김일성종합대학학보 (자연과학) 주체103(2014)년 제60권 제7호
JOURNAL OF KIM IL SUNG UNIVERSITY (NATURAL SCIENCE) Vol. 60 No. 7 JUCHE103(2014).

일정한 제한밀에서의 리보핵산2차구조의 개수평가

우승식, 한선희

론문에서는 조합수학을 리용하여 유전학분야에서의 기본연구대상인 리보핵산2차구조의 개수평가에 관한 연구를 하였다.

데핵산은 리보핵산을 통하여 단백질합성을 지령하기때문에 어떤 종류의 단백질이 몇개 생기는가 하는것은 그것에 대한 정보를 가지고있는 리보핵산이 몇개 있는가에 관계된다.

따라서 리보핵산의 개수를 평가하는것은 유전학에서 매우 중요하다.

선행연구[2]에서는 정점수와 염기쌍수, 블로크수, 머리빈침수가 주어진 기초우에서 리보핵산2차구조의 개수에 대한 반복관계식을 제기하였으며 선행연구[3]에서는 2진나무의 개수렬에 대한 생성함수로부터 리보핵산2차구조의 개수렬에 대한 생성함수를 유도하였다.

또한 선행연구[4]에서는 포화된 리보핵산2차구조와 리보핵산합성2차구조의 개수평가에 대한 반복관계식을 제기하였다.

우의 결과들은 n이 충분히 커질 때 계산량이 방대해지며 어떤 특수한 제한조건밀에서는 리보핵산2차구조의 개수평가를 쉽게 할수 없다는것을 보여준다.

론문에서는 우와 같은 제한성을 극복하기 위하여 리보핵산2차구조의 몇개 파라메터들이 주어진 기초우에서 리보핵산2차구조개수에 관한 닫긴공식을 제기하였다.

1. 일정한 제한을 가진 나무의 개수평가

먼저 일정한 제한을 가지는 순서있는 표식뿌리나무의 개수를 평가하였다.

정의 1 다음과 같은 세가지 조건을 만족시키는 이웃행렬 $A=(a_{i,\ j})_{n\times n}$을 가지는 정점표식된 그라프를 리보핵산2차구조라고 부르고 RS로 표시한다.

① $a_{i,\ i+1}=1$, $1\le i\le n-1$

② 매 i에 대하여 $a_{i,\ k}=1$인 $k\ne i\pm1$이 하나 존재한다.

③ $a_{i,\ j}=a_{k,\ l}=1$이고 $i<k<j$이면 $i\le l\le j$이다.

정의 2 RS를 이웃행렬 $A=(a_{i,\ j})_{n\times n}$를 가지는 리보핵산2차구조라고 하자.

이때 릉 $(i,\ k)$ $(k-i>1)$를 RS의 염기쌍이라고 부르며 두 염기쌍 $(i,\ j)$, $(k,\ l)$에 대하여 릉 $(i,\ k)$ 혹은 릉 $(j,\ l)$이 존재할 때 두 염기쌍은 서로 이웃이라고 부른다.

또한 정점 $i-1$과 $i+1$과만 련결된 정점 i를 쌍짓지 않은 정점이라고 부르며 $a_{i,\ j}=1$인 정점 $i\ne j\pm1$이 존재하면 j를 쌍지은 정점이라고 부른다.

염기쌍 $(k,\ l)$에 대하여 $k<i<l$인 정점 i를 이 염기쌍에 대한 아낙정점이라고 부른다. 특히 염기쌍 $(k,\ l)$에 대한 아낙정점 i에 대하여 $k<p<i<q<l$인 염기쌍 $(p,\ q)$가 존재하지 않는다면 정점 i를 염기쌍 $(k,\ l)$에 대한 강한 아낙정점이라고 부른다.

정의 3 RS의 염기쌍렬 $(p-k,\ q+k)$, $(p-k+1,\ q+k-1), \cdots,$ $(p,\ q)$ 에 대하여 $(p-k-1,\ q+k+1)$ 과 $(p+1,\ q-1)$ 이 둘 다 염기쌍이 아닐 때 이 염기쌍렬을 RS의 사다리라고 부르고 이때 염기쌍 $(p-k,\ q+k)$ 를 이 사다리의 말단염기쌍이라고 부른다.

