

Full Length Research Paper

A combinatorial enumeration problem of RNA secondary structures

Wenwen Wang^{1*} and Ming Zhang²

¹School of Sciences, China University of Mining and Technology, Xuzhou, 221008, P.R. China.

²School of Management, China University of Mining and Technology, Xuzhou, 221008, P.R. China.

Accepted 2 June, 2009

The total number of the RNA secondary structures with a limited length m in each hairpin loop, that is, $S_m(n)$ has already been studied. In this paper, the number of this kind of RNA secondary structures is discussed with the technique of combinatorics. As a result, the recurrence relations, generating functions, explicit expressions, and some other properties were obtained. Furthermore, a new method to compute $S_m(n)$ is given in the end.

Key words: Combinatorial enumeration, generating function, RNA secondary structure.

INTRODUCTION

In 1978, M.S. Waterman firstly gave a mathematical definition of RNA secondary structures (Stein and Waterman 1978). And in order to satisfy the need of research, RNA secondary structures are usually modeled by some discrete mathematic objects, which establish a connection between Discrete Mathematics and Computational Molecular Biology (Chen et al., 2005; Hofacker et al., 1998; Liu, 2006; Nebel, 2004; Nkwanta, 1997; Schmitt and Waterman 1994; Vernizzi et al., 2005). On the one hand, the skills of combinatorics enumeration have been successfully applied to the enumeration problems of RNA secondary structures (Doslic et al., 2004; Howell et al., 1980; Liao and Wang 2004; Nebel, 2002; Reginer, 1998; Vauchassade and Viennot, 1985; Waterman et al., 1978). On the other hand, the enumeration problems of RNA secondary structures have inspired some new interesting combinatorial questions (Chen et al., 2005; Liu, 2006; Schmitt and Waterman 1994; Stein and Waterman 1978). The research on the enumeration of RNA secondary structures becomes one of the hot topics in Computational Molecular Biology.

An RNA molecule is described by its sequences of bases, which is a string of A , U , G and C . Each entry in

such a string is called a base. However, it has been known for some time that bases can interact with one another: A can pair with U and C can pair with G , which is called Watson-Crick pairing rules. Of course, there exists GU base pair, which is neglected here. The fact that RNA is single-stranded makes the molecule fold. Furthermore, the nucleotide sequence uniquely determines the way the molecule folds and this is why we can attempt to predict RNA's structure by analyzing its sequence.

Generally, we consider an RNA molecule as a string of n characters $R = r_1 r_2 \cdots r_n$, such that $r_i \in \{A, U, G, C\}$. A secondary structure of the molecule is a collection S of pairs (r_i, r_j) of bases such that $1 \leq i \leq j \leq n$. If $(r_i, r_j) \in S$, in principle we should require that r_i be a complement to r_j and that $j - i \geq m$, for m a certain threshold (because it is known that an RNA molecule does not fold too sharply on itself (Waterman, 1995).

Here, a mathematical definition of *secondary structure* is given (Stein and Waterman 1978). RNA secondary structure need satisfy the following conditions: (1) each base can be paired with at most one other non-adjacent base; (2) r_i and r_j form a pair only if they satisfy the Watson-Crick pairing rules; (3) if r_i and r_j are paired,

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

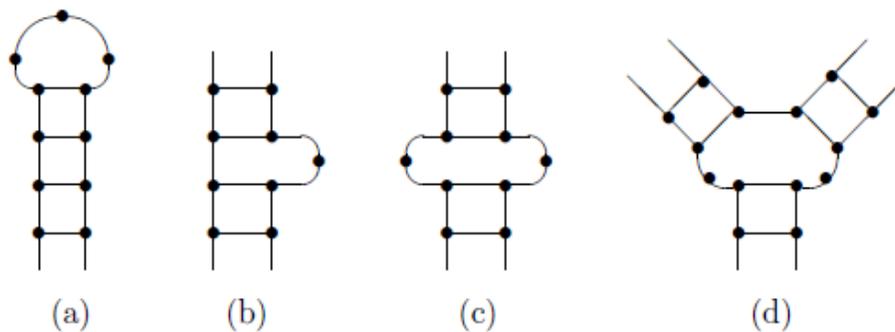


Figure 1. (a) A hairpin loop; (b) A convex loop; (c) A interior loop (d) A multiple loop.

then any pairing of r_k , $i < k < j$, must be with r_l , $i < l < j$. This last condition prohibits certain knot structures. We will call an edge (i, j) , $|i - j| \neq 1$ a *bond* (a base pair). A vertex i connected only to $i-1$ and $i+1$ will be called *unpaired* (a free base). 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, q) , $(p+1, q-1)$, \dots , $(p+k, q-k)$ such that neither $(p-1, q+1)$ nor $(p+k+1, q-k-1)$ is a base pair. $k+1$ is the length of the stack. (p, q) is the terminal base pair of the stack.

