

DNA 分子の濃度と反応速度の関係解析

塩崎 真史 (Masashi Shiozaki)*
定兼 邦彦 (Kunihiko Sadakane)†

小野 廣隆 (Hirotaka Ono)†
山下 雅史 (Masafumi Yamashita)†

* 九州大学大学院システム情報科学府 † 九州大学大学院システム情報科学研究院

* † Department of Computer Science and Communication Engineering,
Graduate School of Information Science and Electrical Engineering, Kyushu University

1 はじめに

0, 1 のビット演算で処理を行う現在の電子計算機の計算能力向上にはマイクロ化が不可欠であるが, そのマイクロ化にも限界が存在することが指摘されている。そこで近年, 生体分子の組み換え規則を利用することにより計算速度, エネルギー効率, 情報格納量の点で革新的な向上が期待できる, 『分子計算』なる新たな計算パラダイムが注目されている。その分子計算の中でも DNA 分子を取り扱うものを特に DNA 計算と呼ぶ。

DNA 計算の原理は, DNA 分子がワトソン・クリック相補性に基づいて選択的に水素結合する構造変化を利用して DNA 分子に計算を行わせることにある。1994 年, Adleman はハミルトンパス問題を, この特徴を利用して解くことに成功した [1]。これにより DNA 計算の研究が盛んに行われるようになった。

DNA 計算が実用上有用であるためには, 構造変化を自在に取り扱うことが必要であり, 高速かつロバストな計算が可能でなければならない。そこで著者らは, DNA 分子の構造変化を解析することに注目している。この解析が正しく行われれば, どのような DNA 配列が速く正確に反応するかが判定可能となり, DNA 計算のさらなる発展が期待できるであろう。

近年, DNA 分子に類似の特徴を持つ RNA 分子のモデル化の研究がなされている [2, 3, 9, 11]。これらのモデルはマルコフ過程に基づいており, RNA の構造変化が各構造の自由エネルギーによって定まる状態遷移確率に依存すると仮定している。結果として, RNA の構造変化は良く模倣されている。しかし, RNA は一本の配列間でのみ塩基対を形成し構造を変化させるため, 2 本鎖を形成する DNA の動作と異なること

が予想される。著者らの知る限りでは, これまでに DNA の 2 本鎖構造の解析に関する研究はなされていない。

これらを考慮し, 著者らは, 試験管内で自由に動き回り, 2 本鎖を形成する DNA 配列の構造変化を計算機上で模倣し, その結果と化学実験結果を比較した [7]。本発表では, このシミュレーションモデルの理論的説明を得るため, 簡易化した状態遷移モデルを用いて DNA 反応速度の解析を行う。

2 準備

以下では, DNA 構造変化の基本となるワトソン・クリック相補性とそれにより形成される DNA 配列の 2 次構造, 及び我々のシミュレーションモデルを構築する上で重要なマルコフ過程・ギブス分布と, RNA の振る舞いの模倣に成功した過去の研究について述べる。

2.1 ワトソン・クリック相補性

DNA 分子は, 糖, リン酸, 4 種類の塩基からなるヌクレオチドが一本鎖上に結合してできる生体高分子である。塩基のアデニン (A), とチミン (T), シトシン (C) とグアニン (G) は選択的に水素結合する性質を持ち, これ以外の組合せでは結合しない。この組合せの原理をワトソン・クリック相補性と呼び, これにより結合された 2 つの塩基を塩基対と呼ぶ。

RNA 分子の場合にはチミン (T) の代わりにウラシル (U) が用いられる。また, グアニン (G) とウラシル (U) の間にも弱い結合が生じる。

2.2 配列の2次構造

ある DNA 配列 a に対し, その i 番目の塩基と j 番目の塩基が水素結合して形成されている塩基対を (i, j) で表すことにする (ただし, $i < j$ とする). 3つの塩基の位置 i, j, r に対して, $i < r < j$ が成り立ちかつ塩基対 (i, j) が存在する時, (i, j) は r を囲むという. 同様に塩基対 (p, q) に対し, (i, j) が p と q を囲む時, (i, j) は (p, q) を囲むという.

