General Combinatorics of RNA Hairpins and Cloverleaves

Bo Liao* and Tian-ming Wang

Department of Applied Mathematics, Dalian University of Technology, Dalian 116024, China

Received November 2, 2002

For an abstract single-strand RNA, a combinatorial analysis is given for two important structures, hairpins and cloverleaves. The total number of hairpins and cloverleaves of a given length with minimal hairpin loop length $m(m > 0)$ and with minimal stack length $l(l > 0)$ is computed, under the assumption that all base pairs can occur.

1. INTRODUCTION

Determining the shape a single-stranded RNA takes in solution is an important problem in computation biology; see ref 20. In ref 21 the problem of prediction is shown to be polynomial. The prediction algorithms are combinatorial, and the enumeration of secondary structure is a natural problem (see refs 4, 5, and 7). In these enumeration studies, the specific identities of the bases are ignored, and in effect all possible base pairs are allowed. In this way, the entire set of possible RNA secondary structures is studied.

Here we will focus on enumeration problems which are related to the secondary structure of RNA. This sort of study has a long history starting with the investigations of Waterman[7] who gave the first formal framework for the topic.[5] As shown in ref 4 the sequences arising in the enumeration of secondary structures which can occur under various reasonable restrictions may be considered as natural generalizations of the Catalan and Motzkin numbers. For most of the problems considered one finds similar decomposition patterns. This observation is used in ref 16 to describe algorithms and techniques for computing generating functions for certain RNA configurations. Based on the traditional approach, i.e. based on the determination of recurrence relation from decomposition properties of the objects, the authors of ref 1 obtain several enumeration results for restricted configurations of secondary structures. But all results were obtained previously under the assumption that the minimal loop length is 1 or the minimal stack length is 1. We will consider general combinatorics of the RNA secondary structure with minimal hairpin loop length $m(m > 0)$ and with minimal stack length l $(l > 0)$.

In this paper, we explore in detail the number of the two major secondary structures: hairpins and cloverleaves as a generation of the above papers.[1,7] The number of possible hairpins is $2^{n-m-2l} - 1$, the number of structures with b hairpins is $O(8/2^{b(m+2l+2)}b!(b\text{-}1)!\ n^{2(b-1)}2^n)$, and the number of cloverleaves with g hairpins is $O(n^g/g\ 2^{n-g(m+2l)-2})$.

In section 2 we introduce the definition of secondary structure and structural elements. Section 3 presents the recursion formulas of hairpins and their asymptotics. In section 4 we present the asymptotics of cloverleaves.

* Corresponding author e-mail: dragonbw@163.com.

2. SECONDARY STRUCTURES AND STRUCTURAL ELEMENTS

Definition 2.1 (Waterman[2]). A secondary structure is a vertex-labeled graph on n vertices with an adjacency matrix A fulfilling the following:

(i) $a_{i,i+1} = 1$ for $1 \le i \le n - 1$

(ii) For each i there is at most a single $k \ne i - 1, i + 1$ such that $a_{i,k} = 1$

(iii) If $a_{i,j} = a_{k,l} = 1$ and $i < k < j$ then $i < l < j$

We will call an edge (i, k), $|i - k| \ne 1$ a base pair. A vertex i connected only to $i - 1$ and $i + 1$ will be called unpaired. A vertex i is said to be interior the base pair (k, l), if $k < i < l$. If, in addition, there is no base pair (p, q) such that $k < p < i < q$, we will say that i is immediately interior to the base pair (k, l).

Definition 2.2.[1] Suppose A is the adjacency matrix for a secondary structure of size n. A secondary structure consists of the following structure element:

(i) 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. $k + 1$ is the length of the stack, and $(p - k, q + k)$ is the terminal base pair of the stack.

(ii) The sequence $i + 1, i + 2, ..., j - 1$ is a loop, if $i + 1, i + 2, ..., j - 1$ are all unpaired and $a_{i,j} = 1$. The pair (i, j) is said to the foundation of the loop, and the loop length is $j - i$.

(iii) A hairpin is the longest sequence $i + 1, i + 2, ..., j - 1$ containing exactly one loop such that $a_{i+1,j-1} = 1$ and $a_{i,j} = 0$. The paired points $i + 1$ and $j - 1$ will be called the foundation of the hairpin.