A *hairpin loop* consists of all unpaired vertices that are immediately interior to some base pair $(p+k, q-k)$. The number of unpaired vertices is the length of the hairpin loop.

A stack $[(p, q), \dots, (p+k, q-k)]$ 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 will be called a *component* of secondary structure. We will say that a structure on n vertices has a terminal base pair if $(1, n)$ is a base pair.

An *external vertex* is an unpaired vertex which does not belong to a loop, including hairpin loop, convex loop, interior loop, multiple loops, which are shown in the following Figure 1 respectively. A collection of adjacent external vertices is called an *external element*. If it contains the vertex 1 or n , it is called a *free end*; otherwise it is called a *joint*. An *internal vertex* is an unpaired vertex, which is interior to a base pair.

In Biology, the prediction of the shapes of biological molecules is an important problem. And its algorithms are combinatorial. The enumeration of RNA secondary structures is an interesting topic (Nebel, 2002; Stein and Waterman 1978; Waterman, 1978), and many recurrence formulas obtained in this topic play an important role in guiding for the predictions of RNA secondary structures (Rivas and Eddy 1999; Waterman and Smith, 1978,

1986; Zuker and Sankoff, 1984). Here we are only concerned with the enumeration problem of RNA secondary structures. First research on the enumeration of various constrained secondary structures of RNA molecules are due to M.S. Waterman (Howell et al, 1980; Schmitt and Waterman 1994; Stein and Waterman 1978; Waterman, 1978, 1995; Waterman and Smith, 1978, 1986). And about twenty years later, I.L. Hofacker with his coworkers established a framework about the enumeration of RNA secondary structures (Hofacker et al., 1998). They are mainly concerned with the number of secondary structures with limited length in each hairpin loop, in which the influence of pseudoknots are neglected. Of course, there are many other authors 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). Here, we discuss the enumeration problem based on Waterman's definition without pseudoknots.

In this paper, we will present the recursion formulas and the generating functions for the corresponding enumerations of constrained RNA secondary structures. Some properties and explicit formulas of them are also given there. Finally, a new method to compute the number of RNA secondary structures with a limited length in each hairpin loop is given, which is a generalization of the result given by M.S. Waterman.

MAIN RESULTS

From previous investigations, we know that all of recurrence relations about RNA secondary structures are established based on the following rule: a secondary structure on n vertices may be obtained from two cases: either n pairs with k base or n is a free base. According to this rule, M.S. Waterman with his coworkers obtain the recurrence relation about the number of secondary structures on n vertices with k base pairs and minimum size m of unpaired bases in each hairpin loop, denoted by $S_m(n, k)$. In the following paper, the parameter m has the same meaning.

Lemma 1. (Waterman, 1978)

$$S_m(n, k) = S_m(n-1, k) + \sum_{j=2}^{m+1} S_m(n-j, k-1) + \sum_{j=1}^{n-2} \sum_{i=0}^{k-1} S_m(j-1, i) S_m(n-j-1, k-1-i),$$

for $n \geq m+2, 0 \leq k \leq \lfloor \frac{n-m}{2} \rfloor$. (1)

Proof. It is clear that $S_m(0, k) = S_m(1, k) = \dots = S_m(m-1, k) = 0$, and

$$S_m(n, 0) = \begin{cases} 1, & \text{if } n \geq m; \\ 0, & \text{if } n < m. \end{cases}$$

Now we consider the sequence $[1; n]$, and there are two cases to be considered:

Case 1: n is a free end, the number of secondary structures is $S_m(n-1, k)$;

Case 2: n is paired with j , where $1 \leq j \leq n-2$. We'll consider three subcases:

If $1 \leq j \leq m$, there is $S_m(n-j-1, k-1)$ secondary structures; if $m+1 \leq j \leq n-m-1$, the number is $\sum_{i=0}^{k-1} S_m(j-1, i) S_m(n-j-1, k-1-i)$; if $n-m \leq j \leq n-2$, it doesn't satisfy the condition. Then a simple calculation completes the proof.

