. If the unpaired base belonging to $\{G,C\}$, we can get the same number. So if we add an unpaired base, we get the total number $2E_n(b-1,l)$;

ii. Inserting an additional pair and making the bracketed part length k. Based on the fact that there are no external digits in the newly bracketed part of length k, we have $\phi_{n-k}^*(l)$ times all the structures with b external digits in the reminder of the sequence. Summing over k, we can get the desired result.

For the special case $l = 1$, we consider above recurrence relation. Submitting $l = 1$, we can get;

$$E_{n+1}(b,1) = 2E_n(b-1,1) + \sum_{k=1}^{n-m} \phi_{n-k}^*(1)E_{k-1}(b,1), n \geq m + 1.$$

Application of the identity $\phi_n^*(1) = 2S_n$, we can immediately get the following relation.

$$E_{n+1}(b,1) = 2E_n(b-1,1) + 2\sum_{k=1}^{n-m} S_{n-k}E_{k-1}(b,1), n \geq m + 1.$$

Here, $E_n(b,1)$ is the same as $E_n(b)$ used in Wang et al. (2008a). It is obvious that the result coincides with the Theorem 5.1.4 in Wang et al. (2008a).

Theorem 4

Let $E_n(l)$ denote the number of external digits in the set of all structures with n bases, then;

$$E_{n+1}(l) = 2[E_n(l) + \phi_n(l)] + \sum_{k=1}^{n-m-2l+2} \phi_{n-k}^*(l)E_{k-1}(l), n \geq m+2l-1.$$

$$E_n(l) = n2^n, n \leq m+2l-1.$$

Proof

The number $E_{n+1}(l)$ of external digits can be computed as follows:

i. Adding an unpaired base belonging to $\{A,U\}$ to a structure on *n* digits, we obtain $E_n(l)$ external digits plus the $\phi_n(l)$ newly added ones. In the same way, if the unpaired base belonging to $\{G,C\}$, we can get the same number. So if we add an unpaired base, we can get the total number $2[E_n(l) + \phi_n(l)]$,
ii. Inserting an additional pair, because there are no external digits in the newly bracketed part of length *k*, we have $\phi_{n-k}^*(l)$ times all the structures in the reminder of the sequence. Summing over *k*, we can get the desired result.

For the special case *l* = 1, we consider above recurrence relation. Submitting *l* = 1, we can get;

$$E_{n+1}(1) = 2[E_n(1) + \phi_n(1)] + \sum_{k=1}^{n-m} \phi_{n-k}^*(1)E_{k-1}(1), n \geq m+1;$$

Application of the identity $\phi_n^*(1) = 2S_n$, we can immediately get the following relation.

$$E_{n+1}(1) = 2[E_n(1) + \phi_n(1)] + 2\sum_{k=1}^{n-m} S_{n-k}(1)E_{k-1}(1), n \geq m+1.$$

Here $E_{k-1}(1)$ is the same as E_{k-1} used in Wang et al. (2008a), which coincides with the result obtained in Wang et al. (2008a).

Theorem 5

Let $U_{n+1}(l)$ denote the total number of unpaired bases in the set of all secondary structures on n vertices, then;

$$U_{n+1}(l) = 2[U_n(l) + \phi_n(l)] + \sum_{k=1}^{n-m-2l+2} [\phi_{n-k}^*(l)U_{k-1}(l) + 2\phi_{k-1}(l)U_{n-k}(l)], n \geq m+2l-1.$$

$$U_n(l) = n2^n, n \leq m+2l-1.$$

Proof

The number $U_{n+1}(l)$ of external digits can be computed as follows:

i. Adding an unpaired base belonging to $\{A,U\}$ to a structure on *n* digits, we obtain $U_n(l)$ unpaired bases plus the $\phi_n(l)$ newly added ones. In the same way, if the unpaired base belonging to $\{G,C\}$, we can get the same number. So if we add an unpaired base, we can get the total number $2[U_n(l) + A_n(l)]$,
ii. Inserting an additional pair, because there is no unpaired bases in the newly bracketed part of length *k*, we have $\phi_{n-k}^*(l)$ times all the unpaired digits in the reminder of the sequence plus all the unpaired digits in the newly bracketed part of length *k* times the number of structures that can be formed the reminder of the structure. Summing over *k*, we can get the desired result.

