総説

システムバイオロジーにおける
シミュレーション技術と実装

舟橋 啓, 広井 賀子

要約: 生命現象の理解には個別のコンポーネントについての理解に加え, それらの統合されたふるまいを知る必要がある。このような様々な分子のダイナミクスの背景にあるメカニズムを理解する上で強力なツールとなるのが数理モデルである。システムバイオロジーでは, 数理モデルを用いて生命現象を司るメカニズムを理解する手法を頻用する。歴史的には解析的に解くことが可能なモデルが用いられたが, コンピュータの発達により, 常微分方程式モデル, 偏微分方程式モデル, 確率モデル, これらのハイブリッドモデルなど多岐に渡る形式で記述されたモデルを数値計算によって解くようになった。こうした背景から, 作成したモデルの正しい挙動を数値的に得るための, 正確な数値計算技術が不可欠の要素となっている。本稿では, 各モデル化手法の特徴を概説した後, システムバイオロジー研究において広く利用されている代表的なアプリケーション (シミュレータ) について, ソフトウェアの全体像から実装されているアルゴリズムまで, 原典を挙げて解説する。数値計算手法の実装内容を正確に知ることは, ユーザにとってはモデルが備えるべき内容を理解し, 結果を正しく解釈するために有用となる。また開発者にとっては, 不必要な重複を避け, 新しいブレークスルーを実現するための技術が何であるか, 自らの創造性の発端を得るためにより重要な知識となるだろう。

1. はじめに

近年, 生命現象を司る動作メカニズムの理解を目標としたシステム生物学的アプローチによる研究が隆盛を誇っている。生命現象を司る動作メカニズムの理解には, 対象とする生命現象にかかわる分子の特定, およびそのダイナミクスを理解する必要がある。一般的

にダイナミクスは遺伝子制御, シグナル伝達, 代謝などの分子間反応, 細胞内分子の移流・拡散として表現され, ダイナミクスの特性の記述には反応 (移流・拡散) 方程式が用いられる。これら反応方程式のシミュレーションは生命現象のシステム的理解に必要不可欠であり, 多様なシミュレーションアルゴリズムに対応した細胞シミュレータ, および解析環境を提供することがシステムバイオロジーの発展において重要である。本稿では, システムバイオロジー研究において用いられているシミュレーション技術とその実装について, 既存のアプリケーションを例示しながら解説する。

上述した通り, システムレベルのダイナミクスを理解する際には分子動態を数理モデルとして記述することが一般的である。分子動態は反応方程式として記述され, それらは①分子濃度の時間変化を記述した常微分方程式 (図 1), ②分子濃度の時空間的变化を記述した偏微分方程式, ③分子数の変化を確率的な変動として記述した確率モデル, に大別される。実際には適用したい生命現象に依存してこれら異なる記述方法を採用することになる。1990 年以降, 数理モデルを用いた細胞シミュレーションに関する研究事例が多数報告されてきており, シミュレーションと実験の両輪による生命現象のシステムレベルでの理解は 21 世紀の生物学として脚光を浴びてきているが, その歴史は意外と古くに遡ることができる。

1943 年に Britton Chance は実験により計測された酵素反応における基質-酵素間の関係と, 実測されたパラメータを用いたミカエリス-メンテン式 (4 分子, 3 反応) のシミュレーション結果との比較を行った (1)。Chance によるこの研究は, 反応素過程の組み合わせにより記述された数理モデルが実際の生化学反応を正しく表現できることの可能性, そしてシミュレーショ

キーワード: システムバイオロジー, シミュレーション, 数値計算, 生化学ネットワーク, 数理モデル
慶應義塾大学 理工学部 生命情報学科 (〒223-8522 神奈川県横浜市港北区日吉 3-14-1)

