

代謝工学におけるバイオインフォマティクスの利用

戸谷 吉博*・松田 史生

はじめに

化石資源の枯渇やさまざまな環境問題への対策として、化成品原料や燃料を生産するバイオプロセスの確立が求められている。生物工学の発展に伴い、微生物の代謝を利用した有用化合物の生産が可能になってきた。特に安価なバルクケミカル生産を実用化するには、さらなる収率および反応速度の向上が課題となっている。

代謝工学は、宿主の代謝を「合理的に」改変し、目的化合物の生産に最適化することを目指している。近年では design – build – test – learn (DBTL) のサイクルを回す開発スタイルが注目されている(図1)。すなわち、利用可能な情報を基に目的物質の生産に適した代謝経路を設計し (design)、遺伝子組換え技術によって現実の細胞に実装し (build)、その細胞を培養・評価し (test)、期待した成果が得られなければ評価に基づいて問題点を特定する (learn)。本レビューでは、design, test, learn の各パートにおけるバイオインフォマティクス技術の利活用を紹介する。

代謝経路の設計におけるバイオインフォマティクス

Design 生物の代謝は数千の化学反応が複雑に相互作用するネットワークを構成している。このため、ある反応の追加や削除がシステム全体にどのように影響を及ぼすかを直感で予測することが困難な場合も多い。代謝フラックス分布のシミュレーションは、経路全体における代謝の流れ(フラックス分布)の変化を予測するための有効なアプローチであり、フラックス分布を予測するためのシミュレーション技術が開発されている。ゲノ

ムスケールモデル (GSM) を用いたフラックスバランス解析 (FBA) 法は、酵素反応速度論を無視し、化学量論のみを利用する思い切った単純化を行うことで、経路全体のフラックス分布を高速に予測できることから、代謝経路の設計に広く用いられている¹⁾。

GSMは対象とする生物が有する化学反応を網羅的に実装したものである。たとえば、ヘキソキナーゼの反応は、 $\text{glucose} + \text{ATP} \rightarrow \text{glucose-6-phosphate} + \text{ADP} + \text{Pi}$ という化学量論式で記述される。大腸菌GSMのiMO1515は、1515遺伝子、1192代謝物と2719反応を考慮して構築されたモデルである²⁾。すでに主要な有用物質生産宿主についてはGSMが構築されている。FBAはGSMに物質収支や反応の可逆性などの制約条件を課し、増殖最大化などの目的関数を設定することで、線形計画法によりフラックス分布を予測する。この手法は栄養条件など培養環境に対する応答や、反応の欠失や導入に対するフラックス分布の変化を予測する際に威力を発揮する。たとえば、大腸菌GSMのiMO1515を使った遺伝子破壊のシミュレーションでは、16種類の異なる炭素源の最小培地を用いた実験における遺伝子破壊株の表現型を93.4% (23,617中22,077が正解) の精度で予測することに成功している²⁾。

アミノ酸や有機酸から二次代謝物まで、さまざまな化合物を目的とした代謝工学において、FBAによる成功例が報告されている。特に、細胞の増殖と目的物の生産が連動した、増殖連動型の生産に適した代謝経路を設計するときに破壊すべき遺伝子の同定を得意としている。たとえば、エタノール生産シアノバクテリアの代謝工学では、呼吸鎖のNADPH脱水素酵素を破壊することで、エタノール生産量が145%に増加したことが報告されている³⁾。多くの場合、細胞の構成部品を生産するには、目的物質を生産しなくては補酵素の物質収支が保てないように設計される。上記の例では、遺伝子破壊によりNADPHの消費が妨げられ、余剰なNADPHを消費するためエタノール合成が促進された³⁾。

FBAの予測精度を向上させるために、さまざまなモデリングの技術が用いられている。転写や翻訳の過程を考慮したモデル (ME-model)⁴⁾ や酵素量や酵素の回転数を考慮したモデル (GECKO)⁵⁾ が開発されている。

