

## 仮想核酸配列インタラクション反応系の平衡状態解析

松本典子<sup>†1</sup> 小林 聡<sup>†1</sup>

核酸配列の設計問題は、注意深く設計された配列は望まない構造を形成しないので、DNA 計算アルゴリズムの実験を成功させるために非常に重要である。配列設計の問題を解くためには、核酸配列のインタラクション反応の平衡状態を解析するアルゴリズムが必要である。本編では、配列インタラクション反応系にある制限を加えることにより、平衡状態を計算するための従来手法よりも効率のよい列挙アプローチを提案する。

### Equilibrium Analysis on Virtual Nucleic Acid Strands Interaction

NORIKO MATSUMOTO <sup>†1</sup> and SATOSHI KOBAYASHI <sup>†1</sup>

The problem of nucleic acid strand design is very important for DNA computing algorithms to be successfully executed in bio-lab experiments, since carefully designed strands will not form unwanted structures. In order to solve strand design problem, we need an efficient algorithm for analysing equilibrium of interacting nucleic acid strands. In this paper, by imposing some restrictions on reaction systems, we will propose an enumeration approach to compute equilibrium of interacting strands, which is more efficient than previous methods.

#### 1. はじめに

分子計算アルゴリズムを実現するためには、塩基配列設計が非常に重要である。この塩基配列設計には、実験を行い予測通りの反応が起きるかどうかを検証する必要があるが、実際に実験を行うには費用も手間もかかってしまう。そこで、設計した塩基配列が予測通りに反応を行うかどうか計算機で解析することが重要となる。本研究では、仮想塩基配列がインタ

ラクションする場合に高速に平衡状態を解析する手法を提案する。

#### 2. 関連研究と本研究の目的

Dirks らの研究では、複数の RNA 分子がインタラクションし、構造を形成する反応系の平衡状態を計算する手法を提案している ([1])。彼らの手法では、会合する全ての配列の組み合わせに関して分配関数を計算し、分配関数の値を係数に用いた目的関数を最小化する凸計画問題を解くことにより、平衡状態を高速に計算することに成功している。しかし、彼らの手法は配列の本数に関して組み合わせ爆発を起こしてしまうという問題がある。

この問題を数え上げの仕方を工夫することで回避しているのが川上らの研究である ([2])。川上らは、マルチプルループとヘアピン構造が現れない構造に限定することで凸計画問題を解く際の変数の個数を配列の本数に関して多項式個に抑えることに成功している。本研究では、川上らの結果をもとに、さらに変数の個数を削減できる場合があることを示す。

さらに、本研究では核酸配列を一般化した仮想塩基配列のインタラクション反応系を解析の対象とする ([3])。仮想塩基を用いた反応系の設計は、多くの DNA 計算の研究で採用されている。仮想塩基の導入により研究成果がより広い応用に用いられる可能性があると考ええる。

仮想塩基を導入している関連研究には、萩谷らの Virtual DNA に対するシミュレーションがある ([3])。しかし、彼らの研究においても、仮想塩基配列のインタラクション反応において組み合わせ爆発が起きるといった課題がある。

#### 3. 対称列挙法による平衡状態解析

本研究では、対称列挙法 ([4]) によって平衡状態解析を行う。

対称列挙法では、 $M$  を分子の有限集合とし、 $M$  の要素が結合してさまざまな複合体を形成する反応を考える。この時、生成されるすべての複合体の集合を有限集合  $\mathcal{A}$  とする。分子  $x \in M$  と複合体  $X \in \mathcal{A}$  に対し、 $\#_x(X)$  によって  $X$  に出現する  $x$  の個数を表す。濃度分布は  $[\ ]$ ,  $[\ ]_1$ ,  $[\ ]_2$  などの記号で表される。例えば、 $\mathcal{A}$  の濃度分布  $[\ ]$  と  $X \in \mathcal{A}$  に対し、 $[X]$  は  $X$  の濃度を表す。反応の開始時点においては、分子の集合  $M$  が初期濃度分布  $[\ ]_0$  で提供されるものとする。また、 $X \in \mathcal{A}$  に対し、 $E(X)$  は  $X$  の自由エネルギーを表すものとする。このように定義される化学反応系  $\mathcal{P}$  において、次の FEMP1 という最小化問題を考える。

<sup>†1</sup> 電気通信大学

The University of Electro-Communications

自由エネルギー最小化問題 1(FEMP1)

$$\text{最小化: } FE_1(P) = \sum_{X \in A} E(X) \cdot [X] + \sum_{X \in A} [X](\log[X] - 1) \quad (1)$$

$$\text{制約条件: } \begin{aligned} [X] &\geq 0 & (\forall X \in A) \\ \sum_{X \in A} \#_x(X) \cdot [X] &= [x]_0 & (\forall X \in A) \end{aligned}$$

FEMP1 の最小値を与える濃度分布  $[\ ]$  が平衡状態を与えることが示されている ([4]).