For the special case *l* = 1, we consider above recurrence relation. Submitting *l* = 1, we can get;

$$U_{n+1}(1) = 2[U_n(1) + \phi_n(1)] + \sum_{k=1}^{n-m} [\phi_{n-k}^*(1)U_{k-1}(1) + 2\phi_{k-1}(1)U_{n-k}(1)], n \geq m+1;$$

Application of the identity $\phi_n^*(1) = 2S_n$, we can immediately get the following relation.

$$U_{n+1}(1) = 2[U_n(1) + \phi_n(1)] + \sum_{k=1}^{n-m} [2S_{n-k}U_{k-1}(1) + 2\phi_{k-1}(1)U_{n-k}(1)], n \geq m+1;$$

Based on the fact that $\phi_n(l)$ is equal to S_n given in Wang et al. (2008a), $U_{k-1}(1)$ is the same as U_{k-1} used in Wang et al. (2008a). It is obvious that the result coincides with the Theorem 5.2.1 in Wang et al. (2008a).

The total number of internal vertices is denoted by V_n. It is clear that *Vn + En = Un*. By combining Theorems 5 and 6, we obtain;

$$V_{n+1}(l) = 2V_n(l) + \sum_{k=1}^{n-m-2l+2} [\phi_{n-k}^*(l)U_{k-1}(l) + 2\phi_{k-1}(l)U_{n-k}(l)], n \geq m+2l-1;$$

$$V_n(l) = 0, n \leq m+2l-1, V_0(l) = 0.$$

ASYMPTOTIC ENUMERATIONS

Based on the fact that the recursions stated above are complicated, it is hard to give simple closed form expressions for them. According to the asymptotic methods given in Bender (1974); Canfield, (1984) and Meir and Moon (1989), we now consider the asymptotic enumeration in terms of the generating functions. In this section, the symbol ~ has its usual meaning:

$$a(n) \sim b(n) \text{ means } \lim_{n\to\infty} \frac{a(n)}{b(n)} = 1.$$

Lemma 2

Suppose $y_n \geq 0$ *and* $y(x) = \sum_{n=0}^{\infty} y_n x^n$ *is of form* $y(x) = \beta(x) + g(x)(1-\frac{x}{\alpha})^{\omega},$ *where* $\alpha > 0$ *is real,* $\beta(x)$ *and* $g(x)$ *are analytic near* $\alpha,$ *and* ω *is real but not a nonnegative integer. If y(x) is analytic for* $|x| < \alpha$ *and* $x = \alpha$ *is the only singularity of y on its circle of convergence, then;*

$$y_n \sim \frac{g(\alpha)}{\Gamma(-\omega)} n^{-1-\omega} (\frac{1}{\alpha})^n.$$

By Lemma 2 which is a simple version of Darboux' theorem, I.L. Hofacker, P. Schuster and .F. Stadler give us the following proposition in order to be applied directly to the counting of RNA problem.

Lemma 3

Let $\Phi(x, y)$ be a polynomial in y and analytic in x for $|x| < \alpha + \delta$, $\delta > 0$. Suppose y fulfills the conditions of Lemma 2 with;

$$y(x) = \beta(x) + g(x)(1-\frac{x}{a})^{\frac{1}{2}},$$

Let the generating function $z(x) = \sum_{n=0}^{\infty} z_n x^n$ be of the form

$$z = \Phi(x, y).$$ Then, $$\lim_{n\to\infty} \frac{z_n}{y_n} = \Phi_y(\alpha, \beta(\alpha))$$ yields

$$z_n \sim \frac{g(\alpha)\Phi_y(\alpha, \beta)}{\Gamma(-\frac{1}{2})} n^{-\frac{3}{2}}(\frac{1}{\alpha})^n.$$

Denote by

$$\phi(x) = \sum_{n=0}^{\infty} \phi_n(l)x^n, \ \phi^*(x) = \sum_{n=0}^{\infty} \phi_n^*(l)x^n, \ \phi^{**}(x) = \sum_{n=0}^{\infty} \phi_n^{**}(l)x^n,$$

The generating functions. And we use the notations

$$\tau_m(x) = \sum_{k=0}^{m-1} (2x)^k, \ x\frac{d\tau_m(x)}{dx} = x\tau_m'(x) = \sum_{k=1}^{m-1} k(2x)^k.$$

Lemma 4 The generating functions $\phi(x), \phi^*(x)$ and $\phi^{**}(x)$ fulfill the following recursion.