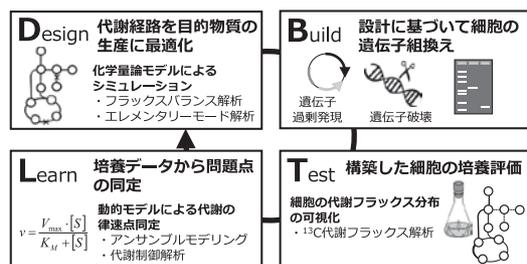


図1. DBTLサイクルにおけるバイオインフォマティクスの役割

*著者紹介 大阪大学大学院情報科学研究科バイオ情報工学専攻 (准教授) E-mail: ytoya@ist.osaka-u.ac.jp

GECKOでは、各酵素の存在量と回転数 (k_{cat}) の積が、その反応のフラックスの上限を超えないように制約を課すことで、出芽酵母のさまざまな炭素源における増殖速度の予測を実測値に近づけることに成功している⁵⁾。細胞内の転写物、タンパク質、低分子化合物の分析技術が飛躍的に発展した近年、これらの知見をモデルに統合し、より現実に即したシミュレーションを実施することが求められている。上記のGECKOはプロテオームのデータと相性がよく、これを制約条件に加えることでフラックスの実現可能範囲を狭めることができる。同様に、遺伝子発現や転写制御を考慮する手法⁶⁾、代謝物濃度から熱力学的に反応の方向を予測して制約を課す手法⁷⁾などが提案されている。

GSMは応用範囲の広さから、微生物からヒトまでさまざまな生物種を対象に開発されており⁸⁾、KEGGやBioCycなどデータベースから自動的にモデルを構築するツールも公開されている^{9,10)}。一方で、アノテーションされていない反応による経路の穴埋め、反応の補酵素特異性の選択、細胞組成、増殖・維持に必要なエネルギーの見積もりなどは自動化が難しい作業であり、実用に足るGSMの開発には地道なキュレーションが不可欠である。今後、モデル構築技術が発展し、モデルが存在しない産業微生物についても簡便にGSMが利用できるようになると、さらに応用範囲が広がると期待される。

代謝経路の設計に役立つもう一つの手法は、エレメンタリーモード解析 (EMA) である。EMAは、代謝経路を物質収支が満たされる最小の反応セット (EFM) に分割する手法であり、あらゆるフラックス分布はEFMの線形和で表現される。また、各EFMに含まれる反応を1つ以上欠失することで、そのEFMを削除することができる。EMAの長所はFBAでは適用が困難な増殖非

連動型の物質生産に適した経路も設計できる点にあり、フラックスの実現範囲 (解空間) に着目した設計方法が提案されている¹¹⁾。図2に典型的な細胞増殖と目的物質生産のフラックスの解空間を示した。EMAで求めたEFMは解空間のさまざまな箇所が存在し、解空間の外郭の端点には必ずEFMが存在する。増殖非連動型の物質生産には、増殖停止時に目的物質を必ず生産する経路が適しており、解空間の左下の領域に含まれるEFMをすべて削除するために欠失すべき反応を探索することで設計できる¹¹⁾。

EMAは、代謝フラックスの解空間の形状を指標に検討できるため、FBAよりも自由に代謝経路を設計できる。一方で、EMAは計算コストの制約から現状ではGSMのような大規模なモデルに適用することが難しく、数十反応の主要経路についての解析がほとんどである。今後の計算アルゴリズムの発展によってGSMを扱うことができるようになれば、EMAの活躍の機会が広がるであろう。

代謝状態の評価におけるバイオインフォマティクス

Test 代謝経路の設計をもとに、改変株を構築して物質生産能をテストしても、期待通りの結果が得られないこともしばしばである。改変株の代謝フラックス分布を測定できれば、FBAやEMAで予測した最適なフラックス分布との比較ができる。¹³C代謝フラックス解析は、安定同位体標識実験を行い、代謝物の標識情報から細胞内のフラックス分布を推定する技術である。詳細は過去のレビューで紹介されているので¹²⁾、ここでは要点のみを記述し、最新の技術の発展を紹介する。

¹³C代謝フラックス解析では、特定部位の炭素原子を標識した基質を用いて細胞を培養する。細胞に取り込まれた基質はさまざまな経路を通して代謝される。代謝物の標識パターンにはフラックス分布の情報が反映されているため、これを質量分析計によって測定する。この測定値をフラックスに変換するため、各反応における炭素原子の移動を記述したモデルを構築する。このモデルは任意のフラックス分布を入力すると代謝物の標識パターンを出力する。このシミュレーションで予測される代謝物の標識パターンが実測値に近づくようにフラックス分布を最適化することで実際の細胞のフラックス分布を推定することができる。

