

分子構造変化のモデル化と反応速度の理論的解析

塩崎 真史 (Masashi Shiozaki)*

小野 廣隆 (Hirotaka Ono)[†]定兼 邦彦 (Kunihiko Sadakane)[†]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

概要

DNA 計算とは、DNA 分子が組み換え規則に従って反応し、タンパク質を形成していくプロセスを計算処理と見立てた新しい計算パラダイムである。DNA 計算を自由に取り扱うためには、その構造変化を解析することが必要不可欠であるが、複雑な構造をとる DNA 分子の振る舞いを予測することは困難であり、これまでのところ解明されていない。本稿では、DNA 分子の構造変化をマルコフ過程とみなすことにより、その解析を試みる。そこで、計算機上でシミュレーションモデルを作成し、化学実験の結果と比較することにより、我々の試みの正当性を検証する。また、この複雑なシミュレーションモデルに対する理論的解析の第一歩として、構造に制約を加えることにより単純化したモデル上で、DNA 分子の反応速度を近似計算する。

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、また高速かつロバストに計算可能であることが不可欠である。つまり、DNA 分子の構造変化を解析することが DNA 計算の飛躍的な進歩に繋がると考えられる。近年、DNA 分子に類似の特徴を持ち、1 本の配列間でのみ塩基対を形成する RNA 分子に対して、様々な解析がなされている。[2, 3, 9, 11]

これらの研究では、各構造が持つ自由エネルギーから定まる状態遷移確率によって RNA 配列の構造が変化するという仮定の下、RNA 分子の構造変化をマルコフ過程とみなしてシミュレーションモデルを構築している。結果として、RNA の振る舞いは上手く模倣されている。

本研究では、同様のアプローチで DNA 分子の構造変化をマルコフ過程とみなし、その解析を行う。この正当性を証明するために、ある DNA 配列とその相補配列を同量同じ試験管に入れ、一定の温度下で反応させるというシミュレーションモデルを計算機上で作成し、化学実験の結果と比較する。

さらに、解析の第一歩として、複雑なシミュレーションモデルに対する理論的説明を得るため、動作を限定した簡単なモデル上での DNA 分子の反応速度を近似計算する。

本稿の構成は以下の通りである。第 2 節では、DNA 計算の基本となるワトソン・クリック相補性、配列の

2次構造について説明した後、マルコフ課程及びギブス分布について述べる。第3節では、計算機シミュレーションモデルの説明の後、予備実験の結果を紹介する。第4節では、2本鎖間の塩基対を考慮したモデルについて説明し、その後実験結果について述べる。第5節ではシミュレーションモデルの理論的解析の第一歩として、簡単化したモデル上での反応速度の近似計算について述べる。最後に第6節では3,4,5節より得られた結果について議論する。

2 準備

以下では、DNA 計算の基本となるワトソン・クリック相補性、配列の2次構造及びマルコフ課程、ギブス分布について述べる。

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1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。

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つの塩基対 i, j, r に対して、 $i < r < j$ が成立かつ塩基対 (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。閉路 $c(i, j)$ が持つ塩基対の数 k により、配列の2次構造を以下のように分類する。

1. $k = 1$ の時、ヘアピンと呼ぶ。

2. $k = 2$ でかつ、 (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
3. $k = 2$ でかつ、 (i, j) に近接する塩基対がある p ($i < p < j$) に対して、 $(i + 1, p)$ or $(p, j - 1)$ となる場合で、バルジループと呼ぶ。
4. $k = 2$ で上記以外の場合、内部ループと呼ぶ。
5. $k > 2$ となる場合、マルチループと呼ぶ。

スタックした塩基対は安定するが、それ以外の構造は不安定であることが知られている [7]。

また、塩基対が全く形成されていない構造を開鎖と呼ぶこととする。

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。

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

(X_i : 確率変数, x_j : ある2次構造)

DNA の構造変化が過去の状態に依存しないことと、ランダムな動作をすることを考慮すると、マルコフ過程を用いてモデル化することが妥当であると考えられる。DNA 配列の各々の構造をマルコフ課程の状態に対応させることによって、DNA の構造変化を状態遷移図で表すことができる。また、各構造が独立に持つ自由エネルギーによって定められる状態遷移確率に基づいてDNA配列が構造を変化することとする。

2.4 ギブス分布

分布関数は以下のように与えられる。[12]