3. HAIRPINS

In this section, we derive the number of hairpins for an RNA of length n and the number of structures with b hairpins. As Figure 1 indicates, there are two major components to a hairpin: the loop (hairpin loop) and the stack. The constraint we impose is that any loop have m or more unpaired bases and any stack have l base pairs or more base pairs.

10.1021/ci020071c CCC: $25.00
Published on Web 06/17/2003

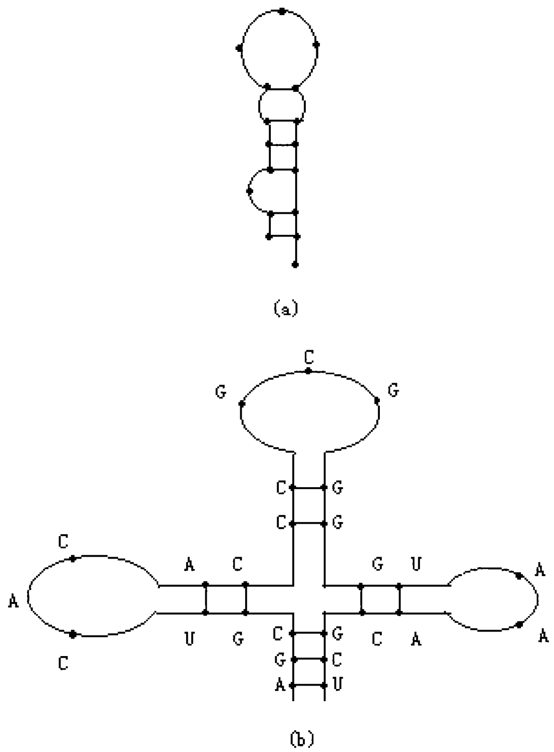

Figure 1. (a) Hairpin and (b) cloverleaf.

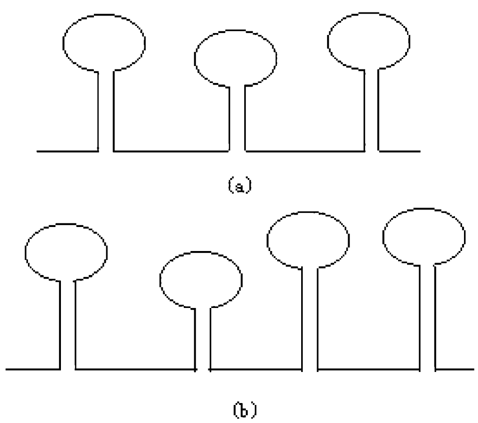

Figure 2. Cloverleaf "loops": (a) three hairpins and (b) more than three hairpins.

Lemma 1.[1] Suppose $y_n \geq 0$ and $y(x) = \Sigma_{n=0}^{\infty} y_n x^n$ is of the form

$$y(x) = \beta(x) + \left(1 - \frac{x}{\alpha}\right)^{\omega}$$

where $\alpha > 0$ is real, $\beta(x)$ is analytic near α, 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

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

where $f(n) \sim g(n)$ means $lim_{n\to\infty} f(n)/g(n) = 1$.

Lemma 2.[7] Let $K(a, b)$ be the number of stems with a bases on one side, b bases on the other, and the lead (or topmost) bases bound, then

$$K(a, b) = \binom{a + b - 2}{a - 1}$$

Theorem 1. Let $H_n(b)$ denote the number of structures with exactly b hairpins where the RNA secondary structures with minimal hairpin loop length $m (m > 0)$, $H_n(b)$ fulfills

$$H_{n+1}(b) = H_n(b) + \Sigma_{k=m}^{n-1} [\Sigma_{t=1}^{b} H_k(t) H_{n-k-1}(b - t) + H_{n-k-1}(b - 1)],\ n \geq m + 1$$

$$H_n(b) = \begin{bmatrix} 1 & b = 0, n \leq m \\ 0 & b \neq 0, n \leq m \end{bmatrix}$$

Proof. The number $H_{n+1}(b)$ of structures on $n + 1$ vertices with exactly b hairpins can be computed as follows: (i) adding an unpaired digit to any structure on n vertices, we obtain $H_n(b)$ structures with exactly b hairpins; (ii) inserting an additional pair $(1, k + 2)$ we have $H_k(t)$ times all the structures with exactly $b - t$ hairpins in the reminder of the sequence plus $H_{n-k-1}(b - 1)$, where the number of hairpins is unchanged by an enclosing substructure which exactly contains a base pair in an additional base pair. Summing over k, we can get the recursion.

