

構造的分子計算理論

-- 自律的計算系の解析と設計のための基礎理論

横森 貴, 上田和紀, 楠元範明(早稲田大学)

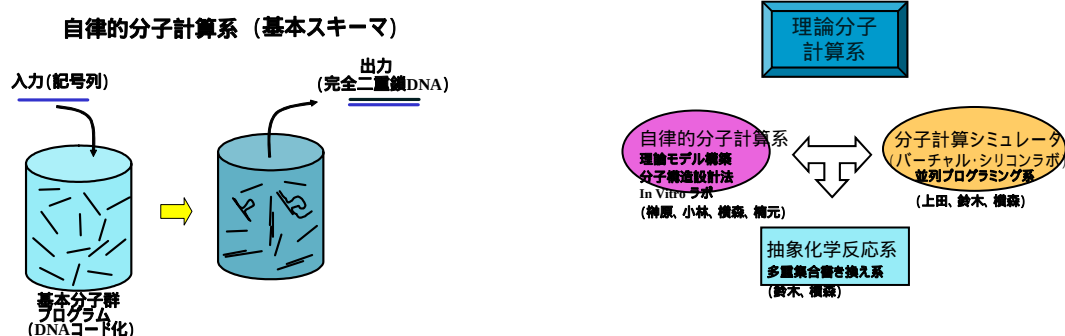
榊原康文(慶応義塾大学)

小林 聡(電気通信大学)

鈴木泰博(名古屋大学)

1. 研究の背景・目的

【研究の背景】我々はこれまでの研究において、核酸塩基対の相補性による自律的会合、分子構造の形態変化などの様々な生体分子反応を解析し、それらの知見を基礎に情報処理のための新たな基本的計算機構の開発と**自律的分子計算モデル**の提案を行ってきた。提案した多くの計算モデルは従来の計算機と同等の計算能力(チューリング計算可能性)を有することが証明され、実際に**分子生物学的な基礎実験**によりその実装可能性の一部が検証されたものもある。また、細胞内の生体分子反応をモデル化した**抽象化学反応系**にコンパートメントを導入することにより細胞計算モデルを構築しその計算能力を理論的に解析してきた。一方、上述の新規な計算モデルの実装可能性を検証するためにも計算機シミュレーションシステムを開発し、人工的な仮想分子などを用いることによりPCRなどDNAによる様々な生体分子反応や**分子計算モデルの挙動をシミュレーション**することが必要となっている。この意味において、分子反応系のための確率的あるいは近似的計算モデルとその解析手法に関する研究の必要性が強く示唆される。(各テーマ毎の主な組織分担については下図を参照されたい。)



【研究の目的】これらの研究成果を背景として、本研究の目的は『生体構造分子と生体化学反応系に関わる情報処理能力を計算理論的、アルゴリズム論的な視点から解析し設計することである。すなわち、生体分子の構造とその機能との関係を計算能力(あるいは情報処理能力)という視点から解析することにより、基本計算ユニットの集合(基本構造分子群)と基礎となる計算原理(生体反応系)を明らかにして、それらの組み合わせを基本とする体系的な分子計算の理論モデルを構築することである(上図を参照)。さらに、ある問題を解く(ある機能を実現する)ために基本構造分子群を構成する各分子計算ユニット上にその問題をどのように符号化し制御すればよいかという分子計算におけるプログラミング技法を研究する。具体的には、本研究では以下の3つの研究項目、すなわち「自律的分子計算系の基礎的考察」、「抽象化学反応計算モデルの研究」、および「並列計算系による分子計算シミュレータの設計と試作」を設定し、研究を実施した。

2. 研究成果の概要

本年度の研究成果の概要を各研究項目について以下にまとめる。

2.1 自律的分子計算系の基礎的考察

(1). **DNA相補性に基づく計算モデルの統一的特徴付け**: 分子計算の理論において核酸の相補性を利用する計算モデルの提案は多く報告されているが、それらの間の関係についてはほとんど議論されていなかった。そこで1次元の記号列の集合と二重鎖の集合との間を関係付ける2つの写像 linearizer と doubler を導入することにより、これまで独立に提案され研究されてきた Watson-Crick オートマトンモデル、Sticker 計算モデルなどの種々の計算モデルに対する統一的特徴付け定理を得ることができた。また、これらの写像を通して、帰納的可算集合に対する新たな表現定理など幾つかの結果を得ることにより、二重鎖による相補性のもつ計算論的な意味付けと計算可能性に関する新しい知見を導くことができた。

