

超ロバスト分子計算

萩谷 昌己

情報理工学系研究科コンピュータ科学専攻

概要

DNA 等の生体分子を用いて情報処理機能を有する分子システムを構築することを目標として、個々の分子反応をロバストに制御するとともに、分子システム全体としてロバストな情報処理機構を実現するためのモデルを提案し、併せてモデルの分子による実装方法について研究を進める。特に、個々の分子を離散的側面と連続的側面の融合したハイブリッド・システムと捉え、さらにそれらが集合したシステムのロバスト性を保障する原理を確立することを最終目標とする。

はじめに

本研究は、DNA 等の生体分子を用いて情報処理機能を有する分子システムを構築することを目標としている。そのためには、分子内および分子間の個々の反応をロバストに制御するとともに、分子システム全体としてロバストな情報処理機構を実現することが必要である。

分子反応をロバストに制御するためには、分子反応の適切なモデル化とそれに基づく分子の設計方法が必要である。特に、DNA 等の生体分子は、塩基配列や二次構造のような離散的な側面と、温度やエネルギーのような連続的な側面を併せ持っているため、ハイブリッド・システムとしてモデル化することが適切であると考えられる。

しかしながら、分子反応は本質的に確率的に進むため、個々の分子は一定の誤差を含む不安定な構成要素と考えられる。従って、ロバストな分子システムを実現するには、多数の不安定な構成要素を不均質に集合させて信頼性の高い情報処理を実現するためのモデルが必要になる。これがアモルファス・コンピューティングである。本プロジェクトの目標は、アモルファス・コンピューティングを分子によって実現することであるということができる。

以上の目標を達成するために、本研究では常に、ロバスト性のモデルによる解析と分子による実現の両面から研究を行っている。

本研究は、代表者が行っている以下のプロジェクトと深く関連している。これらのプロジェクトはそれぞれの目標を持っているが、本研究の目標とオーバーラップするところが多いので、これらのプロジェクトと協調して本研究を進めている。

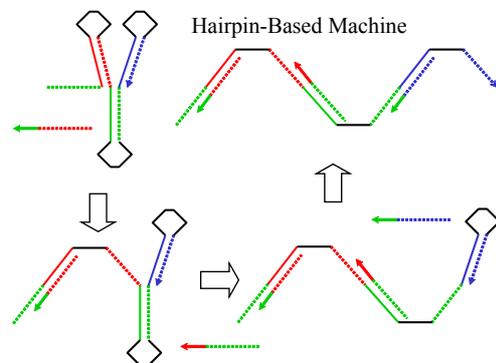
- 特定領域研究：分子プログラミング
- 特定領域研究：形態変化する分子を用いた並行計算と分散計算
- 萌芽研究：ハイブリッド・セル・オートマトンを用いた生物系と化学系の解析と検証
- JST CREST：多相的分子インタラクションによる大容量メモリの構築
- 特定領域研究(C)：抽象モデル検査のためのグラフ探索アルゴリズムの形式化と検証

本年度の成果の概要

状態が分岐する分子機械の設計と実装

本年度は、分子による実装に関しては、昨年度に引き続き DNA のヘアピン構造の開裂を利用した状態機械の設計と実現を行った。

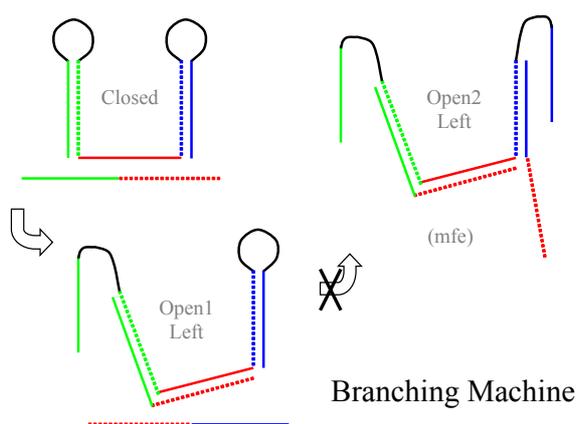
昨年度より、DNA の連続したヘアピン構造が逐次的に構造変化を起こすことによって多状態を実現する分子機械の二次構造設計に関する研究を進めている。以下は、JST の CREST のプロジェクトにおいて分子メモリとして開発された分子機械である。