塩基対 (p, q) あるいは塩基 r は, (i, j) に囲まれていて, かつ (i, j) に囲まれている任意の塩基対 (k, l) に囲まれていない時, (i, j) に近接するという.

塩基対 (i, j) に対し, 閉路 $c(i, j)$ を塩基対 (i, j) と (i, j) に近接する塩基及び塩基対からなる部分構造と定義する. DNA 配列の自由エネルギーは, この閉路 $c(i, j)$ によって定義される [6].

(i, j) に近接する塩基対が $(i+1, j-1)$ である場合, スタックした塩基対と呼ぶ. スタックした塩基対は安定していて自由エネルギーが小さいが, それ以外の閉路は不安定であり, エネルギーが高い.

2.3 マルコフ過程

マルコフ過程とは, 状態遷移確率が直前の状態のみに依存する確率過程のことであり, 一般に以下のように表される.

$$P_r(X_{t+1} = x_{t+1} | X_1 = x_1, X_2 = x_2, \dots, X_t = x_t) \\ = P_r(X_{t+1} = x_{t+1} | X_t = x_t)$$

DNA の構造変化が過去の状態に非依存であること, ランダムに振る舞うことを考慮すると, マルコフ過程に基づくモデル化は妥当であると言える. ここでは, DNA の各構造をマルコフ過程の各状態に対応させる. また, DNA が各構造の自由エネルギーにより定められる遷移確率に基づいて構造を変化させることとする.

2.4 ギブス分布

定常状態の DNA の分布関数は, 以下のように定義される. [12].

$$Z = \sum_{x \in X} \exp \left\{ -\frac{E(x)}{RT} \right\}. \quad (1)$$

ここで X は全構造の集合, $E(x)$ はエネルギー関数, R は気体定数, T は絶対温度である. この関数では, 自由エネルギーが小さくなれば, 重みが高くなる. 定常分布 $\pi(x)$ はギブス分布を満たし, 以下のように定義される.

$$\pi(x) = \frac{\exp \left\{ -\frac{E(x)}{RT} \right\}}{Z}. \quad (2)$$

2.5 RNA に関する過去の研究

Wolfinger ら [11] は, 以下の状態遷移確率を用いることで, RNA の振る舞いを模倣することに成功した.

$$P(X_{t+1} = y | X_t = x) = \frac{r_{xy}}{\sum_{z \in N(x)} r_{xz}}, \quad (3)$$

ここで $N(x)$ は状態 x の近傍であり, r_{xy} は以下に定義される遷移割合である.

$$r_{xy} = \exp \left\{ -\frac{E_{yz}^* - E(x)}{RT} \right\} \\ E_{yz}^* = \max\{E(y), E(z)\}.$$

しかし, この状態遷移確率では, 2 本鎖を形成する DNA の振る舞いは模倣できない.

3 シミュレーションモデル

この節では, モデルの対象としている DNA 反応と, 我々の提案するシミュレーションモデルについて説明する.

3.1 問題と問題例

モデル化する DNA 反応として, ある DNA 配列とその相補配列を同量, 同じ試験管の中に同時に入れ, 一定温度下で反応させる状況を考える. この時点では, 全ての DNA は塩基対を形成していない. 試験管の中に投入された後, 各 DNA 配列は構造を変化させ, 時間が経過すると完全 2 本鎖を形成する.

この反応の例として, 同じ配列長でほぼ同じ MFE を持つ 3 種類の DNA 配列を用いる. ここで MFE とは, 塩基対を全く形成していない構造を 0 とした時の, その配列と相補配列を用いて形成される構造の相対最小自由エネルギーである. 表 1 に, その

DNA 配列と、上嶋ら [10] により部分的に改良された ViennaRNA-1.4 package [4, 5] によって求められた MFE (kcal/mol) の値を示す。

表 1: 実験に使用した DNA 配列とその MFE

ID	Sequence	MFE
60	TTCGCTGATTGTAGTGTTCACA	-36.16
171	CGCGATTTCCTATTGATTGATCCC	-34.78
176	GGGATCAATCAATAGGAATCGCG	-37.10

図 1 は、上記の反応の生化学実験の結果である¹。横軸は実時間であり、縦軸の蛍光強度は 2 本鎖間のスタックした塩基対の数に比例する。この実験は、温度 298(K)、DNA 濃度 10nM/l で行われた。

