

最大クリーク問題に関する DNA コンピューティングと行列解析による最適化

大阪大学工学部* 丸山 健悟 (Kengo Maruyama)

大阪大学大学院情報科学研究科** 齋藤 誠慈 (Seiji Saito)

大阪大学大学院情報科学研究科** 小倉 裕介 (Yusuke Ogura)

大阪大学大学院情報科学研究科** 谷田 純 (Jun Tanida)

* : Faculty of Engineering, Osaka University

** : Graduate School of Information Science and Technology, Osaka University

1 はじめに

最大クリーク問題は組合わせ最適化問題の一つであり, グラフやネットワークをはじめとする様々な分野における重要な問題である. この問題はクラス NP 完全に含まれる複雑な問題であり, 既存の電子計算機を使って多項式時間で解くアルゴリズムは未だ発見されていない.

また最大クリーク問題は種々の 2 次計画問題に定式化することが可能であり, 2 次計画問題を考えることにより解の上・下界を与える様々な条件が得られ, 多項式時間で解の上・下界を計算することができる.

一方, 1994 年に Adleman[5] が有向ハミルトン経路問題に対する DNA 分子計算を報告して以来, 近年では DNA 分子計算により分子反応の並列性, 自律性を用いて様々な組み合わせ最適化問題に対するアプローチが行われ, 現在ではこのような NP 完全問題は分子計算の能力を計るためのベンチマーク という位置付けで研究が進められている.

本研究では 2 次 0-1 計画問題で表現した最大クリーク問題に対する分子計算法を考案し, 解の上・下界の適用・DNA ヘアピン計算という自律計算法の適用により計算量の軽減が可能な計算アルゴリズムを与える.

2 離散型最大クリーク問題

重みのない無向連結グラフを $G = (V, E)$ と表す. グラフの n 個の頂点には 1 から n までの数字が割り当てられ, その頂点の集合を $V = \{1, 2, \dots, n\}$ とする. また頂点 i と頂点 j に枝が存在する場合, その枝を (i, j) と表し, 枝の集合を $E = \{(i, j) \mid i, j \in V, i \neq j\}$ とする.

グラフ $G = (V, E)$ に対し, 補グラフを $\bar{G} = (V, \bar{E})$ とする. ここで $\bar{E}$ は補グラフの枝の集合であり, $\bar{E} = \{(i, j) \mid i, j \in V, i \neq j, (i, j) \notin E\}$ と表される. グラフ $G = (V, E)$ が完全グラフであるとは, 任意の 2 頂点間に枝が存在するようなグラフであり, $E = \{(i, j) \mid \forall i, j \in V, i \neq j\}$ となる場合である. 頂点の部分集合 S がクリークであるとは, G の部分グラフ $G(S) = (S, E')$ が完全グラフ, つまり $E' = \{(i, j) \mid \forall i, j \in S, i \neq j\}$ となるような部分グラフの頂点の集合である. クリークの大きさとはクリークに含まれる頂点数のことをいい, $|S|$ で表す. 最大クリーク C とは大きさが最大となるクリークであり, グラフ G の最大クリークの大きさを $\omega(G)$ で表す. 最大クリーク問題とは, 与えられたグラフの最大クリークとその大きさ $\omega(G)$, または大きさ $\omega(G)$ のみを求める問題である. n 頂点のグラフ $G = (V, E)$ に対し, n 次正方の隣接行列 $A_G = (a_{ij}), i, j \in V$ を $a_{ij} = 1$ for $(i, j) \in E; a_{ij} = 0$ for $(i, j) \notin E$ として表す. 本論文で扱うグラフは無向グラフなので (i, j) と (j, i) は同じ枝を表し, $a_{ij} = a_{ji}$, すなわち隣接行列 A_G は対称行列である. また, 自己ループ (i, i) は存在しないので隣接行列 A_G の対角成分は $a_{ii} = 0$ ($i = 1, 2, \dots, n$) である.