$$\phi(x) = 1 + 2x\phi(x) + x^2\phi(x)\phi^*(x);$$

$$\phi^*(x) = \frac{2(2x^2)^{l-1}}{1-2x^2}(\phi^{**}(x) - \tau_m(x));$$

$$\phi(x) = x^2\phi^*(x) + \phi^{**}(x).$$

Lemma 5 The generating function ϕ is analytic at 0 and fulfills

$$2(2x^2)^l\phi(x) = b(x) - \sqrt{b^2(x) - 4(2x^2)^l[1-2x^2+(2x^2)^l]},$$

Where

$$b(x) = (1-2x)[1-2x^2+(2x^2)^l] + (2x^2)^l\tau_m(x).$$

Lemma 6

$$\phi_n(l) \sim \frac{-g(\alpha)}{2\sqrt{\pi}} n^{-3/2}(\frac{1}{\alpha})^n,$$

Where, α is the smallest positive solution of
$$p(x) = b^2(x) - 4(2x^2)^l[1-2x^2+(2x^2)^l] = 0$$

And

$$g(\alpha) = \frac{-1}{2(2\alpha^2)^l}\sqrt{-\frac{1}{\alpha}\frac{dp(x)}{dx}\bigg|_{\alpha}} \neq 0$$

The proofs of Lemmas 4, 5 and 6 are given in Wang et al. (2008b) in detail. Throughout the remainder of this paper we will assume that α denotes the solution of the following equation;

$$p(x) = 0 \text{ and } \beta = \frac{\sqrt{1 - 2\alpha^2 + (2\alpha^2)^l}}{(\sqrt{2}\alpha)^l}.$$

Theorem 6

$$\lim_{n \to \infty} \frac{\phi_n^*(l)}{\phi_n(l)} = \frac{1}{\alpha^2 \beta^2}.$$

Proof. From Lemma 4 we obtain

$$\phi^*(x) = \frac{(1 - 2x)\phi(x) - 1}{x^2 \phi(x)}.$$

Application of Lemma 3 completes the proof.

Theorem 7

$$\lim_{n \to \infty} \frac{\phi_n^{**}(l)}{\phi_n(l)} = \frac{1 - 2\alpha^2}{(2\alpha^2)^l \beta^2}.$$

Proof

From Lemma 4 we obtain;

$$\phi^{**}(x) = \frac{(1 - 2x^2)\left[(1 - 2x)\phi(x) - 1\right]}{(2x^2)^l \phi(x)} + \tau_m(x).$$

Application of Lemma 3 immediately yields the desired result.

Theorem 8

The number of structures with b components $J_n(b,l)$ fulfills;

$$\lim_{n \to \infty} \frac{J_n(b,l)}{\phi_n(l)} = \frac{b\left[\beta(1 - 2\alpha) - 1\right]^{b-1}}{\left[\beta(1 - 2\alpha)\right]^{b+1}}.$$

Proof

Let $J_b(x) = \sum_{n=0}^{\infty} J_n(b,l)x^n$ be the generating function for the number of secondary structures with exactly b components. We can get;

$$J_b = \frac{x^2}{(1 - 2x)}\phi^*(x)J_{b-1} = \left[\frac{x^2}{(1 - 2x)}\phi^*(x)\right]^b J_0,$$

and from $J_n(0,l) = 2^n$, we obtain $J_0 = \frac{1}{1 - 2x}$. Application of Lemma 3 and Lemma 4 we obtain the desired result.

Especially, let $l = 1$, and application of $\beta = \frac{1}{\sqrt{2}\alpha}$, then we can obtain;

$$\lim_{n \to \infty} \frac{J_n(b,1)}{\phi_n(1)} = \frac{b\left[\frac{1}{\sqrt{2}\alpha}(1 - 2\alpha) - 1\right]^{b-1}}{\left[\frac{1}{\sqrt{2}\alpha}(1 - 2\alpha)\right]^{b+1}}.$$

Simplifying above relation, we can obtain the result which is identical with the Theorem 3.3.2 in Wang et al. (2008b).

Theorem 9

The number of structures with b components, $E_n(b,l)$ fulfills;

$$\lim_{n \to \infty} \frac{E_n(b,l)}{\phi_n(l)} = (b+1)\frac{(2\alpha\beta)^b}{(1 + 2\alpha\beta)^{b+2}}.$$

Proof

Let $E_b(x) = \sum_{n=0}^{\infty} E_n(b,l)x^n$ be the generating function of the number of secondary structures with exactly b external digits. According to Theorem 3, we get the function equation;

$$E_b(x) = \left(\frac{2x\phi(x)}{1 + 2x\phi(x)}\right)^b E_0, \text{ where } E_0(x) = \frac{\phi(x)}{1 + 2x\phi(x)}.$$

By Lemma 3, the desired expression is yielded.