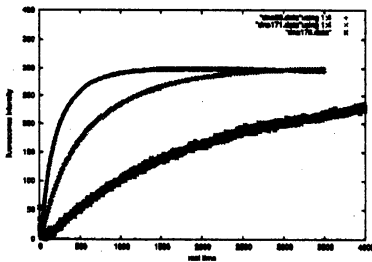


図 1: 生化学実験結果 (60, 171, 176)。

3.2 シミュレーションモデル：提案

DNA 配列のランダムな振舞いを模倣するため、 n 本の DNA 配列とその相補配列を同数用意する。全 DNA 配列の集合を S 、 i 番目の DNA 配列を s_i ($\in S$) とする。また、 N を 2 本鎖間で塩基対を形成していない DNA 配列の集合とする。 P を試験管中で出会った 2 本の DNA 配列の対 $\{s_i, s_j\}$ の集合とする。

初期状態では、 $\{s_1, s_2, \dots, s_{2n}\} \in N$, $P = \phi$ である。以下の操作を各ステップで行う。

1. $N \neq \phi$ なら、等確率でランダムに 2 本の DNA 配列 $s_i, s_j \in N$ を選び、 $\{s_i, s_j\}$ を P に追加する。同時に、 s_i, s_j を N から削除する。この操作を、 $N = \phi$ になるまで繰り返す。

2. P の各要素 $\{s_i, s_j\}$ ごとに、 s_i と s_j を合わせて一つの状態とし、各構造の自由エネルギーによって定まる状態遷移確率 (後述) に従って次状態に遷移する。

3. この時点で s_i と s_j ($\{s_i, s_j\} \in P$) の間に塩基対が存在しなければ、 $\{s_i, s_j\}$ を P から削除する。同時に、 s_i, s_j を N に追加する。

1.~3. の操作を繰り返し、各ステップで 2 本鎖間のスタックした塩基対の総和を評価する。

近傍

DNA の構造変化において、複数の塩基対が同時に結合・解離することが起こり得る。ここでは簡単化のため、現在の構造から一つの塩基対を変化させて形成される構造の集合を近傍と定義する。また、現在の構造も近傍に含まれることとする。

衝突係数 p

DNA 配列は試験管内を自由に動き回るので、ある配列と他の配列の間の塩基対は、2 本の配列が出会わなければ形成されない。つまり、この反応を模倣するためには、どの程度の頻度で 2 本の配列が出会うかを考慮する必要がある。これを反映させるため、我々のシミュレーションモデルに状態遷移を制御する衝突係数 p を導入する。

この係数は DNA の濃度に依存し、理論的に一意に決定されるべきである。現段階では、適当な p の値が得られていないので、いくつかの p の値を用いて計算機シミュレーションを行った。3.3 節で、それらの結果の中の一つを示す。

状態遷移確率

我々のシミュレーションモデルでは、2.5 で紹介した遷移確率に、上述の衝突係数 p ($0 < p \leq 1$) を組み合わせた以下の遷移確率を用いる。

$$P(X_{t+1} = y | X_t = x) = \frac{R_{xy}}{\sum_{z \in N(x)} R_{xz}}, \quad (4)$$

¹このグラフは、東京大学の陶山教授に提供して頂いた。

ここで R_{xy} は以下に定義される新しい遷移割合である。

$$R_{xy} = \begin{cases} p \cdot r_{xy} & (2 \text{ 本鎖間の最初の塩基対}) \\ r_{xy} & (\text{その他}) \end{cases}$$

状態遷移確率を上記のように定義すると、 $p = 1.0$ の時には遷移エネルギー $E_{yx}^\ddagger$ は、 $E_{yx}^\ddagger = E_{xy}^\ddagger$ を満たし、式 (2) を満たす。