E-mail: funa@bio.keio.ac.jp 原稿受領日: 2015 年 10 月 29 日, 依頼原稿

Title: Simulation technology and its application in Systems Biology Author: Akira Funahashi, Noriko Hiroi

$$\frac{d[G]}{dt} = V_{in} - k[G][ATP]$$

$$\frac{d[ATP]}{dt} = 2k[G][ATP] - \frac{V_m[ATP]}{K_m + [ATP]}$$

図1 酵母の解糖系モデルの生化学ネットワーク記述と常微分方程式モデル

ンの有用性を示すことに成功した。当時は電子計算機（今で言うコンピュータ）が一般の研究施設で使用可能な状況ではなかったため、Chance はアナログ・機械式の微分解析機を用いてシミュレーションを行った。1940 年代初頭は細胞シミュレーションの夜明けであったとともに、計算機科学も飛躍的に発展した時代であった。世界初のプログラム可能な電子計算機である ENIAC(2) の開発が成功し、研究者らは大型計算機を用いて大規模な生化学反応のシミュレーションを行うことが可能となった。1960 年には Chance らは世界で初めて電子計算機上で動作する数理モデルを発表した(3)。この数理モデルは腹水がん細胞における糖代謝と呼吸の関係を表したものであり、49 種の化学物質と 22 種の生化学反応によって構成されていた。当時としては大規模な数理モデルであり、商用のスーパーコンピュータである UNIVAC I を用いたシミュレーションに 500 時間以上かかったと記されている。

1995 年には Venter のグループにより *Mycoplasma genitalium* の 525 個の全ゲノムが解読され(4, 5)、細胞内のダイナミクスを構成するすべての要素（遺伝子）の同定が可能だということが示された。一方で計算機の演算性能も飛躍的に向上し、現在では Chance らが 1960 年にシミュレーションを実行した UNIVAC I と比較して 1 兆倍の演算性能を持つスーパーコンピュータが開発されている。顕微鏡技術やシーケンス技術などの測定技術の向上および計算機科学の進歩により、精密な数理モデルを構築しそのふるまいをコンピュータ上で予測することが可能となった現在、多くの研究者がシステムレベルでの理解を目指し、シミュレーションに興味を持つのは当然の流れだと考えられる。次章

では近年注目を集めている細胞シミュレーション技術とその実装例（アプリケーション）について述べる。

2. 数理モデルとシミュレータ

第 1 章で述べた通り、細胞シミュレーションを行う上で分子動態を数理モデルとして記述することが一般的である。分子動態は反応方程式として記述され、常微分方程式モデル、確率モデル、偏微分方程式モデルに大別される。現在、様々な細胞シミュレーション手法が提案されており、細胞シミュレーションを行う際にどのようなアプローチで数理モデルを構築すべきか、その選択は非常に重要なプロセスとなる。一般的に数理モデルはある現象を抽象化したものであることが多く、それゆえ、研究者自身がどのような現象に注目しているのか、またどのような仮説を立てているかが選択において重要となる。以降、これら数理モデルとそのシミュレーション技術について順に紹介する。

1) 常微分方程式モデル

常微分方程式モデルは分子濃度の時間変化を記述する際に用いられ、そのモデルの性質上、分子の空間的局在については考慮していない。例えば、細胞質内に存在するタンパク質 A は細胞質内では均一に分布していることを仮定している際に用いられる。これは注目している生化学反応の反応速度はその反応機構によって支配され、各分子の拡散は反応速度に影響を及ぼさないことを意味している（反応律速）。反応律速の系では分子の拡散は反応と比較して十分に速いことを仮定している。逆に言えば、多数の分子が狭い空間内に存在するような混雑した環境では分子の拡散が遅くなり、反応律速であるための条件が満たされなくなるので注意が必要である。

常微分方程式モデルは生化学反応の数理モデルでも古くから採用されており、Chance による 1943 年の論文(1) や Sel'kov による細胞周期モデルの論文(6) でも常微分方程式モデルが採用されている。また、常微分方程式モデルはモデルの解析手法が他の手法と比較して多数提案されているため数理モデルで採用されることが多い。Kummer のグループはここ 10 年間に刊行されたシステムバイオロジー研究に関する論文内で使用されていた数理モデルについて調査し、常微分方程式モデルが最も多く利用されていると報告している(7)。常微分方程式を用いたシミュレーションでは細胞内の各分子反応を生化学反応方程式の形で表し、微分方程式から各分子反応の濃度変化を一つずつ計算し、数値積分を行う。そのため常微分方程式モデルのシミュレータでは採用されている数値積分のアルゴリズム