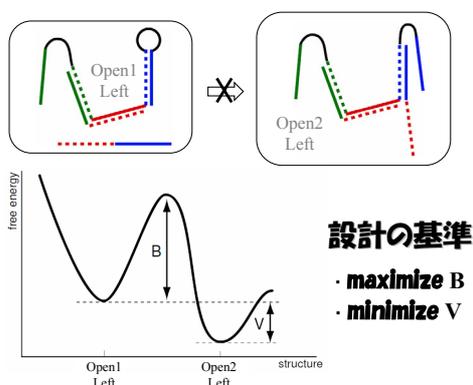
この分子機械は複数のヘアピンが連続的に結合した構造をしており、オープナーと呼ばれる DNA 分子がヘアピンのステム部分にハイブリダイズし、ヘアピン

構造がブランチ・マイグレーションによって開裂する。

以下は、上の分子機械と同様に、オープナーによって左右の二つのヘアピンのうちの一方が開く分子機械である。オープナーには左右のヘアピンに対応した二種類がある。一方のオープナーが左か右のヘアピンを開くと、もう一方のオープナーを与えてもハイブリダイズできない。すなわち、初期状態からオープナーによって二つの状態のどちらかに分岐する機械となっている。



実は、両方のオープナーを入れた場合、両方のヘアピンが開いた形態が最もエネルギーが低い状態 (mfe すなわち minimum free energy の状態) になる。しかし、この状態に到達するにはエネルギーの障壁を乗り越えなければならないので、通常は片方のヘアピンが開いた状態に留まるわけである。従って、この分子機械をロバストに動作させるためには、エネルギーの障壁がなるべく高くなるような分子設計 (配列設計) を行わなければならない。



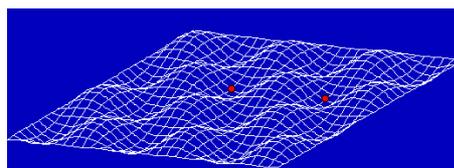
我々は、分子の形態のエネルギー地形、特に、ある形態から別の形態に変化する際の形態変化経路を解析することにより、エネルギー障壁を求めたり、そのエネルギー障壁が高くなるような配列を探索したり

する技術の開発を行っている。

地形グラフ (energy landscape)

- 地形グラフ $G(V, E, f)$.
- 頂点の集合 $V \cdots$ 構造を表す。
- 接続関係の集合 E (塩基対の形成/解離)。
- 各頂点のコスト関数 $f: V \rightarrow \mathbb{R} \cdots$ エネルギー。

→ 各頂点に実数値の重みを持つグラフ



2つの二次構造間にあるエネルギー障壁の高さは構造変化経路に依存する値だが、それは一意には決まらない。そこで本研究では、下記のような3種類に特徴的な構造変化経路を定義し、その定義に基づいた経路予測のアルゴリズムを解析し、実装した。

- 局所最適最短経路：最短経路のうち局所的なエネルギー増加が最小に保たれた経路。
- 大域最適最短経路：最短経路のうちエネルギーのピークが最も低い経路。
- 大域最適経路：あらゆる経路のうちエネルギーのピークが最も低い経路。

昨年度は、大域最適最短経路をヒューリスティクスによって近似的に求める効率のよいアルゴリズムの開発を行った。今年度は、以下のように大域最適経路を求めるアルゴリズムを考案し実装した。

Minimum Basin Algorithm (最小流域アルゴリズム)

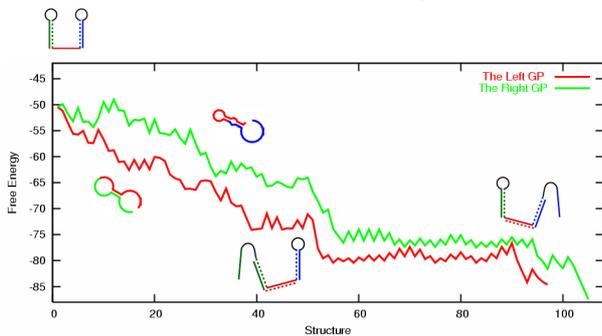
- 2点を含む最小の流域を求めるグラフ探索法。
- 始点から終点までのエネルギー地形 (集合 B) を、低い点を選択しながら探索・生成。(最小の流域)
- Prim-Dijkstra法の一様。

