

ロバスト分子計算

二次構造の解離に基づく汎用分子システムと そのロバストな設計

萩谷昌己

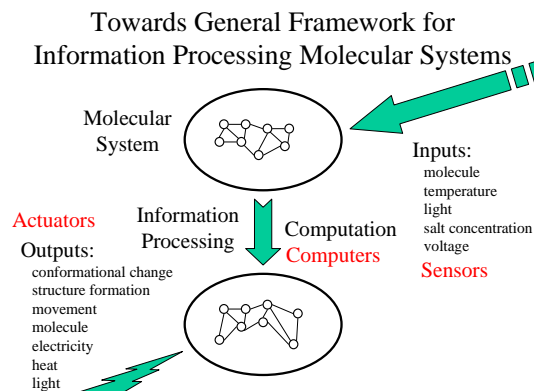
東京大学大学院情報理工学系研究科 および JST CREST

0. はじめに

我々は DNA 分子の作る二次構造を利用した分子コンピューティングの様々な計算モデルを提案し、それらの計算モデルの設計と実装を行って来た[1,2]。最近では、ヘアピンやバルジなどの二次構造の解離によって形態を変化させる DNA マシンの作るネットワークにより、計算能力を持った汎用的な分子システムを構築することを目指して、各種の DNA マシンの設計と実装を進めている[3,4,5]。本発表では、分子メモリ素子、論理ゲート、光センサ、ウォーカなどの具体的な DNA マシンの例を挙げつつ汎用的分子システムの枠組みについて紹介した後、DNA マシンを構成する DNA 分子の設計方法について簡単に説明した。目的の構造のみを安定的に形成し望まれる形態変化を効率よく起こすとともに、望まれない形態変化をできり限り抑制するように DNA 分子の塩基配列を設計することが目標である。具体的には、DNA 分子の塩基配列を構成する要素配列の配列セットを生成する方法、DNA 分子の作る二次構造の予測方法、二次構造の作るエネルギー地形を解析し特にエネルギー障壁を求めて形態変化の起こり易さを予測する方法について述べた。

1. 二次構造の解離に基づく汎用分子システムと分子マシンの例

下図は計算能力を持った汎用的な分子システムのイメージを表現したものである。

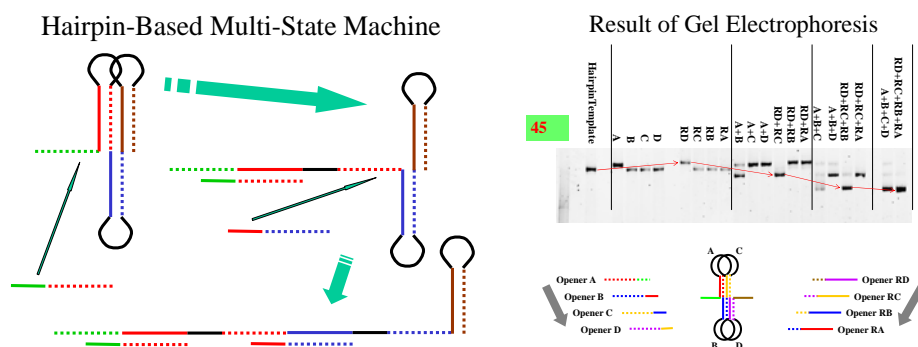


分子システムは相互作用する分子マシンのネットワークから成り立っている。個々の分子マシンはその形態によって状態を表現する。一つの分子マシンの形態変化が他の分子マシンの形態変化を惹き起こし、全体として分子マシンから成る並行計算系を実現する。分子マシンは、環境の状況に従って状態を変化させるセンサ、メモリや論理ゲートなどの情報処理を行うコンピュータ、外界に出力を行うアクチュエータに分類される。以下では具体