与えられたグラフ $G = (V, E)$, $|V| = n$ に対する最大クリーク問題は幾通りかの定式化があり,

その1つとして以下の離散型2次計画問題に定式化できる ([2]).

$$\text{minimize} \quad f(\mathbf{x}) = -\sum_{i=1}^n x_i + 2 \sum_{i>j, (i,j) \in \bar{E}} x_i x_j = \mathbf{x}^T A \mathbf{x} \quad (2.1)$$

$$\text{subject to} \quad \mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \{0, 1\}^n \quad (2.2)$$

ただし, $A = A_{\bar{G}} - I$ である. このとき, 最適解 $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*)^T$ は次の関係をみたす ($i = 1, 2, \dots, n$): $x_i^* = 1$ for $i \in C$; $x_i^* = 0$ for $i \notin C$. さらに, 最適値は $\omega(G) = -f(\mathbf{x}^*)$ のように表される. ここで i, j はグラフの頂点番号, $\bar{E}$ は補グラフ $\bar{G}$ の枝集合であり, C は最大クリーク, $A_{\bar{G}}$ は補グラフ $\bar{G}$ の隣接行列, I は n 次単位行列である.

2つの変数 x_i と x_j がともに1であるためには枝 (i, j) が与えられたグラフ G の枝の集合 E に含まれなければならない. したがって, 値が1となる変数に対応する頂点間にはグラフ G において枝が存在するので, そうした頂点は G のクリークとなる. 故に, 最大クリークでは1を割り当てた変数の数 (1次項の値の絶対値) が最大となり, その時の2次項の値は各項で全て0となる.

3 最適値の上・下界計算

与えられた連結グラフ $G = (V, E)$ の頂点, 枝の数, 密度, 隣接行列などにより, 最大クリークの大きさ $\omega(G)$ の上・下界を求めることができる. ここでは幾つかある上・下界のうち, 実際に計算が容易かつ良好な条件のものを示す.

(a) **上界** 最大クリークの大きさ $\omega(G)$ の上界は次の条件を用いて求めることができる ([2]). M.Budinich, P.Budinich は複素数空間における幾何学的な定式化により次の関係を示した.

上界条件 1 補グラフの固有値に関する条件

$$\omega(G) \leq \frac{n + \bar{N}_0}{2} \quad (3.3)$$

ここで n はグラフ G の頂点数, $\bar{N}_0$ は補グラフ $\bar{G}$ の隣接行列 $A_{\bar{G}}$ の固有値のうち, 0 の個数を表す.

各頂点に連結する枝数から, 次のようにして上界を与える条件が導かれる. 連結している枝の数の多い頂点から順に $1, 2, \dots, n$ とし, 各頂点 i に連結している枝数を $k(i)$ とする. すると $k(1) \geq k(2) \geq \dots \geq k(n)$ となる. 最大クリークの大きさを $\omega(G)$ とすると, 大きさ $\omega(G)$ のクリークを作るには, 連結する枝を $\omega(G) - 1$ 本以上持った頂点が $\omega(G)$ 個必要であることがわかる. このことから次の条件が得られる.

上界条件 2 頂点の度数に関する条件

$$\omega(G) \leq \max_{i \in V} i \quad \text{s.t.} \quad k(i) \geq i - 1 \quad (3.4)$$

(b) **下界** 最大クリークの大きさの下界を与える条件として, 次の条件式が与えられる [2].

下界条件 1 隣接行列の成分1の密度に関する条件

$$\omega(G) \geq \frac{1}{1 - \delta} \quad (3.5)$$

ここで δ は隣接行列 A_G における成分1の密度 $\delta = \frac{2m}{n^2}$ (m はグラフの枝数) を表す. また最大クリークの大きさは整数であるので, それも上・下界条件として用いることができる.

これまでに [10] では、ランダムグラフを用いたこれらの上・下界条件の計算実験により、密度の低いグラフに対しては上界条件 2 で与えられる上界がより小さくなり、密度の高いグラフに対しては上界条件 1 で与えられる上界のほうが小さくなることが分かった。また縮小定数 (頂点数に対する解候補の数の割合) を考えると、密度約 0.25 のグラフでは約 25%、密度約 0.50 から 0.75 のグラフでは 50% 弱、密度 0.90 のグラフでは約 30~40% になっている ([10])。このことを用いると、次の章で述べる DNA 分子計算において解候補となる配列を $1/4$ から $1/2$ に減少させることができると思われる。

