

研究終了報告書

「転移学習を用いた非モデル生物の無細胞タンパク質合成系開発」

研究期間: 2021年10月~2024年3月

研究者: 西田 暁史

1. 研究のねらい

これまでにヒト腸内や環境中の微生物生態系をメタゲノム解析によって明らかにし、統計・機械学習モデルによって環境に特徴的な微生物が推測されてきた。しかし dry 研究だけではその微生物の機能までは明らかにできず限界があった。微生物学の分野ではこの数年のメタゲノム解析技術の向上によりゲノムデータベースは大幅に拡充され、次のステップとしてデータベース上の莫大な遺伝子を発現しテストするハイスループットな系が必要になっている。そこで注目されているのが細胞抽出液に DNA を加えることでタンパク質を合成する無細胞タンパク質合成系（以下、無細胞系）である。無細胞系は遺伝子組換えや細胞培養が必要ないため短い時間で遺伝子パーツをテストでき、*in vivo* 実験のためのブレッドボードとして利用されている。実際にバイオ燃料やバイオプラスチックなどを生産するための代謝系構築に役立っている。これからは難培養性細菌由来の遺伝子の解析も増えてくるであろうが、遺伝子は由来となった宿主細菌にとって系統的に近い細菌の抽出液を用いた方がより正確にタンパク質が合成されるという仮定のもと、非モデル細菌の無細胞系が開発されている。この仮定は、10 種 (Proteobacteria 7 種, Firmicutes 2 種, Actinobacteria 1 種) の細胞抽出液を用いた転写反応によって検証された (Sung Sun Yim *et al.*, *molecular systems biology*, 2019)。しかし、無細胞系の組成最適化ができておらず *Lactobacillus* や *Bacteroides* の抽出液でのタンパク質合成に失敗した。多種類の試薬で構成されている無細胞系には多次元連続変数の最適化が必要となり、そのためには機械学習モデルが有効である。Borkowski らはデータを測定しながら繰り返し学習する機械学習の1つであるアクティブラーニングにより、モデル生物である大腸菌の無細胞系の組成最適化に成功した (Olivier Borkowski *et al.*, *nature communications*, 2020)。しかし既存手法では多くの学習データが必要となり、様々な細菌由来の無細胞系を作製するには時間的・金銭的コストが多くかかる。そこで本研究では、大腸菌の学習済みモデルを転移学習に利用することによって、少ない学習データで他の細菌の無細胞系の組成最適化を達成する手法を開発し、無細胞系のレパートリーを爆発的に拡充する。そして、多様な細菌の無細胞系の有効性を検証する。

2. 研究成果

(1) 概要

無細胞系の組成最適化のために、自動分注機 OT-2 (Opentrons 社) とベイズ最適化を組み合わせたシステムを開発し、少ない学習データで効率的に無細胞系を最適化させる。さらに大腸菌由来の無細胞系モデルを他の微生物由来の無細胞系開発に転用することで、更に少ない学習データで多様な微生物由来の無細胞系を開発するシステムを構築するのが目的である。そのためにもまず、自動分注機 OT-2 による無細胞系調製の自動化を行った。最小可能分注量が 1 μ L と制約はあるが 10,000 ドル以下で手に入るためよく普及しており、OT-2 で無細胞系の組成最適化を達成できることで、様々な機関での利用が期待できる。そのためにもまず

OT-2 の分注誤差を計測し、ソフトウェアによる補正処理を行った。さらに無細胞系を調製するときにはアミノ酸を分注前に攪拌することが重要であることが分かり、その機能を追加した。次に最適化手法としてベイズ最適化を採用し、無細胞系組成を効率的に最適化するシステムを構築したが、6試薬(Mg、K、アミノ酸、スペルミジン、3-PGA、NTP)の濃度を最適化することとした。OT-2 で6試薬の濃度を調整しようとする1サンプルあたりの無細胞系調製に6分ほどかかったため、1つの作業では 30 サンプルを扱うこととした(分注時間3時間強)。はじめは各試薬の上限と下限を設定し、その連続値範囲内で各試薬濃度を最適化しようとしたが、OT-2 の分注速度や配置できるチューブ数の制約などから連続値での最適化は断念し、各試薬濃度条件を3つ設定した上での離散値の組合せの最適化を行った。ベイズ最適化では次に実験する試薬濃度条件を 30 条件提示し、OT-2 でそれらの無細胞系組成を自動調製し、遺伝子発現として GFP を発現させ蛍光強度を測定、測定結果をモデルに学習させる。この工程を繰り返すことで無細胞系の組成を効率的に最適化できるシステムを構築した。大腸菌由来の無細胞系モデルを用いて、他の微生物の無細胞系を効率良く開発することが目的のため、大腸菌 BL-21 株以外に *Pseudomonas putida* KT2440 株由来の無細胞系も作製した。その際に、大腸菌由来の無細胞系と同じ組成で *Pseudomonas putida* 由来の無細胞系も遺伝子発現できることを確認できた。