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

X : 全状態の集合 $E(x)$: エネルギー関数 R : 気体定数 T : 絶対温度
--

この関数では、自由エネルギーが小さくなると重みが大きくなることに注意する。統計力学によると、このボルツマンの重みづけは全ての構造の確率密度を与える。つまり、定常状態 $\pi(x)$ は以下ようになり、ギブス分布を満たす。

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

3 マルコフ過程に基づいたシミュレーションモデル

この節では、本研究で対象とする問題、及び構築したシミュレーションモデルについて説明する。

3.1 対象とする問題

ある DNA 配列とその相補配列を同量、同じ試験管に入れ、一定の温度下で反応させた時の分子の反応速度について考える。

以下では、同じ配列長で、その配列と相補配列から形成される構造の最小エネルギー値がほとんど同じであるにも関わらず、化学実験の結果で反応速度がそれぞれ大きく異なる 3 つの配列に関して比較する。

ID	配列	mfe
60	TTCGCTGATTGTAGTGTTCACACA	-30.59
171	CGCGATTTCCTATTGATTGATCCC	-29.21
176	GGGATCAATCAATAGGAATCGCG	-31.17

※ mfe : 開鎖に対する相対最小自由エネルギー (kcal/mol).

図 1 に、この問題に対する化学実験の結果を示す。

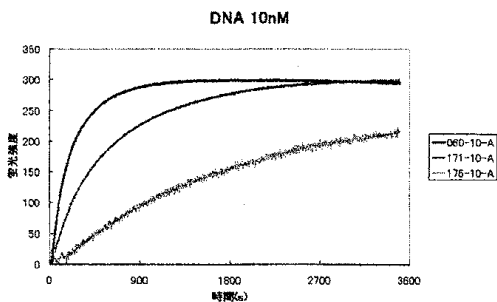


図 1: 化学実験の結果

※ 縦軸の蛍光強度は 2 配列間のスタックした塩基対の数に比例する。

3.2 シミュレーションモデル

試験管内を自由に動き回る DNA 配列を実装するため、それぞれ異なる構造を持つ n 本の DNA 配列を用意し、以下のようなシミュレーションモデルを構築した。

各ステップで以下の操作を行う。

1. 他の配列と塩基対を形成していない配列を等確率でランダムに 2 つ選び出し、仮想的なペアを形成する。
2. それぞれのペア間で近傍への状態遷移確率を求め、その確率に従いランダムに状態遷移する。
3. この時点で 2 配列間で塩基対を形成していれば、そのまま次のステップでも同じ配列とペアを形成する。そうでなければ次のステップでは別のペアを形成する。

これらを繰り返し、各ステップで 2 配列間のスタックした塩基対の総和を評価する。

・近傍の定義

現実には起こり得る現象としては、塩基対が同時に複数生じる、または解離することが考えられる。しかし本稿では単純化のため、ある構造から塩基対が一つ生じる、または一つ解離することにより形成される構造を近傍とする。また、現在の状態も近傍に含まれることとする。

・状態遷移確率

分子が反応し、別の状態に遷移する際には、各状態間に存在する活性化エネルギーより大きなエネルギーが必要である。故に、DNA の構造変化に活性化エネルギーが密に関係していると考えられる。しかし、任意の状態間の活性化エネルギーはエントロピーやエントルピーに依存しており、これを求めることは困難である。そこで本稿では、活性化エネルギーを 0 と仮定した上で各構造が持つ自由エネルギーのみを考慮し、以下の状態遷移確率を用いることとする。

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

また、遷移割合 r_{xy} は以下のものを用いる [8].

$$r_{xy} = \exp\left(-\frac{E_{yx}^\# - E(x)}{RT}\right)$$

$$E_{yx}^\# = \max\{E(x), E(y)\}$$

$E_{yx}^\#$ は詳細平衡 $E_{yx}^\# = E_{xy}^\#$ を保証し、この状態遷移確率は式 (2) の定常分布を満たす.

3.3 予備実験の結果と考察

計算機シミュレーションにより得られた結果を図 2 に示す. (横軸: ステップ数, 縦軸: 2 配列間のスタックした塩基対の総和)

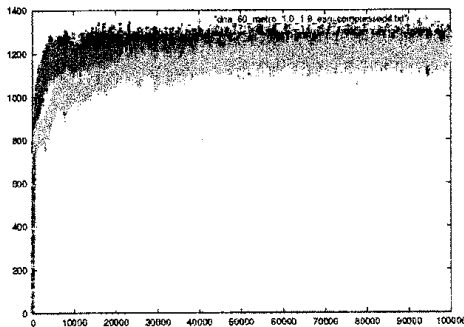


図 2: 計算機シミュレーションの結果

配列長: 23 (mer)
 配列: 主配列, 相補配列それぞれ 100 個ずつ
 RT: 2.0
 エネルギー計算: ViennaRNA-1.4 パッケージ

※ ViennaRNA-1.4 パッケージ [4, 5] は, DNA 配列の自由エネルギーを計算するため, 上嶋らによって改良されている [10].