### 3.1 グラフによる数え上げ

RNA や DNA の二次構造の自由エネルギーは、その局所構造（ヘアピンやバルジループなど）の自由エネルギーの総和を求めることにより得られる。このような局所性は、平衡状態を求める問題においても、何か役立つであろうという考えから、対称列挙法では局所性を「グラフによる構造の列挙」という考え方で定式化する。

まず、アサイクリックな有向グラフ  $G = (V, E_g)$  を考える。ここで、 $V$  は頂点集合、 $E_g$  は有向辺の集合である。入る辺を持たない頂点の集合  $V_0$ 、出る辺を持たない頂点の集合を  $V_f$  で表す。 $V_0$  の要素を初期頂点、 $V_f$  の要素を最終頂点と呼ぶ。各頂点  $v \in V$  に対して、初期頂点から  $v$  への経路と  $v$  からの最終頂点への経路が存在するとき、 $G$  は無駄が無いという。

グラフ  $G$  初期頂点から最終頂点への経路集合を  $PT(G)$  で表す。グラフによる複合体の列挙とは、「 $PT(G)$  から  $A$  への上への写像  $\psi$  を考える」ということである。つまり、 $G$  の初期頂点から最終頂点への経路を数え上げることで複合体をすべて数え上げる。このとき、異なる経路により同じ複合体を重複して数え上げることも許される。ただし、その場合、3.2 節で説明する「対称性」に関する条件を満たさなければならない。

このような数え上げを行うことで、局所的な構造は  $G$  の辺として捉えることができる。しかし、辺を局所構造として捉えるためには、局所構造を構成する分子の種類とその自由エネルギーを辺に割当しなければならない。したがって、 $G$  は次に述べるような重み付きグラフにする必要がある。

$G$  には 2 種類の関数（重み関数）が付随している。1 つは  $e$  という  $E_g$  から  $\mathbf{R}$  (実数の集合) への関数である。辺  $e \in E_g$  に対して、 $\epsilon(e)$  は辺  $e$  が対応する局所構造の自由エネルギーを表す。2 つ目の関数は、 $\sigma_x$  という  $E_g$  から  $Z_+$  への関数である。ここで、 $x$  は  $M$  の要素であり、 $Z_+$  は非負整数の集合である。分子  $x \in M$  と辺  $e \in E_g$  に対して、 $\sigma_x(e)$  は  $e$  が対応する局所構造に含まれる  $x$  の個数を表す。

パス  $\gamma \in PT(G)$  および  $X \in A$  に対し、 $r_\gamma = |\psi^{-1}(\psi(\gamma))|$  および  $r_X = |\psi^{-1}(X)|$  と定義する。 $r_X$  は  $X$  が対応しているパスの個数を表し、 $r_\gamma$  はパス  $\gamma$  が表す構造に対応しているパスの個数を表す。反応系  $P$ 、グラフ  $G$ 、写像  $\psi$  の組  $S = (P, G, \psi)$  が与えられたとき、修正された自由エネルギー  $E_r : A \rightarrow \mathbf{R}$  を以下のように定義する。

$$E_r(X) = E(X) + \log r_X \quad (2)$$

この関数  $E_r$  は数え上げの重複による影響を補正する項  $\log r_X$  を加えた自由エネルギー関数である。

以下に、「グラフによって複合体を列挙する」ことの定義を形式的に述べる。会合反応系  $P$ 、重み付きでアサイクリックな無駄のない有向グラフ  $G$ 、 $PT(G)$  から  $A$  への上への写像  $\psi$  を考える。 $S = (P, G, \psi)$  は、以下の条件をすべての  $\gamma \in PT(G)$  に対して満たすとき列挙スキームであるという。

$$E_r(\psi(\gamma)) = \sum_{e \in E_g \text{ s.t. } e \in \gamma} \epsilon(e) \quad (3)$$

$$\#_x(\psi(\gamma)) = \sum_{e \in E_g \text{ s.t. } e \in \gamma} \sigma_x(e) \quad (4)$$