Considered a special case, that is, let $k = 1$ in above Lemma, that is to say, we consider the number of structures on n vertices with only one base pairs, then we can get a explicit formula.

Corollary 2.

$$S_m(n, 1) = \frac{1}{2}(n-m)(n-m-1), \quad n \geq 2m+2.$$

Proof. According to (1), we can easily obtain

$$S_m(n, 1) = S_m(n-1, 1) + \sum_{j=1}^m 1 + \sum_{j=m+1}^{n-m-1} 1 = S_m(n-1, 1) + (n-m-1)$$

It can be brought to the form

$$S_m(n, 1) = S_m(m+2, 1) + (m+2-m) + \dots + (n-m-1).$$

Observing $S_m(m+2) = 1$ yields the desired expression.

Theorem 3. Let $\phi(x, y) = \sum_{n, k \geq 0} S_m(n, k) x^n y^k$, then

$\phi(x, y)$ is analytic in a neighborhood of $(0, 0)$ and fulfills $x^2 y \phi^2 + (x + x^2 y + x^3 y + \dots + x^{m+1} y - 1) \phi + x^m = 0$.

Proof. By Lemma 1, we get

$$S_m(n, k) = S_m(n-1, k) + \sum_{j=2}^{m+1} S_m(n-j, k-1), \quad \text{for } m+2 \leq n \leq 2m+2.$$

Substituting the recursion for $S_m(n, k)$, we can get

$$\begin{aligned} \phi(x, y) &= \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} S_m(n, k) x^n y^k \\ &= \sum_{n=m}^{m+1} S_m(n, 0) x^n + \left(\sum_{n=1}^{\infty} \sum_{k=0}^{\infty} S_m(n-1, k) x^n y^k - x^{m+1} \right) \\ &\quad + \sum_{n=1}^{\infty} \sum_{k=0}^{\infty} \left[\sum_{j=m+1}^{n-m-1} \sum_{i=0}^{k-1} S_m(j-1, i) S_m(n-j-1, k-1-i) \right] x^n y^k, \end{aligned}$$

By the technique of generating function, it is straightforward to yield

$$\phi(x, y) = x^2 y \phi^2(x, y) + (x + x^2 y + x^3 y + \dots + x^{m+1} y) \phi(x, y) + x^m.$$

from which we get

$$x^2 y \phi^2(x, y) + (x + x^2 y + x^3 y + \dots + x^{m+1} y - 1) \phi(x, y) + x^m = 0.$$

then

$$\phi(x, y) = \frac{1-x - \sum_{i=2}^{m+1} x^i y - \sqrt{\left(1-x - \sum_{i=2}^{m+1} x^i y\right)^2 - 4x^{m+2} y}}{2x^2 y}, \quad (2)$$

where the correct sign of the solution follows from $S_m(0, 0) = \phi(0, 0) = 0$. Taylor expansion shows that $\phi(x, y)$ has an analytic continuation at the origin. By the use of the generating function of (2), we can get the exact value of $S_m(n, k)$ for any $m, n, k \geq 0$. For example; we obtain Tables 1, 2 and 3 by calculation.

A nice expression about $S(n, k)$ has already been treated by Schmitt and Waterman (1994). Here, by virtue of multinomial identity, we can get another explicit expression for $S(n, k)$. And the advantage of this method is that for any given m in $S_m(n, k)$, we can also obtain the explicit expression in the same way. However, according to the method given by Waterman, we can't obtain the expression about $S_m(n, k)$ except for $m = 1$. Our aim is to present a new way used to simplify the recurrence relations.

Table 1. $S_1(n, k)$

| | | | | | | | | | | | | | | | | |
|-------|---|---|---|---|---|----|----|----|-----|-----|-----|------|------|------|-------|-------|
| k/n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | | | 1 | 3 | 6 | 10 | 15 | 21 | 28 | 36 | 45 | 55 | 66 | 78 | 91 | 105 |
| | | | | | 1 | 6 | 20 | 50 | 105 | 196 | 336 | 540 | 825 | 1210 | 1716 | 2366 |
| | | | | | | | 1 | 10 | 50 | 175 | 490 | 1176 | 2520 | 4950 | 9075 | 15730 |
| | | | | | | | | | 1 | 15 | 105 | 490 | 1764 | 5292 | 13860 | 32670 |
| | | | | | | | | | | | 1 | 21 | 196 | 1176 | 5292 | 19404 |
| | | | | | | | | | | | | | 1 | 28 | 336 | 2520 |