Theorem 2. The number of structures with b hairpins fulfills

$$H_n(b) \sim \frac{8}{2^{b(m+4)} b!(b\text{-}1)!} n^{2(b-1)} 2^n,\ (n \to \infty)$$

Proof. Let $h_b(x) = \Sigma H_n(b) x^n$ denote the generating function. From recursion as above, we obtain

$$h_b(x) = x h_b(x) + x^2 \Sigma_{i=1}^{b} h_i(x) h_{b-i}(x) + \frac{x^2}{1 - x} h_{b-1}(x) - x^2 t_m(x) h_{b-1}(x)$$

and $h_0(x) = 1/1 - x$, and we obtain

$$h_b(x) = x h_b(x) + x^2\ \Sigma_{i=1}^{b-1}\ h_i(x) h_{b-i}(x) + \frac{x^2}{1 - x} h_b(x) + \frac{x^2}{1 - x} h_{b-1} - x^2 \frac{x^2(1 - x^m)}{1 - x} h_{b-1}(x)$$

Collecting all terms containing $h_b(x)$ yields

$$h_b(x)(1 - 2x) = x^2(1 - x) \Sigma_{i=1}^{b-1} h_i(x) h_{b-i}(x) + x^{m+2} h_{b-1}(x)$$

With the ansatz

$$h_b(x) = \left(\frac{x^{m+2}}{1 - x}\right)^b \frac{1}{(1 - 2x)^{2b-1}} \eta_b(x)$$

we find the following recursion for the polynomials $\eta_b(x)$

$$\eta_b(x) = x^2(1 - x)\ \Sigma_{i=1}^{b-1}\ \eta_i(x) \eta_{b-i}(x) + (1 - 2x)(1 - x) \eta_{b-1}(x),\ \eta_1(x) = 1$$

Lemma 1 now implies that the relevant singularity occurs at $x = {}^1/_2$ leaving us with

$$\eta_b\left(\frac{1}{2}\right) = \frac{1}{8}\ \Sigma_{i=1}^{b-1} \eta_i\left(\frac{1}{2}\right) \eta_{b-i}\left(\frac{1}{2}\right)$$

$$\eta_b\left(\frac{1}{2}\right) = \frac{1}{2^{3(b-1)}} C_{b-1}$$

From Lemma 1 we find now that

$$H_n(b) \sim \frac{C_{b-1}}{2^{3(b-1)}2^{b(m+1)}\Gamma(2b\text{-}1)} n^{2(b-1)}2^n, \ (n \to \infty)$$

A simple calculation completes the proof.

Theorem 3. Let $H(n)$ denote the number of hairpins for a string of length n with m or more bases in the loop, then

$$H(n) = 2^{n-m-1} - 1$$

Proof. Obviously $H(n) = H_n(1)$, from Theorem 1 we can obtain Suppose $h(x) = \Sigma_{n=m+1}^{\infty} H_n(1)$, for $H_n(1) = 0$, $n \le m$.

$$H_n(1) = H_{n-1}(1) + \Sigma_{k=m}^{n-2} [H_k(1) + 1]$$

$$= H_{n-1}(1) + \Sigma_{k=m}^{n-2} H_k(1) + n - 1 - m$$

$$= \Sigma_{k=m}^{n-1} H_k(1) + n - 1 - m$$

From the above-mentioned recursion we can obtain

$$h(x) = \frac{xh(x)}{1-x} + \Sigma_{n=m+1}^{\infty}(n-1-m)x^n$$

$$= \frac{h(x)x}{1-x} + \Sigma_{n=m+1}^{\infty} nx^n - \frac{(1+m)x^{m+1}}{1-x}$$

$$= \frac{h(x)x}{1-x} + \frac{x[(m+1)x^m(1-x) + x^{m+1}]}{(1-x)^2} - \frac{(1+m)x^{m+1}}{1-x}$$

$$= \frac{h(x)x}{(1-x)} + \frac{x^{m+2}}{(1-x)^2}$$

So we can obtain $h(x) = x^{m+1}[1/(1-2x) - 1/(1-x)]$. Comparing the coefficient of x^n, we can obtain

$$H_n(1) = 2^{n-m-1} - 1$$

So $H(n) = 2^{n-m-1} - 1$.

