

有用タンパク質高生産に向けた配列設計技術の現状と展望



研究ノート

山崎 将太郎*

Current Status and Future Perspectives of Sequence Design Technologies
for High-Yield Production of Useful Proteins

Key Words: Protein production, Sequence optimization,
Plant-based vaccine production, High-quality data generation

はじめに

生物の体を構成する要素の多くは多種多様なタンパク質であり、その機能は生命活動を支える根幹をなしている。医薬・産業分野においても、酵素、抗体、ワクチンなどさまざまな有用タンパク質が活用されている。一方で、タンパク質は巨大分子であるため、化学的手法による合成は依然として困難である。そのため、有用タンパク質は、生物が本来持つ「遺伝子発現 (DNA の情報からタンパク質を作り出す仕組み)」を利用し、遺伝子組換え技術によって細胞に作らせる方法が一般的である。

タンパク質は、その設計図である遺伝子の DNA 配列に基づき、DNA の情報が RNA に写し取られ (転写)、それをもとにタンパク質が作られる (翻訳) という過程を経て合成される (図 1)。したがって、有用タンパク質の設計図である DNA を細胞に導入すれば、目的とするタンパク質を同様に生産できる。遺伝子の DNA 配列には、「何を」「どの条件で」「どの細胞で」「どの程度の量」作るのかという、設計図と生産指示の両方がコードされている。そのため、有用タンパク質をいつでも、どこでも、効率よく大量に生産できるように DNA 配列を最適化することが、生産効率を最大化する鍵となる。本稿では、筆者らが取り組んできた高生産技術の開発について、これまでの成果と現在の挑戦を紹介する。

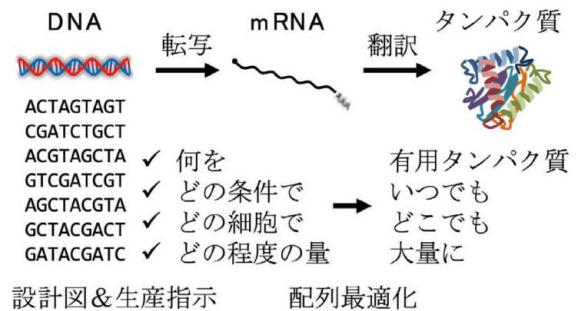


図 1 遺伝子発現過程と有用タンパク質生産

生産技術の現状と課題 — 植物由来ワクチンを例に —

植物を利用した組換えタンパク質生産では、動物細胞と同様に、タンパク質合成後に糖鎖付加などの化学的修飾 (翻訳後修飾) が行われ、高品質なタンパク質を生産できる。加えて、病原体やエンドトキシンの混入リスクが低く、高い安全性を有する。さらに、植物は栽培が容易で、低コストかつ大規模生産にも適している。このため、医薬品やワクチン製造において注目されており、RNA ワクチンに抵抗感を持つ人々にとっても、新たな選択肢を提示する技術となり得る。

実際にカナダの Medicago 社は、タバコ科植物 *Nicotiana benthamiana* を用いた、感染性を持たないウイルス外殻のみを模したウイルス様粒子 (VLP: Virus-Like Particle) ワクチン「COVIFENZ」を開発した¹⁾。COVIFENZ は COVID-19 に対して高い有効性と安全性を示し、カナダ保健省から承認を受けた世界初の植物由来ワクチンである。その生産系には、筆者が加藤晃教授 (奈良先端科学技術大学院大学) および Medicago 社と共同開発した“翻訳エンハンサー配列”が導入されている^{2,3)}。この配列は、Polysome 解析と Cap Analysis of Gene Expression (CAGE) 法を組み合わせた新規解析手法により⁴⁾、タバコ内在性遺伝子の mRNA 翻訳効率を詳細に評価する過程で発見されたものである。効率化の対象



* Shotaro YAMASAKI

1988年2月生まれ
奈良先端科学技術大学院大学 バイオサイエンス研究科 博士後期課程 (2016年)
現在、大阪大学 微生物病研究所附属バイオインフォマティクスセンター 准教授
博士 (バイオサイエンス)
TEL : 06-6879-8297
E-mail: shotaro-yamasaki@biken.osaka-u.ac.jp