また下界条件を用いて、得られた下界よりも少ない数の枝で結ばれている頂点は最大クリークになり得ないことから、それらの頂点をグラフから削除し、グラフを縮小することができるが、上に示した計算実験によると、下界条件 1 で与えられる下界は 100 頂点のグラフに対して 2, 3 程度にしかならないことが分かっている。

4 DNA 分子計算

DNA 分子計算とは問題を DNA 分子上に符号化し、分子生物学的実験手法を用いることによって問題を解くという計算手法である。その利点として、試験管内の少量の溶液中に濃縮された大量数の DNA 分子が Watson-Crick 相補性 [9] により自律的かつ並列に反応することが挙げられる。こうした分子反応の自律性、並列性は NP 問題のような計算困難な問題に対して利用できるのではないかと考えられており、1994 年に L. Adleman [5] によって NP 完全問題である有向ハミルトン経路問題に対する DNA 分子計算が発表されて以来、様々な問題に対する実験が報告されている。最大クリーク問題に対しては 1997 年に Q. Ouyang ら [6] によって分子計算アルゴリズムとその実験結果が報告されている。また近年では坂本、萩谷ら [7] によって 1 本鎖 DNA がヘアピン構造を形成することにより自律計算を行う計算モデルが提案されている。

本研究では 2 次 0-1 計画問題で表現した最大クリーク問題を考えることによる自律的な分子計算の可能性、またその分子計算手法を述べ、3 頂点の簡単なグラフについて、実際に考案した分子計算手で最大クリーク問題を解くための実験手順を示す。

DNA 分子の構造 DNA はデオキシリボ核酸の略称であり、ヌクレオチドがホスホジエステル結合により 1 列に連なったポリマーである。DNA 分子を形成するヌクレオチドは 4 種類あり、その塩基部分がアデニン、グアニン、シトシン、チミンであるヌクレオチドをそれぞれ A, G, C, T と表す。4 種類のヌクレオチドの配列により DNA 分子は情報を蓄えることができ、生体分子内では遺伝情報を蓄えている。同様に DNA 計算においてもヌクレオチドの配列によって DNA 分子に問題の情報を持たせることができる。また 1 本鎖 DNA は方向性を持ち、一方の端を 5' 末端、他方を 3' 末端とし、 $5' \rightarrow 3'$ のように表現する。

またヌクレオチドの塩基部分は、Watson-Crick 相補性により A と T, G と C のみが水素結合し、塩基対を形成する。このような相補的な塩基配列により、お互いに相補的な 1 本鎖 DNA は結合して 2 本鎖 DNA になる (図 1)。

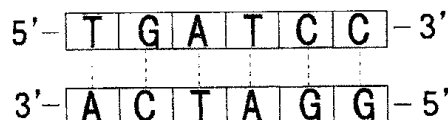


図 1: 水素結合による 2 本鎖 DNA の形成

5 分子計算アルゴリズム

本研究で提案する分子計算では、グラフ $G = (V, E)$ に対する最大クリーク問題を式 (2.1) の 2 次 0-1 計画問題へと定式化し、その目的関数を DNA 分子に符号化することを考える。その利点としては、2 次 0-1 計画問題に定式化可能な他のクラス NP 問題に対する汎用性、上・下界の適用と後に示す DNA ヘアピン構造による自律計算の適用による計算時間の短縮が挙げられる。

まず目的関数

$$f(\mathbf{x}) = -\sum_{i=1}^n x_i + 2 \sum_{i>j, (i,j) \in \bar{E}} x_i x_j$$

の 1 次項 x_i ($i = 1, 2, \dots, n$) と 2 次項 $x_i x_j$ ($i > j, (i, j) \in \bar{E}$) はそれぞれ別々に 1 本鎖 DNA 分子に符号化される。ここで 1 次項を符号化した 1 本鎖 DNA は解候補を表す部分であり、解候補の分類や上・下界適用による計算量の軽減に用いられる。一方 2 次項を符号化した 1 本鎖 DNA 分子は解候補の評価を行う部分であり、不適な解を除去する際に用いられる。

以下に最大クリーク問題に対する分子計算法の流れを示す。この計算法は 2 つのステップ（解候補の分類を行うステップと解候補の評価を行うステップ）からなる。

Step1 解候補の分類

- (1) 1 次項を符号化した 1 本鎖 DNA 分子 (1 次項配列) の作成
- (2) 1 次項配列の分類
- (3) 上・下界の適用による不適な配列の除去