Theorem 4. Let $H_n(b, l)$ denote the number of structures with exactly b hairpins where the RNA secondary structures with minimal hairpin loop length $m(m > 0)$ and with minimal stack length l ($l \ge 1$), $H_n(b, l)$ fulfills

$$H_{n+1}(b, l) = H_n(b, l) + \Sigma_{k=m+2l-2}^{n-1} [\Sigma_{i=1}^{b} H_k(i, l) H_{n-k-1}(b-i, l) + H_{n-k-1}(b-1, l)], \ n \ge m+1$$

$$H_n(b, l) = \begin{bmatrix} 1 & b = 0, n \le m+2l \\ 0 & b \ne 0, n \le m+2l \text{ or } 0 < b < l \end{bmatrix}$$

Theorem 5. The number of structures with b hairpins where the RNA secondary structures with minimal hairpin loop length $m(m > 0)$ and with minimal stack length l ($l \ge 1$) fulfills

$$H_n(b, l) \sim \frac{8}{2^{b(m+2l+2)} b!(b\text{-}1)!} n^{2(b-1)}2^n, \ (n \to \infty)$$

Proof. Let $h_{b,l}(x) = \Sigma H_n(b, l)x^n$ denote the generating function. From recursion as above, we obtain

$$h_{b,l}(x) = xh_{b,l}(x) + x^2\Sigma_{i=1}^{b} h_{i,l}(x)h_{b-i,l}(x) + \frac{x^2}{1-x} h_{b-1,l}(x) - x^2 t_{m+2l-2}(x) h_{b-1,l}(x)$$

and $h_{0,l}(x) = 1/1 - x$, and we obtain

$$h_{b,l}(x) = xh_{b,l}(x) + x^2 \ \Sigma_{i=1}^{b-1} h_{i,l}(x)h_{b-i,l}(x) + \frac{x^2}{1-x} h_{b,l}(x) + \frac{x^2}{1-x} h_{b-1,l} - \frac{x^2(1-x^{m+2l-2})}{1-x} h_{b-1,l}(x)$$

Collecting all terms containing $h_{b,l}(x)$ yields

$$h_{b,l}(x)(1-2x) = x^2(1-x) \ \Sigma_{i=1}^{b-1} h_{i,l}(x)h_{b-i,l}(x) + x^{m+2l}h_{b-1,l}(x)$$

With the ansatz

$$h_{b,l}(x) = \left(\frac{x^{m+2l}}{1-x}\right)^b \frac{1}{(1-2x)^{2b-1}} \eta_{b,l}(x)$$

we find the following recursion for the polynomials $\eta_b(x)$

$$\eta_{b,l}(x) = x^2(1-x) \ \Sigma_{i=1}^{b-1} \eta_{i,l}(x)\eta_{b-i,l}(x) + (1-2x)(1-x)\eta_{b-1,l}(x), \eta_{1,l}(x) = 1$$

Lemma 1 now implies that the relevant singularity occurs at $x = {}^1/_2$ leaving us with

$$\eta_{b,l}\left(\frac{1}{2}\right) = \frac{1}{8} \ \Sigma_{i=1}^{b-1} \eta_{i,l}\left(\frac{1}{2}\right)\eta_{b-i,l}\left(\frac{1}{2}\right)$$

$$\eta_{b,l}\left(\frac{1}{2}\right) = \frac{1}{2^{3(b-1)}} C_{b-1}$$

From Lemma 1 we find now that

$$H_n(b, l) \sim \frac{C_{b-1}}{2^{3(b-1)}2^{b(m+2l-1)}\Gamma(2b\text{-}1)} n^{2(b-1)}2^n, \ (n \to \infty)$$

A simple calculation completes the proof.

Theorem 6. Let $H(n, l)$ denote the number of hairpins for a string of length n with m or more bases in the loop and with minimal stack length l, then

$$H(n, l) = 2^{n-m-2l+1} - 1$$

Proof. Obviously $H(n, l) = H_n(1, l)$, and from Theorem 4 we can obtain

$$H_n(1, l) = H_{n-1}(1, l) + \Sigma_{k=m+2l-2}^{n-2} [H_k(1, l) + 1]$$

$$= H_{n-1}(1, l) + \Sigma_{k=m+2l-2}^{n-2} H_k(1, l) + n - 2l - m + 1$$

$$= \ \Sigma_{k=m+2l-2}^{n-1} H_k(1, l) + n - m - 2l + 1$$

Suppose $h'(x) = \Sigma_{n=m+2l-1}^{\infty} H_n(1, l)$, for $H_n(1, l) = 0$, $n \le m + 2l - 2$. From the above-mentioned recursion we can obtain

