

研究終了報告書

「DNA 配列のボトムアップ型自動設計技術の開発」

研究期間：2020年11月～2024年3月

研究者：前田和勲

1. 研究のねらい

自然界の生物は複雑な化合物を効率的に合成できる。例えば、ある海水魚が持つマイトキシンという毒素は 32 個ものエーテル環を持つ。このような複雑な化合物は医薬品、プラスチック、繊維としての利用が期待されている。しかし、現在の技術レベルでは細胞を用いた発酵生産が可能なのは、アミノ酸やバイオ燃料のような比較的単純な化合物に限られている。この原因は、生物の化学工場としての潜在能力を引き出せていないことにある。

既存の遺伝子組換え技術は、遺伝子の欠損、挿入、または過剰発現といった「粗い」レベルの制御に限定されている。複雑な化合物を効率的に生産するには、適切な遺伝子を適切なタイミングで適切な量だけ発現させなければならない。遺伝子回路、すなわち細胞内の転写制御ネットワークによる情報処理システムを細胞に導入することが、この課題に対する鍵となる。ただし、遺伝子回路は、多数の遺伝子とタンパク質から構成されるので、実験による試行錯誤で設計することは難しい。このため、遺伝子回路を計算機で設計する技術が必要である。

本研究では、期待する生物機能を実現する遺伝子回路とそれに対応する DNA 配列を自動的に設計する計算技術を開発する。特に、DNA パーツを組み合わせて新たな DNA 配列を設計するボトムアップアプローチをとる。本研究の完成により、遺伝子発現の強さやタイミングの細かい制御が可能になり、遺伝子組換え微生物を使った有用物質生産に貢献できる。

2. 研究成果

(1) 概要

遺伝子回路とそれに対応する DNA 配列の自動設計を実現するために、「遺伝子回路設計の自動化」、「塩基配列設計と DNA 構築プロセス設計の自動化」、「実験による検証」の 3 つの研究項目を実施した。

遺伝子回路設計は、混合整数最適化問題として定式化できる。本さきがけ研究者は、その問題を解くために進化計算アルゴリズムのソフトウェアを開発した。また、大規模言語モデルと機械学習を使って、遺伝子回路設計とシミュレーションを自動化できることを示した。領域内共同研究では、遺伝子回路が大腸菌の増殖速度に与える影響を予測することに取り組んだ。

短時間に多数の遺伝子回路を設計することが可能になったが、その 1 つ 1 つに対して塩基配列を手作業で作るのは非現実的であるため、塩基配列設計と DNA 構築プロセスを自動化した。

最後に、生物実験によって、遺伝子回路、塩基配列、DNA 構築プロセスの自動設計技術の有効性を確認した。具体的には、論理演算 (OR、AND) を行う遺伝子回路を自動設計し、大

腸菌に導入して検証した。

このように本研究では、遺伝子回路の設計から DNA 構築プロセスの設計まで自動化した。既存技術は、遺伝子回路設計あるいは塩基配列設計までしかサポートしていないので、DNA 構築プロセスまでカバーしている点は本研究の大きなアドバンテージである。また、時間的に遺伝子発現を制御する遺伝子回路を設計できる点も本研究の強みである。

(2) 詳細

遺伝子回路とそれに対応する DNA 配列の自動設計を実現するために、主に次の 3 つの研究を実施した。

1. 遺伝子回路設計の自動化
2. 塩基配列設計と DNA 構築プロセス設計の自動化
3. 実験による検証

1. 遺伝子回路設計の自動化

遺伝子回路設計は、混合整数最適化問題として定式化できる。現実的な時間内でこの問題を解くには進化計算アルゴリズムが適している。研究期間の初期には、遺伝子回路設計を行うための進化計算アルゴリズムのソフトウェアを開発した (Maeda et al., 2021)。このソフトウェアは制約付きの最適化問題を効率的に解くことができる。これを利用して、既存の DNA パーツを組み合わせただけでは理想とする挙動を生み出すことが不可能な場合には、どのような DNA パーツが追加で必要なか提案できる。これは、遺伝子回路設計に失敗した際に、必要な改善点を提案しない既存の手法とは対照的である。

大規模言語モデルを応用することで、遺伝子回路のネットワーク構造を自然言語で指示して、そのシミュレーションモデルを簡単に構築できるようになった (Maeda and Kurata, 2023)。その他に、動力学パラメータのランダムサンプリングによって、設計した遺伝子回路が細胞内で機能する確率を評価する手法を開発した。また、機械学習を用いて速度定数を予測する手法も開発した (Maeda et al., 2022)。

領域内共同研究では、海津一成上級研究員(理研)と遺伝子回路が大腸菌の増殖速度に与える影響を予測する技術の開発に取り組んだ。

2. 塩基配列設計と DNA 構築プロセス設計の自動化

短時間で多数の遺伝子回路を設計できるようになった。しかし、その遺伝子回路の 1 つ 1 つに対して、それを実現するための塩基配列を手作業で設計することは現実的ではない。そこで、塩基配列設計とその DNA 構築プロセス設計を自動化した。PCR や DNA アセンブリといった DNA 構築に関する実験の手順を自動的に生成することで、実験者の負担を大幅に減らすことができた。