### 3.2 列挙スキームの対称性

$V \cup E_g$  から  $V \cup E_g$  への 1 対 1 写像  $\phi$  で、 $\phi(V) = V$  および  $\phi(E_g) = E_g$  を満たすものを考える。ここで、 $\phi(PT(G)) = PT(G)$  を満たすとき、 $\phi$  をパス写像と呼ぶ。また、 $\Phi$  をパス写像の集合とする。 $\Phi$  が関数の合成に関して群をなすとき、 $\Phi$  はパス写像群であるという。

辺  $e$  の始点を  $t(e)$ 、終点を  $h(e)$  で表す。パス写像  $\psi$  は、すべての  $e \in E_g$  に対して、 $\phi(t(e)) = t(\phi(e))$  かつ  $\phi(h(e)) = h(\phi(e))$  が成り立つとき、同型写像であるという。また、すべての  $e \in E_g$  に対して、 $\phi(t(e)) = h(\phi(e))$  かつ  $\phi(h(e)) = t(\phi(e))$  が成り立つとき、逆同型写像であるという。列挙スキーム  $S = (P, G, \psi)$  に対して、以下の条件を満たすパス写像群  $\Phi$  が存在するとき、 $S$  は準自己同型であるという。

$\Phi$  の全ての要素は同型写像であるか、逆同型写像である。

$$\Phi(\gamma) = \psi^{-1}(\psi(\gamma)) \quad (\forall \gamma \in PT(G))$$

**定理 1.** 列挙スキーム  $S = (P, G, \psi)$  が準自己同型であるならば、FEMP1 を  $G$  の辺の個数だけ変数を持つ自由エネルギー最小化問題 FEMP2 に帰着することができる。

帰着される最小化問題は以下で与えられる。ここで、 $\Phi$  は  $S$  のパス写像群であり、 $v_{in}$  で

頂点  $v$  に入る辺の集合,  $v_{out}$  で頂点  $v$  から出てゆく辺の集合を表す.

自由エネルギー最小化問題 2(FEMP2)

最小化:  $FE_2(P, (\omega_e | e \in E_g)) \stackrel{\text{def}}{=}$

$$\sum_{e \in E_g} \epsilon(e) \cdot \omega_e + \sum_{e \in E_g} \omega_e (\log \omega_e - 1) - \sum_{v \in V - V_0 - V_f} \omega_v (\log \omega_v - 1) \quad (5)$$

$$\sum_{e \in E_g} \#_x(e) \cdot \omega_e = [x]_0 \quad (\forall x \in M)$$

$$\text{制約条件: } \sum_{e \in v_{in}} \omega_e = \sum_{e \in v_{out}} \omega_e \quad (\forall v \in V - V_0 - V_f)$$

$$\omega_e = \omega_{\phi(e)} \quad (\forall \phi \in \Phi, \forall e \in E_g)$$

$$\omega_e \geq 0 \quad (\forall e \in E_g)$$

ここで,  $\omega_v = \sum_{e \in v_{in}} \omega_e$  である.

反応系に対して, 定理 1 の条件を満たすような列挙スキームを構成することが目標となる.

#### 4. 川上の研究の概要 [2]

川上らは, 対称列挙法を, 複数の RNA 鎖がインタラクションしてマルチプルループ, ヘアピンループ, シュードノットが存在しない RNA 二次構造を形成するような反応系に対して, 列挙スキームを構成することを提案した.

川上らの列挙グラフ  $G = (V, V_0, V_f, E_g, \epsilon, \sigma)$  において,  $V$  はノードの有限集合,  $V_0$  は初期ノードの集合,  $V_f$  は最終ノードの集合,  $E_g$  は辺の有限集合,  $\epsilon$  は局所構造の自由エネルギーを与える辺の重み関数,  $\sigma$  は配列の本数を与える辺の重み関数である.