Especially, let $l = 1$, and application of $\beta = \frac{1}{\sqrt{2}\alpha}$, then we can obtain;

$$\lim_{n\to\infty}\frac{E_n(b,1)}{\phi_n(1)} = (b+1)\frac{(2\alpha\frac{1}{\sqrt{2\alpha}})^b}{(1+2\alpha\frac{1}{\sqrt{2\alpha}})^{b+2}} = (3-2\sqrt{2})(b+1)(2-\sqrt{2})^b,$$

which is identical with the Theorem 3.3.3 in (Wang et al., 2008b in Wang et al. (2008a).

Theorem 10

The number of external digits $E_n(l)$ fulfills;

$$\lim_{n\to\infty}\frac{E_n(l)}{\phi_n(l)} = 4\alpha\beta .$$

Proof

Let

$$E(x) = \sum_{n=0}^{\infty} E_n(l)x^n$$

be the generating function of the total number of external digits. The recursion can be brought to the form.

$$E_{n+1}(l) = 2\left[E_n(l)+\phi_n(l)\right] + \sum_{k=0}^{n-1}\phi_{n-k-1}^* E_k(l) - \sum_{k=0}^{m-1}\phi_k(l)E_{n-k-1}(l) .$$

Multiplying by x^{n+1} and summing over n yields

$$E(x) = 2xE(x) + 2x\phi(x) + x^2\phi^*(x)E(x).$$

According to Lemma 4, we get

$$E(x) = 2x\phi^2(x)$$

Application of Lemma 3 completes the proof.

Let

$$l = 1$$

and application of

$$\beta = \frac{1}{\sqrt{2\alpha}},$$

then we can obtain;

$$\lim_{n\to\infty}\frac{E_n(l)}{\phi_n(l)} = 4\alpha\beta = \frac{4\alpha}{\sqrt{2\alpha}} = 2\sqrt{2},$$

which is identical with the Theorem 3.3.5 in (Wang et al., 2008b).

Theorem 11

The number of components $I_n(l)$ fulfills;

$$\lim_{n\to\infty}\frac{I_n(l)}{\phi_n(l)} = 2\beta(1-2\alpha)-1$$

Proof

Let
$$I(x) = \sum_{n=0}^{\infty} I_n(l)x^n$$
be the generating function of the total number of components. We can get

$$I(x) = (1-2x)\phi^2(x) - \phi(x).$$

Application of Lemma 3 immediately yields the desired result.

Especially, let $l=1$, and application of $\beta = \dfrac{1}{\sqrt{2\alpha}}$, then we can obtain;

$$\lim_{n\to\infty}\frac{I_n(l)}{\phi_n(l)} = \frac{2}{\sqrt{2\alpha}}(1-2\alpha)-1 = \frac{\sqrt{2}}{\alpha}-2\sqrt{2}-1,$$

which is identical with the Theorem 3.3.6 in Wang et al. (2008a).

CONCLUSIONS

It is known that the stack structure is as important as the bonding loop structure in any *RNA* secondary structure. Most previous researches are only concerned with the influence of minimum length of bonding loop for the secondary structures. This paper, we consider not only the length of each bonding loop but the length of each stack based on a new arc representation given in (Wang et al., 2008a). The number of various elements of *RNA* secondary structures was also discussed. Furthermore, according to these recurrence relations, we make asymptotic analysis with the combinatorial technique of generating function. Considering the special case $l=1$,

Table 1. Comparison of asymptotic formulas with obtained results.

The number of various secondary structures	The asymptotic formulas of elements for any l	The asymptotic formulas for special case $l = 1$
$J_n(b)/S_n$	$\dfrac{b\left[\beta(1-2\alpha)-1\right]^{b-1}}{\left[\beta(1-2\alpha)\right]^{b+1}}$	$\dfrac{2\alpha^2 b}{(1-2\alpha)^2}(\dfrac{1-(2+\sqrt{2})\alpha}{1-2\alpha})^{b-1}$
$E_n(b)/S_n$	$(b+1)\dfrac{(2\alpha\beta)^b}{(1+2\alpha\beta)^{b+2}}$	$(3-2\sqrt{2})(b+1)(2-\sqrt{2})^b$
E_n/S_n	$4\alpha\beta$	$2\sqrt{2}$
I_n/S_n	$2\beta(1-2\alpha)-1$	$\dfrac{\sqrt{2}}{\alpha}-2\sqrt{2}-1$

the results obtained all coincided with previous results in (Wang et al., 2008ab). Finally, we present the asymptotic formulas compared with some results obtained in (Wang et al., 2008b) in Table 1.