3.3 実験結果と考察

計算機シミュレーションと生化学実験の結果を比較するために、表 1 の 3 種類の DNA 配列を用いる。計算機シミュレーションでは、100 本の主配列と同数の相補配列を合わせた 200 本の DNA 配列を入力として与える。また、生化学実験の結果に合わせるため $RT = 2.0$ となるように温度を設定する。自由エネルギーの計算には、前述の ViennaRNA-1.4 package を使用する。

図 2 は、衝突係数 $p = 0.001$ とした時のシミュレーション結果である。横軸はステップ数、縦軸は 2 本鎖間のスタックされた塩基対の総和である。

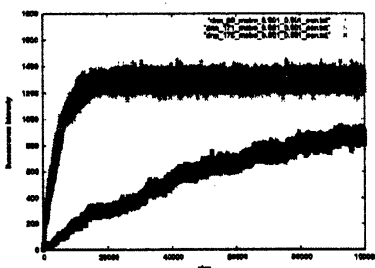


図 2: 実験結果 ($p = 0.001$)

図 1 と図 2 のグラフの形状は、多くの共通点を示している。例えば、どちらのグラフにおいても、60 番と 171 番の DNA 配列は、176 番の配列と比べて非常に速く反応している。この結果は、衝突係数を導入する前では得られなかった。また、すべてのグラフの形状は指数関数的に増加している。しかし、60 番と 171 番の反応速度の関係が逆になっている等、異なる部分も見受けられる。

シミュレーションを通じて、 p の値を小さくすれば、176 番の反応が他の配列に比べて劇的に遅くなると

ということが分かった。これは、176 番の配列は 1 本の配列内で塩基対を形成し易い傾向にあるということが考えられる。表 2 は、例の 3 本の DNA 配列の、主配列と相補配列の一本鎖内の LMFE(kcal/mol) を示したものである。ここで LMFE とは、塩基対を全く形成していない構造を 0 とした時の、相対局所最小自由エネルギーのことである。この表は、176 番は、主配列・相補配列ともに一本鎖内で安定な構造を形成するということを意味する。一方、60 番と 171 番の配列は一本鎖間内では安定な構造を取らないことを意味する。これが、176 番と他の 2 つで反応速度に大きな違いがある原因であろう。

生化学実験との結果の比較は、まだ多くの問題が残されているとはいえ、我々のモデルが DNA の構造変化の良い近似となっていることを示している。次節では、このモデルを理論的に解析する。

表 2: 問題例の一本鎖内の LMFE

	LMFE (kcal/mol)	
ID	主配列	相補配列
60	0.0	0.0
171	0.0	0.0
176	-4.92	-3.88

4 提案モデルの理論的解析

前節のシミュレーション結果は、我々のモデルが DNA の分子構造変化を上手く模倣できていることを示している。本節では、そのモデルを通して、DNA の反応速度を理論的観点から解析する。しかし、我々のモデルは、厳密に解析するには非常に複雑過ぎるという問題がある。この困難を避けるために、単純化したモデルを用いて反応速度を近似計算する。このために、我々は 3 状態のみを持つ簡易マクロ状態遷移モデルを提案する。これは、塩基対を全く形成していない状態から、平衡状態で完全 2 本鎖に貢献する塩基対が一つ以上形成されるまでの平均ステップ数を見積もるためのモデルである。そのような塩基対が形成されると、短時間で完全二本鎖を形成することを [8] で示したので、これは提案モデルの近似として有用であると考えられる。以下、そのモデルの説明及びそれを

用いた反応速度の解析を行う。

4.1 簡易マクロ状態遷移モデル

DNA の反応を単純にモデル化するために、簡易マクロ状態遷移モデルを導入する。このモデルでは、全ての DNA の構造を 3 状態、 X_1, X_2, X_3 に分ける。初期状態である X_1 は全ての一本鎖の構造の集合であり、これらは他の配列と塩基対を形成していない。 X_2 は完全 2 本鎖を即座に形成することが期待される構造であり、平衡状態で完全 2 本鎖に貢献する塩基対が一つ以上形成されている構造の集合である。 X_3 は他の全ての構造であり、不安定な 2 本鎖の構造等を含む。