$S$  を RNA 鎖の集合とし,  $\alpha \in S$  に対して,  $l_\alpha$  で  $\alpha$  の長さを表す. また,  $1 \leq i \leq l_\alpha$  に対し,  $\alpha[i]$  は  $\alpha$  の 5' 末端側から  $i$  番目の塩基そのものを表す.

川上らの研究においては, 複数の RNA 鎖がインタラクションして構造を形成する. よって, このような構造を考えるために,  $m$  本の RNA 鎖が 5' 末端側から 3' 末端に向かって並べられている配列  $\alpha_1, \alpha_2, \dots, \alpha_m$  を考える. この配列において, \*で示される塩基を表すときは,  $m.p.\alpha.i$  と表す.

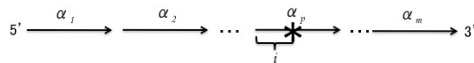


図 1 RNA 鎖の配列  $\alpha_1, \alpha_2, \dots, \alpha_m$

ここで, 塩基  $m.p.\alpha.i$  と  $m'.p'.\alpha'.i'$  に対して,  $m.p.\alpha.i < m'.p'.\alpha'.i'$  であるとは以下を満たすこととして定義する.

$$m.p.\alpha.i < m'.p'.\alpha'.i' \stackrel{\text{def}}{=} \begin{cases} p < p' \\ \text{or} \\ p = p' \wedge \alpha = \alpha' \wedge i < i' \end{cases} \quad (6)$$

これは, 塩基  $m.p.\alpha.i$  が塩基  $m'.p'.\alpha'.i'$  よりも 5' 末端側にあることを示している. 次に, RAN 鎖  $n$  本から構成される RNA 二次構造において, 上側に存在する RNA 鎖が  $s$  本, 下側に存在する RNA 鎖が  $t$  本であるようなものを列挙することを考える. このような構造では, 上側の RNA 鎖の配列  $\alpha_1, \alpha_2, \dots, \alpha_s$  と, 下側の RNA 鎖の配列  $\beta_1, \beta_2, \dots, \beta_t$  にわけて考えることが出来る. ここで, 図 2 の\*で示される塩基対は  $(s.p.\alpha.i, t.q.\beta.j)$  と表される.

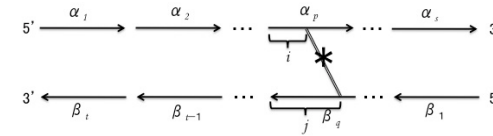


図 2 複数鎖から構成される RNA 二次構造の例

以上から,  $n$  本以下の RNA 鎖が会合してできる塩基対の集合  $BP$  は以下のように表される.

$$BP = \{(s.p.\alpha.i, t.q.\beta.j) \mid \begin{aligned} &1 \leq s < n, 1 \leq t \leq n - s, \\ &1 \leq p \leq s, 1 \leq q \leq t, \\ &\alpha, \beta \in S, 1 \leq i \leq l_\alpha, 1 \leq j \leq l_\beta, \\ &(\alpha[i], \beta[j]) \in WC_v \} \end{aligned} \quad (7)$$

また,  $BP$  上の関係  $\sim$  を以下のように定義する.

$$(s.p.\alpha.i, t.q.\beta.j) \sim (s.p'.\alpha'.i', t.q'.\beta'.j')$$

$$\stackrel{\text{def}}{\equiv} \begin{cases} p = p' \wedge \alpha = \alpha' \wedge 0 \leq q - q' \leq 1 \\ \text{or} \\ q = q' \wedge \beta = \beta' \wedge 0 \leq p - p' \leq 1 \end{cases} \quad (8)$$

このような関係  $\sim$  は塩基対  $(s.p.\alpha.i, t.q.\beta.j)$  と  $(s.p'.\alpha'.i', t.q'.\beta'.j')$  において、上側の RNA 鎖か下側の RNA 鎖がきちんと繋がっていることを意味している。

以上を用いて、塩基対  $(s.p.\alpha.i, t.q.\beta.j)$  と  $(s.p'.\alpha'.i', t.q'.\beta'.j')$  が隣合っていることの必要十分条件は以下のようになる。

$$\begin{aligned} & ((s.p.\alpha.i, t.q.\beta.j) \sim (s.p'.\alpha'.i', t.q'.\beta'.j')) \\ \wedge & ((s.p.\alpha.i < s.p'.\alpha'.i') \wedge (t.q.\beta.j, t.q'.\beta'.j')) \end{aligned} \quad (9)$$