ムがそのシミュレータの精度を評価する上で重要な要素となる。以降、常微分方程式モデルのシミュレーションに対応したアプリケーションを紹介する。

(1) COPASI

COPASI (COmplex PATHway SIimulator) はバージニア工科大学、ハイデルベルク大学、マンチェスター大学の研究グループによって開発が進められているアプリケーションである(8)。COPASIでは生化学反応ネットワークを常微分方程式で記述した数理モデルのシミュレーションおよびダイナミクスの解析を行うことができる。COPASIは常微分方程式モデルだけでなく、後述する確率モデル、およびそれらのハイブリッドモデルのシミュレーションが可能となっている(図2)。COPASIでの数理モデル構築は直感的な操作により容易に行うことができ、表計算ソフトウェアでのセルを埋めていくような操作で細胞内の分子、分子間反応を記述していく。数値積分は高速かつ安定であると知られているLSODA(9)をライブラリとして採用している。構築した数理モデルのシミュレーションを行うことが可能であるだけでなく、時系列の実験データを与えることでモデル中のパラメータ値の推定を行うことも可能である。数理モデル構築の際に障壁となるのがモデルに含まれるパラメータ値の推定であり、COPASIは複数のパラメータ推定アルゴリズムを実装しており、ユーザは異なるアルゴリズムを選択することができる。

(2) CellDesigner

CellDesignerはSBI(システムバイオロジー研究機構)と慶應義塾大学の研究グループによって開発が進められている生化学ネットワーク解析環境である(10)。CellDesignerの最大の特徴は数理モデルを直感的な

GUI操作により生化学反応ネットワークとして構築できる点にある。CellDesignerでは分子反応をネットワークのノード(頂点)とエッジ(辺)として表現し、各ノードには分子名、分子情報(データベースのIDなど)を、各エッジには反応速度方程式を記述することで数理モデルを構築していく(図3)。また、既存の生物学データベースと連動することにより、例えば文献を指定することでBioModels.net(11)から数理モデルのインポートが可能であり、SABIO-RK(12)に問い合わせることで数理モデル中のパラメータ値を文献情報から決定することができる(13)。CellDesignerは3種のシミュレーションエンジン(8, 14, 15)を内包している。これらを用いて多様な数値計算アルゴリズムを選択することができ、常微分方程式に限らず、遅延微分方程式、代数微分方程式などで記述された数理モデルにも対応している。上述したCOPASIもシミュレーションエンジンとして採用されている。

2) 確率モデル

上述したように、システムバイオロジー研究では分子濃度の時間変化を記述した常微分方程式モデルが多く提案されている。一方で、一細胞レベルでの生化学反応は本質的に確率的なふるまいを示すことが確認されている(16)。確率的なふるまいを知るためには、分子動態が決定論的に変化する微分方程式モデルではなく、分子数の変化を確率的な変動として記述した確率モデルを用いる必要がある。確率モデルによるシミュレーションは一細胞レベルでより現実的な分子動態を知ることができるが、一方で計算コストが高く、実験データとのフィッティングが困難であるといった問題がある(17)。

1977年にGillespieにより提案された確率モデルシ

図2 COPASI上で構築した数理モデルとシミュレーション結果

図3 CellDesigner でのモデルの構築およびシミュレーション結果

ミュレーションアルゴリズム (SSA: Stochastic Simulation Algorithm) (18) は計算時間の短縮と実装の容易さを両立させているため、確率モデルシミュレーションで広く利用されている。確率モデルシミュレーションは同じモデルに対して異なる結果を返すため、必然的にシミュレーションの繰り返し回数が増加し計算コストの増加を招く。近年の確率モデルシミュレータではシミュレーションを並列実行することによって計算時間の削減を図っている。特に GPU を用いた並列化による速度向上は効果が高く、100 倍を超える性能向上を達成したという報告もある (19, 20)。以降、確率モデルシミュレーションに対応したアプリケーションを紹介する。