図 3 はその状態遷移図を表している。 q_{ij} を X_i から X_j への遷移確率とすると、以下の式を満たす。 $(i = 1, 2, 3)$

$$q_{i1} + q_{i2} + q_{i3} = 1 \quad (5)$$

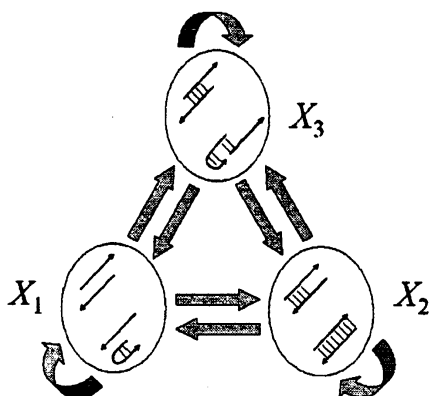


図 3: 状態遷移図。

ここで、時刻 t で状態 X_k である存在確率を $P_{X_k,t}$ と定義すると、以下の漸化式を得ることができる。

$$P_{X_1,t} = q_{11}P_{X_1,t-1} + q_{21}P_{X_2,t-1} + q_{31}P_{X_3,t-1}, \quad (6)$$

$$P_{X_2,t} = q_{12}P_{X_1,t-1} + q_{22}P_{X_2,t-1} + q_{32}P_{X_3,t-1}, \quad (7)$$

$$P_{X_3,t} = q_{13}P_{X_1,t-1} + q_{23}P_{X_2,t-1} + q_{33}P_{X_3,t-1}. \quad (8)$$

$P_{X_1,0}=1, P_{X_2,0}=0, P_{X_3,0}=0$ であり、これと上の式より、時刻 t における各状態の存在確率を得ることがで

きる。以下に、 $P_{X_2,t}$ の結果のみ示す。

$$P_{X_2,t} = (Y + Z)A^t - ZB^t - Y \quad (9)$$

ここで、

$$S = q_{12} + q_{23} + q_{31} + q_{13} + q_{32} + q_{21},$$

$$T = q_{12}q_{23} + q_{23}q_{31} + q_{31}q_{12} + q_{13}q_{32} +$$

$$q_{32}q_{21} + q_{21}q_{13} + q_{21}q_{31} + q_{12}q_{32} + q_{13}q_{23},$$

$$A = \frac{2 - S + \sqrt{S^2 - 4T}}{2},$$

$$B = \frac{2 - S - \sqrt{S^2 - 4T}}{2},$$

$$Y = -\frac{q_{13}q_{32} + q_{12}q_{32} + q_{12}q_{31}}{T},$$

$$Z = \frac{1}{\sqrt{S^2 - 4T}} \left\{ q_{12} - \frac{2(q_{13}q_{32} + q_{12}q_{32} + q_{12}q_{31})}{(S + \sqrt{S^2 - 4T})} \right\}.$$

$0 \leq S \leq 3, 0 \leq T \leq 3, Y \leq 0$ であることは容易に示される。この結果は、即座に完全 2 本鎖となる状態の存在確率が指数関数的に増加することを示しており、これは前節のシミュレーション結果と一致する。

4.2 シミュレーション・生化学実験の解析

前述の簡易マクロ状態遷移モデルにより導かれる関数 $f(t) = (a + b) \exp(ct) - b \exp(dt) - a$ を用いて、最小二乗法によりシミュレーションと生化学実験の結果を適合した。どちらの結果も、ほとんどの場合でこの関数を適合することができた。この $f(t)$ を t に関して微分することによって、以下の反応速度式が得られる。

$$\frac{d}{dt}f(t) = (a + b) \cdot c \cdot \exp(ct) - b \cdot d \cdot \exp(dt) \quad (10)$$

$\epsilon > 0$ をある定数とし、 $df(t)/dt < \epsilon$ を満たす最小の整数を収束点と定義する。表 3, 4 はそれぞれ、 $\epsilon = 0.001$ とした時の生化学実験とシミュレーションの収束点の結果である。生化学実験では、60, 171 番に関しては、濃度と反応速度がほぼ比例の関係になっている。また、計算機シミュレーションでは、衝突係数 p の値が小さくなると、反応速度が遅くなるという傾向がある。