3. 実験による検証

これまでに開発した、遺伝子回路・塩基配列・DNA 構築プロセスの自動設計技術を検証するための生物実験を行った。論理演算を行う遺伝子回路と遺伝子発現の時間制御を実現する遺伝子回路を実際に自動設計し、その DNA を大腸菌に導入して、設計通りの挙動が実現されるか検証した。

まず、遺伝子回路 OR の検証を行った。遺伝子回路/塩基配列自動設計システムに対して、「2 つの転写誘導因子の一方または両方が与えられると YFP が発現する遺伝子回路を設計せよ」という指示を与えた(実際には自然言語ではなく、プログラムとして指示を与えている)。その結果、YFP のコード領域の上流に誘導可能な2つのプロモーターが並んだ遺伝子回路が設計できた。そして、それを実現するための具体的な塩基配列(1.3 kb)を設計させた。さらに、その塩基配列を構築するための実験プロセスを自動設計させた。次にこの実験プロセスに従って、実際に生物実験を行った。その結果、実際に、設計通りの塩基配列のプラスミドが設計できた。さらに、そのプラスミドを大腸菌に導入し、期待通りに「2 つの転写誘導因子の一方または両方が与えられると YFP が発現する」という挙動が実現されることを確認した。

次に、遺伝子回路 AND の検証を行った。遺伝子回路/塩基配列自動設計システムに対して、「2 つの転写誘導因子の両方が与えられたときのみ YFP が発現する遺伝子回路を設計せよ」という指示を与えた。その結果、5つのプロモーター配列と3つの転写因子、そしてYFP からなる遺伝子回路とその塩基配列 (3.9 kb)が設計できた。この塩基配列を DNA 合成で構築し、大腸菌に導入すると期待通りの挙動が確認できた。

3. 今後の展開

今後は、遺伝子回路自動設計システムを拡張して、遺伝子回路設計から実験まで一貫した自動化を実現したいと考えている。現在、実験自動化には、高額な装置が必要になることが多い。これに対して、我々は、安価な機器とオープンソースのソフトウェアを組み合わせ、スモールラボスケールで実験自動化を目指す。この部分は残りの研究期間と研究期間終了後の1年間で取り組む。

これまで、最大4遺伝子(3.9 kb)の遺伝子回路までしか実証できていない。そこで、より大きな遺伝子回路を実験で実証する必要がある。今後は、人間では設計が難しい、複雑な遺伝子回路を自動設計し検証する。1つの候補として、転写誘導因子で周期を調節可能な遺伝子発現振動子を設計することを考えている。この部分は研究期間終了後、1年を目処に行う。

最後に、実験自動化まで拡張した遺伝子回路自動設計システムを代謝工学へ応用する。具体的なターゲットとして、重要な医薬品前駆体であるシキミ酸の生産を考えている。遺伝子回路を使って、シキミ酸の生産に関する酵素の発現量を時間的・定量的に制御し、シキミ酸の生産効率を向上させることを目指す。我々の研究グループでは、シキミ酸生産経路を追加した大腸菌の代謝モデルを開発しているため、そのモデルと遺伝子回路自動設計システムを組み合わせる。この部分は研究期間終了後、2年を目処に行う。

4. 自己評価

これまで本さきがけ研究者は、主にバイオインフォマティクスや計算システム生物学の分野で研究

をしており、本研究課題では生物実験の部分が最も困難になることが予想された。しかし、この点は、領域内外の同僚研究者のアドバイスのお陰でスムーズに進めることができた。また、コンピュータ上で設計した遺伝子回路が大腸菌内で期待通りに動作することが確認できたことは、自分の計算技術に対する大きな自信となった。

研究マネジメントの点では反省点もある。実験検証自体は比較的スムーズに進んだが、大規模な遺伝子回路までは到達できなかった。この原因は、遺伝子回路設計技術の改善に取り組むあまり、生物実験の開始時期が遅れたことにある。一部の研究項目が不完全でも別の研究項目を並行して進めるべきであった。この反省点は今後の研究キャリアで活かしていきたい。

未達成の部分がある一方で、DNA 構築プロセスの自動設計や大規模言語モデルの応用など、当初予想していなかった部分で予想以上の進展があった。当初は、遺伝子回路と塩基配列の自動設計のみを行う計画であったが、これに DNA 構築プロセスの自動設計技術を加えることで、合成生物学の DBTL (Design-Build-Test-Learn) サイクルの Design だけでなく Build までカバーしたシステムを構築することができた。研究期間中は ChatGPT が公開されたことで、大規模言語モデル (LLM) を科学研究にどのように応用するのかという関心が世界的に高まった。その中でいち早く、遺伝子回路などの動力学シミュレーションに対して LLM を応用する手法について発表することができた (Maeda and Kurata, 2023)。このように、研究期間中は、自身の研究の進捗状況や世界的な研究の潮流に応じて、柔軟に研究を進めることができた。