(1) StochKit

StochKit はカリフォルニア大学サンタバーバラ校の研究チームによって開発が進められている確率モデルシミュレータである (21, 22)。StochKit は上述した Gillespie による SSA 以外に、計算効率の向上を行ったアルゴリズムや大規模モデルに対応したアルゴリズム (23, 24) の実装も行われており、StochKit がモデルを読み込んだ際にモデルの解析を行い、最適なアルゴリズムを選択する実装となっている。また、計算効率の向上だけでなく、時間の刻み幅を自動的に調整する (tau-leaping) アルゴリズム (25, 26) の実装も行われている。StochKit の最大の長所は、SSA の改良アルゴリズムの提案と実装が同研究チームによって進められている点である。そのため、最新のアルゴリズムがいち早く実装されることが多く、かつユーザが意識せずに最適なアルゴリズムが選択される工夫もあり、初学者にも使いやすいアプリケーションとなっている。

(2) Smoldyn

Smoldyn はフレッド・ハッチンソン癌研究センターの研究チームによって開発が進められている確率モデルシミュレータである (27)。Smoldyn の最大の特徴は粒子モデルをベースとしているため、確率モデルでありながら空間モデルのシミュレーションが実行できる点にある。モデル中の分子は粒子として定義され、1~3 次元空間をブラウン運動により拡散する。反応は拡散する分子同士、もしくは分子と細胞膜のような物体の表面が衝突した際に生起されるとしている。Smoldyn は一細胞レベル (nm スケールの空間解像度) の生化学反応シミュレーションに用いられることが多く、それゆえ生物物理学や細胞生物学の分野にて利用されている。Smoldyn は本稿で紹介している他のアプリケーションのような GUI (グラフィカルユーザーインターフェース) は搭載されていないため、モデルは Smoldyn の設定ファイルとしてユーザが手で編集する必要があるが、モデル記述において自由度が高い点、また他の粒子ベースのシミュレータと比較して高速に動作する点などから空間モデルの確率モデルシミュレーションにおいて多くの利用者を獲得している。

3) 偏微分方程式モデル

上述した Smoldyn 以外のシミュレータでは常微分方程式モデル、確率モデルのシミュレーションにのみ対応しているため、反応区画の形状や領域内での分子濃度の偏りの影響は考慮していない。しかし細胞内の生化学反応は、膜、細胞小器官、細胞質等に空間的に分布している物質が相互作用することによってもたらされる。そのため、細胞内の生化学反応の過程を実際の細胞内における局在も考慮した上で解析するためには、空間情報が組み込まれた偏微分方程式モデルを構

築する必要がある。

偏微分方程式モデルが他のモデル化手法と異なる点は、偏微分方程式モデルは分子濃度の時間および空間に関するダイナミクス（時空間的变化）を記述できる点にある。一個体単位から一細胞単位といった幅広いスケール、そして個体の成長や発生、細胞内シグナル伝達といった様々な段階で反応区画の形状や分子濃度の偏りといった空間情報はダイナミクスに大きな影響を与えることがある(28)。例えば神経細胞が特徴的な形状を有しているように、細胞内の反応区画の形状が多様であることや、細胞内における濃度に偏りがあるタンパク質が存在することが知られており、そのために生じる反応区画各所での反応の進行の違いや、移流・拡散がダイナミクスに重要な影響を与えている場合がある。このような理由から空間モデルの需要は高まっており、今後システムバイオロジーにおいて生命現象をモデル化して解析していく上で、空間モデルはより重要度を増すと考えられる。