Table 2. $S_2(n, k)$

| | | | | | | | | | | | | | | | | |
|-------|---|---|---|---|---|---|----|----|----|----|-----|-----|-----|------|------|------|
| k/n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | | | | 1 | 3 | 6 | 10 | 15 | 21 | 28 | 36 | 45 | 55 | 66 | 78 | 91 |
| | | | | | | 1 | 5 | 16 | 40 | 85 | 161 | 280 | 456 | 705 | 1045 | 1496 |
| | | | | | | | | 1 | 7 | 31 | 105 | 295 | 721 | 1582 | 3186 | 5985 |
| | | | | | | | | | | 1 | 9 | 51 | 219 | 771 | 2331 | 6244 |
| | | | | | | | | | | | | 1 | 11 | 76 | 396 | 1681 |
| | | | | | | | | | | | | | | 1 | 13 | 106 |

Table 3. $S_3(n, k)$

| | | | | | | | | | | | | | | | | |
|-------|---|---|---|---|---|---|---|----|----|----|----|-----|-----|-----|------|------|
| k/n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | | | | | 1 | 3 | 6 | 10 | 15 | 21 | 28 | 36 | 45 | 55 | 66 | 78 |
| | | | | | | | 1 | 5 | 15 | 36 | 75 | 141 | 245 | 400 | 621 | 925 |
| | | | | | | | | | 1 | 7 | 28 | 87 | 231 | 549 | 1177 | 2353 |
| | | | | | | | | | | | 1 | 9 | 45 | 171 | 549 | 1557 |
| | | | | | | | | | | | | | 1 | 11 | 66 | 296 |
| | | | | | | | | | | | | | | | 1 | 13 |

Theorem 4. Let $S(n, k)$ denote the total number of secondary structures on n vertices with k base pairs, then

$$S(n, k) = \frac{1}{2} \sum_{\substack{2v_1 + v_2 \leq n - 2k \\ v_2 + 2v_3 \leq k + 1 \\ v_1, v_2, v_3 \geq 0}} \frac{\binom{1}{2}_{n-v_1-v_2-v_3-k+1} (-2)^{n-k-2v_1-v_2-2v_3+1}}{(v_1!)(v_2!)(v_3)!(n-2k-2v_1-v_2)!(k+1-v_2-2v_3)!} \quad (3)$$

where $(x)_k = x(x-1)\dots(x-k+1)$ for any $x \in \mathbb{Z}$ and $k \in \mathbb{Z}$, $(x)_0 = 1$.

Proof. Let the generating function of $S(n, k)$ be

$$\psi(x, y) = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} S(n, k) x^n y^k. \text{ by setting } m = 1 \text{ in (2),}$$

we get

$$\begin{aligned} \psi(x, y) &= \frac{1-x-x^2-\sqrt{(1-x-x^2)^2-4x^3y}}{2x^2y} \\ &= \frac{1-x-x^2}{2x^2y} - \frac{1}{2x^2y} (1+x-2x-2x^2y-2x^3y+x^4y^2)^{\frac{1}{2}}. \end{aligned}$$

Expand the generating function above and we can find (3) is just the coefficient of $x^n y^k$.

Combining Equations (3) and the explicit expression of Theorem 2.2 (Stein and Waterman 1978), we can obtain the following identity:

$$\frac{1}{2} \sum_{\substack{2v_1+2v_2 \leq n-2k \\ v_2+2v_3 \leq k+1 \\ v_1, v_2, v_3 \geq 0}} \frac{\binom{1}{2}_{n-v_1-v_2-v_3-k+1} (-2)^{n-k-2v_1-v_2-2v_3+1}}{(v_1)!(v_2)!(v_3)!(n-2k-2v_1-v_2)(k+1-v_2-2v_3)!} = \frac{1}{k} \binom{n-k}{k+1} \binom{n-k-1}{k-1}.$$

It is obvious that our expression is less compact than previous expression, however, the connection between two expressions may be useful from combinatorial point of view.

Let $S(n)$ be the number of secondary structures for n points, then it asserts that $S(0) = 1$; and

$$S(n) = S(n-1) + \sum_{j=0}^{n-2} S(j)S(n-2-j), \quad n \geq 1 \tag{4}$$

(Howell et al., 1980). Moreover, $S(n)$ is bounded by a geometric growth rate when $n \geq 2$ (Waterman, 1978). The following explicit expression of $S(n)$ can be calculated in analogy to Theorem 4.

