

DNA 分子計算による新しい計算パラダイムの 基礎的検討

3 J-7

岡田 信人[†] 山本 雅人[†] 大内 東[†] 柴 肇一[‡] 棟方 正信[‡] 滝谷 重治[§]
 北海道大学工学部システム情報工学専攻[†] 北海道大学工学部分子化学専攻[‡]
 北海道大学遺伝子実験施設[§]

1. はじめに

本稿では、1994年に Adleman によって提案された DNA 分子計算の追実験を行うことで、DNA 分子計算の実現の可能性、及び、実際に計算を行う上での問題点に関する検討を行う [1][2]。また DNA のコーディングの際の配列同士の干渉について検討する。

2. 有向ハミルトン経路問題

n 個の頂点をもつ有向グラフにハミルトン経路 (頂点 V_{in} から頂点 V_{out} までの経路で各頂点をちょうど一度だけ含むもの) が存在するかどうかを決定する問題である。この問題は NP 完全問題としてよく知られている。

3. DNA による問題の表現

DNA を用いて有向グラフを実現するために、まず次のように頂点と有向辺を DNA にエンコードする。

頂点 i : 長さ 20 の (一本鎖)DNA O_i

$O_0 = \text{GAGTTCGTTA } \underline{\text{CGGCACAGGA}}$

$O_1 = \underline{\text{ATGGATCAGA}} \text{ CCAGGGGTTTC}$

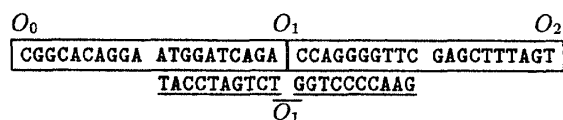
$O_2 = \text{GAGCTTTAGT } \underline{\text{GGCTGTTGGA}}$

i から j の有向辺 : O_i の 3' 末端側の 10 残基と O_j の 5' 末端側の 10 残基を合成した DNA $O_{i \rightarrow j}$

$O_{0 \rightarrow 1} = \underline{\text{CGGCACAGGA}} \underline{\text{ATGGATCAGA}}$

$O_{1 \rightarrow 2} = \text{CCAGGGGTTTC } \underline{\text{GAGCTTTAGT}}$

経路 : 辺 $O_{i \rightarrow j}$ と辺 $O_{j \rightarrow k}$ が存在すれば、 O_j の相補配列 $\overline{O_j}$ が添え木の働きとなって、連結反応が起こる。



4. DNA 分子計算

ハミルトン経路問題を以下の生化学反応で実現する [3]。

Step 1. ランダムな経路を多数生成する 前述のように、問題に存在する辺と頂点の相補配列を用意することで、多数の経路を生成する。

Step 2. V_{in} で始まり、 V_{out} で終わる経路のみを残す O_{in} と $\overline{O_{out}}$ をプライマーとして用いて、PCR (複製連鎖反応) を行うことによって、 V_{in} で始まり V_{out} で終わる経路をエンコードする DNA のみを残す。

Step 3. ちょうど n 個の頂点を含む経路のみを残す ゲル電気泳動によって配列を長さによって分離し、 n 個の頂点を含む DNA ($n=7$ の時、140 塩基対) のみを取り出す。

Step 4. 各頂点を含む経路のみを残す 大腸菌を導入し、各経路を種類ごとに分離して一次的に保存する。各頂点をプローブとしたハイブリダイゼーションを利用し、全ての頂点を含む経路のみを取り出す。

Step 5. もし経路が残っていれば、それはハミルトン経路である 精製した DNA の塩基配列を決定し、ハミルトン経路の存在の有無を確認する。

ただし、ここで経路が残らなかったからといって、ハミルトン経路が存在しないというわけではない。多数の経路を生成する際に、全ての組合せを完全に網羅できるわけではないためである。

5. DNA コーディングの検討

過去の分子計算の実験における問題点として、遺伝子反応には常に誤差を伴うことが指摘されていた [4]。

例えば、Step1 でランダムな経路を多数生成させる場合など、似た配列同士で干渉が起こるため不適切な辺が残り、目的とする結果が得られない。

Development of fundamental aspects of DNA Computation

[†]Nobuhito OKADA, Masahito YAMAMOTO, Azuma OHUCHI

[‡]Toshikazu SHIBA, Masanobu MUNEKATA

[§]Shigeharu TAKIYA

[†]Division of Systems and Information Engineering, Faculty of Engineering, Hokkaido University