(1) VCell

VCell (Virtual Cell) はコネチカット大学医学センターの研究チームによって開発が進められている空間モデルシミュレータである(29)。VCell は GUI による生化学ネットワークモデルの作成、常微分方程式、偏微分方程式、確率モデルの数値シミュレーションを行うことが可能なシミュレータである(図4)。作成されたモデルの保存、および数値シミュレーションはローカル環境もしくは外部のサーバで行うことが可能であり、外部サーバを用いる場合はモデルを公開しユーザー間で共有することも可能である。VCell の常微分方程式シミュレータは CellDesigner と同様に CVODE(30)を採用しており、偏微分方程式には陰的有限体積法の数値積分ライブラリが使用されているため、安定した数値計算の精度が期待できる。

VCell が開発された当初は、最大の特徴である偏微分方程式モデルのシミュレータとして利用される研究事例が多かったが、近年の実装では後述するハイブリッドモデルへの対応が進められている。例えば、VCell では上述した確率モデルシミュレータである Smoldyn が組み込まれており、常微分方程式、偏微分方程式、確率モデルの3種のシミュレーションが実行可能となっている。

4) ハイブリッドモデル

COPASI や VCell が現在提供する機能からもわかるように、近年のシミュレータは単一の数理モデルのシミュレーションを行うだけでなく、複数のモデル化手法を組み合わせた「ハイブリッドモデル」のシミュ

図4 VCell での空間モデルシミュレーション結果

レーション環境を提供する実装に向かっている。この理由は、細胞内の分子動態から多細胞で構成される臓器、個体レベルでの動態を統合的に取り扱うアプローチの研究プロジェクトが多数立ち上がってきているからだと考えられる。このようなハイブリッドモデルのシミュレーションは単一の数理モデルの数値計算と比較して計算時間が膨大となる、数値計算の複雑さが増加するといった問題点があるため、単純に複数種類の数理モデルを同時に数値計算するだけではない工夫が必要となる。

上記問題点を解決する方法として、ハイブリッドモデル内に含まれる異なるモデル間の結合に注目し、モデルの記述方法と数値計算の設計をハイブリッドモデルに特化した設計で実装する手法が提案されている。例えば沖縄科学技術大学院大学の研究グループによって開発が進められている PhysioDesigner では数理モデルをモジュールの集合として取り扱い、複数のモジュールの集合を上位のモジュールとして扱うことで数理モデルの抽象化、階層化を行っており、設計段階からハイブリッドモデルのシミュレーションを想定することで、常微分方程式モデル、偏微分方程式モデルに対応している(31)。PhysioDesigner では PHML と呼ばれる独自のモデル記述言語を採用し、上述した COPASI, CellDesigner, StochKit, Smoldyn, VCell が対応しているモデル記述言語である SBML(32) を PHML 内の一つのモジュールとして読み込むことが可能となっている。

また、ドレスデン工科大学の研究グループによって開発が進められている Morpheus もマルチスケールなシミュレーション環境の提供を目的としており、その実装は PhysioDesigner と同様のアプローチをとって

いる(33). Morpheus は常微分方程式モデル, 確率モデル, 偏微分方程式で記述された空間モデルの数値計算に対応している他, CPM (Cellular Potts Model) やセルオートマトンのシミュレーションに対応することで細胞運動, 細胞接着などの現象をハイブリッドモデルとして記述することが可能である.

PhysioDesigner や Morpheus が多細胞のシミュレーションを志向した設計になっている一方で, ハイブリッドモデルによる一細胞シミュレーションを行うプロジェクトも進行している. 2012 年に Markus Covert の研究グループが世界で初めて *Mycoplasma genitalium* の全遺伝子を含んだ一細胞モデルとそのシミュレーション結果について報告した(5). Covert らのモデルも複数のモジュールに分割され, 微分方程式, 確率モデル, 流速均衡解析が組み合わさったハイブリッドモデルとして構成されている. 国内では理化学研究所と慶應義塾大学の研究グループによる細胞シミュレーションプロジェクトである E-Cell プロジェクトが進行している(34). E-Cell プロジェクトは世界で初めて自活する細胞のシミュレーションに成功している(35). E-Cell プロジェクトは一分子レベルでの分子動態から細胞内の様々な現象をシミュレーションすることを目的としている.

