

## 網羅的翻訳状態解析に基づく外来遺伝子高発現システム

山崎将太郎・加藤 晃\*

## はじめに

組換えタンパク質の市場規模は年々拡大しており、バイオ医薬品に関してJCR Pharmaceuticalsによれば、2020年に約30兆円になると予想されている。たとえば、再生医療で用いられる幹細胞の培養時に添加される細胞増殖因子の中には、mgあたり数十万円以上と非常に高価なものもあり、再生医療製品の高コスト化の原因の一つとなっている。そのため、それらの異種生物での低コスト生産が期待されている。

このような状況の中で、植物を宿主とした組換えタンパク質生産は、その生産の容易さと、アグロバクテリウムやウイルスベクターを用いた一過的な発現システムなどに関わる技術進展により近年注目を集めている。植物分子生物学の目覚ましい発展により、遺伝子発現制御機構が明らかにされるとともに、実際にその知見を活用して植物(細胞)で効率良く目的タンパク質を発現させる試みが多数行われている<sup>1-5)</sup>。

しかし、これら開発された発現システムの多くが個別遺伝子発現に着目した発見的手法に基づいたものである。一方で近年、次世代シーケンサーなどを用いて大規模に細胞内のデータを容易に取得することが可能となり、その情報に基づいた発現システムの開発も行われるようになってきている。本稿では、遺伝子発現制御ステップの一つである「翻訳」の過程に着目し、各mRNAの翻訳状態およびその配列情報を大規模に取得し、*in silico*での数理モデルの構築を通じた導入遺伝子高発現に向けた革新的な試みを紹介する。

## 細胞内mRNAの網羅的な翻訳状態解析と高発現システムの構築

遺伝子の発現量は、さまざまな段階で制御されており、その一つがmRNAの翻訳段階である。哺乳動物細胞を用いた研究では、mRNAの翻訳効率の差が最終的なタンパク質蓄積量に非常に大きな影響を与えることが報告されている<sup>6)</sup>。植物細胞においても翻訳段階の重要性は知られている。植物細胞の中でmRNAの翻訳効率(翻訳状態)はmRNA種によって大きく異なり、成長や組織の発達によって変動し<sup>7)</sup>、高温・乾燥・栄養飢餓など

さまざまな環境ストレス条件下では、多くのmRNAからの翻訳が抑制される<sup>8)</sup>。これは、植物へ導入した外来遺伝子の翻訳についても同様である。導入遺伝子の翻訳状態は、植物の生育条件や発達状態によっては必ずしも良好とは限らず、場合によってはその翻訳状態が抑制されている可能性も考えられる。一方で、細胞内に存在するmRNA種の中には、通常条件に加えて発達した植物体や環境ストレス条件下でも常に活発に翻訳されているmRNA種も存在している。

このようなmRNAの翻訳状態制御に関わる非常に重要な要素として、5'非翻訳領域(5'UTR)があげられる。翻訳は、5'Cap構造に翻訳開始因子(eIF4複合体)が結合し、リボソーム40Sサブユニットが5'UTRにリクルートされることによって開始され、その律速はmRNAへのリボソームのリクルート過程である。そのため、足場となる5'UTRはmRNAの翻訳効率を規定する重要な要因である。植物の内在遺伝子の中で常に活発に翻訳されている(常に翻訳状態が良い)mRNAを探索し、その5'UTRを導入遺伝子発現系に活用することは、高効率な導入遺伝子発現系を開発するうえで理に適ったものといえる。

一般的にmRNAの翻訳状態は、mRNAに多数のリボソームが結合していれば翻訳が活発であり(ポリソームを形成)、リボソームが結合していなければ翻訳が行われていない(ノンポリソーム)というように、mRNAに結合するリボソームの数を指標として判断される<sup>9)</sup>。このような手法はポリソーム解析と呼ばれており、DNAマイクロアレイ解析などと組み合わせることで、mRNAの翻訳状態を網羅的に評価することが可能である。PR(Polysome Ratio)値とは、それぞれのmRNAの全mRNA量に対してのポリソーム画分に存在する量の比を数値化した値であり、各mRNAの翻訳状態を表す指標の一つである。mRNAのPR値が高い程、活発に翻訳されているものと予想される。そこで筆者らは、シロイヌナズナ幼植物体・成長した植物体・未展開葉・展開葉・培養細胞・熱処理した培養細胞などにおいて、全mRNA種の翻訳状態をポリソーム/マイクロアレイ解析により調べた。各試料における全mRNA種のPR値を算出し、すべての条件でPR値が高く、活発に翻訳されているmRNA種のランキ