[‡]Division of Molecular Chemistry, Faculty of Engineering, Hokkaido University

[§]Research Center for Molecular Genetics, Hokkaido University

2つの配列 A, B があると、各配列は二重螺旋構造を構成し、Watson-Click の相補対からなる相補配列 \bar{A} と \bar{B} が形成される。このとき干渉の起きやすさを調べるには、 A と B 、 A と \bar{B} 、また、 A と \bar{A} 、 B と \bar{B} のそれぞれの類似度を調べる必要がある。

類似度の計算は以下のように行う。2つの配列を並べて、同じ文字が現れる回数を数え上げていく。同じ文字が n 回連続して現れたときには、 2^{n-1} としてスコアに加算する。1文字ずつ配列をずらしながらこれを行い、スコアが最大となるものを最終的な類似度とする。

```
;; 類似度計算のアルゴリズム
procedure SCORE(seq1, seq2:array[1..n])
var
  i, j, pos, score, cont, maxscore:integer;
begin
  for j := 1 to n * 2 begin
    score := cont := 0;
    for i := 1 to n begin
      pos := i + n - j;
      if 1 ≤ pos ≤ n & seq1[pos]=seq2[i] then begin
        score := score - cont;
        if cont = 0 then
          cont := 1;
        else
          cont := cont * 2;
        score := score + cont;
      end
    end
    if maxscore < score then
      maxscore := score;
  end
  return maxscore;
end;
```

上の定義を用いて、以下の手続きにより、各配列間で類似度の少ない、すなわち干渉の起こりにくい配列コーディングをデザインする。

```
;; コーディングデザイン
;; 配列  $O(0) \sim O(n-1)$  を作成、
;; 類似度の上限  $sc_{max}$  の初期値を与える。
procedure CODING_DESIGN(O:array[0..n-1],
var
  maxscs:integer)
  i, j, l, sc:integer
begin
  for i := 0 to n - 1 do
    l := 0;
    repeat
      l := l + 1;
      if l > L then begin
        maxsc := maxsc + 1;
        l := 0;
      end
    until
      Create_Random_seq(O[i], n);
    for j := 0 to i do
```

```
      sc := SCORE(O[i], O[j]);
      if sc > maxsc then break;
    end
  until
    sc < maxsc;
  end
  return O;
end;
```

6. 実験

7頂点ハミルトン経路問題に対し、実際に作成した配列データを用いて、現在実験を進行中である。デザインした配列データは以下のようなもので、右は配列の類似度である。

	0	1	2	3	4	5	6
$O_0 = \text{GAGTTCGTTACGGCAGGA}$	0	16					
$O_1 = \text{ATGGATCAGACCAGGGGTTTC}$	1	16	16				
$O_2 = \text{GAGCTTTAGTGGCTGTTGGA}$	2	18	15	16			
$O_3 = \text{CACACGGGAGGCAGATTGCA}$	3	16	17	16	16		
$O_4 = \text{CGATATGAAGAATTGACTGG}$	4	16	17	16	15	16	
$O_5 = \text{GCCCTCTTATTGCACTCT}$	5	17	18	15	17	17	12
$O_6 = \text{CTCGGTGTTTTGTCGGATGA}$	6	18	18	14	16	16	16

7. おわりに

追実験を行うにあたり、特に DNA 配列のコーディングについての検討を行った。また、この生成法の妥当性を検証するため、配列データの1つを干渉が起こりそうなものに変えたコーディングで、同様の実験を行い結果を比較する。

謝辞

今回、実験に協力して頂いた北海道大学工学部分子化学専攻の提香織氏、田中貴美枝氏に深謝します。

参考文献

- [1] L. M. Adleman : Molecular Computation of Solutions to Combinatorial Problems. Science, Vol. 266, No. 11, pp. 1021-1024(1994)
- [2] Elie Sanchez : DNA Biosoft Computing. 4th International Conferece on Soft Computing(1996)
- [3] 有田 正規, 渡辺 真理 : DNA コンピューティング, 情報処理学会プログラミング研究会研究報告, (May 1996)
- [4] 横森 貴, 小林 聡 : 計算論的言語理論と DNA 計算. 情報処理, Vol. 37, No. 10, pp 929-934(1996)
- [5] Watson 他: 遺伝子の分子生物学. トッパン (1988)