1. 始点・終点は与えられている。
2. 集合 B 、集合 N は空集合。(Basin, Neighbor)
3. 始点を集合 B に加える。
4. 集合 B に隣接する点を集合 N に加える。
5. 集合 N の最も低いエネルギーの要素を集合 B に加える。(複数ある時は全て B に加える)
6. Step4に戻る。Step5で、加える点が終点であったら終了。

→ 出力は集合 B 。その最大の要素が障壁になる。
→ 集合 B の解析が形態変化経路予測として有効。

以上のアルゴリズムを用いることにより、実際の DNA 配列に対する実験結果をある程度説明することができた。

最小流域法による経路



→ 陣盤だけでなく経路についての議論ができる。

Mitsuhiro Kubota, Kazumasa Ohtake, Ken Komiya, Kensaku Sakamoto and Masami Hagiya: Branching DNA Machines Based on Transitions of Hairpin Structures, *Proceedings of the 2003 Congress on Evolutionary Computation (CEC'03)*, 2003, pp.2542-2548.

Hiroki Uejima and Masami Hagiya: Secondary Structure Design of Multi-state DNA Machines Based on Sequential Structure Transitions, *DNA9, Ninth International Meeting on DNA Based Computers, Preliminary Proceedings*, 2003, pp.80-91.

Hiroki Uejima and Masami Hagiya: Analyzing Secondary Structure Transition Paths of DNA/RNA Molecules, 同上, pp.92-96.

Atsushi Kameda, Masahito Yamamoto, Hiroki uejima, Masami Hagiya, Kensaku Sakamoto and Azuma Ohuchi, Conformational Addressing using the hairpin structure of single-strand DNA, 同上, pp.197-201.

DNA 計算ロボットの最適スケジューリング

前節で紹介した研究は、DNA 分子の特定の反応に着目し、それをロボストに実現するためのモデル化と配列の設計方法を目指している。

本節で紹介する研究は、同時に並列に進む複数の反応をいかに効率よくスケジュールするかに関するものである。現在のところ、全体の反応時間を短くすることに主眼が置かれているが、将来的には反応のエラー率等も考慮に入れ、ロボスト性も含めて最適化する方法へと展開したいと考えている。

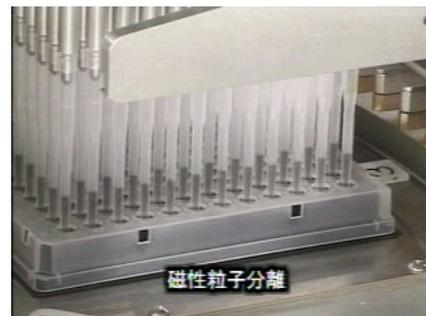
ANP-96 (Algorithmic NA Processor)は DNA 計算の実験を自動的に行うロボットであり、与えられたプログラムに従って複数の実験操作を同時に実行することが可能である。DNA 計算の実験操作は種々の制約のもとで実行され時間もか

かるので、複数の実験操作を自動的かつ効率的に実行することが重要である。

このロボットは、「入力」である複数の与えられた溶液から始まり、以下のような操作を繰り返し適用することで、新たな溶液(これも複数)を作り出す。この最終的な「出力」である溶液を調べることになり、実験が終了する。

1. 特定の溶液を混合して、新しい溶液を作る (Mixing, 懸濁).
2. 溶液から磁気ビーズという仕掛けにより、ある物質(溶液と考えてよい)を取り出すこと (Magtration, 磁性粒子分離).
3. 溶液を特定の温度にまで暖めること (Warming, 温度制御).
4. 溶液を特定の温度で一定時間維持すること (Wait, 維持).

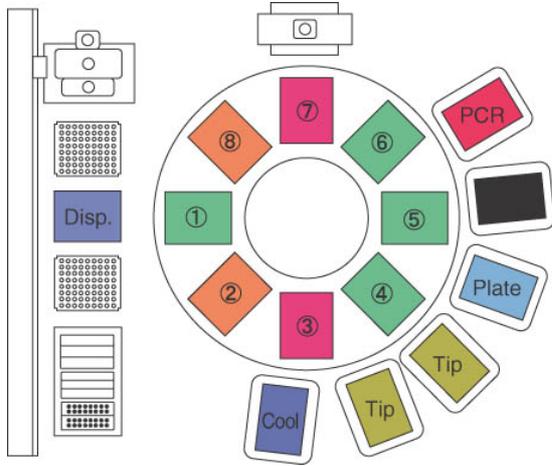
溶液は通常のプログラミング言語における「値」と考えることが出来る。この「値」を保持する「変数」に相当するものとして、プレートとチップがある。プレートは溶液を入れる器である(下図の穴のあいた板)。またチップとはガラス製の先(下)細りの管(下図の複数の管)であり、溶液を保持することが出来る。実際にはチップは複数の管からなり、チップラックというチップを納める複数の穴をもつ容器に置かれており、チップに対する操作はこのチップの 1 セットに対して同時に行われる。プレートも同様に、複数の管に対応する複数の穴からなり、この穴の各々に溶液が入る。