化学実験の結果と同様, どの配列に関しても時間の経過と共に 2 配列間のスタックした塩基対の数が増加し, ほとんど同じ値に収束していく様子が確認できる. しかしこの実験結果は, 収束速度の点では化学実験の結果と大きく異なる. これは, 現在のシミュレーションモデルは 2 配列間の塩基対形成について詳細を考慮していないことが原因と考えられる. 次節では, DNA 分子の動作を正確に模倣するため, この事を考慮した, より精密なシミュレーションモデルを与える.

4 2 配列間の塩基対形成を考慮したモデル

本節では, シミュレーションをより精密に行うため, 2 配列間の塩基対形成を考慮した衝突係数を導入する.

4.1 衝突係数 p の導入

DNA 分子は試験管内を自由に動き回っており, 2 本の配列が衝突しなければそれらの配列同士の間で塩基対は形成されない. この衝突確率と, その時必要なエネルギー [6], 及び他の力学的要因を考慮すると, 2 本の配列間の塩基対は形成され難いと考えられる. そこで, パラメータ係数 p ($0 < p \leq 1$) を導入し, 一本鎖の状態から 2 本鎖の状態に移る時のみ遷移割合に p を掛けることにより, 上記の要素を満たすモデルを構築する.

この係数は DNA 分子の濃度に依存する値であり, 何らかの形で定められる必要がある. しかし, 現時点で妥当な計算方法が解明されていないので, 本稿では p の値を適当に変化させ, その結果を示すことにする.

4.2 実験結果と考察

図 3 に $p = 0.001$ の時の実験結果を示す.

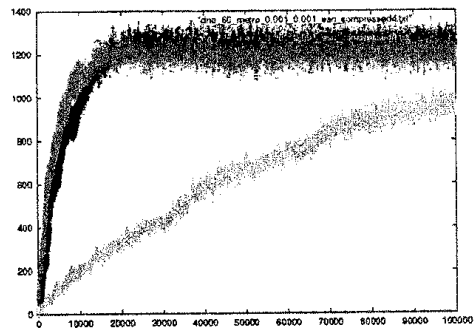


図 3: 係数 $p = 0.001$ を導入した時の結果

一本の配列のみから構成される構造の相対局所最小エネルギー (lmfe) を以下に示す.

ID	lmfe (kcal/mol)
60	0.0
171	0.0
176	-3.6

※ 60番と171番の配列は、1本の配列間で開鎖より小さなエネルギーを持つ構造を形成しない。

実験結果から、係数 p の値を小さくした時、176番のDNA配列の反応速度が極端に遅くなることを確認できた。これは、176番の配列が他の配列と比較すると、一つの配列間で塩基対を形成し、局所最小自由エネルギー状態に陥る傾向があることから、納得できるものである。

しかし、60番と171番の収束速度の関係等、この計算機シミュレーション結果は化学実験の結果と完全には一致していない。故に、配列間の相互作用などを考慮した、より精密なモデルの構築が必要である。

5 理論的解析

本節では、複雑な計算機シミュレーションモデルの正当性を証明するために、開鎖から2本鎖を形成するまでに必要な時間を近似的に計算する。そこで、解析の第一歩として、初歩的なエネルギー関数と2つの制約を導入し、動作を限定した簡単なモデル上での理論的解析を行う。

5.1 エネルギー関数の近似

スタックした塩基対の増加に伴い、各構造が持つ自由エネルギーは小さくなる。この事実から、近似として以下のような初歩的なエネルギー関数を与える。

$$E(x) = -ak + b \quad (x \in M_k) \quad (4)$$

a, b : 定数

M_k : k 個の塩基対を持つ状態の集合

この関数では、塩基対の数に比例して自由エネルギーが減少する。計算機シミュレーションモデルと同じ近傍、状態遷移確率を用いると、遷移割合は以下のようなになる。

$$r_{xy} = \begin{cases} 1 & (E(x) > E(y)) \\ \exp(-\frac{\alpha}{RT}) & (else) \end{cases}$$