以上から、列挙グラフ  $G$  は以下のように構成される。

$$\begin{aligned} G &= (V, V_0, V_f, E_g, \epsilon) \\ V &= BP \cup \{S, F\} \\ V_0 &= \{S\} \\ V_f &= \{F\} \\ E_g &= \{((s.p.\alpha.i, t.q.\beta.j), (s.p'.\alpha'.i', t.q'.\beta'.j')) | \\ & ((s.p.\alpha.i, t.q.\beta.j), (p'.\alpha'.i', t.q'.\beta'.j')) \in BP \\ & \wedge ((s.p.\alpha.i, t.q.\beta.j) \sim (s.p'.\alpha'.i', t.q'.\beta'.j')) \\ & \wedge ((s.p.\alpha.i) < (s.p'.\alpha'.i')) \\ & \wedge ((t.q.\beta.j) < (t.q'.\beta'.j')) \\ & \cup \{(S, bp) | bp = (s.1 - \alpha.i, t.t.\beta.j) \in BP \\ & \cup \{(bp, F) | bp = (s.s.\alpha.i, t.1.\beta.j) \in BP\} \} \end{aligned} \quad (10)$$

川上らの研究における列挙グラフ  $G = (V, E_g)$  の頂点数と辺数の見積もりについて述べる。  $S$  を RNA 鎖の集合とし、  $S$  に含まれる RNA 鎖から出来る二次構造を持つ RNA 鎖の最大本数を  $N$  とする。この時、  $S$  の要素の個数を  $K$  とする。最後に、  $L$  を RNA 鎖の集合  $S$  に含まれる RNA 鎖の最大長とすると、列挙グラフ  $G$  の頂点数  $|V|$  および  $|E_g|$  は以下のように抑えられる。

$$|V| = O(N^4 K^2 L^2) \quad (11)$$

$$|E_g| = O(N^4 K^3 L^4) \quad (12)$$

## 5. 列挙グラフのサイズの縮小

本研究では、川上と同じ二次構造のクラスを対象とする。つまり、ヘアピンループ、マルチプルループ、シュードノットを含まない二次構造のクラスを対象とする。また、塩基配列を仮想塩基配列に拡張した問題を考える。

4.2 節において、異なる経路によって同じ複合体を重複して数え上げることができると述べたが、このことがグラフのサイズを大きくする原因の1つともなっている。このため、グラフの経路と構造が1対1で対応するように反応系を限定すればグラフのサイズはさらに小さくできると予想される。

グラフの経路と構造が1対1で対応する場合と、1つの構造と複数の経路が対応する場合が混在してしまう主な原因は、回転対称な構造が反応系に存在してしまうことである。

よって、本研究では、回転対称な構造が生じないような反応系の条件について考えることとする。つまり、グラフの経路と構造が1対1で対応するように反応系を限定する方法を形式的に議論する。

### 5.1 回転対称な構造が生じないための十分条件

回転対称な構造が生じないための条件を議論するにあたって、いくつかの定義と記法を導入する。

まず、  $\Sigma_v$  を仮想塩基の集合とし、  $\Sigma_v$  が形成する仮想塩基の塩基対の集合を  $WC_v$  とする。また、  $S$  を仮想塩基配列の集合とする。

このとき、  $S$  に含まれる  $\alpha$  と  $\beta$  という仮想塩基配列が共通の仮想塩基を持つとき、  $\alpha \approx \beta$  と定義する。ここで、  $\approx$  の反射的推移的閉包  $\approx^*$  は  $S$  上の同値関係である。また、  $S$  に含まれる  $\alpha$  と  $\beta$  という仮想塩基配列に対して、ある仮想塩基  $A, B$  が存在し、  $A$  が  $\alpha$  に含まれ、  $B$  が  $\beta$  に含まれ、  $A$  と  $B$  が塩基対を形成し得る時、  $\alpha \bowtie \beta$  と定義する。以上の記法を用いて、グラフ  $G = (V_s, E_s)$  を以下のように構成する。

$$V_s = S / \approx^* \quad (13)$$

$$E_s = \{\{v_1, v_2\} | \quad (14)$$

$$v_1, v_2 \in S / \approx, \exists \alpha \in v_1, \exists \beta \in v_2, \alpha \bowtie \beta\}$$