プレートとチップの個数には事実上制限はないが、ロボットは上記の操作を有限(8 個)の実験場所(テーブル)で行うようになっている。また上記の(3)が可能なテーブルは二つしかなく、さらに(1)や(2)を行うための IMU という装置は一つしかない。以下がロボット本体の主要モジュールである。

- IMU (Integrated Magtration Unit)
- ストッカ

- ディスペンサ
- ターンテーブル
- ロボットアーム



このロボットは、与えられたプログラムに従って複数の実験操作を並列に実行することができる。



しかし、現在の ANP-96 のプログラミング環境は、実験に本質的ではない低レベルな記述を必要とするため多くの手間を必要とし、誤ったプログラムにはロボットを破壊する危険もある。特に、プログラマは有限のテーブルを無駄なく割り当てるとともに、温度制御等の時間のかかる命令を効率よくスケジューリングしなければならない。

我々は以上のようなプログラミングを自動化するために、テーブルのアロケーションと命令のスケジューリングを含む問題を、ANP-96 に依存しない一般的な抽象度で記述する枠組みを開発するとともに、その枠組みに基づく制御コード生成系を実装した。

特に、この枠組みでは、各操作におけるリソースの受け渡しに関する記述を簡潔に表現することができる。また、上述の問題に対する最適解を与える手法として、コンパイラ分野で提案された整数線形計画法による手法を利用した。この手法は処理時間が長く、したがって小規模のプロ

グラムにしか適用できないのが欠点であるが、ANP-96 においては、一般にプログラムは比較的小規模であり、その実行回数とコストを考慮すると、時間をかけて最適化する価値があると考えられる。

阿部正佳, 萩谷昌己: 整数線形計画法を用いた DNA コンピュータ制御コードの生成, 情報処理学会プログラミング研究会, 2003年1月.

セル・オートマトンの解析

本稿のはじめに述べたように、我々は最終的には分子の集合体をモデル化する方法を確立したいと考えている。そのために昨年度より、分子の集合体をグラフと捉え、グラフ書き換え系の解析を行う方法について考察している。本年度はその簡単な場合として、グラフの構造が変化せず、各々のノードの状態だけが変化するシステムを、時相論理を用いて解析する方法を定式化した。このようなシステムには、いわゆるセル・オートマトンも含まれる。

我々は従来研究において、時相論理を用いて、互いにポインタで繋がれたセルから成るリンク構造を抽象化する方法を提案した。セルはあらかじめ指定された時相論理式の真偽の組み合わせを表す抽象セルによって抽象化され、リンク構造全体は抽象セルの集合によって抽象化される。

本研究では、この方法を発展させ、時相論理を用いてセル・オートマトンを解析する方法を定式化した。従来研究の場合と違って、すべてのセルが同期して状態を遷移させるので、各抽象セルの次状態を計算することにより、抽象化されたリンク構造の次状態が求まる。この新しい状態を再び抽象化するために、抽象セルの分裂と融合の操作を導入した。この方法により、実際に一次元(および二次元)のセル・オートマトンの簡単なものを解析することができた。

また、本研究では、本方法のベースとなっている時相論理である 2 方向 CTL(computation tree logic)とその充足可能性判定方法の定式化も行った。

萩谷昌己, 高橋孝一, 山本光晴, 佐藤貴洋: 時相論理による抽象化を用いたセル・オートマトンの解析, 日本ソフトウェア科学会第20回大会, 2003.

Masami Hagiya, et al: Analysis of Synchronous and Asynchronous Cellular Automata using Abstraction by Temporal Logic, *FLOPS2004: The Seventh Functional and Logic Programming Symposium*, invited talk, 2004, to appear.