Step2 解候補の評価

- (1) 2 次項を符号化した 1 本鎖 DNA 分子 (2 次項配列) の作成
- (2) 1 次項配列と 2 次項配列の結合
- (3) ヘアピン構造の形成による自律計算
- (4) 解の評価

本論文では次のような表記方法を用いる。

- $\boxed{a}$: a (変数 x_i , または $x_i = 0, 1$ の割り当て) を符号化した 1 本鎖 DNA 配列
 $\boxed{\bar{a}}$: $\boxed{a}$ の相補配列

また特に記していない限り左側を 5' 末端、右側を 3' 末端とする。

Step 1 解候補の分類

(1) 1 次項配列の作成 式 (2.1) より、目的関数の 1 次項部分 $\sum_{i=1}^n x_i$ は x_1 から x_n までの n 個の変数を足し合わせたものである。また、各変数 x_i ($i = 1, 2, \dots, n$) の値は 0 または 1 であるので、 n 個の変数に 0 または 1 を割り当てることは n 桁の 2 進数 $000\dots 0, \dots, 111\dots 1$ を DNA 分子に符号化することに対応する。これを実現する方法として、Q. Ouyang ら [6] の用いた方法が利用できる。

値 0 または 1 を割り当てた各変数を、次のような規則で DNA 分子に符号化する。

$x_i = 0$ のとき $\longrightarrow$ $\boxed{x_i}$

$x_i = 1$ のとき $\longrightarrow$ $\boxed{x_i} \boxed{p_i}$

ただし, $\boxed{p_i}$ は変数 x_i に値 1 が割り当てられていることを示す配列で, それぞれ同じ種類の制限酵素サイトを含む. これらの配列を x_1 から x_n まで結合し, 全ての n 桁の 2 進数を符号化した配列を作成するには次の方法を用いる ([6]). 変数 x_i ($i = 1, 2, \dots, n$) に対して次の配列を用意する.

i が奇数 かつ $i \neq n$ のとき $5' - \boxed{x_i} \boxed{p_i} \boxed{x_{i+1}} - 3' \quad x_i = 1$ に対応

$5' - \boxed{x_i} \boxed{x_{i+1}} - 3' \quad x_i = 0$ に対応

i が偶数 または $i = n$ のとき $5' - \boxed{\bar{x}_{i+1}} \boxed{\bar{p}_i} \boxed{\bar{x}_i} - 3' \quad x_i = 1$ に対応

$5' - \boxed{\bar{x}_i} - 3' \quad x_i = 0$ に対応

これらの配列を, *parallel overlap assembly (POA)* という解法捕生成法によって合成する. この方法は配列の入った溶液の温度を変化させることで変性 (94°C で 30 秒), アニール (60°C で 30 秒), ポリメラーゼ伸長 (72°C で 30 秒) を行い, これを 1 サイクルとしてこのサイクルを繰り返す (300bp の解候補を生成するのに約 30 サイクル) ことで解候補を生成する方法である. 生成溶液内の配列の中には, $\boxed{x_1}$ から始まって $\boxed{x_{n+1}}$ で終わる解候補配列の他に, それと同じ長さであるが $\boxed{x_1}$ から始まっていない, または $\boxed{x_{n+1}}$ で終わっていない, またはどちらも存在しないような不完全な配列が存在する. このような不完全な配列の数を抑え解候補配列の数を増やす為に, 反応溶液を少量取り, この中の配列を鋳型にしてプライマー $\boxed{x_1}$ と $\boxed{x_{n+1}}$ で PCR 増幅を行う. これによって $\boxed{x_1}$ から始まって $\boxed{x_{n+1}}$ で終わる解候補配列のみが優先的に増幅される. こうして得られた DNA 分子には n 桁の 2 進数 000...0, ..., 111...1 の全てが含まれている. 得られた解候補配列は 2 本鎖の状態であり, これを後の 2 次項配列との連結の為に 1 本鎖の状態で保持しておく必要がある. そのために得られた解候補配列を鋳型にして, プライマー $\boxed{x_1}$ を用いて PCR 増幅を行う. これにより溶液中で目的の 1 本鎖のみが増幅され, この反応溶液を電気泳動にかける事で他の配列と分離することができる. この段階における実験を, 頂点 $n = 3$ の場合について継続中である.