$$h'(x) = \frac{xh'(x)}{1-x} + \Sigma_{n=m+2l-1}^{\infty}(n-m-2l+1)x^n$$

$$= \frac{h'(x)x}{1-x} + \Sigma_{n=m+2l-1}^{\infty} nx^n - \frac{(m+2l\text{-}1)x^{m+2l-1}}{1-x}$$

$$= \frac{h'(x)x}{1-x} + \frac{x\,[(m+2l-1)x^{m+2l-2}(1-x) + x^{m+2l-1}]}{(1-x)^2} - \frac{(m+2l\text{-}1)x^{m+2l-1}}{1-x}$$

$$= \frac{h(x)x}{(1-x)} + \frac{x^{m+2l}}{(1-x)^2}$$

So we can obtain $h'(x) = x^{m+2l-1}(1/1 - 2x - 1/1 - x)$. Comparing the coefficient of x^n, we can obtain

$$H_n(1, l) = 2^{n-m-2l+1} - 1$$

So $H(n, l) = 2^{n-m-2l+1} - 1$.

4. CLOVERLEAVES

Notice that a cloverleaf a modified hairpin with a stem as in a hairpin but with a "loop" that is a sequence of $g(g \ge 3)$ hairpins joined by the primary structure.

Theorem 7. The number of general "loop" structures with minimal hairpin loop length m, $L_g^m(\upsilon)$ satisfies

$$L_g^m(\upsilon) \sim 2^{\upsilon-g(m+2)}\upsilon^{g-1}, (\upsilon \to \infty)$$

Proof. Assume the loop of the cloverleaf has υ bases and k_i is bonded to j_i, $i = 1,\cdots, g$ where $1 \le k_1 \le j_1 < k_2 \le j_2 <\cdots < k_g \le j_g \le \upsilon$. Then $H(j_i - k_i - 1) + 1$ structures can form when $n \ge m + 1$, assume $j_i - k_i \ge m + 2$. The number of loop structures, $L_g^m(\upsilon)$, is then seen to be

$$L_g^m(\upsilon) = \Sigma_{j_g=g(m+3)}^{\upsilon}\Sigma_{k_g=(g-1)(m+3)+1}^{j_g-m-2}\cdots\Sigma_{j_t=t(m+3)}^{k_{t+1}-1}\Sigma_{k_t=j_t-1+1}^{j_t-m-2}\cdots \Sigma_{j_1=m+3}^{k_2-1}\Sigma_{k_1=1}^{j_1-m-2}(\Pi_{i=1}^{g}2^{j_i-k_i-m-2})$$

We claim

$$L_g^m(\upsilon) \sim 2^{\upsilon-g(m+2)}\upsilon^{g-1}, (\upsilon \to \infty)$$

The proof is by induction on g and m.

If $g = 3$, we prove $L_3^m(\upsilon) \sim 2^{\upsilon-3(m+2)}\upsilon^{g-1}(\upsilon \to \infty)$ by induction on m

If $m = 1$, after calculation we obtain

$$L_3^1(\upsilon) = \Sigma_{j_3=12}^{\upsilon}\Sigma_{k_3=9}^{j_3-3}\Sigma_{j_2=8}^{k_3-1}\Sigma_{k_2=5}^{j_2-3}\cdots\Sigma_{j_1=4}^{k_2-1}\Sigma_{k_1=1}^{j_1-3}(\Pi_{i=1}^{3}2^{j_i-k_i-3})$$

$$= \upsilon^2 2^{\upsilon-9} - 29\upsilon 2^{\upsilon-9} + 1092^{\upsilon-8} - \frac{\upsilon^3}{6} - 2\upsilon^2 - \frac{65\upsilon}{6} + 19$$

$$\sim \upsilon^2 2^{\upsilon-9}$$

Obviously the claim is correct when $m = 1$. Suppose the claim is true for$m = 2, 3, 4,\cdots, t$, then

