

「シミュレーション技術の革新と実用化基盤の構築」

平成 16 年度採択研究代表者

富田 勝

(慶應義塾大学 教授)

「システムバイオロジーのためのモデリング・シミュレーション環境の構築」

1. 研究実施の概要

生物や細胞をシステムとしてとらえ、その全体のしくみや振る舞いを理解する新しい学問分野は「システムバイオロジー」と呼ばれ、国内外で注目を集めている。システムバイオロジーにおいてはコンピュータシミュレーションが不可欠であり、中心的な役割を果たすことが期待されている。

細胞内の多様な生理現象を対象とする細胞モデリング・シミュレーションは、モデリングの工程が他分野のシミュレーション研究とは異なり、実行時間よりも目的に応じてのチューニングにかなりの時間的コストを要する。これは、細胞のシステムが本質的に複雑であり、コンピュータリソース面で細胞のモデリングをある程度抽象化せざるを得ないために生じる問題である。したがって、この分野で成功するためには、目的に則ったモデリングをいかに効率よく行うかということにある。

そこで本研究では、この課題を解決するために、システムバイオロジーのための統合化されたモデリング・シミュレーション環境を開発する。具体的には、①知識ベースで統合されたデータベースを用いたシミュレーションモデルの構築、②シミュレーション計算の実行、③結果の解析に基づくモデルの評価を、GUI ベースの統合ソフトウェアでインタラクティブに行うことができるシステムを構築する。これにより、コンピュータに精通した一部のユーザだけでなく一般的な生物学者も、直感的かつ実用的な細胞モデリング・シミュレーションの利用が可能となる。また、このモデリング・シミュレーション環境開発の一環として、細胞シミュレーション用のアルゴリズムモジュール群、大規模化するシミュレーションモデルを並列計算機上で高速に計算するためのコアエンジンを開発する。更には、生化学シミュレーションと物理シミュレーションのような異なるスケールの現象を統合して解析できるコアエンジンを開発する。

本プロジェクトでは、これまでに統合環境フレームワークの基礎的な部分を開発した。現在はこれをベースとしたプラグインの開発、改良を継続している。また、要素技術としての並列計算環境への対応、シミュレーションモデルの半自動生成に取り組んでいる。

本研究が目指している実用的な細胞モデリング・シミュレーション環境の完成により、システムバイオロジーの研究が発展する結果、バイオテクノロジーや創薬研究の進展、更にはこれからの生物学研究に対するシミュレーション技術の一般化が大いに期待できる。

2. 研究実施内容

(1) 統合環境ソフトウェアの開発

本年度は、昨年度までの統合環境ソフトウェア基本設計で行った要求定義をベースにして統合環境ソフトウェアを実現するために必要な基本フレームワークの開発を主な課題として取り組み、基本フレームワークの設計・開発を行った。

開発手順に関しては、(1) 要求定義から統合環境に必要なプラグインを選定し、(2) プラグインを実現するためのフレームワークの要求事項を定義した後、(3) フレームワークの要求定義からフレームワークの設計、(4) フレームワークの開発を行った。

この結果、基本フレームワークを「DataManager」、「PluginManager」、「PluginBase」、「MainWindow」、「EcellObject」、「EcellData」、「EcellValue」の7つのクラスで構成することができた。

DataManager クラスは、統合環境ソフトウェアにロードされているモデルの管理を行い、シミュレーションを実行する際には E-Cell の通信を行うクラスである。

PluginManager クラスは、ロードしたプラグインの管理を行い、プラグインの基本クラスである PluginBase クラスと組み合わせることで、MainWindow 等の他プラグインのソースを編集することなくプラグインのロードを実現した。MainWindow クラスは、プラグインの一つで統合ソフトウェアに関する基本的な機能を実現した。

EcellObject クラスは、E-Cell 中のオブジェクト(Process, System, Variable, Stepper など)を表すクラスで、EcellData クラスと EcellValue クラスは EcellObject クラスのプロパティ値を実現するクラスである。

(2) データベースを統一的に取り扱うためのフレームワークの設計

知識ベース駆動型モデリング環境において中心的な役割を果たす、ゲノム・アノテーションデータベース、マイクロアレイによる遺伝子発現データベース、文献データベースなど、全く異なるオントロジーによって設計構築されているデータベースを統合し、統一的に取り扱うためのインターフェースのプロトタイプを設計し実装した。

これまでに構築したデータウェアハウスでは、ゲノム、プロテオーム、メタボローム、インタラクトーム、ジーンオントロジーのデータターを統合し、Web ブラウザおよび、SOAP RPC プロトコルによりアクセス可能な Java インターフェースを開発した。また、これらのフレームワークを利用するモデル生成ツールの開発も行った。

(3) クラスタ環境への対応、並列環境に適した並列化解析モジュールの開発

細胞モデルを用いた実際の仮想実験では、設定条件を微小に変化させ、非常に多数のパターンでシミュレーションを行う。その次に得られた結果を解析してさらに次の実験条件を調整し、シミュレーションを繰り返すといったユースケースが考えられる。