(2) 詳細

「自動分注機 OT-2 による無細胞系調製の自動化」

合成生物学分野の研究では最小可能分注量が 25 nL(Echo 525 の場合)と微量の分注を可能とする Echo アコースティック分注機(Beckman Coulter Life Sciences 社)が自動分注機として使われることが多いが、高額で使える研究チームに限られる。そのため、本研究では最小可能分注量が 1 μ L と制約はあるが 10,000 ドル以下で手に入るためよく普及している OT-2 (Opentrons 社)の自動分注機を用いて、無細胞系の組成調製を自動化した。

まず OT-2 の分注性能を、分注容量 300–20 μ L と 20–1 μ L のピペットマン(それぞれ P300 と P20 と示す)の上限と下限で測った結果を表1に示す。水を P300 での 300 μ L 分注では 4 回、20 μ L 分注では 50 回、P20 での 20 μ L 分注では 50 回、1 μ L 分注では 100 回分注し、重量を測ることによって分注1回あたりの誤差を推定した。無細胞系最適化実験では P20 ピペットマンを使用するため、定格下限の 1 μ L を分注する実験に対してバイアス補正を行った。プログラム上で 1.3 μ L を分注したところ、目標分注量 1 μ L としたときの誤差は -0.02 ± 0.03 SD となり、分注誤差が抑えられた。

表1 OT-2 の各分注量における誤差 (n = 3)

ピペットマン	分注量	誤差 [μ L]	誤差 [%]
P300	300 μ L	-0.26 ± 0.97 SD	0.09 ± 0.32 SD
P300	20 μ L	-1.27 ± 0.55 SD	6.34 ± 2.74 SD
P20	20 μ L	-0.23 ± 0.14 SD	1.16 ± 0.71 SD
P20	1 μ L	-0.27 ± 0.02 SD	26.7 ± 1.84 SD

※ SD: 標準偏差

次に OT-2 により無細胞系を調製し、緑色蛍光タンパク質 (GFP) を発現させ蛍光強度を測定することで、無細胞系調製の評価を行った。無細胞系の調製は基本的には Sun らの手法 (Zachary Z. Sun *et al.*, *journal of visualized experiments*, 2013) に則っているが、今回の実験では Mg、K、アミノ酸、スペルミジン、3-PGA、NTP の濃度を振るため、この6試薬は個別にストック溶液を用意した。その他のストックとして、CoA・NAD・cAMP・フォリン酸・tRNA・PEG-8000 などを混合した溶液ストックや、T7 プロモータ下流に GFP s1 が挿入されたプラスミド DNA のストック (Eiko Seki *et al.*, *analytical biochemistry*, 2008)、大腸菌 BL-21 Star (DE3) 由来の細胞抽出液のストックを用意した。これらのストック溶液を Sun らが報告した濃度で混合し、37°C で 2 時間静置培養後に蛍光強度を測定したところ、手作業で調製した無細胞系は遺伝子発現していたが、OT-2 に調製させた無細胞系は遺伝子発現していなかった。原因はアミノ酸の沈殿であった。アミノ酸溶液は RTS Amino Acid Sampler (biotechrabbit 社) を用いて、ロイシンは 5 mM、その他のアミノ酸濃度は 6 mM となるようにストックを用意していた。OT-2 によるアミノ酸分注前にストック溶液を攪拌することで無細胞系で遺伝子発現を確認できた。

「ベイズ最適化による無細胞系組成最適化」

これまでの無細胞系最適化に使われてきた手法は、多層パーセプトロン等の機械学習モデルを用いた最適化手法であるアクティブラーニングであった。無細胞系組成の最適化という目的を達成するにはアクティブラーニングで問題ないが、モデルがブラックボックスのため最適化の過程の可視化が難しい。そこで本研究ではベイズモデルを用いて最適化を行うベイズ最適化を採用した。本実験では Borkowski らの研究で無細胞系活性に大きな影響を及ぼすと示唆された6試薬 (Mg、K、アミノ酸、スペルミジン、3-PGA、NTP) の濃度を最適化することとした。はじめは各試薬の上限と下限を設定し、その連続値範囲内で各試薬濃度を最適化しようとしたが、OT-2 の分注速度や配置できるチューブ数の制約などから連続値での最適化は断念し、各試薬濃度条件を3つ設定した上での離散値の組合せの最適化を行った。