¹³C代謝フラックス解析の推定結果は、実測データを説明するフラックス分布であり、実測値と推定値の代謝物の標識パターンの適合度は、カイ二乗検定によって統計的に評価できる。しかし、これだけでは推定されたフ

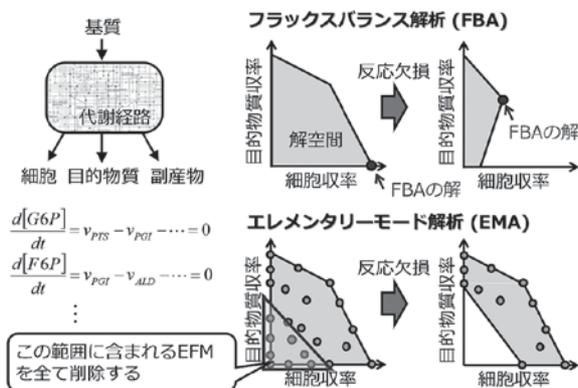


図2. フラックスバランス解析とエレメンタリーモード解析による代謝設計。代謝フラックスの解空間の点の一つひとつがEFMを表す。

ラックス分布がどの程度妥当かを議論するには不十分である。極論を言うと、標識されていない炭素源を基質に用いた場合、どのようなフラックス分布でもカイ二乗検定を通過する。正しく評価するためにはフラックスの95%信頼区間を計算することが必要である。信頼区間は実測値と推定値の代謝物の標識パターンの適合度がカイ二乗検定を通過するフラックスの範囲である。先の非標識基質に用いた場合は、物質収支を満たすフラックスの範囲がすべて信頼区間となる。フラックスの信頼区間の幅は、用いる基質の標識パターンに強く依存する。そのためフラックスの信頼区間の幅が狭くなるように、実験前にシミュレーションにより基質の標識パターンを最適化する技術が発達している^{13,14}。グルコース炭素源とする大腸菌の中核代謝において、さまざまなグルコースの標識パターンについて、フラックスの推定精度が評価されている¹⁴。1つの標識パターンでフラックス分布全体を精度よく推定することは難しいこともわかっており、同一の培養条件について、異なる種類の標識グルコースを用いて並列に培養することで得られる情報を増やし、両方の実験結果を使ってフラックスを推定するパラレルラベリング技術も開発されている¹³。

¹³C代謝フラックス解析の対象とする経路は、中核代謝に限られることが多かったが、ゲノムスケールな代謝ネットワークに適用されることも可能になってきた¹⁵。また、複数種細胞による共培養条件への応用など¹⁶、その適応範囲はますます広がっている。

さらなる改変箇所を予測するための技術

Learn バイオインフォマティクス技術の支援をもっとも必要としているのは、育種した細胞が期待通りの性能を発揮しない原因を探るlearnのパートである。たとえば、¹³C代謝フラックス解析の結果とFBAによる理想状態の比較は、どの経路のフラックスを強化・抑制すればよいかの指針を与える。これをより合理的に行う方法がOptForceというアルゴリズムであり、親株のフラックスに対して、目標のターゲット生産を強いた際のフラックスのとり得る範囲を線形計画法で計算し、その範囲の重なりから強化・抑制すべき反応を予測する。Ranganathanらは、¹³Cフラックス解析の結果を利用し、大腸菌による脂肪酸の生産を増加させるための改変箇所を探索し、*fabZ*の過剰発現と*fadD*の破壊により1gのグルコースあたり0.14gの脂肪酸という高い収率でのグルコースからの脂肪酸生産に成功している¹⁷。