的な分子マシンの例について紹介する。

連続したヘアピンから成る多状態分子マシン

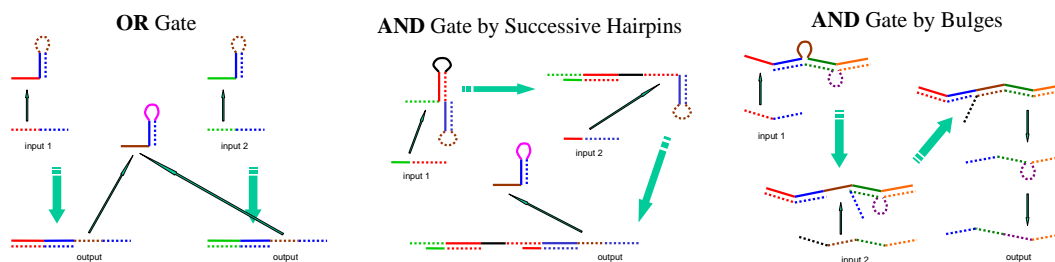
下左図は DNA の二次構造によって表現された複数の状態を持つ分子マシンの例である [3]。この分子マシンは連続したヘアピン構造から成り、ヘアピン構造の開閉によって複数の状態を取る。



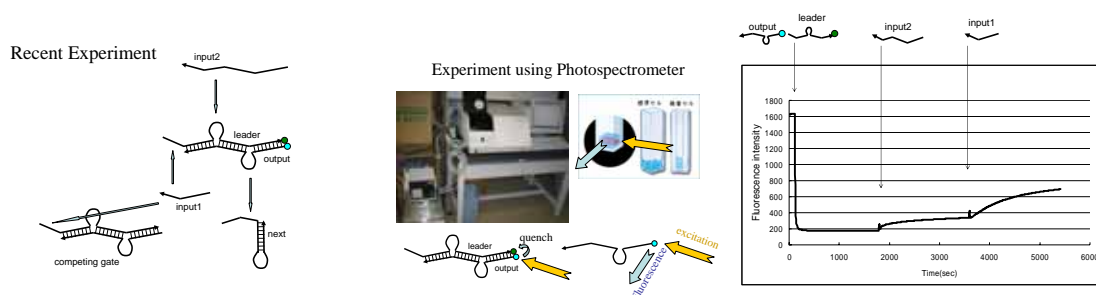
この分子マシンは分子メモリ素子として考案されたものであるが、次項で述べるように論理ゲートなどの様々な用途に用いることができる。オープナと呼ばれる一本鎖 DNA 分子を入力とし、ヘアピンを順に解離させることによって状態を遷移させる。オープナが入力されると、分子マシンの端にある一本鎖部分とハイブリダイズし、ブランチ・マイグレーションと呼ばれる反応によって最初のヘアピンのステム部分に侵入してヘアピンのループを開く。すると、次のヘアピンの隣に一本鎖部分が露出するため、次のオープナがハイブリダイズできるようになる。ヘアピンの解離は定まった順序でしか起こらない。最初のヘアピンが解離しなければ二番目のヘアピンは解離できない。分子メモリ素子として用いる場合は、複数のオープナの組み合わせが階層的なアドレス情報を表す。すべてのアドレス情報が揃ったときにのみ、そのアドレスを持つ分子メモリ素子のすべてのヘアピンが開き、その分子メモリ素子にアクセスすることができる。上右図は筆者が代表を務める JST CREST のプロジェクトにおいて北大の亀田たちが行った実験結果の一部で、四つのヘアピンが順に開いて行く様子をゲル電気泳動によって確認したものである。

論理ゲート

下の三つの図は、汎用分子システムのコンピュータ部分を構成する論理ゲートの例を示している。それぞれ、ヘアピンの解離を用いた OR ゲート、連続ヘアピンの解離を用いた AND ゲート、バルジの解離を用いた AND ゲートを示している。