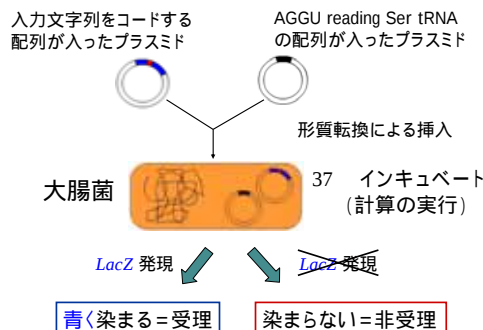
(2). **分子配列設計法の考案**: 前年度に得られた任意の正則な配列集合に対する二次構造予測アルゴリズムの時間計算量は、集合を表現するグラフの頂点数 n に対して $O(n^8)$ で動作する。これでは実用的でないので、対象とする配列集合を有限集合に制限することにより、 $O(n^5)$ の計算時間で動作する二次構造予測アルゴリズムを考案し実装した。Andronesco らは長さが等しい配列集合を有限回連接してできる配列有限集合に対して、効率の良いアルゴリズムを提案しているが、本研究で得られたアルゴリズムは長さが異なる配列集合の連接にも対応できる実用的なアルゴリズムである。また、このアルゴリズムはグラフの形状を変更しない場合は $O(n^5)$ 時間の事前計算の後、 $O(n^4)$ の計算時間で配列評価を行える。従って、配列設計においては配列集合を探索する際には $O(n^4)$ の計算時間で評価を行いながら探索できるという長所を持つ。

(3). **細菌による計算モデルの設計とオートマトンの実装**: 有限オートマトンを大腸菌の細胞内分子メカニズムを用いて実現する研究を行った。はじめに、大腸菌内のタンパク質合成メカニズムを用いて有限オートマトンを計算する方式と実験プロトコルを設計した。この方式は、有限オートマトンという計算モデルを、タンパク質合成メカニズムと拡張コドンを用いて実現し、細胞の中でその計算を実行する。

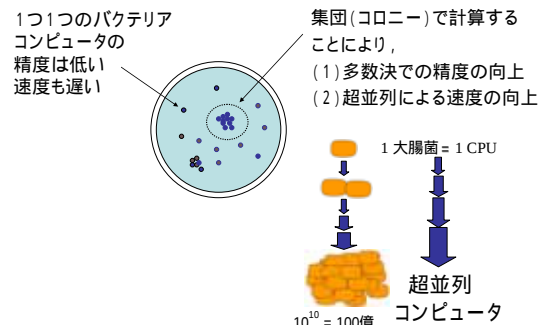
具体的には、入力文字を AGGU という RNA 部分配列にコード化し、有限オートマトンの状態も RNA 部分配列にコード化する。これらのコード化された配列の最後に、入力文字列を受理するか否かを検出するための lacZ レポーター遺伝子を加える。次に、転移 RNA がもつアンチコドンを用いて、有限オートマトンの状態遷移関数をコード化する。このように入力文字列をコード化した RNA 配列がメッセンジャー RNA として発現されるようにプラスミド上に挿入する。また拡張アンチコドンをもつ転移 RNA が転写されるように別のプラスミドを合成して、これらの2つのプラスミドを形質転換により大腸菌内に挿入する。大腸菌をインキュベートすることにより、大腸菌内でタンパク質合成反応が実行される。このとき、入力文字列が正しい場合には、プラスミド上にコード化された入力配列にしたがってタンパク質合成が最後まで実行され、lacZ が発現することによりコロニーが青く染まり、有限オートマトンがその文字列を受理したことが確認される。

実験に用いた有限オートマトンは、2 状態のオートマトンで、入力された文字列中の 1 の個数が偶数倍のときに受理する(すなわち、パリティビットを検出できる)有限オートマトンである。実験の結果、1 の個数が 2, 4, 6 の偶数の場合にはコロニーが青く染まったことが検出され、1, 3, 5 の奇数の場合にはコロニーが染まらなかったことを確認した。このことから、有限オートマトンの計算が正しく実行されたことを確認すると同時に、大腸菌を用いた DNA コンピュータの設計原理が正しかったことを確認できた。この実験結果は、世界で初めてバクテリアを用いて計算が実行できたことを証明するものであり、まだ初等計算機ではあるがバクテリアコンピュータが初めて実現されたことも示すものである。