\*著者紹介 奈良先端科学技術大学院大学先端科学技術研究科(准教授) E-mail: kou@bs.naist.jp

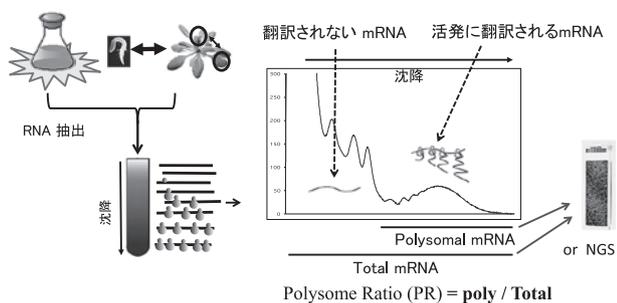


図1. 全mRNAの翻訳状態をPR値として数値化し、ランキング化

ング化を行った。その概念図を図1に示す。

その結果として、ランキング上位のmRNA種から単離した5'UTR配列(データベースの配列情報を利用)が、シロイヌナズナとタバコ植物体において、植物体の成長、発達や熱ストレスによる翻訳の抑制を回避し、恒常的で活発な翻訳に寄与することを示した<sup>10)</sup>。

一方で、遺伝子によっては複数の転写開始点が存在している。これは、5'UTRが異なる複数のmRNAのバリエーションが転写されていることを意味しており、各mRNAのバリエーション間で翻訳状態が異なることが予想される。実際、個別遺伝子を対象として、同一遺伝子由来の異なる5'UTRをレポーター遺伝子に連結し、翻訳状態を解析したところ、5倍以上の差が認められた。そこで筆者らは、ゲノムワイドに転写開始点を決定できるCap Analysis of Gene Expression (CAGE) 解析法<sup>11)</sup>とポリソーム解析を組み合わせて、次世代シーケンサーを用いて転写開始点の異なるmRNAのバリエーション単位でPR値を評価できる新たな解析手法を開発した。これを活用し、さらに高効率な発現に寄与する5'UTRの取得にも成功した<sup>12)</sup>。

### 数理モデルの構築とその活用

細胞内に存在するmRNAの翻訳状態の違いは、mRNAの配列もしくは配列に依存した構造の違いで必ず説明できるはずである。筆者らは、上述のように、全mRNAの翻訳状態データ(PR値)と全mRNAの配列データを取得し、二つのゲノムワイドデータを用いた*in silico*解析を行うことで、mRNAの配列情報からタンパク質への翻訳効率を予測できる数理モデルの構築を行った。具体的には、mRNA上の各塩基の比率、特定の配列パターン、二次構造の形成度合い、rRNAとの相補な配列長、uORFの有無、mRNA長などに関する多くの特徴を、Cap構造やAUGからの距離も考慮したさまざまな領域で評価し、それら膨大な配列的特徴の複合的な影響につ



図2. mRNAの配列情報からタンパク質への翻訳効率を予測できる数理モデルの構築

いて、mRNAの配列情報からPR値を説明できる数理モデルの構築を行った。その概念図を図2に示す。この数理モデルによって示された配列的特徴の多くは、5'UTR内およびmRNAの5'UTRから遺伝子領域(CDS)までを含む5'側の数百塩基の領域に多く存在しており、これらの領域が翻訳状態の決定に特に重要な領域であると考えられた。この結果は、従来の5'UTR配列の重要性を再確認するとともに、これまではあまり重視されていなかった5'UTRからCDSを含む幅広い領域での配列の重要性を示すものであった。

構築した数理モデルを活用することで、以下の取組みが可能になることが期待される。

### 目的遺伝子の高翻訳を可能にする5'UTR配列の選抜

CDSの配列は目的タンパク質によって異なる。構築した数理モデルによって示された、5'UTRからCDSまでの幅広い範囲での配列の重要性は、5'UTR配列とCDS配列の間で協調的に働く特徴が存在していることを意味している。つまり、目的遺伝子の配列ごとに適した5'UTR配列は異なっていることが示唆されている。筆者らが構築した数理モデルは、目的遺伝子に任意の5'UTRを連結したキメラな仮想mRNA配列についても、その翻訳状態を予測することが可能である。そこで、レポーター遺伝子(ホタルルシフェラーゼ)に約4000種の5'UTR配列を*in silico*にて連結した仮想mRNAについて、それぞれの翻訳状態を予測した。レポーター遺伝子を高翻訳できると予測された5'UTRについて、一過性発現実験によって*in vivo*での検証を行った結果、実際に、汎用的な5'UTRとして単離したものよりも翻訳活性の高い5'UTRを選抜することができた。このことから、構築した数理モデルは生体内での翻訳活性、つまり実際のタンパク質の生産効率を十分に説明することが可能であり、使用する発現カセットの生体内における翻訳能力を