これらそれぞれの条件でのシミュレーションを、並列計算環境を利用して実行できるミドルウェア

の設計およびプロトタイピングを行った。具体的には、Globus を利用したグリッド計算環境を構築し、一般的に普及しているバッチジョブマネージャーである、SGE と PBSPro の両方に対応したセッションマネージャーの開発を行った。

(4) シミュレーションエンジンの開発

① 共有メモリ型計算機に対応したマクロレベルでの物質拡散アルゴリズムの開発

大腸菌を始めとする菌体、また細胞一般に言えることだが、その内部はタンパク質等が一杯に詰まっていることから非常に粘度の高いゾルである。そのため、分子の大きさ・種類によっては局在と拡散の影響を大きく受けるため、細胞シミュレーションにおいては単純な拡散方程式を用いることが適当でないケースが多い。そこで、このような比較的粗いレベルでの物質分子局在、拡散を取り扱うために、新たな手法を用いた (lattice-based simulation) アルゴリズムを開発した。

本年度の研究成果として、昨年度までは2次元のみであった実装を3次元格子に対応させ、8 Way プロセッサ並列化対応を含む大幅な拡張を行った。具体的なシミュレーション対象とするモデルとして大腸菌 MinCDE タンパクの振動をモデリングし、細胞膜上のタンパク重合を3次元モデルとしてシミュレーションを行った。

② 複雑な複合体の形成に関わる反応経路のためのアルゴリズムの開発 (MSIチーム)

生化学反応経路のモデリングとシミュレーションを行う際に考慮しなければならない重要な問題の一つに、複雑な複合体形成経路がある。この複合体形成経路の効率的なモデリング/シミュレーション法の E-Cell 上での実装に取り組んだ。

本年度の研究成果として、Molecularizer が自動的に生成する複合体に対して人間が理解しうる程度の命名規則を適用し、出力するアルゴリズムを考案し実装した。これに加えて、E-Cell シミュレーションエンジンの大幅な改良を行い、シミュレーション中のモデルの動的な変更を可能にするプロトタイプ実装を行った。これにより Molecularizer の機能、すなわち複合体形成経路の自動生成機能をシームレスに E-Cell に統合することができるようになった。

3. 研究実施体制

「慶應義塾大学」グループ

①研究分担グループ長：富田 勝（慶應義塾大学、教授）

②研究項目：統合環境フレームワークの開発、モデル構築用の GUI 環境の開発(研究課題 A)、細胞モデル解析・デバッグ環境の開発(研究課題 B、B1)、知識ベース駆動型モデリング環境(研究課題 C)、シミュレーションアルゴリズム群の開発(研究課題 D)、大規模モデルに対応したシミュレーション計算の高速化(研究課題 E)、マルチスケールモデルに対応するマルチフィジックスシミュレーションコアエンジンの開発(研究課題 F)

「MSI」グループ

①研究分担グループ長：Roger Brent (Molecular Sciences Institute、President/Research Director)

②研究項目：知識ベース駆動型モデリング環境(研究課題 C)、シミュレーションアルゴリズム群の開発(研究課題 D)

4. 主な研究成果の発表

(1) 論文(原著論文)発表

In press

- “Parameter estimation for stiff equations of biosystems using radial basis function networks” Matsubara, Y., Kikuchi, K., Sugimoto, M. and Tomita, M. BMC Bioinformatics(2006/03/9)

Published

- “GEM System: automatic prototyping of cell-wide metabolic pathway models from genomes”, Arakawa, K., Yamada, Y., Shinoda, K., Nakayama, Y., and Tomita, M. BMC Bioinformatics(2006 Mar)
- A microarray data-based semi-kinetic method for predicting quantitative dynamics of genetic networks” Yugi, K., Nakayama, Y., Kojima, S., Kitayama, T. and Tomita, M.; *BMC Bioinformatics* 6: 299(2005 Dec)
- Hybrid dynamic/static method for large-scale simulation of metabolism” Yugi, K., Nakayama, Y., Kinoshita, A. and Tomita, M.; *Theoretical Biology and Medical Modeling* 2: 42(2005)
- “Dynamic simulation of red blood cell and its application to pathological analysis of a pathological condition” Nakayama, Y., Kinoshita, A. and Tomita, M.; *Theoretical Biology and Medical Modeling* 2: 18(2005)
- “Space in systems biology of signaling pathways – towards intracellular molecular crowding *in silico*” Takahashi, K., Arjunan, S. N. V. and Tomita, M.; *FEBS Letter* 579: 1783–1788(2005)
- “Reverse engineering of biochemical equations from time-course data by means of genetic programming” Sugimoto, M., Kikuchi, S. and Tomita, M.; *Biosystems* 80(2): 155–164(2005)