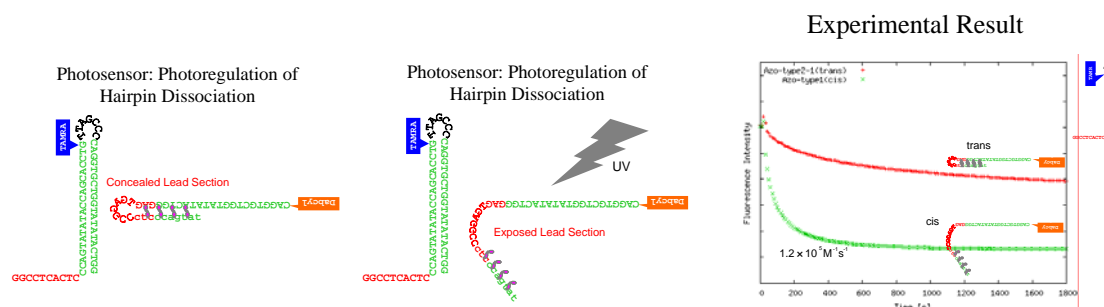


下図は、バルジを用いた AND ゲートに関する実験結果を示している[4]。二本の一本鎖 DNA 分子が二つのバルジ・ループから成る構造を形成している。この構造は入力 1 と入力 2 の両方が存在するときのみ解離し、下の一本鎖 DNA を出力として放出する。



光センサ

下図は名大の浅沼たちと共同で開発を行った紫外光に反応するセンサ分子マシンである [5]。アゾベンゼンが側鎖に挿入されたオープナが小さなヘアピンを形成しており、このままではオープナとして働かない。アゾベンゼン分子は紫外光(UV)を照射するとトランス体からシス体に異性化する。すると、小ヘアピンが不安定になりオープナとして働くことができる。このようにしてヘアピンの解離を光によって制御することが可能である。



2. ロバストな配列設計

以上のような分子マシンを実際に構築するためには、分子マシンを構成する DNA 分子の塩基配列を適切に設計せねばならない。DNA 分子の配列設計問題は大きく二つに分かれる。一つは目的の構造を安定に取る配列を設計する問題である。このためには熱力学に基づく平衡論的解析が行われる。もう一つは、望みの形態変化を起こし易く望まれない形態変化を起こしにくい配列を設計する問題である。このためには何らかの運動論的解析が必要になる。以下の図は DNA マシンの配列設計についてまとめている。

Sequence Design for DNA Machines

- Stable structure → stat-thermo analysis
 - Designing component sequences
 - (Extended) template method
 - Avoiding unwanted secondary structures
 - Free energy prediction of secondary structures
- Dynamic behavior → kinetic analysis
 - Estimation of energy barrier
 - Analyzing energy landscape of secondary structures (measure for unwanted reactions)
 - Estimation of reaction time
 - Probabilistic simulation (measure for expected reactions)

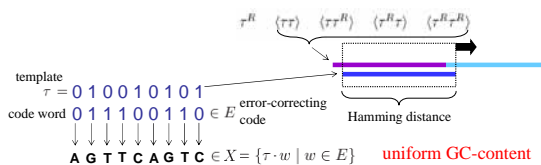
安定構造を保証する配列設計

DNA マシンを構成する DNA 分子の塩基配列は、いくつかの要素配列を連結して求めることが一般的である。個々の要素配列は自分の相補鎖以外の配列とは誤ってハイブリダイズしてはならない。このような要素配列から成る配列セットを設計する問題は DNA コンピューティングの分野で古くから研究されて来た。上述の分子マシンを設計する際にもいくつかの方法を用いているが、以下ではテンプレート法と呼ばれる設計方法とその拡張について紹介する[6]。

下左図のテンプレート法では、テンプレートと呼ばれる 0 と 1 から成る文字列と誤り訂正符号を組み合わせて配列セットを生成する。すなわち、テンプレート中のビット 0 が A か T を表し、テンプレート中のビット 1 が G か C を表す。そして、誤り訂正符号の符号語の各ビットと組み合わせることにより ATGC が確定する。テンプレートとしては、反転してもずらしても、連結したものと重ね合わせても、必ず一定の数以上のミスマッチが保証されるようなものをコンピュータで網羅的に探索する。すると、テンプレートの保証するミスマッチ数と符号語の保証するハミング距離により、相補鎖およびそれらを連結したものも含めて、得られた配列の間には完全に相補的な場合意外は一定の数以上のミスマッチが保証される。なお、以上のようなテンプレートを用いると、要素配列に含まれる G と C の比率が一定となり DNA 分子の融解温度を揃えることができる。