(2) 1 次項配列の分類 1 次項配列の作成で得られた 2^n 種類の DNA 分子は, 1 を割り当てた変数の数が多い配列程長く, 少ない程長くなっている (1 を割り当てた数だけ $\boxed{p_i}$ が含まれているため). したがってゲル電気泳動を用いて長さによる分類を行うことは, 1 次項の値 (値 1 を割り当てた変数の数) により DNA 配列を分類することと等しい. 変数が n 個ある場合, 1 次項の値は $0, \dots, n$ の $n+1$ 通りである. 前段階 (1) のゲル電気泳動によって 1 次項配列を $n+1$ 種類に分類され, 1 次項の値が l ($l = 0, 1, \dots, n$) である配列を試験管 T_l に入れる. 以上の操作により, $n+1$ 本の試験管 T_l が作成される.

(3) 上・下界の適用による不適な配列の除去 与えられたグラフに対して, 3 節で挙げた条件を用いた上・下界を計算によって得られた上界の値が p , 下界の値が q であるとする. 最大クリークの大きさ $\omega(G)$ のとり得る範囲は $q \leq \omega(G) \leq p$ に限定される. つまり, 1 次項の値は q 以上 p 以下であることがわかる. したがって, 最適解になり得る配列を含んでいる可能性のある試験管は $T_q, T_{q+1}, \dots, T_{p-1}, T_p$ の計 r ($r = p - q + 1$) 本であり, それ以外の試験管は後の操作に必要なので廃棄する. この操作により, 最適解になり得ない不適な配列の一部をこの段階で除去することが可能になる. 以上の操作の概略を図 2 に示す.

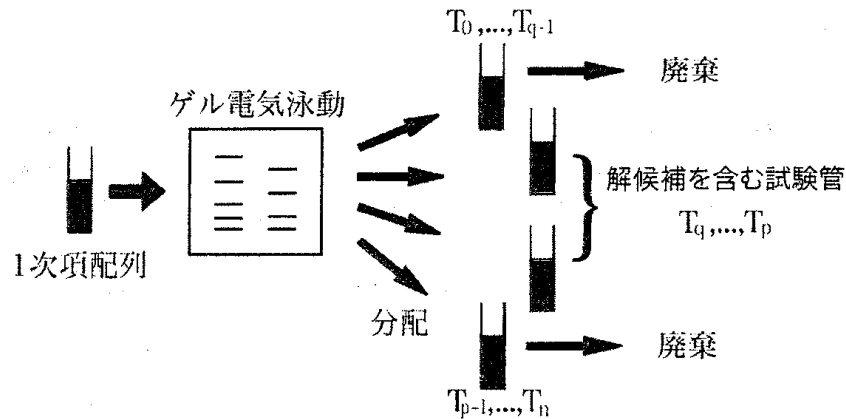


図 2: 1次項配列の分類および不適な試験管の廃棄

Step 2 解候補の評価

(1) **2次項配列の作成** ここでは1次項配列の作成と同様の方法を用いて2次項配列を作成する。式(2.1)より、目的関数の2次項部分は $\sum_{i>j, (i,j) \in \bar{E}} x_i x_j$ である。与えられたグラフ G の補グラフ $\bar{G}$ の枝数を $\bar{m}$ とすると ($|\bar{E}| = \bar{m} = {}_n C_2 - m$), 2次項は全部で $\bar{m}$ 項ある。 t 番目の項 ($t = 1, 2, \dots, \bar{m}$) に変数 x_i と x_j が含まれている場合、次のような規則でDNA符号化を行う。

第 t 項 $y_t = x_i x_j = 0$ のとき $\rightarrow \begin{array}{|c|c|} \hline y_t & \bar{p}_i \\ \hline \end{array} \quad (x_i = 0 \text{ のとき})$

$\begin{array}{|c|c|} \hline y_t & \bar{p}_j \\ \hline \end{array} \quad (x_j = 0 \text{ のとき})$

