

博士論文概要書

**A Study on Computation Capability
of Biochemical Reactions**

生化学反応による計算能力の研究

早稲田大学大学院
教育学研究科

大久保 文哉

2013年11月

近年、理論計算機科学の分野において、生化学反応系の計算モデルの研究が注目を集めている。生化学反応系の計算モデルを研究する主な目的として、以下の二つが挙げられる。一つは、生化学反応系がどのように情報処理を実現しているかを理解することである。試験管内において、化学反応は状態遷移を行うような情報処理とみなすことができる。生化学反応系を計算機構とみなせば、その反応系のもつ性質を計算理論を用いて研究することができる。もう一つの目的は、生化学反応系を人工的に合成するための方法論を確立することである。目標とする生化学的な機能を実現するためには、適切なモデルを設計することが必要とされる。生化学反応系のモデルの設計のための方法論は、以下の三つのアプローチに分類することができる。

(1) **微分方程式**：化学反応のプロセスを解析する際、試験管内の各分子の濃度の測定がしばしば行われる。分子数が膨大であるとき、濃度は連続量で近似される。このとき、微分方程式を用いることによって、連続量の濃度の変化を容易に記述することができる。生化学反応系の微分方程式によるモデルでは、反応系の振る舞いを調べる問題を、対応する微分方程式を解析するという問題に還元できるという利点がある。

(2) **多重集合書き換え系**：反応に関わる分子数が比較的少ないとき、各分子の濃度は離散量として扱うのが適切である。そのような場合、多重集合の概念を用いて分子の濃度を表すことが多い。多重集合書き換え系は、化学反応系を極めて自然な方法でモデリングすることのできる離散状態遷移系である。例えば、計算機科学分野においてよく研究されている、ベクトル加算系やペ

トリネットも多重集合書き換え系に分類することができる。多重集合書き換え系によるモデルを用いることによって、化学反応系の性質を構成的な手法によって解析することができる。

(3) 分子計算：前の二つのアプローチでは、各分子は単なる記号として扱われていたが、分子の構造や生化学的な性質を扱うような形式モデルも考えられる。“分子計算”と呼ばれる研究分野は、DNA 分子などの構造をもつ分子の生化学的な性質を利用して計算を実現することを目標としている。この分野の最初の結果は、1994 年のエイドルマンによる革新的な実験による有向ハミルトンパス問題 (**Hamiltonian Path Problem: HPP**) の解法である。エイドルマンは **NP** 完全問題として知られる **HPP** を、DNA 分子に符号化することによって、実験的に解く方法を与えた。この実験では、DNA の二重鎖構造と”ワトソン・クリック相補性”と呼ばれる性質が重要な役割をもつ。

本論文の目的は、生化学反応系のもつ計算能力を明らかにすることである。そのために、第 3 章では多重集合書き換えによる計算モデルを、第 4 章と第 5 章では DNA の構造と性質に基づく計算モデルを考察する。これらのモデルのもつ計算能力を解析するために、形式言語理論や計算理論を用いる。

本論文は次の章から構成されている。

第 1 章 導入

第 2 章 準備

第 3 章 反応オートマトン (**Reaction automata**)

第 4 章 ヘアピン半完全化演算 (**Hairpin incompleteness**)

第 5 章 挿入システム (**Insertion systems**)

第1章では、研究の背景と位置づけについて述べる。

第2章では、本論文全体において使用する形式言語理論や多重集合理論の基本的な概念や記法を導入する。

第3章では、“反応オートマトン”という多重集合書き換えによる計算モデルを提案する。

生化学反応系に適したモデルを構築するために、2007年に Ehrenfeucht と Rozenberg の導入した **Reaction System** という形式的体系を研究の出発点とする。**Reaction System** は、生体機能は生化学反応間の相互作用によって決定されるというアイデアに基づいて構成されている。相互作用を制御するメカニズムとして、反応物と抑制物という2つの構成要素が重要な役割を担っている。**Reaction System** は、反応過程のシミュレーション解析や細胞内における機能発現のメカニズムの解明などへの貢献が期待されている。