사다리 $(p,\ q)$, $(p+1,\ q-1), \cdots,$ $(p+k,\ q-k)$ 에 대하여 $p=1$ 혹은 $q=n$ 혹은 두 정점 $p-1$ 과 $q+1$ 이 그 어떤 염기쌍의 아낙정점으로도 되지 않을 때 이 사다리를 말단사다리라고 부르며 말단사다리의 말단염기쌍 $(p,\ q)$ 로 둘러막힌 부분구조를 RS의 블로크라고 부르고 $B[p,\ q]$ 로 표시한다.

어떤 염기쌍 $(p,\ q)$ 에 대하여 그것의 강한 아낙정점들로 이루어진 모임으로서 쌍짓지 않은 정점을 적어도 하나 포함하거나 혹은 거기에 속하는 강한 아낙정점들로 이루어진 염기쌍들의 개수가 2이상인 모임을 염기쌍 $(p,\ q)$ 에 대한 고리라고 부른다. 이때 염기쌍 $(p,\ q)$ 를 이 고리의 닫긴 염기쌍이라고 부른다.

고리를 이루는 쌍짓지 않은 정점들의 개수를 고리의 길이라고 부른다.

그 어떤 고리에도 속하지 않는 쌍짓지 않은 정점을 고립정점이라고 부른다.

정의 4 염기쌍 $(p,\ q)$ 를 닫긴염기쌍으로 가지는 고리에서 닫긴염기쌍 $(p,\ q)$ 에 대한 강한 아낙정점들로 이루어진 염기쌍들의 개수보다 하나 큰 수를 이 고리의 차수라고 부른다. 또한 차수가 1인 고리를 머리빈침고리(또는 단순고리), 차수가 2보다 큰 고리를 다중고리, 차수가 2인 고리로서 이 고리의 닫긴염기쌍이 그것에 대한 강한 아낙정점으로 이루어진 염기쌍과 서로 이웃일 때 이 고리를 불룩고리라고 부르고 그렇지 않은 고리를 내부고리라고 부른다.

정의 5 염기쌍 $(p,\ q)$ 에 대하여 $p+1$ 을 시작정점, $q-1$ 을 끝정점으로 하는 리보핵산 부분2차구조를 생각하였을 때 이 부분2차구조의 고립정점을 염기쌍 $(p,\ q)$ 로 둘러싸인 고립정점, 블로크를 염기쌍 $(p,\ q)$ 로 둘러싸인 블로크라고 부른다.

정의 6 뿌리가진 나무에서 매 내부정점의 자식들을 왼쪽에서 오른쪽으로 가면서 순서화한 나무를 순서있는 뿌리나무라고 부르고 순서있는 뿌리나무의 매 정점에 표식을 붙인것을 순서있는 표식뿌리나무라고 부르며 뿌리가 아닌 모든 정점들의 높이가 1인 순서있는 표식뿌리나무를 작은 나무라고 부른다.

정의 7 순서있는 뿌리나무 T에서 어떤 내부정점 v 의 자식정점들중에 아낙정점이 하나도 없으면 정점 v 를 뿌리로 하는 부분나무를 단순가지라고 부른다.

어떤 v 에 대하여 그것의 자식정점중에 내부정점이 꼭 하나 있다고 할 때 그것이 v 의 자식모임에서 제일 오른쪽 또는 제일 왼쪽에 있으면 정점 v 를 뿌리로 하는 부분나무를 T의 불룩가지라고 부르고 그렇지 않으면 내부가지라고 부른다.

또한 v 의 자식정점들중에 내부정점이 둘이상 있으면 정점 v 를 뿌리로 하는 부분나무를 다중가지라고 부른다.

보조정리[1] k개의 내부정점을 가진 n개의 정점우에서의 순서있는 표식뿌리나무는 정점모임 $\{1,\ 2,\ \cdots,\ n+k-1\}$ 우에서 1～n 까지의 번호로 표식된 정점을 뿌리로 가지는 k개 작은 나무들로 이루어진 수림과 1∶1대응된다.

정리 1 정점이 $n-k+1$ 개이고 내부정점이 $k+1$ 개, 단순가지가 p 개(매 단순가지는 적어도 m 개이상의 잎을 가짐.)인 순서있는 표식뿌리나무개수는 다음과 같다.

$$\begin{cases} (n+1)!, & k=0 \\ \dfrac{(n-k+1)!}{k}\dbinom{k}{p-1}\dbinom{k}{p}\dbinom{n-pm}{2k}, & k\ge 1 \end{cases}$$

증명 정점이 $n-k+1$개이고 내부정점이 $k+1$개인 순서있는 표식뿌리나무는 보조정리로부터 정점모임 $\{1, 2, \cdots, n+1\}$ 우에서 $1\sim n-k+1$까지의 번호로 표식된 정점을 뿌리로 가지는 $k+1$개의 작은 나무들로 이루어진 수림과 1 : 1 대응된다.