大腸菌を用いた *in vivo* コンピュータの実装



バクテリアコンピュータと超並列



2.2 抽象化学反応計算モデルの研究

平成17年度は以下のふたつテーマについて主に研究を行なった。すなわち、

ARMS(抽象化学系)を用いて“少数粒子多自由度系”の観点から確率的な立場からの検討

P53(アポトーシス系)のモデル化と数理的特徴付け等、

である。については、2者Lotoka-Vortela(LV)模型において大数粒子系では安定振動を行うのに対し、少数粒子系にすると大数の場合と異なった振る舞いをみせることを確認し、かかる事象の原理を解明するために、状態空間の数え上げを行いその各々について状態遷移確率(反応規則の適用確率)がどのように変化するかを調査した(ARMSの確率場)。その結果、粒子数が少数になるにつれ状態遷移確率に揺らぎが生じ、その結果として確認した現象が生ずることを明らかにした。また、かかる現象につき厳密な解析を試みたが少数粒子系になると通常用いられているFoker-Plank方程式(FPE)による揺らの評価などが使えないために厳密な解析は次年度への課題とすることにした。また、共存解が存在しないことが知られている3者LV模型に情報物質による相互作用を導入し、共存解の存在を示唆した鈴木-高林模型(ST模型)につき、従来のGillespie法に類する方法に加え、生成-消滅過程を以って再定式化し再計算したところやはり共存解の存在が示唆された。また、従来の結果と同様にLV模型と異なり大数でも少数でもその性質に大きな差異がないことが示された。ST模型に関するFPEを用いた解析はその方法が知られていないこともあり解決されず、こちらも次年度への課題としたい。これに加えてについては、P53(アポトーシス系)のARMSを用いたモデル化と数理的特徴づけ、膜計算の化学的実現方法などについての研究を行った。

少数粒子多自由度系としての抽象化学系

ここでの、抽象化学系、Abstract Rewriting System on Multisets (ARMS)とはモンテカルロ法を用いた化学反応の離散モデル(Gillespie法など)である。

○ ARMS (抽象化学系) の確率的動学に関する研究を行った

背景(これまでの成果)

粒子数が多い場合と少ない場合での反応のダイナミクスの変化を考察している。今までにLotoka-Volterra, (LV) (2体) 模型において、初期状態の大小によりダイナミクスが変化することを確認した。
また、一般に3体LV模型は共存解が存在しないことが知られているが、3体LV模型に情報物質による種間相互作用を導入した鈴木-高林(ST)模型は共存解の存在が示唆される。このST模型については粒子数の大小とダイナミクスに特に相関は見られなかった。

本年度は主に確率の立場からの解析を試みた

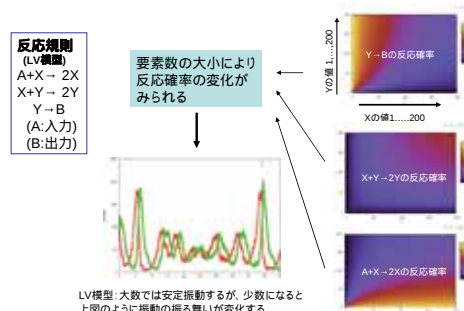
確率場からの考察 状態空間の数え上げを行い、各々の状態について状態遷移確率が粒子数に応じてどのように変化するか調査した。

マスター方程式からの考察を行った

だが、3体のFPEの展開については知られていないため(ST模型は3体系)、線形化近似(均衡点付近での摂動を加えての振る舞いの検討)などはできるが、直接は解けていない(未解決であり、今後の課題の1つ)。

ARMS(抽象化学系)の確率場

$(X, Y) = (x, y)$, $(0 < x, y < 200)$ の全の状態における反応規則の適用確率を求めた



4.3 並列計算系による分子計算シミュレータの設計と試作

階層グラフ書換えモデル LMNtal は、多重集合や並行処理などの概念を持つ計算モデルの統合を目標として開発を進めてきた言語モデルである。従来型の計算モデルにおいても分子計算モデルにおいても、情報や計算の表現においては階層関係と接続関係が重要な役割を果たす。LMNtal が提供する階層グラフはアトム、リンク、膜から構成され、膜によって階層構造を、リンクによって接続構造をダイレクトに表現できるのが特徴である。LMNtal における計算は階層グラフの書換えであり、階層構造と接続構造の動的再構成を柔軟に表現するための機構を提供している。

我々は、単なる計算モデルにとどまらず、実用プログラミングにも供することのできる LMNtal 言語の設計と実装に取り組み、多様な記述例を蓄積することによってその表現力や有用性の検証につとめてきた。計算モデルの理論研究においては計算モデル間のエンコーディングが重要課題である。分子計算における重要概念である多重集合はグラフの特殊な場合であるので、多重集合書換えモデルが LMNtal に自然にエンコードできることは以前から自明であったが、本年度の研究ではそれに加えて、以下のような代表的かつ広範な計算モデルを LMNtal でエンコードして、実際に処理系上で動作させることに成功した。

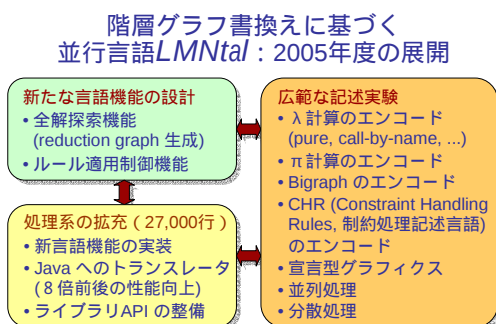
- () 非決定的 λ 計算および名前呼び λ 計算
- () 非同期 π 計算
- () bigraphical reactive systems
- () CHR (Constraint Handling Rules)

