

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）

榎原 康文（慶應義塾大学）

小林 聰（電気通信大学）

鈴木泰博（東京医科歯科大学）

1 研究の背景・目的

【本研究の背景】 Adleman による有向ハミルトン経路問題の DNA を用いた解法が発表されて以来，欧米をはじめ我が国においても非ノイマン型計算モデルの提案や分子計算をはじめとする“ナチュラルコンピュテーション”とよばれる新計算パラダイムの研究が活発に行われている。我々はこれまで，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ヒントとして，情報処理のための新たな基本的計算機構の開発と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，これまで提案された分子生物学実験の技法によるハミルトン経路問題や充足可能性問題に対する解法はハイブリダイゼーションという分子反応や PCR 増幅等の実験操作を含んでいるが，それらの実行過程の信頼性はまだ低く反応特性の詳細な解析や基本操作の問題点については不明な点が少なくない。したがって，必要な情報を分子上へ正しく符号するための分子設計技法が重要になる。また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，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ト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。

【本研究の目的】 DNA, RNA の構造分子やタンパク質のような高次構造を持つ生体高分子は分子認識（分子会合），自己組織化，形態変化などの生体反応・現象によって目的をもった分子系を形成し機能を発現する。そこで，以下のような“自律的計算スキーマ”を考える。まず充分な量の構造分子群からなる溶液（これをプログラムとする）を仮定する。これに対して外部からある情報（これが入力である）が与えられると，その入力に対する計算・演算の結果としてある特定の構造化分子を生じるとする。意図した構造化分子の形成がここで計算の出力結果である。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。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

分子群に属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。また、上述の新規な計算モデルの実装可能性を検証するため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トする。

2 研究成果概要

本年度の研究成果は 4 つの分野に大別される。その概要を以下で述べる。

1. 無細胞タンパク質合成系による計算モデルの研究

無細胞タンパク質合成系を用いたまったく新しい分子コンピュータの方式を提案した ([10])。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系と拡張コドンを用いて実現し、試験管の中でその計算を実行する「新しいタイプの DNA コンピュータ」を提案したものである。タンパク質合成系を用いることにより、従来の DNA コンピュータにおいて最大の問題となっていた、DNA 分子の水素結合を用いたハイブリダイゼーションという演算による精度の低さとエラーの多さを解決することができた。具体的な手順として、まず入力文字のある 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した。これらのコード化された配列の最後に、受理されたか否かを検出するために GFP レポート遺伝子を加えた。次に、転移 RNA がもつアントチコドンを用いて、有限オートマトンの状態遷移関数をコード化した。このように、入力文字列をコード化した RNA 配列をメッセンジャー RNA として、また拡張アントチコドンをもつ転移 RNA を無細胞タンパク質合成系に加えて、タンパク質合成反応を実行した。このとき、入力文字列が正しい場合には、タンパク質合成が最後まで実行され、GFP タンパク質が合成されることにより発光が検出された。このようにして有限オートマトンはその文字列を受理したことが確認できた。

2. 分子配列設計法の考案

DNA 計算で用いる配列集合の設計に役立つアルゴリズムとシステムの開発を行った。まず、配列集合 S が与えられたとき、 S^+ に属するすべての配列が線形な二次構造を取らないための十分条件を求めた。この十分条件は、効率良く評価できるので実際の配列集合の評価に利用できる。また、二次構造を取らない配列の抽出問題を、この十分条件を利用することにより、超グラフ上の最大クリーク抽出問題に還元することに成功した。そして、富田らの彩色数を利用した分枝限定法による最大クリーク探索アルゴリズムを超グラフのクリーク探索に拡張して応用することにより、配列抽出システムを実装し、その有効性を確認した ([14], [15])。

また、その後の研究の進展により、 S^+ が線形な二次構造を取らないための効率良く評価できる必要十分条件を求める成功 ([17])。提案した手法は、与えられた長さ n の配列の有限集合 (m 本の配列) に対し、 S^+ が線形な二次構造を取り得るかどうかを $O(m^6n^6)$ 時間で判定する。このアルゴリズムは、二次構造の非形成判定問題を、グラフの最短経路問題に還元して解く。本年度は、このアルゴリズムを実装して、配列集合を評価するシステムを構築した（補足：この手法によると、配列評価で用いるグラフにおける負の重みの閉路を取り除