研究期間の大半はコロナ禍であったが、人的ネットワークを広げることができた。海津上級研究員と増殖速度の予測に関する領域内共同研究を行うだけでなく、実験自動化に関する共同研究を開始した。山田亮祐准教授(大阪府立大)とは、有用物質生産におけるデータ駆動アプローチとシミュレーション駆動アプローチの比較研究を継続している。また、本領域から多数の参加者がある「細胞を創る」研究会では、遺伝子回路設計に関する講演を行った。また、分子ロボティクス研究会では招待講演を行った。国内での動力学モデリング研究の普及を目指して、杉本昌弘教授(慶応大)と共同で国内最大のバイオインフォマティクスの学会である IIBMP 2022 にて数理モデリングのワークショップを実施した。このように、領域内外の合成生物学コミュニティや数理モデリングコミュニティの中でネットワークが構築できたことは非常に有益であった。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 4件

1. Kazuhiro Maeda*, Fred C. Boogerd, and Hiroyuki Kurata, RCGAToolbox: A real-coded genetic algorithm software for parameter estimation of kinetic models, IPSJ Transactions on Bioinformatics, 14, 30-35, 2021 *Corresponding Author

遺伝子回路や塩基配列の設計は、制約付きの最適化問題として定式化できる。制約付きの最適化問題を解くためのアルゴリズムはこれまでも提案されているが、使い勝手や並列計算の効率の点で問題が残っている。そこで、これらの問題を解決した新しいアルゴリズム/ソフトウェアを開発した。

2. Kazuhiro Maeda*, Shuta Hagimori*, Masahiro Sugimoto, Yasuyuki Sakai, Masaki Nishikawa,

Simulation of the crosstalk between glucose- and acetaminophen metabolism in a liver zonation model, <i>Frontiers in Pharmacology</i> , 13, 2022 *These authors equally contributed to this work.
この研究では、下の機械学習で速度定数を予測する手法を応用して、ヒト肝臓の糖代謝-薬物代謝統合モデルを開発した。このモデルにより、糖代謝の状態によって、薬物の肝臓への影響が大きく変化する可能性があることを予測した。
3. Kazuhiro Maeda*, Aoi Hatae, Yukie Sakai, Fred C. Boogerd, and Hiroyuki Kurata, MLAGO: Machine learning-aided global optimization for Michaelis constant estimation of kinetic modeling, <i>BMC Bioinformatics</i> , 23(1):455, 2022 *Corresponding Author
遺伝子回路などの動力学シミュレーションには、速度方程式だけではなく様々な速度定数の値が必要である。しかし、多くの速度定数は測定されておらず、動力学シミュレーションを行う上でのボトルネックとなっている。そこで、この研究では、機械学習を用いて速度定数を予測する手法を開発した。
4. Kazuhiro Maeda* and Hiroyuki Kurata, Automatic Generation of SBML Kinetic Models from Natural Language Texts using GPT, <i>International Journal of Molecular Sciences</i> , 24(8):7296, 2023 *Corresponding Author
動力学シミュレーションは、遺伝子回路や代謝経路の合理的な設計において重要である。しかし、動力学モデルの構築は手間がかかるという問題がある。そこで、この研究では、自然言語を使って遺伝子回路などの動力学モデルを構築する手法を開発した。

(2) 特許出願

研究期間全出願件数: 0 件

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

- 招待講演: 前田和勲、動力学シミュレーションを使った生化学システムの理解と設計、第2回分子サイバネティクス・第46回分子ロボティクス定例研究会、オンライン、2021年5月10日
- ポスター発表: Kazuhiro Maeda, Fred C. Boogerd, and Hiroyuki Kurata, RCGAToolbox: A real-coded genetic algorithm software for parameter estimation of kinetic models, *ISMB/ECCB 2021*, Online, July 2021
- 口頭発表: 前田和勲、遺伝子回路設計と塩基配列設計の自動化、「細胞を創る」研究会16.0、東京、2023年9月26日
- 招待講演: 前田和勲、倉田博之、動力学シミュレーションを使って生物システムを理解・設計する、第13回日本細胞性粘菌学会例会、九州工業大学戸畑キャンパス、2023年10月21日
- 招待講演: 前田和勲、人工知能とシミュレーションを使って生物を理解・設計する、第54回LaMer特別講演会、愛媛大学、2024年1月26日
- ソフトウェア: RCGAToolbox (<https://github.com/kmaeda16/RCGAToolbox>)
- ソフトウェア: Machine learning-based Km predictor (<https://kmaeda.pythonanywhere.com/>)
- ソフトウェア: KinModGPT (<https://github.com/kmaeda16/KinModGPT>)