またこれらのエンコーディングを実現するために、LMNtal の書換えルール適用制御機能の検討を行い、下記の二つの機能を今年度新たに設計・実装した。

- (a) 特定のデータの組に対して特定のルールの適用を 1 回に限定する機能
(uniq ガード制約)

- (b) 与えられた初期プロセスとルール集合を用いて可能なすべての簡約経路を探索し、結果を簡約グラフの形で求める機能 (非決定的 LMNtal)

(a) は CHR のエンコーディングに本質的な役割を果たし、また (b) は LMNtal の検証への応用に道を開くものである。



<http://www.ueda.info.waseda.ac.jp/lmntal/>

Example: Fullerene (C_{60})

```
/* icosahedron */
dome(L0,L1,L2,L3,L4,L5,L6,L7,L8,L9) :-
  p(T0,T1,T2,T3,T4), p(L0,L1,H0,T0,H4),
  p(L2,L3,H1,T1,H0), p(L4,L5,H2,T2,H1),
  p(L6,L7,H3,T3,H2), p(L8,L9,H4,T4,H3).

dome(E0,E1,E2,E3,E4,E5,E6,E7,E8,E9),
dome(E0,E9,E8,E7,E6,E5,E4,E3,E2,E1).

/* icosahedron -> fullerene */
p(L0,L1,L2,L3,L4) :-
  X=c(L0,c(L1,c(L2,c(L3,c(L4,X))))).
```



【17年度 成果論文リスト】

Kaoru Onodera and Takashi Yokomori: Linearizer and Doubler : Two mappings to unify melocular computing models based on DNA complementarity, *Proceedings of 11th International Meeting on DNA Based Computers*, London, Ontario, pp.139-150, 2005.

Atsushi Kijima and Satoshi Kobayashi : Efficient algorithm for testing structure freeness of finite set of biomolecular sequences, *Proceedings of 11th International Meeting on DNA Based Computers*, London, Ontario, pp.278-288, 2005.

H.Nakagawa, K.Sakamoto and Y.Sakakibara : Development of an in vivo computer based on Escherichia coli, *Proceedings of 11th International Meeting on DNA Based Computers*, London, Ontario, pp.68-77, 2005.

J.Kuramochi and Y.Sakakibara : Intensive in vitro experiments of implementing and executing finite automata in test tube, *Proceedings of 11th International Meeting on DNA Based Computers*, London, Ontario, pp.59-67, 2005.

Y. Suzuki and H. Tanaka : modeling p53 Signaling Pathways by Using Multiset Processing, *Applications of Membrane Computing Series: Natural Computing Series* (Ciobanu, Gabriel; Perez-Jimenez, Mario J.; Paun, Gheorghe Eds.), 2006.

石渡龍輔, 鈴木泰博, 田中博 : 化学反応系のモデル化手法について, 情報研報, 2005-MPS-54-1, 2005年

石渡龍輔, 森岡勝樹, 荻島 創一, 鈴木泰博, 田中博 : アポトーシス誘導因子Cytchrome c放出の数学モデリング, ポスター発表-2P-0912, 日本分子生物学会第28回年会, 2005年.

K. Ueda and N. Kato : LMNtal: a language model with links and membranes. In *Proc. Fifth Int. Workshop on Membrane Computing (WMC 2004)*, LNCS 3365, Springer, pp.110—125, 2005,.

K. Ueda : Constraint-Based Concurrency and Beyond. In *Proc. Workshop on Algebraic Process Calculi, The First Twenty Years(PA'05)*, Aceto, L. and Gordon, A.D. (eds.), BRICS Notes Series, NS-05-3, pp.227-230, 2005.

工藤晋太郎, 加藤紀夫, 上田和紀: LMNtal 処理系におけるグラフ構造の操作機能の設計と実装. 情報科学技術レターズ, pp.9--12, 2005.

櫻井健, 加藤紀夫, 水野謙, 上田和紀 : LMNtal コンパイラにおける並び替えとグループ化を用いた命令列の最適化, 第4回情報科学技術フォーラム (FIT2005), 第一分冊 pp.33--36, 2005.

水野謙, 加藤紀夫, 原耕司, 上田和紀: 階層グラフ書換え言語 LMNtal 処理系における非同期実行の実現. 日本ソフトウェア科学会第22回大会講演論文集, 3A-4, 2005.

乾敦行, 原耕司, 水野謙, 上田和紀: 階層グラフ書き換え言語 LMNtal 処理系とその応用例. 第8回プログラミングおよびプログラミング言語ワークショップ(PPL2006), 2006年3月(発表予定).