有用物質生産においては、収率だけでなく、生産速度の向上も重要な課題である。代謝経路全体の速度の向上

には律速段階、すなわち経路の中で一番遅い反応を特定し、この反応速度を強化することが必要である。FBAでは化学量論式と代謝フラックスのみを考慮していたが、代謝経路の律速段階を解析するには、これに加えて、酵素反応速度式、代謝物濃度、酵素発現量を加味した動的代謝モデルが必要となる。動的モデルが利用できると、代謝制御解析などの強力なアプローチをとることができる。フラックスコントロール係数(FCC)は、各フラックスを微細増加させた際に、目的物質の合成フラックスがどれだけ変化するかを表す指標であり、FCCが大きい反応が律速段階である。Chassagnoleらは大腸菌のスレオニン合成経路について動的モデルを構築し、FCCから律速反応を予測している¹⁸。また、Andreozziらは大腸菌の中核代謝経路の動的モデルを解析して1,4ブタンジオール生産を向上するために強化・抑制すべき反応を予測している¹⁹。このように動的モデルは代謝工学のツールとして大きな可能性を有しているが、その課題はモデル構築に多大な労力を要することである。

大腸菌などの広く研究されている宿主生物については、多数のアロステリック制御や転写因子による遺伝子発現制御を記述することで、外部環境の変化に対する応答を表現可能なモデルが開発されているが^{20,21}、このような生物種は限られている。動的モデルの構築には、個々の酵素反応の反応機構に加えて、ミカエリスメンテン定数(K_m)や k_{cat} などの多数の速度論パラメータを収集する必要がある。BRENDAは K_m や k_{cat} などのパラメータが数多く記載されているデータベースであり²²、動的モデルを構築する際には便利だが、同じパラメータについても多数の値が存在する。たとえば、大腸菌のホスホフルクトキナーゼのATPに対する K_m を調べると、0.008から0.21 mMまでさまざまな値が掲載されている。値の違いは元文献の実験条件の差に起因すると考えられるが、どの値を採用すればよいかは難しい課題である。また、反対に生物種によってはパラメータの値が1つも登録されていない場合も多い。近年、HeckmannらはBRENDAなどのデータベースに登録されている酵素の生化学データ、分析条件、タンパク質構造、代謝ネットワークの情報を入力とした機械学習によって、*in vivo*の k_{cat} の値を予測できることを示している²³。これまでに取得された豊富な実験データを活用できる今、このようなアプローチも重要になってくると考えられる。

近年、注目されている動的シミュレーションの方法にアンサンブルモデリングがある。従来のモデリングでは、*in vitro*で計測された反応速度論定数をベースに、細胞内の代謝挙動を再現できるようにパラメータをチューニン

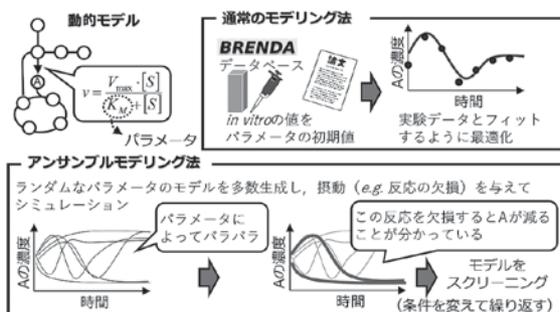


図3. 従来モデリングとアンサンブルモデリングの違い

グし、さまざまな状況を予測することが可能な「1つのモデル」をブラッシュアップして構築するアプローチが多くなされてきた。アンサンブルモデリングでは、リファレンスの状態を説明する反応速度論定数のセットをランダムに生成し、多数のモデルを用意する。このモデルの集団に対して表現型が既知の別条件（たとえば、ある酵素の欠失や過剰発現など）を適用し、シミュレーションを実行する。多くのモデルは現実の細胞の表現型とは異なる挙動を示すが、初期のモデル集団が十分に大きい場合には、表現型と近い挙動を示すモデルも存在する。これらのモデルを抽出し、さらに別の条件におけるスクリーニングを繰り返すことで、より優秀なモデルを選抜することができる。最後まで残った「複数のモデル」について、代謝制御解析などを行うことで、律速段階の予測や目的物質の生合成経路のフラックスを増加させるための改変戦略を提案することができる。このアプローチは、反応速度論定数の実測値がなくてもモデルが構築できることと、選抜された複数のモデルを利用することができるため、よりロバストな解析ができることが特徴である(図3)。試験管内と細胞内とでは、分子の混雑度合いなど環境が大きく異なるため、*in vitro*で測定したパラメータを*in vivo*のモデルに使用してよいのかという問題に対しても、答えることができる優れた手法である。