上で定義したグラフ  $G = (V_s, E_s)$  が2部グラフであれば、与えられた仮想塩基配列のインタラクシオン反応系では、回転対称な構造が生成されないことが証明できる。

**定理 2.** 仮想塩基配列集合  $S$  に対して、  $G = (V_s, E_s)$  が2部グラフならば、  $S$  が形成するマルチプルループ、ヘアピンループ、シュードノットを含まない二次構造は回転対称でない。

## 6. 回転対称な構造が生じない列挙法

仮想塩基配列の集合  $S$  に対して,  $G = (V_s, E_s)$  が 2 部グラフである仮定する. このとき, 2 部グラフの 2 つの頂点集合を  $V_U, V_D$  とする. ここで,  $V_U \cup V_D = S / \approx^*$  である. さらに,  $U = \{\alpha | \alpha \in v, v \in V_U\}$ ,  $D = \{\beta | \beta \in v, v \in V_D\}$  とする. つまり,  $U$  は  $V_U$  の頂点に含まれる仮想塩基配列の集合であり,  $D$  は  $V_D$  の頂点に含まれる仮想塩基配列の集合である.

ヘアピンループ, マルチプルループ, シュードノットを含まない二次構造を持つ複数の仮想塩基配列からなる複合体を列挙するグラフ  $G = (V, V_0, V_f, E_g)$  を以下に構成する. ここで,  $V$  はノードの有限集合,  $E_g$  は辺の有限集合,  $V_0$  は初期ノードの集合,  $V_f$  は終端ノードの集合である.

$S$  を仮想塩基配列の集合とする.  $\alpha \in S$  に対して  $l_\alpha$  で  $\alpha$  の長さを表す. また,  $\alpha[i]$  によって配列  $\alpha$  の  $i$  番目の塩基の種類を表す. 本研究では, 複数の仮想塩基配列がインタラクションして構造を形成する. このような構造を表現するために, 集合  $U$  の要素の列  $\alpha_1, \alpha_2, \dots, \alpha_m$  と, 集合  $D$  の要素の列  $\beta_1, \beta_2, \dots, \beta_n$  を考える.

### 6.1 集合 $U$ の仮想塩基配列の並び

集合  $U$  の要素の列  $\alpha_1, \alpha_2, \dots, \alpha_m$  において, まず図 3 のような \* で示される部分の仮想塩基を,  $p-\alpha-i$  と表す. これは, 5' 末端側から数えて  $p$  番目の仮想塩基配列  $\alpha \in X$  の 5' 末端側から  $i$  番目の仮想塩基という意味である.

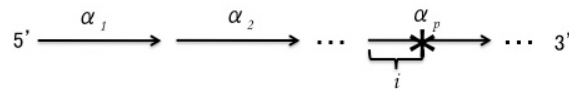


図 3 5' 末端側から 3' 末端方向に並んだ配列

ここで, 仮想塩基  $p-\alpha-i$  と  $p'-\alpha'-i'$  に対して,  $p-\alpha-i < p'-\alpha'-i'$  であるとは, 以下を満たすこととして定義する.

$$p-\alpha-i < p'-\alpha'-i' \stackrel{\text{def}}{\equiv} \begin{cases} p < p' \\ \text{or} \\ p = p' \wedge \alpha = \alpha' \wedge i < i' \end{cases} \quad (15)$$

これは, 塩基  $p-\alpha-i$  が  $p'-\alpha'-i'$  よりも 5' 末端側にあることを示している.

### 6.2 集合 $D$ の仮想塩基配列の並び

次に, 集合  $D$  の要素の列  $\beta_1, \beta_2, \dots, \beta_n$  において, 図 4 のような \* で示される部分の仮想塩基を,  $q-\beta-j$  と表す. これは, 左側の 3' 末端側から数えて  $q$  番目の RNA 鎖  $\beta \in X$  の 3' 末端側から  $j$  番目の塩基という意味である.