Template Method

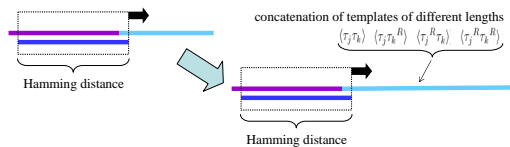
- Designing a DNA sequence set by combining a template and an error-correcting code (Arita and Kobayashi, 2002)
 - Each sequence in the set is guaranteed to have mismatches of a certain length with any sequence or concatenation except its complete complement



Extended Template Method

- Exhaustively search for a optimal template tuple $(\tau_1, \tau_2, \dots, \tau_n)$

by concatenating templates of different lengths



- Generate a DNA sequence set by applying an error-correcting code to each template

DNA マシンを構成するために必要な要素配列は必ずしも長さが一定ではないので、一般に

は複数のテンプレートをを用いる必要がある。実際に、上述した AND ゲートの場合、下左図に示すように、長さ 10 と長さ 15 のテンプレート(後の実験では長さ 10 と長さ 20 のテンプレート)を組み合わせて用いた[4]。テンプレートを探索する際には、それらの組み合わせ方を考慮してミスマッチ数のより多いテンプレートが得られるような工夫をした。

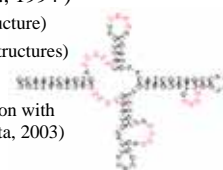
Using Extended Template Method

- Select *good* sequences from the sets and join them to obtain candidate sequences that form the target structures
- Example design of the AND gate
 - Template tuple: (0001001011, 111011101000111)
 - Mismatches guaranteed by the template
 - of length 10 (used as loops): 3
 - of length 15 (used as stems): 4
 - BCH code



Secondary Structure Prediction

- Target structures should have minimal free energy and maximal probability
 - Mismatches do not exclude unwanted secondary structures within a single DNA sequence
 - Need for secondary structure prediction
- Vienna Package (Hofacker, et al., 1994)
 - Fold function (minimal energy structure)
 - Partition function (probability of structures)
- Extension of Vienna Package
 - Multiple sequences by concatenation with virtual bases (Uejima, 2002, Kubota, 2003)
 - Correct treatment of Multi-loops (Takahashi, 2005)



テンプレート法などを用いて配列セットを生成すれば、それに属する要素配列の間で定まった数以上のミスマッチが保証される。しかし、ミスマッチ数を保証するだけでは、要素配列を連結した塩基配列が目的の構造を安定的に形成するとは限らない。これは、一般に配列セットの保証するミスマッチ数が分子内の望まれないハイブリダイゼーションを避けるには十分ではないからである。従って、要素配列を連結して塩基配列を求めた後に、その二次構造を予測することにより、目的構造が自由エネルギーの最も小さい構造であること、さらに、目的構造を取る確率が十分に高いことを確認しなければならない。DNAの二次構造の自由エネルギーの予測には nearest neighbor と呼ばれるエネルギー・モデルが一般に用いられる。このモデルは二次構造の各ループにエネルギー・パラメータを割り当て、二次構造全体の自由エネルギーをループのエネルギーの和として近似するものである。我々は Vienna Package と呼ばれるオープン・ソースのエネルギー予測プログラムを拡張し、特に複数の DNA 分子がハイブリダイズして作る構造の最小エネルギーと分配関数を求めている(上右図)。

配列セットの生成時に構造に関する条件を課すことにより、最初から目的以外の構造を取らないように配列セットを生成する方法も研究されている。このためには、要素配列を連結して得られる塩基配列が二次構造を形成し得るかどうかを調べなければならない。これは構造非形成検証問題と呼ばれている。筆者が代表を務める特定領域「分子プログラミング」において、電通大の小林たちは、シュードノット無しの二次構造に関して、配列セットに属する要素配列を連結して作られる正則言語が(オートマトンとして)与えられたとき、その正則言語に属する塩基配列で二次構造を形成するものが存在するかどうかを、多項式時間で判定するアルゴリズムを与えている[7]。このようなアルゴリズムを走らせながら配列セットの探索を行えば、最初から目的以外の構造を取らないような配列セットを生成することが可能になる(下左図)。