편리상 $n-k+2\sim n+1$까지의 번호로 표식된 정점을 별표붙은 정점이라고 하자.

이때 정리의 조건을 만족시키는 나무들의 개수를 구하는 대신에 아래의 조건을 만족시키는 $k+1$개의 작은 나무들로 이루어진 수림들의 구성방법수를 구하면 된다.

정점개수는 $n+1$(그중 k개는 별표붙은 정점), p개 작은 나무는 매개가 적어도 m개 이상의 자식정점들을 포함하고 나머지 $k+1-p$개 작은 나무에만 별표붙은 정점들이 존재한다.

① $k=0$인 경우 나무개수는 $(n+1)!$이다.

② $k\ge 1$일 때 $k+1$개의 별표붙지 않은 뿌리들의 선택방법수는 $\dbinom{n-k+1}{p}\dbinom{n-k+1-p}{k+1-p}$이며 k개의 별표붙은 정점들을 $k+1-p$개 나무에 비지 않게 하는 배치방법수는 방정식 $x_1+x_2+x_3+\cdots+x_{k+1-p}=k \quad (x_i\ge 1,\ i=\overline{1,\ k+1-p})$를 만족시키는 옹근수풀이개수 $\dbinom{k-1}{k-p}$와 같다. 그런데 표식까지 고려하면 $k!\dbinom{k-1}{k-p}$이다.

③ 나머지 $n-2k$개의 별표붙지 않은 정점들의 배치방법수는 방정식

$$x_1+x_2+x_3+\cdots+x_{2k+1}=n-2k\ \ (x_1,\ x_2,\cdots,\ x_p\ge m,\ \ x_{p+1},\ x_{p+2},\cdots,\ x_{2k+1}\ge 0)$$

을 만족시키는 옹근수풀이개수 $\dbinom{n-pm}{2k}$와 같다. 표식을 고려하면 $(n-2k)!\dbinom{n-pm}{2k}$이다.

따라서 우의 결과들을 종합하면

$$\binom{n-k+1}{p}\binom{n-k+1-p}{k+1-p}k!\binom{k-1}{k-p}(n-2k)!\binom{n-pm}{2k}=\frac{(n-k+1)!}{k}\binom{k}{p-1}\binom{k}{p}\binom{n-pm}{2k},\ k\ge 1$$

따라서 정리의 결과가 나온다.(증명끝)

정리 2 정점이 $n-k+1$개이고 내부정점이 $k+1$개, 단순가지가 p개(매 단순가지는 적어도 m개이상의 잎을 가짐.), 불룩가지가 적어도 r개인 순서있는 표식뿌리나무의 개수는 다음과 같다.

$$\begin{cases} (n+1)!, & k=0 \\ \dfrac{2^r(n-k+1)!}{p}\dbinom{k}{p-1}\dbinom{k-1-r}{k-p-r}\dbinom{n-pm-2r}{2k-r}, & k\ge 1 \end{cases}$$

정리 3 정점이 $n-k+1$개이고 내부정점이 $k+1$개, 단순가지가 p개(매 단순가지는 적어도 m개이상의 잎을 가짐.), 높이 1인 잎이 b개, 가지가 l개인 순서있는 표식뿌리나무의 개수는 다음과 같다.

$$\begin{cases} (n+1)!, & k=p=0 \\ (n-k+1)!\binom{b+k}{b}\binom{n-k-b-km-1}{k-1}, & k=p\geq 1 \\ \dfrac{l(n-k+1)!}{k}\binom{b+l}{b}\binom{k}{p}\binom{k-l-1}{k-p-1}\binom{n-mp-b-l-1}{2k-l-1}, & k\geq p+1\geq 2 \end{cases}$$

증명 정점이 $n-k+1$개이고 내부정점이 $k+1$개인 순서있는 표식뿌리나무는 보조정리로부터 정점모임 $\{1, 2, \cdots, n+1\}$ 우에서 $1\sim n-k+1$까지의 번호로 표식된 정점을 뿌리로 가지는 $k+1$개의 작은 나무들로 이루어진 수림들과 1∶1 대응된다.