以下、簡単化のため $\exp(-\frac{\alpha}{RT}) = A$ とおく。

5.2 2つの状態遷移の制約

以下では、簡略化のため、主配列と相補配列が常に隣接しており、他の配列から影響を受けない状態を仮定したモデルを考える。

5.2.1 結合する塩基対を限定したモデル

主配列と相補配列で、対称的な位置にある塩基同士のみが塩基対を形成すると仮定する。つまり、最終的に2重螺旋構造を形成する塩基対のみを考慮する。このように単純化した場合、配列長を n とすると任意の状態で近傍の数は n となり、 M_k を一つのマクロ状態とみなすことにより、 M_0, M_1, \dots, M_n を頂点とするパスグラフで状態遷移図が表される。この時、状態遷移確率は以下のようなになる。

$$P(X_{t+1} = M_j | X_t = M_i) = \begin{cases} \frac{1}{1+A} & (i < j) \\ \frac{A}{1+A} & (\text{その他}) \end{cases} \quad (5)$$

$H(s, t)$ を M_s から M_t に到達するまでの平均ステップ数と定義すると、以下の漸化式を満たす。

$$H(k-1, k) = 1 + \frac{A(k-1)}{n-(k-1)} H(k-2, k-1) \quad (6)$$

よって、初期状態を塩基対が全く形成されていない状態、最終状態を2重螺旋構造とすると、初期状態から最終状態に到達するまでの平均ステップ数は以下のようなになる。

$$H(0, n) = \begin{cases} O(2^{3n}) \\ \Omega(n^{-\frac{1}{2}} 2^n) \end{cases} \quad (7)$$

これは、配列長 n に対して指数的に到達時間が遅くなることを意味するが、実際の化学実験では長い配列を用いた場合にでもある程度の速度で反応し、2重螺旋構造を形成する。よって、この単純化したモデルは、実際のDNA配列の反応の様子を十分に模倣できていないと言える。

5.2.2 順次的結合・解離モデル

DNA配列がジッパーのように連続的に塩基対を形成するという事実から、塩基対の結合・解離の順番が

決まっているという制約を与えたモデルを与える。ここではこの制約と、結合する塩基対を限定したモデルを組み合わせ、対称の位置に端から順番に結合、解離していくモデルの2重螺旋構造を形成するまでのステップ数を考える。この時、状態数は $n+1$ となり、パスグラフで状態遷移図が表される。そして、初期状態から最終状態に到達するまでの平均ステップ数は以下ようになる。

$$H(0, n) = \frac{1}{(1-A)^2} \{n(1-A) + A(A^n - 1)\} \quad (8)$$

よって、以下の結果が得られる。

$$H(0, n) = O(n)$$

現段階では、この計算結果の正当性を解析していない。今後の課題として、これらのモデルでの結果と計算機シミュレーション結果の比較を行う。加えて、構造を考慮したエネルギー関数、及び状態・近傍の拡張などを導入した、より精密なモデル上での理論的解析を行う。

6 結論

本稿では、DNA 分子構造変化を解析するために、マルコフ課程に基づいた計算機シミュレーションモデルを構築した。また、濃度に依存すると考えられる衝突確率を考慮した、2本鎖形成係数を導入し、その結果を観察した。化学実験の結果と比較したところ、提案するモデルは DNA 分子構造変化を模倣できていると思われる。

しかし、未解決の問題はいくつか残されている。現段階では、濃度に依存すると考えられる衝突係数 p の値を変化させることにより DNA 分子の反応を模倣できているが、この値の妥当な決定方法を未だ得られていない。また、 RT の値に 2.0 という値を代入したが、これに関しても明確な根拠が無い。今後の課題として、著者はこれらの問題に取り組み、適切なシミュレーションモデルを構築する予定である。

また、構造に制約を加え、簡単化したモデルを用いて反応速度の近似計算を理論的に行った。この結果は、より精密なモデルに対する解析の必要性を示唆している。

謝辞

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, Walter Fontana, Ivo L. Hofacker and Peter Schuster "RNA Folding at Elementary Step Resolution", *Santa Fe Institute* 12-078, 1999
- [3] Alexander P. Gulyaev, F. H. D. van Batenburg and Cornelis W. A. Pleij "The Computer Simulation of RNA Folding Pathways Using a Genetic Algorithm", *Academic Press Limited*, 1995
- [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] Alexander J. Hartemink and David K. Fifford "Thermodynamic Simulation of Deoxyribose Nucleotide Hybridization for DNA Computation" *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*
- [7] Satoshi Kobayashi "Strand Design for Molecular Computation", *Sciences*, 2003
- [8] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. "Equations of state calculations by fast computing machines." *J. Chem. Phys.*, 21, 1087-1092
- [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