

数理モデルに触れてみよう

高橋 弘喜

1995年のインフルエンザウイルスのゲノム解読以来、ゲノム科学の進歩は凄まじい。近年では、次世代シーケンサーという超ハイスループット技術の登場によって生物学は大きな変換点を迎えている。米国ではカリフォルニア大学を中心に、食物関連の病原菌10万株ものゲノム決定を目指した「100K Foodborne Pathogen Genome Project」が発表されるなど、非モデル生物の膨大なゲノム情報が容易に入手できる時代が到来している。さらに次世代シーケンサーを用いれば、ゲノム情報がない生物種での遺伝子発現解析が可能であり、Meyerらは²⁾、ゲノム情報の乏しい珊瑚において遺伝子カタログの作成に成功している。

この大量ゲノム情報をもたらすものとしては、生物が保有する全遺伝子セットが明らかにできることが挙げられる。つまり、生命の構成要素である遺伝子のカタログを手に入れることができる。機能未知遺伝子が大半を占めるとしても、各生物種の大体の基本構成に関しては整備できたと言ってもいいかもしれない。

では一方で、大容量データに基づいて構成要素をすべて列挙できた先に、どうやって生命システムを理解するのか、という問いには、さまざまなアプローチがあり、どれも必要だと思うが、ここでは数理モデル研究に言及したい。数理モデルを用いて研究対象を表現する目的としては、大きく二つ挙げられる³⁾。

- **知識の溝を理解する**：対象となる生命システムを数理モデルで表現し、実データとの整合性を検証することで、我々の理解がどの程度なのかを明らかにできる。
- **挙動を予測する**：構築した数理モデルを元にして、システムの挙動を予測できる。特に、実現困難な条件下で対象の挙動が予測できることは数理モデルならではの

特徴である。

数理モデル化において使用されるのは、微分方程式である。本稿では、「大腸菌でのタンパク質Aの合成量を測定した実験」(図1)を例にして数理モデル化に触れてみたい。数理モデル化においては、なるべくシンプルなモデルが良いとされている。タンパク質Aの細胞内での合成速度が一定だと仮定すると、次式が得られる。

$$\frac{dA}{dt} = k - \gamma A$$

第一項の k はタンパク質の合成速度、第二項は大腸菌の増殖によってタンパク質Aが希釈・分解されることを考慮した項になる。この微分方程式の解は、

$$A(t) = \frac{k}{\gamma} + (P_0 - \frac{k}{\gamma}) \exp(-\gamma t)$$

となる。 P_0 は時間 $t=0$ の時のタンパク質Aの量である。解法については、微分方程式に関する入門書やWebを参照していただくとして、これを図示してみると図のようになる($k=100$, $\gamma=0.5$, $P_0=540$)。ただこの場合、手元にある実験データに明らかにフィットしていない。では、実験データに合う最適なパラメーター値は存在するか？

実験データがあれば、パラメーターのもっともらしい値を統計的に推論することが多くの場合では可能である。その手法の一つとして、マルコフ連鎖モンテカルロ法であるメトロポリス・ヘイスティングス法が挙げられる。この手法は、非常に汎用性が高くパラメーター推定法としてあらゆる分野で使用されている。端的にいうと、最初に k , γ , P_0 を適当な値で始めて、値を少しずつ変化させていった時に、尤度、つまり実験データへの適合度を指標にして、もっともらしいパラメーターの分布を推定しようというものである⁴⁾。そうやって推定した k , γ , P_0 の値を用いて図示すると、手元にある実験データにフィットする数理モデルを構築できたことになる。したがって、タンパク質の合成速度 $k=300$ 、細胞分裂の速度 $\gamma=0.14$ と見積もれたことになる。

以上簡単に紹介したが、実験データがあれば信頼性の高い数理モデルが構築でき、挙動解析の幅が大いに広がることが期待できる。

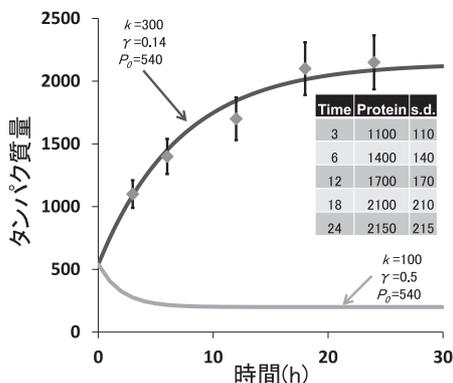


図1. 大腸菌でのタンパク質Aの合成量を測定した実験

- 1) <http://100kgenome.vetmed.ucdavis.edu/>
- 2) Meyer, E. et al: *BMC Genomics*, **10**, 219 (2009).
- 3) Wilkinson, D. J.: *Stochastic Modelling for Systems Biology*, CRC Press (2011).
- 4) 久保拓弥：データ解析のための統計モデリング入門, 岩波書店 (2012).