따라서 정리의 조건을 만족시키는 나무들의 개수를 구하는 대신에 아래의 조건을 만족시키는 $k+1$개의 작은 나무들로 이루어진 수림들의 구성방법수를 구하면 된다. 즉 정점개수는 $n+1$개(그중 k개는 표식있는 정점), 그중 p개 작은 나무는 매개가 적어도 m개이상의 자식정점들을 포함하고 나머지 $k+1-p$개 작은 나무에만 별표붙은 정점들이 존재하는데 그중 1개 나무는 제일 큰 별표붙은 정점을 꼭 포함하며 l개의 별표붙은 정점들과 b개의 별표붙지 않은 정점들이 존재한다.

① $k=0$인 경우 나무개수는 $(n+1)!$이다.

② $k=p\geq 1$일 때는 별표붙은 작은 나무가 하나밖에 존재하지 않으며 그 나무에는 $l+b$개의 잎이 존재하여야 한다.

③ $k+1$개의 별표붙지 않은 뿌리들의 선택방법수는 $(n-k+1)\binom{n-k}{k}$이다.

④ $k+1$개 뿌리들에 대한 자식정점들의 배치방법수는

$$(n-2k-b)!\binom{n-2k}{b}\binom{n-2k-b-pm+p-1}{p-1}(b+k)!.$$

따라서 우의 결과들을 종합하면

$$(n-k+1)\binom{n-k}{k}(n-2k-b)!\binom{n-2k}{b}\binom{n-2k-b-pm+p-1}{p-1}(b+k)!=(n-k+1)!\binom{b+k}{b}\binom{n-k-b-km-1}{k-1}.$$

또한 $k\geq p+1\geq 2$일 때 $k+1$개의 별표붙지 않은 뿌리들의 선택방법수는 다음과 같다.

$$(n-k+1)\binom{n-k}{p}\binom{n-k-p}{k-p}$$

⑤ 마지막위치에 있는 제일 큰 별표붙은 정점을 포함하는 뿌리에 대한 자식정점들의 배치방법수에 대하여 고찰하자.

제일 큰 별표붙은 정점은 이미 선택되였다고 하면 $l-1$개의 별표붙은 정점들과 b개의 별표붙지 않은 정점들을 선택할수 있는 방법수는 표식까지 고려하여

$$\binom{k-1}{l-1}\binom{n-2k}{b}(b+l)!.$$

⑥ 나머지 k개 뿌리에 대한 자식정점들의 배치방법수를 보면 $k-l$개의 별표붙은 정점들을 $k-p$개 뿌리들에 비지 않게 배치하는 방법수는 표식까지 고려하여

$(k-l)!\binom{k-l-1}{k-p-1}$ 이며 $n-2k-b$ 개의 별표붙지 않은 정점들을 k 개 뿌리에 배치하는 방법수는 방정식

$$x_1+x_2+x_3+\cdots+x_{2k-l}=n-2k-b-l\ \ (x_1,\ x_2,\ \cdots,\ x_p\ge m,\ \ x_{p+1},\ x_{p+2},\ \cdots,\ x_{2k-l}\ge 0)$$

을 만족시키는 옹근수풀이개수 $\binom{n-mp-b-l-1}{2k-l-1}$ 과 같다.

따라서 우의 결과들을 종합하면 다음과 같다.

$$(n-k+1)\binom{n-k}{p}\binom{n-k-p}{k-p}\binom{k-1}{l-1}\binom{n-2k}{b}(b+l)!(k-l)!\binom{k-l-1}{k-p-1}\binom{n-mp-b-l-1}{2k-l-1}=$$

$$=\frac{l(n-k+1)!}{k}\binom{b+l}{l}\binom{k}{p}\binom{k-l-1}{k-p-1}\binom{n-mp-b-l-1}{2k-l-1}$$

따라서 정리의 결과가 증명된다.(증명끝)

2. 몇가지 리보핵산2차구조의 개수평가

1) 순서있는 뿌리나무모임과 리보핵산2차구조모임사이의 대응관계

정리 4 모든 리보핵산2차구조들의 모임은 순서있는 뿌리나무들의 모임과 1：1대응된다.

증명 먼저 2차구조 RS를 순서있는 뿌리나무 T에 대응시키는 과정은 다음과 같다.

① 점 하나를 새로 취하고 그것을 뿌리 x로 놓는다.

② 2차구조 RS의 고립정점, 블로크들을 결정한다.