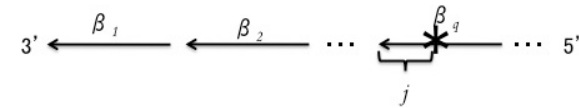


図 4 3' 末端側から 5' 末端方向に並んだ配列

ここで, 塩基  $q-\beta-j$  と  $q'-\beta'-j'$  に対して,  $q-\beta-j < q'-\beta'-j'$  であるとは, 以下を満たすこととして定義する.

$$q-\beta-j < q'-\beta'-j' \stackrel{\text{def}}{\equiv} \begin{cases} q < q' \\ \text{or} \\ q = q' \wedge \beta = \beta' \wedge j < j' \end{cases} \quad (16)$$

これは, 塩基  $q-\beta-j$  が  $q'-\beta'-j'$  よりも 3' 末端側にあることを示している.

### 6.3 集合 $U$ と集合 $D$ からなる二次構造の列挙

次に, 集合  $U$  と集合  $D$  から成る仮想塩基配列の二次構造を列挙することを考える. 列挙される構造は, 上側と下側にわけて考えられ, 上側は先に説明した集合  $U$  から構成され, 下側は集合  $D$  から構成されると取り決める. このとき, 集合  $U$  と集合  $D$  の間で塩基対ができ, 二次構造が形成される. ここで, 図 5 の \* で示される塩基対は塩基  $p-\alpha-i$  と塩基  $q-\beta-j$  の間で形成されるので,  $(p-\alpha-i, q-\beta-j)$  のように表すこととする. このように,  $-$  (ハイフン) と  $\_$  (下線) を使い分けるのは,  $U$  が上側の配列であり,  $D$  が下側の配列であることを意識している.

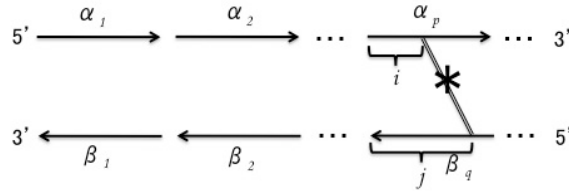


図5  $\alpha_p$  と  $\beta_q$  の間で形成される塩基対

以上から、仮想塩基配列が会合してできる塩基対の集合  $BP$  は以下のように表される。

$$BP = \{(p-\alpha-i, q-\beta-j) \mid \alpha \in U, \beta \in D, \\ 1 \leq p < n, 1 \leq q \leq n-p, \\ 1 \leq i \leq l_\alpha, 1 \leq j \leq l_\beta, \\ (\alpha[i], \beta[j]) \in WC_v\} \quad (17)$$

また、 $BP$  上の関係  $\sim$  を以下のように定義する。

$$(p-\alpha-i, q-\beta-j) \sim (p'-\alpha'-i', q'-\beta'-j') \\ \stackrel{\text{def}}{\equiv} \begin{cases} p = p' \wedge \alpha = \alpha' \wedge 0 \leq q - q' \leq 1 \\ \text{or} \\ q = q' \wedge \beta = \beta' \wedge 0 \leq p - p' \leq 1 \end{cases} \quad (18)$$

このような関係  $\sim$  は塩基対  $(p-\alpha-i, q-\beta-j)$  と  $(p'-\alpha'-i', q'-\beta'-j')$  において、上側の仮想塩基配列か下側の仮想塩基配列がきちんと繋がっていることを意味している。

以上を用いて、塩基対  $(p-\alpha-i, q-\beta-j)$  と  $(p'-\alpha'-i', q'-\beta'-j')$  が隣合って存在することの必要十分条件は以下ようになる。

$$((p-\alpha-i, t-q-\beta-j) \sim (p'-\alpha'-i', q'-\beta'-j')) \\ \wedge ((p-\alpha-i < p'-\alpha'-i') \wedge (q-\beta-j, q'-\beta'-j')) \quad (19)$$

以上から、列挙グラフ  $G$  は以下のように構成される。

$$\begin{aligned} G &= (V, V_0, V_f, E_g, \epsilon) \\ V &= BP \cup \{S, F\} \\ V_0 &= \{S\} \\ V_f &= \{F\} \\ E_g &= \{((p-\alpha-i, q-\beta-j), (p'-\alpha'-i', q'-\beta'-j')) \mid \\ & (p-\alpha-i, q-\beta-j) \in BP, (p'-\alpha'-i', q'-\beta'-j') \in BP, \\ & \wedge ((p-\alpha-i, q-\beta-j) \sim (p'-\alpha'-i', q'-\beta'-j')) \\ & \wedge ((p-\alpha-i < (p'-\alpha'-i')) \wedge ((q-\beta-j) < (q'-\beta'-j')))\} \\ & \cup \{(S, bp) \mid bp = (1-\alpha-i, 1-\beta-j) \in BP \\ & \cup \{(bp, F) \mid bp = (p-\alpha-i, q-\beta-j), p+q \leq n\} \end{aligned} \quad (20)$$