Rizkらは大腸菌の芳香族アミノ酸生産について、いくつかの酵素の過剰発現時や欠失によって目的物質の生産速度が増加・減少したかという知見をスクリーニングに用い、1500個のモデルから195個のモデルを選抜した²⁴⁾。これらのモデルアンサンブルのシミュレーションから、中枢代謝経路の律速段階を明らかにしている²⁴⁾。同様のアプローチがシアノバクテリアでも試みられ、エタノール生産能向上に有効な改変点の同定に成功した²⁵⁾。モデルのスクリーニングに用いる表現型は、目的物質の

生産速度の増加・減少のようなデータに限らない。メタボロームやプロテオーム、代謝フラックスなどさまざまなデータを利用することができるため、これらオミックスデータとの相性も抜群である。そのため、マルチオミックスデータの統合利用という観点からも今後ますます発展が期待されるアプローチであろう。

おわりに

本稿では、代謝工学におけるDBTLサイクルのフレームワークに沿って、バイオインフォマティクスがどのように活用されているか紹介した。ゲノム編集技術の発展により、buildの開発速度が加速している昨今、構築された細胞の性能をさらなる改善するため有効な戦略を立案することが課題となっている。モデルを用いた情報学的手法は、データを統合・解釈するための効果的な技術であり、今後益々の利活用が期待されている。

文献

- 1) Orth, J. D. *et al.*: *Nat. Biotechnol.*, **28**, 245 (2010).
- 2) Monk, J. M. *et al.*: *Nat. Biotechnol.*, **35**, 904 (2017).
- 3) Yoshikawa, K. *et al.*: *Bioprocess Biosyst. Eng.*, **40**, 791 (2017).
- 4) O'Brien, E. J. *et al.*: *Mol. Syst. Biol.*, **9**, 693 (2013).
- 5) Sánchez, B. J. *et al.*: *Mol. Syst. Biol.*, **13**, 935 (2017).
- 6) Machado, D. and Herrgård, M.: *PLoS Comput. Biol.*, **10**, e1003580 (2014).
- 7) Salvy, P. *et al.*: *Bioinformatics*, **35**, 167 (2019).
- 8) King, Z. A. *et al.*: *Nucleic Acids Res.*, **44**, D515 (2016).
- 9) Karlsen, E. *et al.*: *BMC Bioinformatics*, **19**, 467 (2018).
- 10) Latendresse, M. *et al.*: *Bioinformatics*, **28**, 388 (2012).
- 11) Toya, Y. *et al.*: *Biotechnol. Bioeng.*, **112**, 759 (2015).
- 12) 清水 浩ら: 化学と生物, **53**, 455 (2015).
- 13) Crown, S. B. *et al.*: *Metab. Eng.*, **28**, 151 (2015).
- 14) Maeda, K. *et al.*: *Metab. Eng. Commun.*, **3**, 187 (2016).
- 15) Gopalakrishnan, S. and Maranas, C. D.: *Metab. Eng.*, **32**, 12 (2015).
- 16) Gebreselassie, N. A. and Antoniewicz, M. R.: *Metab. Eng.*, **31**, 132 (2015).
- 17) Ranganathan, S. *et al.*: *Metab. Eng.*, **14**, 687 (2012).
- 18) Chassagnole, C. *et al.*: *Biochem. J.*, **356**, 433 (2001).
- 19) Andreozzi, S. *et al.*: *Metab. Eng.*, **36**, 148 (2016).
- 20) Usuda, Y. *et al.*: *J. Biotechnol.*, **147**, 17 (2010).
- 21) Matsuoka, Y. and Shimizu, K.: *J. Biotechnol.*, **168**, 155 (2013).
- 22) Placzek, S., *et al.*: *Nucleic Acids Res.*, **45**, D380 (2017).
- 23) Heckmann, D. *et al.*: *Nat. Commun.*, **9**, 5252 (2018).
- 24) Rizk, M. L. and Liao, J. C.: *PLoS One*, **4**, e6903 (2009).
- 25) Nishiguchi, H. *et al.*: *Metab. Eng.*, **52**, 273 (2019).