REFERENCES

Bender EA (1974). Asymptotic methods in enumeration. SIAM Rev.16:485-515.
Canfield ER (1984). Remarks on an asymptotic method in combinatoric. J. Combin. Theory A 37:348-352.
Couzin J (2002). Breakthrough of the year: Small RNAs make big splash. Science 298:2296-2297.
Darboux G (1878). Memoir sur l'approximation des fonctions de tres grande nombres, et sur une classe etendu de developpements en serie. J. Math. Pure Appl. 4:5-56.
Doslic T, Svrtan D, Veljan D (2004). Enumerative aspects of secondary structures. Disc. Math. 285:67-82.
Doudna JA (2000). Structural genomics of RNA. Nat. Struct. Biol. 7:954-956.
Haslinger C, Stadler PF (1999). RNA structures with pseudo-knots: Graph-theoretical and combinatorial Properties. Bull. Math. Biol. 61:437-467.
Hofacker IL, Schuster P, Standler PF (1998). Combinatorics of RNA secondary structures. Disc. Appl. Math. 88:207-237.
Holley RR, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH, Penswich JR, Zamer A (1965). Structure of a ribonucleic acid. Science 147:1462-1465.
Howell JA, Smith TF, Waterman MS (1980). Computation of generating function for biological molecules. SIAM J. Appl. Math. 39:119-133.
Jin EY, Qin J, Reidys CM (2008a). Combinatorics of RNA structures with pseudoknots. Bull. Math. Biol. 70:45-67.
Jin EY, Reidys CM (2008b). Asymptotic enumeration of RNA structures with pseudoknots. Bull. Math. Biol.70:951-970.

Liao B, Wang TM (2004). General combinatorics of RNA secondary structure. Math. Biosci. 191:69-81.
Meir A, Moon JW (1989). On an asymptotic method in enumeration. J. Combin. Theory A 51:77-89.
Nebel M (2002). Combinatorial properties of RNA secondary structures. J. Comput. Biol. 9:541-573.
Nebel ME (2004). Investigation of the Bernoulli model for RNA secondary structures. Bull. Math. Biol. 66:925-964.
Nkwanta A (1997). Lattice Paths and RNA Secondary Structures. DIMACS Series in Discrete Mathematics and Theoretical Computer Science 34:137-147.
Rivas E, Eddy S (1999). A dynamic programming algorithm for RNA structure with pseudoknots. J. Mol. Biol. 285:2053-2068.
Rodland EA (2006). Pseudoknots in RNA secondary structures: representation, enumeration and prevalence. J. Comput. Biol. 13:1197-213.
Schmitt WR, Waterman MS (1994). Linear trees and RNA secondary structure. Disc. Appl. Math. 12:317-323.
Stein PR, Waterman MS (1978). On some new sequences generalizing the Catalan and Motzkin numbers. Disc. Math. 26:261-272.
Wang WW, Wang TM (2010). RNA secondary structures, Schroder paths and abab-free partitions. Utilitas Math. 81:41-52.
Wang WW, Zhang M, Wang TM (2008a). A new enumeration problem of RNA secondary structure. Utilitas Math. 77:155-165.
Wang WW, Zhang M, Wang TM (2008a). Asymptotic enumeration of RNA secondary structure. J. Math. Anal. Appl. 342:514-523.
Waterman MS (1978). Secondary structure of single-stranded nucleic acids. Adv. Math. Suppl. Stud. 1:167-212.
Waterman MS (1995). Introduction to Computational Biology: Maps, Sequences and Genomes. Chapman and Hall, London.
Waterman MS, Smith TF (1978). RNA secondary structure: A complete mathematical analysis. Math. Biosci. 42:257-266.
Waterman MS, Smith TF (1986). Rapid dynamic programming algorithms for RNA secondary structure. Adv. Appl. Math. 7:455-464.
Zuker M, Sankoff D (1984). RNA secondary structures and their prediction. Bull. Math. Biol. 46:591-621.