くことにより、配列設計問題へ接近できるので、来年度以降に、このアイデアに基づいた配列設計システムを構築する予定である。)

3. 細胞内シグナル伝達系・化学反応系モデルの設計と実装

分子コンピューティングの基礎的な研究として、化学反応の抽象モデル Abstract Rewriting System on Multisets, ARMS の研究を応用と理論の両面から行った。応用研究としては癌関連タンパクとしてよく知られている P53 に関する細胞内シグナル伝達系のモデル化を中心に行った。このモデルは細胞質と核の 2 つの部分からなり、その各々に分子生物学的な知見を基にして化学反応を定義し、DNA の損傷を活性化された P53 の 4 量体が修復する過程をモデル化した。計算機シミュレーションによりこのモデルが分子生物学的な知見と一致するよう振舞うことを確認した。また、生物を含めた自己発展系にみられるランダムな構造から相互作用の履歴により構造が出現・発展する系のモデルとして自己強化反応系を構築しその振舞を計算機実験を通して調査した。この系では最初はランダムに相互作用が生じるが、相互作用の履歴を持つ分子とはより反応が生じやすくする”自己強化反応 (self-reinforcement reaction)”を仮定すると、分子間相互作用内に構造が出現し、またカタストロフィックな構造崩壊を繰り返す動的な現象が生じることを確認した。理論的な側面からは化学反応の計算論的な特徴付けを目指し、ミラノ大学と共同研究を行った。ARMS と P Systems をある確率的 P Systems を構築することで融合させ、非線形化学反応のモデルであるブルセレータのモデル化とシミュレーションを行った。

4. 分子計算シミュレータ LMNtal の設計と実装

階層的グラフ書換えに基づく並行言語モデル LMNtal について下記の研究開発を行った。

(a) 言語仕様の詳細検討：2002 年度に設計した基本言語仕様について、セル（膜で囲まれた分子）の複製および消去を実現するための拡張を行った。この拡張を実現するために、プロセス文脈（膜内の不特定多数のアトムとマッチする言語機能）に関する仕様を詳細に再検討し、新たな言語規則の形でまとめた。また上記の拡張に伴い、言語の操作的意味論（構造合同規則および遷移規則）も詳細に再検討し再定義した（これらの結果は「コンピュータソフトウェア」誌に採録され、2004 年前半に掲載予定である。）

(b) 処理系の開発：2002 年度の試作処理系をベースに、2003 年度には、今後開発予定の本格的並列分散処理系開発のベースとなる強固な基準実装の構築を進めることとした。Ruby で全体の枠組を試作したのち、Java で処理系を作成して最初の版を公開した (<http://www.ueda.info.wased.ac.jp/lmnatal/>)。本処理系はまだ開発過程にあるが、LMNtal 言語仕様のうち、プロセス文脈の拡張機能などを除く大多数の機能を提供しており、さらに数値演算などを記述するために書換え規則に条件部（ガード）を指定する機能を設計実装している。また、Java で記述したインラインコードを呼び出す機能および簡単なモジュール化機能を持ち、ライブラリの蓄積による発展の礎を与えることができた。

(c) プログラミング例の蓄積：LMNtal の重要な応用分野として自己組織化原理に基づくプログラミングがあり、この領域の開拓は分子プログラミングの発展のみならず、プログラミングパラダイム全般にとっても重要である。そこで、自己組織化と関連の深い題材として PCR、セルオートマトン、二

分子生物ロボット、平衡二分木などをとりあげ、例題の記述実験を通じて、LMNtal が自己組織化反応のプログラミングのための記述力と簡潔性の両方を備えていることを確認した。