第 t 項 $y_t = x_i x_j = 1$ のとき $\rightarrow$ 割当なし

ただし、 y_t は第 t 番目の項を表す配列である。また、 $\bar{p}_i$ は1次項配列で用いた配列 p_i の相補配列である。また最大クリーク問題の場合、解候補の評価に使用する2次項配列は、各項で $x_i x_j$ の値が全て0である配列のみなので(式(1)より)、各項で $x_i x_j = 1$ のときの符号化は行う必要が無い。これらの配列を y_1 から $y_{\bar{m}}$ まで結合し、2次項配列を作成するために次のような配列を用意する。

t が奇数かつ $t \neq \bar{m}$ のとき $5' - \begin{array}{|c|c|c|} \hline y_t & \bar{p}_i & y_{t+1} \\ \hline \end{array} - 3' \quad (x_i = 0 \text{ のとき})$

$5' - \begin{array}{|c|c|c|} \hline y_t & \bar{p}_j & y_{t+1} \\ \hline \end{array} - 3' \quad (x_j = 0 \text{ のとき})$

t が偶数または $t = \bar{m}$ のとき $3' - \begin{array}{|c|c|c|} \hline \bar{y}_{t+1} & p_i & \bar{y}_t \\ \hline \end{array} - 5'$

$3' - \begin{array}{|c|c|c|} \hline \bar{y}_{t+1} & p_j & \bar{y}_t \\ \hline \end{array} - 5' \quad x_i x_j = 0 \text{ に対応}$

これらの $2\bar{m}$ 種類の配列を試験管に入れ1次項配列の作成と同様の操作を行うことで、 $2\bar{m}$ 種類の2次項配列が作成される。POAで合成される配列の中で最長なので、他の不完全配列が混ざっていることは無い。

(2) **1次項配列と2次項配列の結合** 解候補の分類で得られた1次項配列と2次項配列の作成で得られた2次項配列の結合を行うが、各項の1本鎖配列をライゲーション反応により結合する

場合、1次項と2次項の各1本鎖内に長い1本鎖DNAを作成する。1次項配列と2次項配列の中には互いに相補的な部分が多数存在し、どちらも1本鎖のままでは連結時にそのような相補的な部分も結合してしまい連結を妨げるおそれがある。そこでPOAによって生成した2次項配列の2本鎖を鋳型にして、プライマー $\overline{y_t}$ のみを用いてPCR増幅を行う。そうすると2次項配列の相補鎖である1本鎖が抽出される。この配列は1次項配列との相補的な部分を含まず、目的外の結合を起こす事はない。1次項配列を含む試験管 $T_q, T_{q+1}, \dots, T_p$ に分配し、各配列の連結部分をアニーリングさせ、DNAポリメラーゼによって伸長反応を行う事で、1次項配列と2次項配列が結合配列の2本差鎖を得ることができる。

(3) ヘアピン構造の形成による自律計算 1次項配列と2次項配列の結合で作成した目的関数を符号化した配列は、それらを単純に連結したものである。1次項で現れる変数の値と2次項で現れる項の値が一致していないものも多く含まれる(例えば、 $x_i = 1, x_i = 1$ に対して $x_i x_j = 0$ となっている)。こうした配列は不適であるので除去されなければならない。このような不適な配列を削除するために、坂本、萩谷ら[7]が提案した分子のヘアピン構造の形成による自律計算を用いる。ヘアピン構造の形成による自律計算法では、1本の1本鎖DNA分子の中に互いに相補的な配列が存在するとき、1分子内でハイブリダイゼーションが起こりヘアピン構造を形成する。するとハイブリダイゼーションにより2本鎖になっている部分は制限酵素の影響を受けやすくなり、切断される。

目的関数を符号化した配列において、1次項配列での変数の値は配列 P_i 、2次項配列での項の値は配列 $\overline{p_i}$ によって互いに照合することができる。1次項で用いた配列 P_i と2次項で用いた配列 $\overline{p_i}$ は互いに相補的であるので、1つの分子内にこれらの配列が共に存在する場合はヘアピン構造が形成され切断される(図3)。

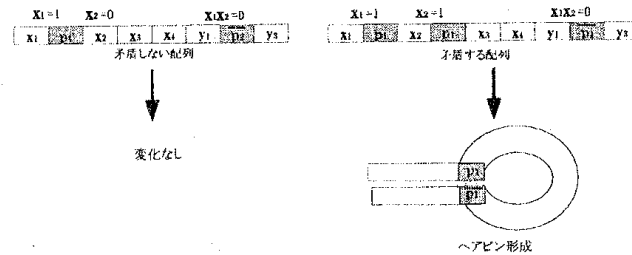


図3: ヘアピン構造の形成による不適な配列の除去を表し、左図は $x_i = 1, x_j = 0, y_t = x_i x_j = 0$ の場合で、右図は $x_i = 1, x_j = 1, y_t = x_i x_j = 0$ の場合である。