Designing a Sequence Set Excluding Unwanted Secondary Structures

- General setting: $SFTP(\mathcal{C}, \mathcal{T})$ (Condon, 2003)
 - Given a language class \mathcal{C} over Σ and a class \mathcal{T} of secondary structures, answer **yes** if a language $R \in \mathcal{C}$ contains a sequence that forms a structure in \mathcal{T} , and **no** otherwise
- Positive answer (Kobayashi, 2004)
 - Reg : class of regular languages
 - T_{pnf} : pseudo-knot-free secondary structures
 - $SFTP(Reg, T_{pnf})$ can be solved in $O(m^8)$ -time, where m is the number of states in the automaton of R

Kinetic Analysis

- Estimates how difficult or easy is a transition from one conformation (secondary structure) to another
 - Difficulty \rightarrow energy barrier
 - Easiness \rightarrow probabilistic simulation
- Energy barrier
 - Analysis of the energy landscape (Kubota, 2004, Ono, et al., 2005)
- Probabilistic (Monte-Carlo) simulation
 - Random walk over the energy landscape (Sadakane, et al., 2005)

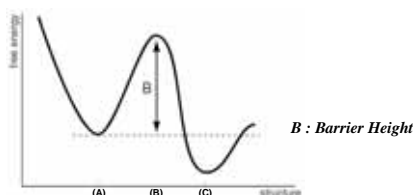
形態変化の予測

以上で述べたような方法を用いることにより、目的の構造を安定的に取る DNA 分子の設計ができたとしても、設計した DNA 分子が望まれる形態変化を効率よく行い、望まれない形態変化を起こさないとは限らない。形態変化の起こり易さ、起こりにくさを予測するためには、いわゆる運動論的な解析が必要となる(上右図)。

一般に、化学反応における生成物と被生成物の最終的な(平衡状態での)割合はそれらの自由エネルギーの差に依存するが、生成物が得られる速度は、生成物と被生成物の間にある中間物のエネルギーなど、被生成物から生成物が得られる反応経路に依存する。このような反応経路の解析が運動論的解析である。DNA 分子の形態変化に対しては、初期の二次構造が目的の二次構造に変化して行く二次構造の遷移過程を解析しなければならない。二次構造から二次構造への遷移の基本単位として、一つの塩基対の生成もしくは消滅を考えることができる。もちろん、DNA 分子の実際の形態変化反応に照らしてこれが適切なモデル化かどうかは議論の余地があるが、DNA の形態変化を解析するための第一近似として広く受け入れられている。すると、下右の概念図にあるように、与えられた(複数の)DNA 分子が作り得る二次構造の全体は、一つの塩基対の生成もしくは消滅によってグラフ構造を成す。各二次構造にはその自由エネルギーが定まっているので、二次構造の作るエネルギー地形が得られる。すると、下左図にあるように、二次構造から別の二次構造に変化する際のエネルギー障壁を定義することができる。

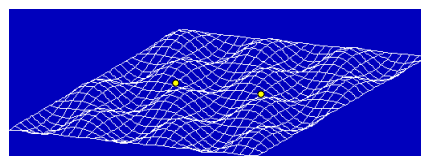
Energy Barrier

- The structure (B) has the highest energy on a transition path from a structure (A) to another structure (C)
- The energy barrier between (A) and (C) is the structure of the lowest energy among such (B)



Energy Landscape

- Graph $G(V, E, f)$ for a sequence
- V : set of secondary structures
- E : adjacency relation (breaking/formation of a single base pair)
- Energy mapping $f: V \rightarrow \mathbf{R}$



二次構造間のエネルギー障壁が高ければ、その間の形態変化は起こりにくい。従って、望まれない形態変化に対しては、エネルギー障壁を高くするような設計を行えばよい。そのためには、エネルギー障壁を効率よく計算する必要がある。我々はグラフ探索のアルゴリズムを用いて、初期構造と目的構造を含む「最小流域」を計算することにより、エネルギー障壁を効率よく計算するアルゴリズムを実装し、実際の DNA マシンの実験結果とエネルギー障壁の関係を調べた[8]。また、上記の特定領域においては、九大の小野たちが局所探索の技術を用いて、近似的ではあるが非常に長い塩基配列に対してもエネルギー障壁を効率よく計算する手法を提案し実装している[9]。現在、前節で挙げた分子マシンの実験結果と計算によって予測したエネルギー障壁の関係を仔細に調べているところである。