$$L_3^{t+1}(\upsilon) = \Sigma_{j_3=3(t+4)}^{\upsilon}\Sigma_{k_3=2(t+4)+1}^{j_3-t-3}\Sigma_{j_2=2(t+4)}^{k_3-1}\Sigma_{k_2=t+5}^{j_2-t-3}\cdots\Sigma_{j_1=t+4}^{k_2-1}\Sigma_{k_1=1}^{j_1-t-3}(\Pi_{i=1}^{3}2^{j_i-k_i-t-3})$$

$$= 2^{-3}\Sigma_{j_3=3(t+4)}^{\upsilon}\Sigma_{k_3=2(t+4)+1}^{j_3-t-3}\Sigma_{j_2=2(t+4)}^{k_3-1}\Sigma_{k_2=t+5}^{j_2-t-3}\cdots\Sigma_{j_1=t+4}^{k_2-1}\Sigma_{k_1=1}^{j_1-t-3}(\Pi_{i=1}^{3}2^{j_i-k_i-t-2})$$

$$\sim 2^{-3}\Sigma_{j_3=3(t+3)}^{\upsilon}\Sigma_{k_3=2(t+3)+1}^{j_3-t-2}\Sigma_{j_2=2(t+3)}^{k_3-1}\Sigma_{k_2=t+4}^{j_2-t-2}\cdots\Sigma_{j_1=t+3}^{k_2-1}\Sigma_{k_1=1}^{j_1-t-2}(\Pi_{i=1}^{3}2^{j_i-k_i-t-2})$$

$$\sim 2^{-3}2^{\upsilon-3(t+2)}\upsilon^2$$

$$= 2^{\upsilon-3[(t+1)+2]}\upsilon^2$$

So we proved $L_3^m(\upsilon) \sim 2^{\upsilon-3(m+2)}\upsilon^2$, $(\upsilon \to \infty)$.

Suppose the claim is true for $g = 3, 4,\cdots, s - 1$, then

$$L_s^m(\upsilon) = \Sigma_{j_s=s(m+3)}^{\upsilon}\Sigma_{k_s=(s-1)(m+3)+1}^{j_s-m-2}2^{j_s-k_s-m-2}L_{s-1}^m(k_s-1)$$

$$\sim 2^{-(m+2)}\Sigma_{j_s=s(m+3)}^{\upsilon}\Sigma_{k_s=(s-1)(m+3)+1}^{j_s-m-2}2^{j_s-k_s}2^{k_s-1-(s-1)(m+2)}(k_s-1)^{s-2}$$

$$\sim 2^{-s(m+2)-1}\Sigma_{j_s=s(m+3)}^{\upsilon}\Sigma_{k_s=(s-1)(m+3)+1}^{j_s-m-2}2^{j_s}(k_s)^{s-2}$$

$$\sim 2^{-s(m+2)-1}(\Sigma_{j_s=s(m+3)}^{\upsilon}2^{j_s})(\Sigma_{k_s=(s-1)(m+3)+1}^{\upsilon-m-2}(k_s)^{s-2})$$

$$\sim 2^{-s(m+2)-1}[2^{\upsilon+1} - 2^{s(m+3)}]\upsilon^{s-1}$$

$$\sim 2^{\upsilon-s(m+2)}\upsilon^{s-1}$$

and the claim is proved.

Theorem 8. The number of general cloverleaves with minimal hairpin loop length $m > 0$, $C_g^m(n)$ satisfies

$$C_g^m(n) \sim \frac{n^g}{g}2^{n-g(m+2)-2}, (n \to \infty)$$

Proof.

$$C_g^m(n) = \Sigma_{k_2=1}^{n-2}\Sigma_{k1=0}^{n-k_2}L_g^m(k_2)K(k_1, n-k_1-k_2)$$

$$\sim \Sigma_{k_2=0}^{n}\Sigma_{k_1=0}^{n-k_2}L_g^m(k_2)K(k_1, n-k_1-k_2)$$

$$\sim \Sigma_{k_2=0}^{n}\Sigma_{k1=0}^{n-k_2}2^{k_2-g(m+2)}k_2^{g-1}2^{n-k_2-2}$$

$$= 2^{-g(m+2)-2}\Sigma_{k_2=0}^{n}2^n k_2^{g-1}$$

$$\sim 2^{n-g(m+2)-2}\Sigma_{k_2=0}^{n}(k_2)^{\underline{g-1}}$$

$$= \frac{n^g}{g}2^{n-g(m+2)-2}$$

where $x^{\underline{m}} = x(x-1)\cdots(x-m+1)$.