また x_i, x_j に関して全ての組合わせの例を(i)-(iv)に示す、

(i) $x_i = 0, x_j = 0, y_t = x_i x_j = 0$ のとき、矛盾は無くいずれも適切である：

x_i	x_j	y_t	$\overline{p_i}$
x_i	x_j	y_t	$\overline{p_j}$

ヘアピン構造を作らない($\overline{p_i}$ 有)

ヘアピン構造を作らない($\overline{p_j}$ 有)

(ii) $x_i = 1, x_j = 0, y_t = x_i x_j = 0$ のとき、矛盾は無くどちらか一方が残れば良い：

x_i	$\overline{p_i}$	x_j	y_t	$\overline{p_i}$
x_i	$\overline{p_i}$	x_j	y_t	$\overline{p_j}$

ヘアピン形成・切断($\overline{p_i}$ 有)

ヘアピン構造を作らない($\overline{p_j}$ 有)

(iii) $x_i = 0, x_j = 1, y_t = x_i x_j = 0$ のとき、矛盾は無くどちらか一方が残れば良い：

x_i	x_j	$\overline{p_j}$	y_t	$\overline{p_i}$
-------	-------	------------------	-------	------------------

ヘアピン構造を作らない($\overline{p_i}$ 有)

x_i	x_j	p_j	y_t	$\bar{p}_j$
-------	-------	-------	-------	-------------

ヘアピン形成・切断 ($\bar{p}_j$ 有)

(iv) $x_i = 1, x_j = 1, y_t = x_i x_j = 0$ のとき, いずれも矛盾だがヘアピン計算により切断される:

x_i	p_j	x_j	p_j	y_t	$\bar{p}_i$
-------	-------	-------	-------	-------	-------------

ヘアピン形成・切断 ($\bar{p}_i$ 有)

x_i	$\bar{p}_i$	x_j	p_j	y_t	$\bar{p}_j$
-------	-------------	-------	-------	-------	-------------

ヘアピン形成・切断 ($\bar{p}_j$ 有)

以上の実験過程・計算において次のことに注意しなければならない。温度を調節することで分子中の互いに相補的な配列どうしがハイブリダイゼーションにより2本鎖になるが、このとき分子間のハイブリダイゼーションも起こり得る。こうした分子間ハイブリダイゼーションを防ぐことは、DNA分子を含む溶液の濃度を低く保つことで実現される。また切断後、ノンヘアピン分子の存在比率を高める為に、各試験管でプライマー x_1 と $\bar{x}_{n+1}$ を使ってPCRを行う。

(4) 解の評価 ヘアピン構造の形成による自律計算までの操作により r ($r = p + q - 1$, p : 上限候補 q : 下限候補) 本の試験管 $T_q, T_{q+1}, \dots, T_p$ において、1次項配列の変数 x_i の値の割り当てに矛盾しない2次項配列のみが結合した目的関数を符号化した配列が残っている。これらの試験管から最適値 $\omega(G)$ を求める。

最適値 $\omega(G)$ を符号化している配列は、2次項が全て0、かつ1次項での1の数が最大、換言すると、残った配列の中で最も長い配列である。ここで、1次項配列の分類において値1を割り当てた変数の数によって試験管を分ける操作を行い、試験管 T_l ($l = q, q+1, \dots, p$) に含まれる配列の1次項では値1を割り当てた変数の数が l 個ということが分かっている。したがって最適値 $\omega(G)$ を符号化している配列は、2次項と連結している配列を含む試験管 T_l のうち、 l が最大である試験管に含まれ、またそのときの最大クリークの大きさ $\omega(G)$ は l に等しい。