として、従来注目されてきたプロモーターやコーディング領域ではなく、翻訳制御に関わる5'非翻訳領域(5'UTR)を改良する点に特徴を持つ。その結果、当初Medicago社の生産系で使用が予定されていた既存のエンハンサー配列の数倍の発現促進効果を示した。遺伝子組換えによるタンパク質生産において、DNA配列の改良は化学的・物理的なコスト増加を伴わない。そのため、この翻訳エンハンサー配列の導入は、同じ原料・手間・時間でより多くの産物を得ることを可能にする技術的革新といえる。残念ながら、COVIFENZはWHOの承認方針(タバコ関連企業との資本関係)やmRNAワクチンの急速な普及など、政治的・商業的要因により生産終了となった。しかし本プロジェクトを通じて、植物由来医薬品の大規模工場生産、品質管理体制、安全性・有効性の実証など、多くの技術的基盤が確立されたことは大きな成果である。今後、植物発現系が再び社会実装されるためには、これまで得られた経験を生かしつつ、コスト競争力を強化することが求められる。そのためには、生産性向上技術のさらなる発展が重要な鍵となる。

生産システムは、人にとって最適化されているのか」という点である。生物は進化の過程で、各タンパク質が必要な量だけ産生されるよう精緻に調節機構を発達させてきた。過剰な生産はエネルギーの浪費や細胞機能の阻害を招くため、むしろ抑制的に制御されている。一方、人が有用タンパク質を大量に得ようとする場合には、進化的には不利な、長期的には細胞機能に障害を起こすほどの「高発現」を目指す必要がある。ウイルスや一部の細菌はそのような高発現機構を備えており、これらに由来する発現制御要素は実際に生産系で広く応用されている。しかし、宿主生物は進化の過程でこれらの高発現を抑制する複数の防御機構を形成しており、理論的な最高効率で発現する配列は自然界には存在しない可能性が高い。

では、どのようにして高発現配列を得ることができるのか。遺伝子発現量は転写、翻訳、mRNA安定性など複数の過程によって制御されており、それぞれに多様な制御因子が関与している。理論的には、効率化に寄与する因子を同定し、それらを最適に組み合わせれば良いが、実際には因子間の相互作用や副次的効果が複雑に絡み合い、単純な組み合わせでは期待通りの高発現は得られない。この複雑性を克服する手段の一つとして注目されているのが、AIを用いた配列最適化である(図2)。この手法では、まず注目する発現制御過程に関して、発現効率とそれに関与する塩基配列の大規模データを取得する。データ取得には、次世代シーケンシング(NGS)技術を用いた網羅的解析が一般的に使用される。次に、得られたデータをもとに、AIが塩基配列と発現効率の関係を学習し、予測モデルを構築する。最後に、得られたモデルを用いて配列の発現効率を予測し、その結果をもとに配列を改変・再評価する操作を繰り返すことで、より効率的な配列を探索する。最適化過程では、遺伝的アルゴリズムをはじめとする進化的探索法がしばしば用いられ、いわば「高発現」という人工的な目的関数に基づく計算機内での進化シミュレーションといえる。筆者らもこれまでに複数の制御過程を対象とした最適化システムを構築し、既存の発現系を上回る配列設計に成功してきた⁵⁾。現在は、植物、哺乳類細胞、無細胞タンパク質合成系など、多様な発現プラットフォームを対象に開発を進めている。

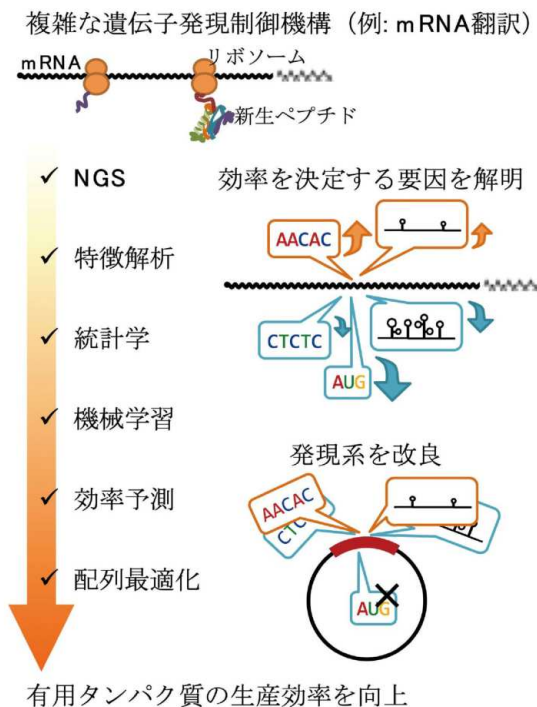


図2 配列最適化の流れ

AIを用いた配列最適化

ここで改めて問いたいのは、「生物のタンパク質

精度向上のためのデータ戦略

最適化技術の性能を左右する最大の要因は、塩基配列がもつ発現能力を評価する予測モデルの精度である。一方で、AI アルゴリズム自体はすでに成熟段階にあり、最新手法間での性能差は限定的である。また、いかに高度なアルゴリズムを用いても、学習に使用するデータの「量」と「質」が不足していれば、高精度な予測は実現できない。したがって、重要なのは学習に供するデータの設計である。