ベイズ最適化では次に実験する試薬濃度条件を複数提示し、OT-2 でそれらの無細胞系組成を自動調製し、遺伝子発現量を測定、測定結果をモデルに学習させる。本実験では GFP を発現させ蛍光活性を測定することで遺伝子発現量を推定している。この実験条件の提示から測定結果の学習までの工程をラウンドと呼び、ラウンドを複数回まわすことで効率的に高発現量を達成する無細胞系組成を探索する。1 サンプルあたりの無細胞系調製に6分ほどかかったため、1ラウンドあたりのサンプル数は 30 とした (全体で3時間強)。各試薬の3つの終濃度条件は、Mg { 8.0/4.4/0.8 } mM、K { 80.0/44.0/8.0 } mM、アミノ酸 { 1.5/1.0/0.2 } mM、スペルミジン { 1.0/0.6/0.1 } mM、3-PGA { 30.0/16.5/3.0 } mM、NTP { 1.3/0.7/0.2 } mM である。2ラウンド目までまわした段階で、Mg 0.8 mM、K 80.0 mM、アミノ酸 1.5 mM、スペルミジン 1.0 mM、3-PGA 16.5 mM、NTP 1.3 mM の終濃度条件が最も無細胞系の発現量が大きかった。

「*Pseudomonas putida* の細胞抽出液を用いた無細胞系開発」

大腸菌 BL-21 Star (DE3) 株由来の細胞抽出液を用いた無細胞系のモデルを開発し、それを他の微生物由来の無細胞系開発に利用することで、少ない実験試行回数で無細胞系組成の最適化ができないかと考えた。その実証のために *Pseudomonas putida* KT2440 株由来の細胞抽出液を利用した無細胞系を開発した。DNA は T7 プロモータ下流に sfGFP の遺伝子が挿入

された Michael Jewett 提供の pJL1-sfGFP プラスミド (Addgene plasmid # 102634 ; <http://n2t.net/addgene:102634> ; RRID:Addgene_102634) を用いた。試薬組成は Sun らのプロトコルに則り、大腸菌由来の無細胞系と同条件で行ったが、*P. putida* 細胞抽出液を用いた無細胞系で遺伝子発現していた(図1)。

本プロジェクトの当初の到達目標は、ベースとなる無細胞系のモデルを利用することで、他の微生物由来の無細胞系を開発するとき、無細胞系組成を最適化するのに必要な実験データ数を減らせることを実証することである。研究期間において、自動分注機とベイズ最適化を組み合わせることで無細胞系の組成を最適化するシステムを構築できた。また、大腸菌以外にも *P. putida* の無細胞系を大腸菌無細胞系と同じ組成で開発することに成功した。材料はそろっているため、今後大腸菌の無細胞系モデルを他の微生物由来の無細胞系開発に転用する手法を実証する予定である。

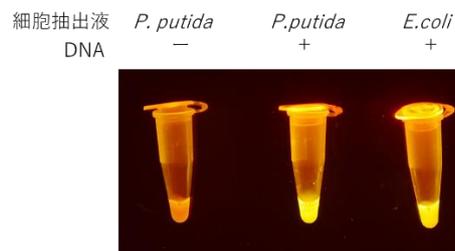


図1 *P. putida* 由来無細胞系での GFP 発現結果

3. 今後の展開

今後の展開として、まず本プロジェクトの成果を利用し、東京農業大学の微生物リソースセンターが保有する多様な乳酸菌由来の無細胞系を、2年ほどのスケールで開発する予定である。その他に基礎研究では微生物を遺伝子組換えした *in vivo* 系と無細胞系の *in vitro* 系での発現量の違いが何に由来するのか翻訳に着眼しながら2年ほどのスケールで解明する予定である。応用研究としては無細胞系による有用物質生産に着目しており、ファージの合成に2年ほどのスケールで取り掛かっている。

4. 自己評価

本プロジェクトの目標は、まず自動分注機とベイズ最適化を用いて大腸菌由来の無細胞系組成を最適化し、次に大腸菌由来の無細胞系モデルを転用することで他の微生物由来の無細胞系開発を効率的に行うことである。達成状況は、前者の大腸菌由来無細胞系の最適化については自動分注機とベイズ最適化を組み合わせることで無細胞系組成を最適化する仕組みは完成し、無細胞系の最適化を達成できる見込みである。一方後者の、モデル転用を利用した他の微生物由来の無細胞系開発に関しては、*Pseudomonas putida* 由来の無細胞系での遺伝子発現を確認できた状況で、今後モデル転用を利用した最適化実験を行う予定である。期間内に研究成果を出せなかったため、引き続き研究を行い、成果を出す予定である。

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: 0件

(2)特許出願

研究期間全出願件数： 0件(特許公開前のものは件数にのみ含む)

(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)

1. 西田暁史、山本渚生、志波優、田中尚人、「自動分注機 OT-2 による無細胞系最適化」、デザイン生命工学会 第8回大会、2023 年 3 月 9 日(東京都 文京区)