《本年度 成果論文リスト》

- [1] Akihiro Takahara and Takashi Yokomori:
On the computational power of Insertion-Deletion Systems, *Natural Computing*, Vol.2, No.4, pp.321-336, 2003.
- [2] Kazunori Ueda and Norio Kato :
The Language Model LMNta, In Proc. Nineteenth Int. Conf. on Logic Programming (ICLP'03), LNCS 2916, Springer-Verlag, pp.517-518, 2003.
- [3] 上田和紀，加藤紀夫：
言語モデル LMNtal . コンピュータソフトウェア , Vol.21, No.2 (2004), to appear .
- [4] 稲垣 良一, 上田 和紀：
SMP クラスタにおける効率的な MPI 集団通信関数の実装. 先進的計算基盤システムシンポジウム SACSIS2003, pp.171-172, 2003 年 5 月.
- [5] 加藤紀夫，上田和紀：
LMNtal プロセスの振舞いの定式化 . 日本ソフトウェア科学会第 20 回大会論文集 , 2003 年 9 月 , pp.16-20 .
- [6] 矢島伸吾，永田貴彦，加藤紀夫，上田和紀：
LMNtal プロトタイプ処理系の設計と実装 . 日本ソフトウェア科学会第 20 回大会論文集 , 2003 年 9 月 , pp.21-25 .
- [7] 坪 弘明，加藤 紀夫，上田 和紀：
LMNtal におけるルールセット不变性の検査 . 第 6 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 (PPL2004) , 日本ソフトウェア科学会 , 2004 年 3 月 .
- [8] 加藤 紀夫，上田 和紀：
並行言語モデル LMNtal におけるプロセス構造の解析 . 第 6 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 (PPL2004) , 日本ソフトウェア科学会 , 2004 年 3 月 .
- [9] 水野 謙，永田 貴彦，加藤 紀夫，上田 和紀：
LMNtal ルールコンパイラにおける内部命令の設計 . 情報処理学会第 66 回全国大会 , 2004 年 3 月 .
- [10] Yasubumi Sakakibara and Takahiro Hohsaka :
In Vitro Translation-based Computations, *Proceedings of 9th International Meeting on DNA Based Computers*, Madison, Wisconsin, 175–179, 2003.
- [11] Yasubumi Sakakibara :
DNA-based Algorithms for Learning Boolean Formulae, *Natural Computing*, Vol. 2, No. 2, 153–171, 2003.
- [12] 榊原康文:
分子コンピュータを用いた計算と学習アルゴリズム, 第 6 回情報論的学習理論ワークショップ (IBIS2003), 京都, 2003.

- [13] 榊原康文：
「ナノテクノロジーハンドブック」(第3編5章(DNA・分子コンピュータ)分担執筆),
オーム社, 2003.
- [14] Satoshi Kobayashi, Tomohiro Kondo, Kohei Okuda and Etsuji Tomita :
Extracting Globally Structure Free Sequences by Local Structure Freeness, in *Preliminary Proceedings of 9th International Meeting on DNA Based Computers*, pp.206, 2003.
- [15] Satoshi Kobayashi, Tomohiro Kondo, Kohei Okuda, and Etsuji Tomita :
A method for extracting globally structure free set of sequences, Technical Report of IEICE, COMP2003-16, pp.1-6, 2003.
- [16] Satoshi Kobayashi :
Structure Free Sequences for Biomolecular Computing, The 990th American Mathematical Society Fall Eastern Sectional Meeting, Binghamton University, Oct. 11-12, 2003, p.80, 2003 (invited).
- [17] Satoshi Kobayashi, Takashi Yokomori, Yasubumi Sakakibara :
An Algorithm for Testing Structure Freeness of Biomolecular Sequences, in *Aspects of Molecular Computing — Essays dedicated to Tom Head on the occasion of his 70th birthday*, Springer-Verlag, LNCS **2950**, pp.266-277, 2004.
- [18] 小林 聰：
分子計算のための配列設計, 第6回情報論的学習理論ワークショップ予稿集, pp.307-314, 2003.
- [19] 小林 聰：
DNA コンピューティングのための配列設計, 情報処理, Vol.45, No.2, pp.164-169, 2004.
- [20] Yasuhiro Suzuki, Peter Davis, Hiroshi Tanaka :
Emergence of auto-catalytic structure in stochastic self-reinforcing reaction networks,
J. of Artificial Life and Robotics, Vol.7, Springer-Verlag Tokyo, (in press).
- [21] Yasuhiro Suzuki, Junji Takabayashi, Hiroshi Tanaka :
Investigation of tritrophic system in ecological systems by using an artificial chemistry,
J. of Artificial Life and Robotics, Vol. 6, 129-132, Springer-Verlag, Tokyo, 2002.
- 【附記】(2003年度受賞)
網代育大, 上田和紀：人工知能学会2002年度全国大会ベストプレゼンテーション賞,
2003(発表題目「反復深化 A*探索によるもっともらしいプログラムの効率的生成」)