Reaction System に基づいて、反応オートマトンを文字列の言語を受理する計算機構として導入する。**Reaction System** と同様に、反応オートマトンにおいて各反応は反応物、抑制物、生成物によって定義される。ただし、**Reaction System** では反応物、抑制物、生成物が通常の集合であったのに対して、反応オートマトンでは反応物と生成物を多重集合として扱う。もう1つの **Reaction System** との相違点は、反応オートマトンは入力として文字列を受け取り、計算することである。つまり、反応オートマトンは多重集合書き換えに基づいて言語を受理する計算モデルである。

この章では、反応オートマトンの理論における最初の結果として、反応オートマトンがチューリング完全であること、つまり任意の帰納的可算言語が反応オートマトンによって受理されることを示す。また、反応オートマトンの領域計算量を制限したクラスを導入し、それらのクラスと対応するチューリング機械のクラスについても考察する。

第4章では、DNAヘアピン構造と呼ばれる分子生物学における現象をもとに、“ヘアピン半完全化演算”という形式言語理論における演算を導入する。

DNAヘアピン構造は、ワトソン・クリック相補性とアニーリングによって、ある条件下で一本鎖のDNA分子（またはRNA）がとる二次構造のことをいう。DNAヘアピン構造は、分子計算の分野における新しい計算メカニズムにおいて多くの応用をもつ。“ヘアピン半完全化演算”は、既に研究されているDNAヘアピン構造に基づく演算と以下のような関係にある：

- ヘアピン半完全化演算は、2012年にManeaらによって提案された“hairpin lengthening”の制限付きの演算と見なすことができる。hairpin lengthening演算はDNA分子がヘアピン構造を構成するよう伸長させる演算である。ただし完全なヘアピン構造を構成する前に伸長を止めることができる。完全な伸長のみを許す場合、“hairpin completion”と呼ばれる。
- ヘアピン半完全化演算は、2011年にItoらによって提案された“bounded hairpin completion”の自然な拡張である。bounded hairpin completionは、hairpin completionにおいて伸長される接頭辞（接尾辞）の長さが定数で制限された演算である。
- ヘアピン半完全化演算は、2002年にHagiyaらによって提案された分子生物学における重要な実験におけるテクニックである“Whiplash PCR”の形式的なモデルを提供することができる。

この章では、ヘアピン半完全化演算やその反復の演算について、どのような言語族が閉包性をもつかについて述べる。まず、言語族がある複数の演算について閉じている場合、ヘアピン半完全化演算についても閉じていることを示す。そして反復演算に関して、すべての抽象言語族 (abstract family of

languages: AFL) が反復片側ヘアピン半完全化演算について閉じていることを示す. さらに一般的な両側の反復演算に関して, ある演算について閉じている線形言語は反復ヘアピン半完全化演算について閉じていることを示す. その系として, 反復ヘアピン半完全化演算に関する文脈自由言語の閉包性が得られる.

第5章では, 形式言語理論において長く研究されてきた, 挿入・削除演算に基づく計算モデルについて考える. DNA 配列の挿入・削除は, 遺伝子組み換えで分子内において実際に行われている演算であるため, 近年, 分子計算理論の観点からも注目を集めている. 計算モデルの生体分子による実装という観点から, 解析が煩雑になる文脈依存の演算よりも, より単純な文脈自由挿入・削除演算を研究することは重要であると考えられる.

この章では, 可能な限り短い文字列の挿入を用いた文脈自由挿入システムと, 形式言語理論における写像や演算を用いて, 文脈自由言語の特徴付けを得ている. より詳細には, 任意の文脈自由言語 L に対して, 以下の等式を成り立たせる射影 h , 文脈自由挿入システム γ , スター言語 F^+ が存在することを示した:

$$L = h(L(\gamma) \cap F^+).$$

ここで, γ は高々3記号の挿入が許された文脈自由挿入システム, F は各要素の長さが高々2であるような有限集合である. さらに, 文脈依存挿入システム γ と有限集合 F を用いて, 上の結果と同様の $h(L(\gamma) \cap F^+)$ という形による帰納的可算言語の特徴付けも得ている. これらの結果は2008年のPăunらによる特徴付けの改善になっている.