Theorem 5.

$$S(n) = \sum_{\substack{4v_1+3v_2+2v_3 \leq n+2 \\ v_1, v_2, v_3 \geq 0}} \frac{\binom{1}{2}_{n-3v_1-2v_2-v_3+2} (-2)^{n-4v_1-2v_2-v_3+1}}{(v_1)!(v_2)!(v_3)!(n+2-4v_1-3v_2-2v_3)!2^{v_3}}.$$

Proof. Let the generating function of $S(n)$ be $y(x) = \sum_{n \geq 0} S(n)x^n$. the algebraic function is

$x^2 y^2 - (1-x-x^2)y + x = 0$. by the initial condition, we can deduce that

$$y(x) = \frac{1-x-x^2}{2x^2} - \frac{1}{2x^2} (1+x^4-2x^3-x^2-2x)^{\frac{1}{2}}$$

Hence, we obtain the coefficient of x^n , which is just the explicit value of $S(n)$. After some simple calculation, it is easy to verify the recursion. In the following part, we will give a method to compute $S_m(n)$ with i base pairs. Let t be the number of secondary structures with i base pairs and limited length m for a molecule of length n , and

$$S_m^n = (S_{m0}^n, S_{m1}^n, \dots, S_{mi}^n, \dots),$$

Where

$$S_{mi}^n = \begin{cases} 1, & \text{if } i = 0; \\ t, & \text{if } i > 0. \end{cases}$$

We also define a shift operator ζ by $\zeta(a_0, a_1, a_2, \dots) = (0, a_0, a_1, a_2, \dots)$ and the Cauchy

product of a and b , by $c = a * b$, where $c_m = \sum_{i=0}^m a_i b_{m-i}$.

Theorem 6. S_m^n Satisfies the following relations:

If $1 \leq n \leq m+1$, $S_m^n = (1, 0, 0, \dots)$;

If $m+2 \leq n \leq 2m+1$, $S_m^n = S_m^{n-1} + \sum_{j=1}^m \zeta S_m^{n-j-1}$;

If $n \geq 2m+2$, $S_m^n = S_m^{n-1} + \sum_{j=1}^m \zeta S_m^{n-j-1} + \sum_{j=m+1}^{n-m-1} \zeta (S_m^{j-1} * S_m^{n-j-1})$.

Proof. (1) It can be easily obtained by the definition. (2) If $m+2 \leq n \leq 2m+1$, according to Lemma 1, we have

$$S_m(n, k) = S_m(n-1, k) + \sum_{j=2}^{m+1} S_m(n-j, k-1),$$

That is, $S_{mk}^n = S_{mk}^{n-1} + \sum_{j=1}^m S_{m, k-1}^{n-j-1}$. And by the definition, we

can see that $S_{m, k-1}^{n-j-1}$ is the $(k-1)$ th

Coordinate of S_m^{n-j-1} . Hence $S_m^n = S_m^{n-1} + \sum_{j=1}^m \zeta S_m^{n-j-1}$.

(3) When $n \geq 2m+2$, according to Lemma 1 we can obtain

$$S_m(n, k) = S_m(n-1, k) + \sum_{j=2}^{m+1} S_m(n-j, k-1) + \sum_{j=1}^{n-2} \sum_{i=0}^{k-1} S_m(j-1, i) S_m(n-j-1, k-1-i)$$

A simple calculation confirms

$$S_{mk}^n = S_{mk}^{n-1} + \sum_{j=2}^{m+1} S_{m, k-1}^{n-j-1} + \sum_{j=m+1}^{n-m-1} \left(\sum_{i=0}^{k-1} S_{mi}^{j-1} S_{m, k-i-1}^{n-j-1} \right).$$

The term in parentheses is the $(k-1)$ th coordinate of $S_m^{j-1} * S_m^{n-j-1}$ and the result follows.

By setting $m = 1$, we immediately obtain

$$S_1^0 = S_1^1 = S_1^2 = (1, 0, 0, 0, \dots),$$

$$S_1^3 = S_1^2 + \zeta S_1^1 = (1, 1, 0, 0, \dots),$$

and when $n \geq 4$,

$$S_1^n = S_1^{n-1} + \zeta S_1^{n-2} + \sum_{j=2}^{n-2} \zeta (S_1^{j-1} * S_1^{n-j-1}) = S_1^{n-1} + \sum_{j=1}^{n-2} \zeta (S_1^{j-1} * S_1^{n-j-1}),$$

Which coincides with the result in (Howell et al, 1980)?