③ 2차구조 RS의 왼쪽부터 오른쪽으로 가면서 순서를 고려하여 고립정점들과 블로크들을 뿌리 x의 자식정점으로 놓는다.

이때 고립정점들은 잎에 대응시키고 블로크들은 나무의 아낙정점에 대응시킨다.

④ x의 내부자식정점 y에 대응되는 블로크 $B[i,\ i+k]$에 대하여 점 $i,\ i+k$와 그에 붙은 릉을 제거한 그라프를 RS′로 놓고 이에 대하여 과정 ②를 반복한다.

⑤ 2차구조 RS의 모든 정점들이 선택되였으면 끝내며 이때 얻어지는 나무 T는 2차구조 RS에 대응되는 순서있는 뿌리가진 나무이다.

반대로 순서있는 T를 리보핵산 2차구조 RS에 대응시키는 과정은 다음과 같다.

① 나무 T의 뿌리 x의 자식정점모임 $S(x)$를 결정한다.

② 자식정점모임 $S(x)$를 왼쪽에서부터 오른쪽으로 훑으면서 임의의 $v\in S(x)$에 대하여 v가 잎이면 2차구조 RS의 고립정점에 대응시키고 v가 내부마디점이면 블로크 $B[p,\ q]$에 대응시킨다.

③ v가 내부마디점일 때 v의 자식정점모임 $S(v)$를 ②에 의하여 얻어진 고립정점들과 블로크들을 염기쌍 $(p,\ q)$로 둘러싸인 고립정점과 블로크들에 대응시킨다.

④ 나무의 모든 정점들이 취해질 때까지 반복한다.

이때 얻어지는 리보핵산2차구조RS는 나무 T에 대응되는 2차구조이다.(증명끝)

2) 몇가지 리보핵산2차구조의 개수평가

정리 5 정점이 n 개이고 k 개의 염기쌍과 p 개의 머리빈침고리(매 머리빈침고리의 길이는 적어도 m 이상)를 가지는 리보핵산2차구조수는 다음과 같다.

$$\begin{cases} 1, & k=0 \\ \dfrac{1}{k}\begin{pmatrix} k \\ p-1 \end{pmatrix}\begin{pmatrix} k \\ p \end{pmatrix}\begin{pmatrix} n-pm \\ 2k \end{pmatrix}, & k \ge 1 \end{cases}$$

정리 6 정점이 n 개이고 k 개의 염기쌍, p 개의 머리빈침고리(매 머리빈침고리의 길이는 적어도 m 이상)와 적어도 r 개의 불룩고리를 가지는 리보핵산2차구조수는

$$\begin{cases} 1, & k=0 \\ \dfrac{2^r}{p}\begin{pmatrix} k \\ p-1 \end{pmatrix}\begin{pmatrix} k-1-r \\ k-p-r \end{pmatrix}\begin{pmatrix} n-pm-2r \\ 2k-r \end{pmatrix}, & k \ge 1 \end{cases}$$

정리 7 정점이 n 개이고 k 개의 염기쌍, p 개의 머리빈침고리(매 머리빈침고리의 길이는 적어도 m 이상), b 개의 고립정점, l 개의 블로크를 가지는 리보핵산2차구조수는

$$\begin{cases} 1, & k=p=0 \\ \begin{pmatrix} b+k \\ b \end{pmatrix}\begin{pmatrix} n-k-b-km-1 \\ k-1 \end{pmatrix}, & k=p \ge 1 \\ \dfrac{l}{k}\begin{pmatrix} b+l \\ b \end{pmatrix}\begin{pmatrix} k \\ p \end{pmatrix}\begin{pmatrix} k-l-1 \\ k-p-1 \end{pmatrix}\begin{pmatrix} n-mp-b-l-1 \\ 2k-l-1 \end{pmatrix}, & k \ge p+1 \ge 2 \end{cases}.$$

참 고 문 헌

[1] Y. C. Chen et al.; Proc. Natl. Acad. Sci., 87, 9635, 1990.

[2] I. L. Hofacker et al.; Discrete Appl. Math., 88, 207, 1998.

[3] M. E. Nebel; J. Comput. Biol., 9, 541, 2002.

[4] Y. Emma; Discrete Appl. Math., 158, 25, 2010.

주체103(2014)년 3월 5일 원고접수

A Research about Computing the Number of RNA Secondary Structure

U Sung Sik, Han Son Hui

We compute the number of RNA secondary structure that is very important object in genetics by using ordered trees.

Key words: RNA secondary structure, ordered tree