以下の手順で $\omega(G)$ を求める。各試験管内の配列をゲル電気泳動を用いて長さにより分離する。各試験管内の1次項配列の長さは全て同じで、また2次項が全て0に対応する2次項配列の長さも既知である。したがってゲル電気泳動で長さによる分離を行ったときに対応する長さの位置にバンドが現れるか否かの確認を行うことで1次項と2次項が連結した配列に対応する2次項配列の有無が分かる。この操作を試験管 $T_q, T_{q+1}, \dots, T_p$ と順次行っていく、試験管 T_ω までは目的となるバンドが現れ、試験管 $T_{\omega+1}$ では現れなかったとする。このとき最大クリークの大きさは $\omega(G) = \omega$ と求まる。試験管 $T_p, T_{p-1}, \dots, T_q$ と順次操作を行った場合は、試験管 T_ω で初めて目的となるバンドが現れる。概略を図4に示す。

最大クリークの決定 以上の操作から得られた、最適解に該当する配列のバンドをゲルから取り出し、DNAシーケンサーによってその塩基配列を調べる。1次項配列部分から、どの頂点に値1が割り当てられているかが分かるので、それに該当する頂点が最大クリーク C となる。

また与えられたグラフの中に最大クリークが2つ以上ある場合、最適解の長さは全て等しいため、各種類で分離をして1種類ごとにDNAシーケンサーにかけないと配列デザインを読み取ることはできない。そこで配列の中の A, T 含有量によって配列を分離できる試薬をゲルに添加して再び電気泳動を行う事で、複数種類存在する最適解配列を分離することができる。

参考文献

- [1] R. Horst, Panos M. Pardalos and Nguyen V. Thoai, "Introduction to Global Optimization", Kluwer Academic Publishers(1995).
- [2] Immanuel M. Bomze, Marco Budinich, Panos M. Pardalos, Marcello Pellilo, "The Maximum Clique Problem" Handbook of Combinatorial Optimization Supp. Vol.A, Kluwer Academic Publishers(1999).

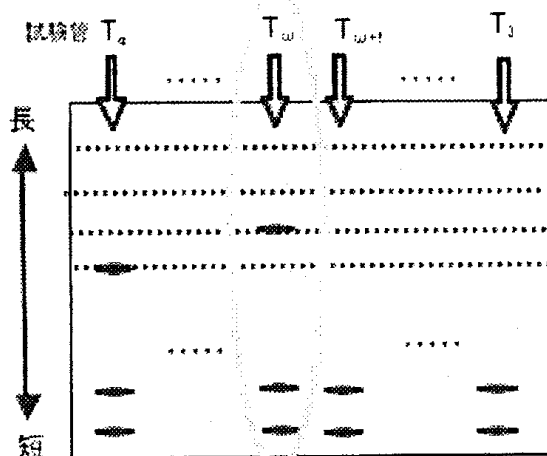


図 4: ゲル電気泳動を用いた最適解の検知

- [3] A. T. Amin, S. L. Hakimi, "Upper Bounds on the Order of a Clique of a Graph" SIAM Journal on Applied Mathematics, Vol.22, No.4, 569-573 (1972).
- [4] Marco Budinich, "Exact bounds on the order of the maximum clique of a graph" Discrete Applied Mathematics, Vol.127, 535-543 (2003).
- [5] Leonard M. Adleman, "Molecular Computation of Solution to Combinatorial Problems" Science, Vol.266, 1021-1024 (1994).
- [6] Qi Ouyang, Peter D. Kaplan, Shumao Liu, Albert Libchaber, "DNA Solution of the Maximal Clique Problem" Science, Vol.278, 446-449 (1997).
- [7] K. Sakamoto, H. Gouzu, K. Komiya, D. Kiga, S. Yokoyama, T. Yokomori and M. Hagiya, "Molecular Computation by DNA Hairpin Formulation" Science, Vol.288, 1223-1226 (2000).
- [8] Y. Sakakibara, "Solving Computational Learning Problems of Boolean Formulae on DNA Computers" Proc. of the 6th International Meeting on DNA Based Computers(DNA6), Lecture Notes in Computer Science, Vol.2054, 220-230 (2001).
- [9] 萩谷昌己, 横森貴 "DNA コンピュータ", 培風館 (2000).
- [10] 藤原祐介, "最大クリーク問題から導出される 2 次 0-1 計画問題に対する固有値解析と分子計算アルゴリズム" 大阪大学工学部応用自然科学科応用物理卒業論文 (2004).