第一に、遺伝子発現の各段階（転写、翻訳、mRNA 安定化など）において特に重要な制御機構や制御因子に焦点を当てたデータ取得が重要である。これまで見逃されてきた制御機構や因子を特定・改良できれば、新たな発現制御技術の開発につながる可能性がある。第二に、多様で高品質なデータを取得することが精度向上に不可欠である。筆者らは、CAGE 法やロングリード cDNA シーケンシングなどの専門的な NGS 技術を活用し、転写アイソフォーム（同一遺伝子から生じる複数の mRNA 配列）単位での詳細かつ網羅的な転写後調節解析を実施している。これにより、従来は把握が難しかった発現制御過程を高解像度で評価することが可能となった。また、超並列レポーターアッセイ法は、数千から数百万種類に及ぶ人工配列を一度に合成し、その発現能力を評価できる画期的な手法である（図3）。この方法の最大の特長は、単なるランダム配列ではなく、すべての合成配列を設計段階で指定できる点にある。例えば、数個の制御因子の全組み合わせを実際に合成し、その影響や相互作用を検証することが可能である。従来のゲノムワイド解析に基づく手法では、学習データは生物が本来もつ配列パターンのみであった。つまり、自然界に存在しない高効率配列を探索しながらも、学習データ自体は自然配列に依存していたのである。超並列レポーターアッセイ法はこの制約を克服し、今後のデータ取得における中心的技術となると筆者は考えている。筆者は、これらの技術を組み合わせることで、自然界に存在する配列の高精度データと、人工的に設計された大規模データの双方を収集し、高効率配列をより正確に設計・合成できる新しい技術の開発を進めている。

おわりに

有用タンパク質の高生産を実現するうえで、配列

最適化技術は欠かせない要素である。近年では、Web 上で簡便かつ短時間で遺伝子配列を設計できるツールも普及している。しかし、遺伝子発現の制御機構は極めて複雑であり、既存のツールもそのすべてを予測できる段階には至っていない。配列最適化技術は現在も進化を続けており、それに伴って生産プロセスの効率化も着実に進展している。今後、こうした技術の高度化により、より安定的で高効率なタンパク質生産が可能になるだろう。この機会に、みなさんが使用されている遺伝子配列を見直し、再設計してみるのも有意義ではないだろうか。

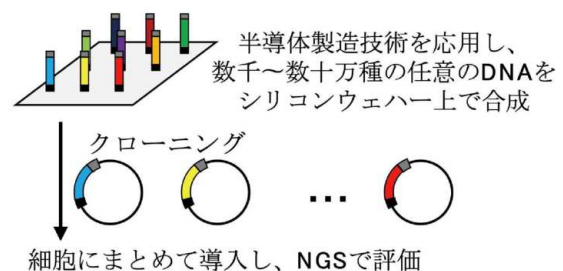


図3 超並列レポーターアッセイ法の概要

参考文献

- 1) Hager, K. J., Pérez Marc, G., Gobeil, P., et al. (2022) Efficacy and safety of a recombinant plant-based adjuvanted COVID-19 vaccine. *New England Journal of Medicine*, 386(22), 2084–2096.
- 2) Lavoie, P.-O., D'Aoust, M.-A., Kato, K., and Yamasaki, S. (2019) *Endogenous plant expression enhancer*. WO2019173924A1; Patent JP7512546B2 (granted 2024).
- 3) Lavoie, P.-O., D'Aoust, M.-A., Kato, K., and Yamasaki, S. (2020) *Plant expression enhancer*. WO2020181354A1; Patent JP7423875B6 (granted 2024).
- 4) Yamasaki, S., Suzuki, A., Yamano, Y., Kawabe, H., Ueno, D., Demura, T., and Kato, K. (2018) Identification of 5' -untranslated regions that function as effective translational enhancers in monocotyledonous plant cells using a novel method of genome-wide analysis. *Plant Biotechnology*, 35(4), 365–373.
- 5) 加藤 晃, 山崎 将太郎. (2022) 植物へ導入する遺伝子の人工配列の設計システム, 特開 2022-155412