予測することができる。また、目的遺伝子配列に特化した5'UTRの選抜も可能である(テラーメイド)。

#### 目的タンパク質の翻訳量を自在にコントロール

細胞内で複数のタンパク質(酵素)を発現させ、新たな代謝経路を創設する場合、必ずしもすべての酵素を高発現させる必要はなく、それぞれの酵素の単位活性量に見合う発現が求められる。構築した数理モデルにより、多数の5'UTRを目的遺伝子に連結した仮想mRNAを評価することで、発現量が異なる5'UTRを選抜することができ、代謝流束に応じた合目的な発現システムを構築できると期待される。

**数理モデルを利用した配列の準最適化** 筆者らが構築した数理モデルには、各特徴が翻訳状態に与える影響が正負の情報として含まれている。たとえば、翻訳状態に負の影響を与える特徴を限りなく少なくし、正の影響を限りなく多くすれば、これまでになく高翻訳を可能にする5'UTR配列を人工的に設計できる可能性がある(準最適化による極大化)。

#### 細胞内mRNAの網羅的な内部切断部位解析と数理モデルの構築

また筆者らは、mRNAの安定性についても網羅的な解析を行っている。これまでに、導入した遺伝子由来のmRNA蓄積量が意図に反し、きわめて低い場合がしばしば報告されている。mRNAの分解機構は、ポリA短縮、キャップ除去に依存する分解とエンドヌクレアーゼによる内部配列の切断に起因する分解に大別できる<sup>13)</sup>。一般的に、導入遺伝子が発現させる際は、両末端配列がある程度同一の発現カセット(5'UTRおよび3'UTR)を使用することが多い。導入遺伝子によって蓄積mRNA量が大きく異なる要因には、特にmRNAの内部切断に依存する分解機構が大きく関与すると考えられる。そこで筆者らは、既存の解析系の問題点を改善し、mRNAの内部切断部位を網羅的に同定できるTruncated RNA end sequencing (TREseq)法をシロイヌナズナにおいて確立した<sup>14)</sup>。この実験手法を用いて、シロイヌナズナに加え、

イネ・バラ・レタスなどで同様の解析を行い、特定の配列パターンがmRNAの内部切断に強く関与することを見いだしている<sup>15)</sup>。このmRNAの内部切断についても、mRNA配列情報から切断部位と切断のされやすさを予測可能な数理モデルの構築を行っており、最終的には内部切断が起きにくい人工配列の設計につなげていきたいと考えている。

#### 今後の展望

現在、深層学習などの画期的な手法の登場によって、さまざまな分野でAIの構築やモデリングといった試みが行われている。今後は、バイオテクノロジーの分野においても、本稿で紹介したような生命現象の数理モデル化と、実際の有用タンパク質生産といった産業への活用が期待される。

#### 謝 辞

本研究は、経済産業省プロジェクト「密閉型植物工場を活用した有用物質高発現システム基盤技術開発およびNEDOプロジェクト「遺伝子発現制御および栽培環境制御の融合による代謝化合物高生産基盤技術開発」により支援を受け実施された。

#### 文 献

- 1) Desai, P. N. *et al.*: *Biotechnol. Adv.*, **28**, 427 (2010).
- 2) Sugio, T. *et al.*: *J. Biosci. Bioeng.*, **109**, 170 (2010).
- 3) Xu, J. *et al.*: *Biotechnol. Adv.*, **29**, 278 (2011).
- 4) Matsui, T. *et al.*: *Plant Biotechnol.*, **29**, 319 (2012).
- 5) Matsui, T. *et al.*: *Plant Biotechnol.*, **31**, 191 (2014).
- 6) Schwanhäusser, B. *et al.*: *Nature*, **473**, 337 (2011).
- 7) Yamasaki, S. *et al.*: *Plant Cell Physiol.*, **56**, 2169 (2015).
- 8) Matsuura, H. *et al.*: *Plant Cell Physiol.*, **51**, 448 (2010).
- 9) Bailey-Serres, J. *et al.*: *Trends Plant Sci.*, **14**, 443 (2009).
- 10) Yamasaki, S. *et al.*: *J. Biosci. Bioeng.*, **125**, 124 (2018).
- 11) Murata, M. *et al.*: *Methods Mol. Biol.*, **1164**, 67 (2014).
- 12) Yamasaki, S. *et al.*: *Plant Biotechnol.*, **35**, 365 (2018).
- 13) Parker, R.: *Genetics*, **191**, 671 (2012).
- 14) Ueno, D. *et al.*: *J. Biosci. Bioeng.*, **125**, 723 (2018).
- 15) Ueno, D. *et al.*: *Plant Cell Physiol.*, **61**, 53 (2020).