これらに関して、前述の簡易マクロ状態遷移モデルを用いて考察する。生化学実験では濃度が高くなると、状態遷移確率 q_{12}, q_{13} が比例して大きくなる。また、それに伴い q_{11} が小さくなる。一方、計算機シミュ

レーションの衝突係数 p を k 倍することは、以下に対応する。

$$\begin{aligned} q_{11} &\rightarrow \frac{q_{11}}{q_{11} + k(q_{12} + q_{13})}, \\ q_{12} &\rightarrow \frac{kq_{12}}{q_{11} + k(q_{12} + q_{13})}, \\ q_{13} &\rightarrow \frac{kq_{13}}{q_{11} + k(q_{12} + q_{13})}. \end{aligned}$$

よって、我々のモデルの衝突係数は、生化学実験の濃度の部分を制御する係数であると言える。

表 3: 生化学実験の収束点

ID	濃度 (nmol/l)	収束点 (秒)
60	5	3,217
	10	1,962
	20	750
171	5	10,391
	10	4,994
	20	2,069
176	10	15,420
	20	-

表 4: 計算機シミュレーションの収束点

ID	衝突係数 p	収束点 (ステップ)
60	0.01	7,625
	0.001	27,240
	0.0001	95,632
171	0.01	6,975
	0.001	24,401
	0.0001	137,254
176	0.01	70,713
	0.001	132,407
	0.0001	-

5 結論と今後の課題

本稿で紹介した簡易マクロ状態遷移モデルを用いることにより、[7] で我々が提案したシミュレーションモデルの反応速度の解析を行った。これをシミュレーション結果、生化学実験結果に適合させたところ、このモデルはそれらの良い近似となっていた。

今後の課題としては、衝突係数 p と濃度の関係や、それらと反応速度の詳細な関係を解析する予定である。

謝辞

生化学実験データを提供して下さった東京大学陶山教授に感謝致します。この研究の一部は文部科学省科学研究費の援助を受けた。

参考文献

- [1] L. Adleman *Molecular Computing of Solutions to Combinatorial Problems* Science 266, pp 1021-1024, 1994
- [2] Christoph Flamm *Kinetic Folding of RNA* Dissertation, 1998
- [3] Christoph Flamm, Walter Fontana, Ivo L. Hofacker and Peter Schuster *RNA Folding at Elementary Step Resolution* Santa Fe Institute 12-078, 1999
- [4] I. L. Hofacker *The Vienna RNA Secondary Structure Server* Nucleic. Acids Res. 31 3429-31, 2003
- [5] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer M. Tacker and P. Schuster *Fast Folding and comparison of RNA Secondary Structured* Monatsh. Chem. 125 167-88, 2003
- [6] Satoshi Kobayashi *Strand Design for Molecular Computation Sciences*, 2003
- [7] Masashi Shiozaki, Hirotaka Ono, Kunihiko Sadakane and Masafumi Yamashita *Modeling DNA Conformation Change and Theoretical Analysis on the Reaction Rate*, DNA11, p408, 2005
- [8] Masashi Shiozaki, Hirotaka Ono, Kunihiko Sadakane and Masafumi Yamashita *分子構造変化のモデル化と反応速度の理論的解析*, LA, p232 ~237, 2005
- [9] Manfred Tacker, Walter Fontana, Peter F. Stadler and Peter Schuster *Statistics of RNA Melting Kinetics* Eur. Biophys. J., 23:29-38, 1993
- [10] H. Uejima, M. Hagiya *Analyzing Secondary Structure Transition Paths of DNA/RNA molecules*, Preliminary Proc. of 9th Int. Meeting of DNA Based Computers(DNA9) pp.4731-4741, 2004
- [11] Michael T Wolfinger, W Andreas Svrcek-Seiler, Christoph Flamm, Ivo L Hofacker and Peter F Stadler *Efficient Computation of RNA Folding Dynamics* J. Phys. A: Math. Gen. 37: 4731-4741, 2004
- [12] Michael Zucker *The Equilibrium Partition Function*, 2003