Theorem 9. The number of general "loop" structures with minimal hairpin loop length m and minimal stack length l, $L_{g,l}^{m}(v)$, satisfies

$$L_{g,l}^{m}(v) \sim 2^{v-g(m+2l)} v^{g-1}, \ (v \to \infty)$$

Proof. Assume the loop of the cloverleaf has v bases and k_i is bonded to j_i, $i = 1, \cdots, g$ where $1 \leq k_1 \leq j_1 < k_2 \leq j_2 < \cdots < k_g \leq j_g \leq v$. Then $H(j_i - k_i - 1) + 1$ structures can form when $n \geq m + 2l$, assume $j_i - k_i \geq m + 2l$. The number of loop structures, $L_{g,l}^{m}(v)$, is then seen to be

$$L_{g,l}^{m}(v) = \Sigma_{j_g=g(m+2l+1)}^{v} \Sigma_{kg=(g-1)(m+2l+1)+1}^{j_g-m-2l} \cdots \Sigma_{j_t=t(m+2l+1)}^{k_{t+1}-1} \Sigma_{k_t=j_{t-1}+1}^{j_t-m-2l} \cdots \Sigma_{j_1=m+2l+1}^{k_2-1} \Sigma_{k_1=1}^{j_1-m-2l} (\Pi_{i=1}^{g} 2^{j_i-k_i-m-2l})$$

We claim $L_{g,l}{}^{m}(v) \sim 2^{v-g(m+2l)} v^{g-1}$, $(v \to \infty)$, the proof is by induction on l.

If $l = 1$, $L_{g,1}{}^{m}(v) = L_{g}^{m}(v) \sim 2^{v-g(m+2)} v^{g-1}$, obviously, the claim is correct when $l = 1$. Suppose the claim is true for $l = 1, 2, 3, ..., s$, then

$$L_{g,s+1}{}^{m}(v) = \Sigma_{j_g=g(m+2s+3)}^{v} \Sigma_{kg=(g-1)(m+2s+3)+1}^{j_g-m-2s-2} \cdots \Sigma_{j_t=t(m+2s+3)}^{k_{t+1}-1} \Sigma_{k_t=j_{t-1}+1}^{j_t-m-2s-2} \cdots \Sigma_{j_1=m+2s+3}^{k_2-1} \Sigma_{k_1=1}^{j_1-m-2s-2} (\Pi_{i=1}^{g} 2^{j_i-k_i-m-2s-2})$$

$$= 2^{-2g} \Sigma_{j_g=g(m+2s+3)}^{v} \Sigma_{k_g=(g-1)(m+2s+3)+1}^{j_g-m-2s-2} \cdots \Sigma_{j_t=t(m+2s+3)}^{k_{t+1}-1} \Sigma_{k_t=j_{t-1}+1}^{j_t-m-2s-2} \cdots \Sigma_{j_1=m+2s+3}^{k_2-1} \Sigma_{k_1=1}^{j_1-m-2s-2} (\Pi_{i=1}^{g} 2^{j_i-k_i-m-2s})$$

$$\sim 2^{-2g} \Sigma_{j_g=g(m+2s+1)}^{v} \Sigma_{k_g=(g-1)(m+2s+1)+1}^{j_g-m-2s} \cdots \Sigma_{j_t=t(m+2s+1)}^{k_{i+1}-1} \Sigma_{k_t=j_{t-1}+1}^{j_t-m-2s} \cdots \Sigma_{j_1=m+2s+1}^{k_2-1} \Sigma_{k_1=1}^{j_1-m-2s} (\Pi_{i=1}^{g} 2^{j_i-k_i-m-2s})$$

$$\sim 2^{-2g} 2^{v-g(m+2s)} v^{g-1}$$

$$= 2^{v-g[m+2(s+1)]} v^{g-1}$$

and the claim is proved.