反応を起こし易くするためにはエネルギー障壁を低くする必要があるが、それだけでは不十分な場合もある。エネルギー障壁が低くとも反応経路が細ければ反応はなかなか進まない。単純にエネルギー障壁を求めるだけでなく、二次構造の作るエネルギー地形上で確率的なシミュレーションを行えば、より精密に反応の起こり易さを予測できると考えられる。上記の特定領域では、九大の定兼たちが二次構造の作るエネルギー障壁上で確率的なシミュレーションを行い、実際の実験結果との関連を調べている[10]。

参考文献

- 1 . Kensaku Sakamoto, Hidetaka Gouzu, Ken komiya, Daisuke Kiga, Shigeyuki Yokoyama, Takashi Yokomori and Masami Hagiya: Molecular Computation by DNA Hairpin Formation, *Science*, Vol.288, 2000, pp.1223-1226.
- 2 . Ken Komiya, Kensaku Sakamoto, Hidetaka Gouzu, Shigeyuki Yokoyama, Masanori Arita, Akio Nishikawa and Masami Hagiya: Successive State Transitions with I/O Interface by Molecules, *DNA Computing, 6th International Workshop on DNA-Based Computers, DNA 2000*, (Anne Condon and Grzegorz Rozenberg Eds.), Lecture Notes in Computer Science, Vol.2054, 2001, pp.17-26.
- 3 . Atsushi Kameda, Masahito Yamamoto, Hiroki Uejima, Masami Hagiya, Kensaku Sakamoto and Azuma Ohuchi: Hairpin-based state machine and conformational addressing: Design and experiment, *Natural Computing*, Vol.4, No.2, 2005, pp.103-126.
- 4 . Keiichiro Takahashi, Satsuki Yaegashi, Atsushi Kameda and Masami Hagiya: Chain Reaction Systems Based on Loop Dissociation of DNA, *DNA11, Eleventh International Meeting on DNA Based Computers, Preliminary Proceedings*, 2005, pp.343-353.
- 5 . Keiichiro Takahashi, Satsuki Yaegashi, Hiroyuki Asanuma and Masami Hagiya: Photo- and Thermoregulation of DNA Nanomachines, *DNA11, Eleventh International Meeting on DNA Based Computers, Preliminary Proceedings*, 2005, pp.147-156.
- 6 . Masanori Arita and Satoshi Kobayashi: DNA Sequence Design Using Templates, *New Generation Computing*, Vol.20, 2002, pp.263-277.
- 7 . Satoshi Kobayashi: Testing Structure Freeness of Regular Sets of Biomolecular

Sequences (Extended Abstract), *DNA Computing: 10th International Workshop on DNA Computing, DNA10*, Lecture Notes in Computer Science, Vol.3384, 2005, pp.192-201.

8 . Mitsuhiro Kubota and Masami Hagiya: Minimum Basin Algorithm: An Effective Analysis Technique for DNA Energy Landscapes, *DNA Computing: 10th International Workshop on DNA Computing, DNA10*, Lecture Notes in Computer Science, Vol.3384, 2005, pp.202-214.

9 . Tsutomu Takeda, Hirotaka Ono, Kunihiko Sadakane, and Masafumi Yamashita: A Local Search Based Barrier Height Estimation Algorithm for DNA Molecular Transitions, *DNA11, Eleventh International Meeting on DNA Based Computers, Preliminary Proceedings*, 2005, pp.354-365.

10 . Masashi Shiozaki, Hirotaka Ono, Kunihiko Sadakane and Masafumi Yamashita: Modeling DNA Conformation Change and Theoretical Analysis on the Reaction Rate, *DNA11, Eleventh International Meeting on DNA Based Computers, Preliminary Proceedings*, 2005, p.408.