

# 植物の遺伝情報を 人工知能で読み解き、 有用なタンパク質の 量産システムを設計する

バイオサイエンス領域 バイオエンジニアリング研究室



加藤 晃 教授



山崎 将太郎 助教

## 最適な状態を予測する

医薬品などに使われる付加価値が高いタンパク質は、量の確保や品質の安定化、コスト削減のために動植物、微生物の細胞に作らせている。生命の設計図である長い鎖状のDNA分子の中から、目的のタンパク質の情報が指定された部分（遺伝子）など合成に関わる遺伝情報を抽出し、別の細胞の核内に導入して生産を肩代わりさせるという手法だ。

バイオエンジニアリング研究室の加藤教授らは、こうした手法の生産効率を高めるため、遺伝情報を膨大な実験データと突き合わせてコンピュータ解析し、人工知能（AI）の機械学習を使って、最適な遺伝情報の組み合わせを予測する研究に挑んでいる。企業との共同研究を進め、特許化された成果も多い。

「バイオテクノロジーでの社会貢献をめざし、植物の細胞をメーン材料に、目的のタンパク質を効率よく生産するための技術開発にシフトした研究を行っています」と加藤教授。デジタル技術を駆使してバイオ研究を手掛ける本学のデジタルグリーンイノベーションセンター（CDG）のバイオエコノミー部門長を務める。

## 最強のエンハンサーを発見

タンパク質の遺伝情報は、鎖状のDNA分子の一部の区間にアデニン（A）、グアニン（G）、シトシン（C）、チミン（T）の4種類の核酸塩基により暗号文字のように書き込まれている。その塩基の並び順（配列）が、タンパク質を構成するアミノ酸の種類、結合の順番などを指定するのだ。この遺伝情報の暗号文はmRNA（メッセンジャーRNA）という分子にコピー（転写）されたあと、mRNAに結合したりボソームという構造体により「翻訳」されて基の設計図通りにアミノ酸が連結したタンパク質が出来上がる。

加藤教授は、「研究室では、DNAの遺伝子暗号を基にタンパク質が合成される過程で、転写後のmRNA分子の安定性、

mRNAからタンパク質への翻訳の効率など、すべての段階を最適化し、最終生産量を増やすことをめざしています」と説明する。例えば、mRNAが持つ遺伝情報をリボソームが翻訳してタンパク質を合成する場合、それぞれの遺伝情報に特有の塩基(A、G、T、C)の配列パターンと翻訳の効率との関係について、細胞の実験でデータを取り、そこから数理モデルをつくって、コンピュータ解析する。ただ、タンパク質を構成する数百のアミノ酸それぞれが複数の異なるパターンの塩基配列により指定されるので、膨大な数の塩基の組み合わせパターンになる。そのため、人工知能の機械学習の手法などを使って特徴点を捉え、有力なパターンに絞って、最適な塩基配列を探り当て、それを入力して高効率の翻訳を予測するシステムを構築した。その結果、DNA上で遺伝子に隣接した翻訳の開始に重要な領域(非翻訳領域の5'UTR)である「AtCOR47-5'UTR」という塩基配列が、植物細胞での翻訳の効率を高め、タンパク質の合成量を増加する翻訳エンハンサーとして最強であることを発見。この成果はすでに特許化されている。

### 完全人工遺伝子のシステムめざす

「植物に本来持たないタンパク質を作らせる場合、その細胞内には対応するmRNAが備わっていません。そこで、コンピュータに植物のmRNAのパターンを学習させながら、数理モデルを使って解析し、植物細胞でのタンパク質の翻訳効率の向上にふさわしい塩基配列の変更の情報を提示するという最適化の技術を開発しています」と説明する。

現在のターゲットは、高付加価値の医療用タンパク質。再生医療の細胞培養に必要な成長ホルモンやワクチンの抗原などで「植物細胞では、ウイルス感染のチェックなどの必要がなく、コストを下げることができます」という。「最終的には、優れた性質のタンパク質を安価につくるために、DNAの転写の開始から翻訳の終了まで全工程の塩基配列を設計する完全人工遺伝子のシステムをめざしています」と抱負を語る。

加藤教授は、発酵工学の出身で、微生物や単細胞の緑藻を材料に遺伝子の発現を研究していたが、1996年に本学に赴任して多細胞植物の研究を始めた。当時は植物の全ゲノム情報すらなかったが「超高速で塩基配列を解読する次世代シーケンサーとAIの登場が精度の高い数理モデルの研究の後押ししてくれた」と振り返った。

### 効率を精密に予測

一方、山崎助教は、機械学習の手法を取り入れたコンピュータ解析によって新たな予測・最適化システムを構築した。このシステムを用いれば、mRNAの配列から翻訳効率を高精度で予測することが可能なので、翻訳効率を高めるmRNAの「5'UTR」領域の塩基配列を入れ替えて遺伝子の領域につなげたときの予測に基づいて、目的の遺伝子に特化して発現力を向上させる最適の翻訳エンハンサーの塩基配列を設計できる。実際に最高レベルの効率を発揮するとみられる塩基配列を取得できた。

山崎助教は、「突然変異と自然淘汰を繰り返して進化するような手法で最適化を進める遺伝的アルゴリズムと、大規模データから自動的に塩基配列と翻訳効率の関係を学ぶ機械学習の導入により、研究にかかる時間を大幅に短縮して結果を出せます。今後、予測精度を高めるために大規模データの改良や計算手法の変更を重ねて、翻訳効率を最大化したエンハンサーなどを設計していきたい」と意欲を見せる。加えて、植物は成長や環境ストレスなど条件に対応して、巧妙に遺伝子の発現量を調節している。この調節にも塩基配列が重要であることから、塩基配列と効率の研究を通して、そうした生命現象の解明も視野に入れている。

本学の博士後期課程の出身である山崎助教は、博士研究員のときに、コンピュータ解析の研究に本格的に取り組み始めた。「自分の興味を大事にして、本当に好きになるまで頑張ることが大事です。その点、本学の奨学金などさまざまなサポート体制には感謝しています」と語る。研究のストレス解消は、最近生まれた子供と遊ぶこと。大爆笑させるのが得意だそう。

### 壁は乗り越える

新型コロナウイルス禍の影響で、生物実験が必要なバイオ分野は3密を避けるなど対策に追われているが、その中でも学生は研究に励んでいる。

博士前期課程2年の照山拓海さんのテーマは、植物細胞に遺伝子を導入し、ワクチンの抗原になるタンパク質を安定的に発現させる研究だ。「実験の無い日は自宅待機だったので、研究は少し遅れ気味ですが、就活を終えたので、あとは着実に研究に進捗するだけです。将来的に、植物細胞で新型コロナウイルスのワクチンができるような研究に結びつけばいいと思っています」と期待する。

同1年の梅田健人さんは、「学部時代にイネの有用な遺伝子を探索する研究だったので、植物細胞を扱う研究室を選びました。実験だけでなく、コンピュータ解析も学べるところがうれしい。これから、かなりの確率で、研究に行き詰まるときがあるでしょうが、そのような壁は絶対に乗り越えていきたい」と張り切っている。



▲照山拓海さん



▲梅田健人さん

▶ バイオサイエンス領域 バイオエンジニアリング研究室  
<https://bsw3.naist.jp/courses/courses116.html>