As an illustration, we calculate $S_2^0, S_2^1, \dots, S_2^8$,

$$S_2^0 = S_2^1 = S_2^2 = S_2^3 = (1, 0, 0, 0, \dots); S_2^4 = (1, 1, 0, 0, \dots); S_2^5 = (1, 3, 0, 0, \dots); \\ S_2^6 = (1, 6, 1, 0, \dots); S_2^7 = (1, 10, 5, 0, \dots); S_2^8 = (1, 15, 16, 1, 0, \dots).$$

Hence, $S_2^8 = \sum_{i=0}^8 S_{2,i}^8 = 1 + 15 + 16 + 1 = 33$, which is the same as the result in (Hofacker et al, 1998).

Conclusions

The aim of this paper is to give the enumeration of RNA secondary structure by the technique of combinatorics. A remarkable formula about $S(n, k)$ has been obtained by Waterman. In this paper, we can get another explicit expression for $S(n, k)$ by virtue of multinomial identity. Although the result is not very informative, the method is worthy. Besides, for a given m in $S_m(n, k)$ we can also obtain the explicit expression in the same way. Moreover, the method which is pointed out in Theorem 6 may be useful to the prediction of RNA secondary structure. Here we give it just from the combinatorial point of view.

ACKNOWLEDGMENTS

We are grateful to the anonymous referees for their various valuable comments. This work is supported financially by "the Fundamental Research Funds for the Central Universities" and the Fundamental Research Funds for the Central Universities (2010LKX03).

REFERENCES

- Chen WYC, Deng EYP, Du RRX (2005). Reduction of m -regular noncrossing partitions. *Eur. J. Combin.* 26: 237-243.
- Doslic T, Svrtnan D, Veljan D (2004). Enumerative aspects of secondary structures. *Disc. Math.* 285: 67-82.
- Haslinger C, Stadler PF (1999). RNA Structures with Pseudo-Knots. *Bull. Math. Biol.* 61: 437-467.
- Hofacker IL, Schuster P, Stadler PF (1998). Combinatorics of RNA secondary structures. *Disc. Appl. Math.* 88: 207-237.
- 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.
- Liu CHL (2006). RNA secondary structure and bioloured ordered trees. *Ars Combin.* 81: 305-309.
- Nebel ME (2002). Combinatorial properties of RNA secondary structures. *J. Comput. Biol.* 9: 541-573.
- Nebel ME (2004). Investigation of the Bernoulli-model of RNA secondary structures. *Bull. Math. Biol.* 66: 925-964.
- Nkwanta A (1997). Lattice paths and RNA secondary structures. *DIMACS Ser. Disc. Math. Theor. Comput. Sci.* 34: 137-147.
- Reginer M (1998). Generating functions in computational biology, (invited to MABS'97).
- Rodland EA (2006). Pseudoknots in RNA secondary structures: representation, enumeration and prevalence. *J. Comput. Biol.* 13: 1197-213.
- Rivas E, Eddy S (1999). A dynamic programming algorithm for RNA structure with pseudoknots. *J. Mol. Biol.* 285: 2053-2068.
- Schmitt WR, Waterman MS (1994). Linear trees and RNA secondary structure. *Disc. Appl. Math.* 51: 317-323.
- Stein PR, Waterman MS (1978). On some new sequences generalizing the Catalan and Motzkin numbers. *Disc. Math.* 26: 261-272.
- Vernizzi G, Orland H, Zee A (2005). Enumeration of RNA structures by matrix models. *Phys. Rev. Lett.* 94: 168103.
- Vauchaussade de Chaumont M, Viennot XG (1985). Enumeration of RNA's secondary structures by complexity. *Math. Med. Biol. Lecture Notes Biomath.* 57: 360-365.
- Waterman MS (1978). Secondary structure of single-stranded nucleic acids. *Adv. Math. Suppl. Stud.* 1: 167-212.
- 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 methods for RNA secondary structure. *Adv. Appl. Math.* 7: 455-464.
- Waterman TF (1995). *Introduction to Computational Biology: Maps, Sequences and Genomes*, Chapman & Hall, London.
- Zuker M, Sankoff D (1984). RNA secondary structures and their prediction. *Bull. Math. Biol.* 46: 591-621.