Theorem 10. The number of general cloverleaves with minimal hairpin loop length m and minimal stack length l, $C_{g,l}{}^{m}(n)$, satisfies

$$C_{g,l}^{m}(n) \sim \frac{n^{g}}{g} 2^{n-g(m+2l)-2}, \ (n \to \infty)$$

Proof.

$$C_{g,l}^{m}(n) = \Sigma_{k_2=1}^{n-2} \Sigma_{k_1=0}^{n-k_2} L_{g,l}^{m}(k_2) K(k_1, n-k_1-k_2)$$

$$\sim \Sigma_{k_2=0}^{n} \Sigma_{k_1=0}^{n-k_2} L_{g,l}^{m}(k_2) K(k_1, n-k_1-k_2)$$

$$\sim \Sigma_{k_2=0}^{n} \Sigma_{k_1=0}^{n-k_2} 2^{k_2-g(m+2l)} k_2^{g-1} 2^{n-k_2-2}$$

$$= 2^{n-g(m+2l)-2} \Sigma_{k_2=0}^{n} k_2^{g-1}$$

$$\sim 2^{n-g(m+2l)-2} \Sigma_{k_2=0}^{n} k_2^{\underline{g-1}}$$

$$= \frac{n^{g}}{g} 2^{n-g(m+2l)-2}$$

REFERENCES AND NOTES

(1) Ivo, L.; Hofacker, P. S.; Stadler, P. F. Combinatorics of RNA secondary structures. *Discr. Appl. Math.* **1998**, *88*, 207−237.
(2) Penner, R. C.; Waterman, M. S. Spaces of RNA secondary structures. *Adv. Math.* **1993**, *101*, 31−49.
(3) Schnitt, W. R.; Waterman, M. S. Linear trees and RNA secondary structure. *Discr. Math.* **1994**, 12.
(4) Stein, P. R.; Waterman, M. S. On some new sequences generalizing the Catalan and Motzkin numbers. *Discr. Math* **1978**, *26*, 261−272.
(5) Waterman, M. S. Secondary structure of single-standed nucleic acids. In *Advanced Mathematical Supplement Studies 1*; Hall: London, 1995.
(6) Waterman, M. S. *Introduction to Computational Biology: Maps, Sequences and Genomes*; Chapman & Hall: London, 1995.
(7) Waterman, M. S.; Smith, T. F. Combinatorics of RNA hairpins and cloverleaves. *Studies Appl. Math* **1978**, *60*, 91−96.
(8) Waterman, M. S.; Smith, T. F. RNA secondary structure: a complete mathematical analysis. *Math. Biosci.* **1978**, *42*, 257−266.
(9) Hsich, W. N. Proportions of irreducible diagrams. *Studies Appl. Math* **1973**, *52*, 227−283.
(10) Kleitman, D. Proportions of irreducible diagrams. *Studies Appl. Math* **1970**, *49*, 297−299.
(11) Stein, P. R. On a class of linked diagrams, I. Enumeration. *J. Combin. Theory A* **1978**, *24*, 357−366.
(12) Stein, P. R.; Everett, C. J. On a class of linked diagrams, II. Asymptotics. *Disc. Math* **1978**, *22*, 309−318.
(13) Touchard, J. Surune problene de configurations et sur les fractions continues. *Can. J. Math* **1952**, *4*, 2−25.
(14) Odlyzko, A. M. Asymptotic enumeration methods. In *Handbook of combinatorics*; Graham, R. L., Grotschel, M., Lovasz, L., Eds.; Elsevier: Amsterdam, 1995; Vol. II, pp 1021−1229.
(15) Meir, A.; Moon, J. W. On an asymptotic method in enumeration. *J. Combin. Theory A* **1989**, *51*, 77−81.
(16) Howell, J. A.; Smith. T. F.; Waterman, M. S. Computation of generating functions for biological. *SIAM. J. Appl. Math* **1980**, *39*, 119−133.
(17) Bender, E. A. Asymptotic methods in enumeration. *SIMA Rev.* **1974**, *16*, 485−515.
(18) Regnier, M.; Tahi, F. Enumeration and asymptotics in computational biology. In *Mathematical Analysis for Biological Sequences Workshop;* Trondheim, Norway, 1996.
(19) Vienmt, X. G.; de chaumont, M. v. Enumeration of RNA's secondary structures by complexity. In *Mathematics in Medicine and Biology, Lect. Notes in Biomath*; Capasso, V., Grosso, E., Paver-Fontana, S. L., Eds.; Springer: Berlin, 1985; Vol. 57, pp 360−365.
(20) Zuker, M.; Sankoff, D. RNA secondary structures and their prediction. *Bull. Math. Biol.* **1984**, *46*(4), 591−621.
(21) Waterman, M. S.; Smith, T. F. Rapid dynamic programming algorithms for RNA Secondary structure. *Adv. Appl. Math.* **1986**, *7*, 455−464.

CI020071C