例として、図6のような二次構造を列挙スキームによって表すと、この二次構造はグラフの  $s \rightarrow (1-\alpha-4, \gamma-1-2) \rightarrow (2-\beta-3, 1-\gamma-5) \rightarrow f$  のような経路と1対1に対応する。

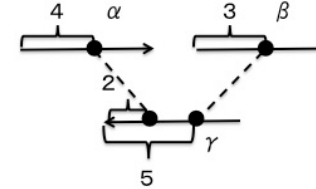


図6  $\alpha, \beta, \gamma$  から構成される二次構造の例

このグラフを用いて列挙スキームを構成するには、以下の式を満たす重み  $\epsilon$  と  $\sigma_x (x \in \mathcal{M})$  を割り当てなければならない。

$$E_\gamma(\psi(\gamma)) = \sum_{e \in E_g \text{ s.t. } e \in \gamma} \epsilon(e) \quad (21)$$

$$\#_x(\psi(\gamma)) = \sum_{e \in E_g \text{ s.t. } e \in \gamma} \sigma_x(e) \quad (22)$$

(21) を満たす重み  $\epsilon$  を構成する方法を以下に示す。

左辺の  $E_\gamma(\psi(\gamma))$  は  $E(\psi(\gamma)) + \log \gamma_{\psi(\gamma)}$  で定義される。本研究では、回転対称な構造を列挙しないため、 $\psi$  は1対1対応である。よって、すべての  $\gamma \in PT(G)$  に対して、 $\gamma_{\psi(\gamma)} = 1$  である。また、パス写像群  $\Phi$  としては、恒等写像  $\phi_0$  のみからなる群  $\Phi = \phi_0$  を考えればよい。

一方、 $\sigma(e)$  の割り当ては次のように定義出来る。辺  $e$  が対応する局所構造に仮想塩基配

列  $x$  の 5' 末端が含まれるとき,  $\sigma(e) = 1$  とし, それ以外の時  $\sigma(e) = 0$  とする.

定理 3. 上記で定義される列挙スキームは定理 1 の条件を満足する.

構成されたグラフ  $G$  の  $|V|$  と  $|E_g|$  は以下のように抑えられる.

$$|V| = O(N^2 K^2 L^2) \quad (23)$$

$$|E_g| = O(N^2 K^3 L^4) \quad (24)$$

つまり, 川上の手法の場合と比べて,  $|E_g|$  の  $N$  に関する次数が 2 つ下げられている.

## 7. ま と め

本研究では, 与えられた仮想塩基配列から構成されるグラフが 2 部グラフであるならば, 列挙するグラフの経路と構造が 1 対 1 に対応し, グラフのサイズを小さくできることを示した. 今後は, まず回転対称な構造が生じないような反応系に着目し, その後回転対称な構造も考慮した列挙グラフの提案および実装を目標としている. これにより, [2] により示されたグラフよりもグラフのサイズを小さくすることで, 高速な平衡状態解析システムの実装を目指す.

## 参 考 文 献

- 1) Dirks, R., Bois, J., Schaeffer, J., Winfree, E. and Pierce, N.: Thermodynamic Analysis of Interacting Nucleic Acid Strands, SIAM Review, Vol.49, No1, pp.65-88(2007)
- 2) Kobayashi, S. and Kawakami, T.: Enumeration Approach to the Analysis of Interacting Nucleic Acid Strands, a book chapter in "Molecular and Supramolecular Information Processing - From Molecular Switches to Logic Systems", edited by Evgeny Katz, Wiley-VCH, to appear.
- 3) Nishikawa, A., Hagiya, M., Yamamura, M.: Virtual DNA Simulator and Protocol Design by GA, Proceedings of the Genetic and Evolutionary Computation Conference, GECCO'99, Vol.2, pp.1810-1816. (1999)
- 4) Kobayashi, S.: Symmetric Enumeration Method: A New Approach to Computing Equilibria, Technical Report CS 08-01, University of Electro-Communications, Department of Computer Science(2008)