3. おわりに

本稿では, システムバイオロジー研究にて用いられているシミュレーション技術とその実装例について概説した. アプリケーションについては当該分野にて開発の歴史が長く, 多くのユーザを獲得しているものを中心にピックアップして紹介した. 開発期間が長く, 利用者数が多いアプリケーションはそれだけユーザからのフィードバックが多く, 数値計算のミスなどの不具合が取り除かれた, 良い意味で「枯れた」アプリケーションであることが多い. 特に数値計算アルゴリズムのバグ(プログラム上の誤り)はユーザには気づきにくいものが多く, 研究目的で利用する際には注意

が必要である. シミュレータの計算精度は数値計算アルゴリズムに強く依存しており, どのライブラリを使用しているか, どのような数値計算アルゴリズムが実装されているかを知っておくことも合わせて重要である. 後半部で紹介したハイブリッドモデルのシミュレーションは今後のシステムバイオロジー研究において重要な役割を果たすことが期待されており, 今後の進捗に注目したい.

著者の利益相反: 開示すべき利益相反はない.

文 献

- 1) Chance B. J Biol Chem. 1943;151:553-577.
- 2) Goldstine HH, et al. Mathematical Tables and Other Aids to Computation. 1946;2:97-110.
- 3) Chance B, et al. J Biol Chem. 1960;235:2426-2439.
- 4) Fraser CM, et al. Science. 1995;270:397-403.
- 5) Karr JR, et al. Cell. 2012;150:389-401.
- 6) Sel'kov EE. Biofizika. 1970;15:1065-1073.
- 7) Huber K, et al. FEBS J. 2011;278:2767-2857.
- 8) Hoops S, et al. Bioinformatics. 2006;22:3067-3074.
- 9) Petzold L. SIAM J Sci Stat Comput. 1983;4:136-148.
- 10) Funahashi A, et al. Biosilico. 2003;1:159-162.
- 11) Chelliah V, et al. Nucleic Acids Res. 2015;43:D542-D548.
- 12) Wittig U, et al. Nucleic Acids Res. 2012;40:D790-D796.
- 13) Funahashi A, et al. In Silico Biol. 2007;7:81-90.
- 14) Machne R, et al. Bioinformatics. 2006;22:1406-1407.
- 15) Keller R, et al. BMC Syst Biol. 2013;7:55.
- 16) McAdams HH, et al. Trends Genet. 1999;15:65-69.
- 17) Wilkinson DJ. Nat Rev Genet. 2009;10:122-133.
- 18) Gillespie DT. J Phys Chem. 1977;81:2340-2361.
- 19) Komarov I, et al. PLoS One. 2012;7:e46693.
- 20) Sumiyoshi K, et al. Front Physiol. 2015;6.
- 21) Hong L, et al. Biotechnol Prog. 2008;24:56-61.
- 22) Kevin R, et al. Bioinformatics. 2011;27:2457-2458.
- 23) Cao Y, et al. J Chem Phys. 2004;121:4059-4067.
- 24) Slepoy A, et al. J Chem Phys. 2008;128:205101.
- 25) Cao Y, et al. J Chem Phys. 2006;124:044109.
- 26) Gillespie DT. J Chem Phys. 2001;115:1716-1733.
- 27) Andrews SS. Methods Mol Biol. 2012;804:519-542.
- 28) Schmick M, et al. Cell. 2014;156:1132-1138.
- 29) Cowan AE, et al. Methods Cell Biol. 2012;110:195-221.
- 30) Hindmarsh AC, et al. ACM Trans Math Software. 2005;31:363-396.
- 31) Asai Y, et al. Adv Biomed Eng. 2014;3:50-58.
- 32) Hucka M, et al. Syst Biol (Stevenage). 2004;1:41-53.
- 33) Jörn Starrau, et al. Bioinformatics. 2014;30:1331-1332.
- 34) Takahashi K, et al. Bioinformatics. 2004;20:538-546.
- 35) Tomita M, et al. Bioinformatics. 1999;15:72-84.