

物理化学を駆使した半経験的汎用細胞シミュレーション環境の開発

A General-Purpose Cell Simulator —Semiempirical Application of Physical Chemistry to *In Silico* Whole Cell Reconstruction—

長島 優¹⁾ 相阪 有理²⁾ 北野 健太郎³⁾ 朽名 夏磨⁴⁾
Yu NAGASHIMA Yuri AISAKA Kentaro KITANO Natsumaro KUTSUNA

- 1) 東京大学医学部医学科 (〒 113-0033 東京都文京区本郷 7-3-1 E-mail: yunaga-ky@umin.ac.jp)
2) 東京大学大学院 総合文化研究科 (〒 153-8902 東京都目黒区駒場 3-8-1 E-mail: yuri@hep1.c.u-tokyo.ac.jp)
3) 東京大学医学部医学科 (〒 113-0033 東京都文京区本郷 7-3-1 E-mail: ken-ky@umin.ac.jp)
4) 東京大学大学院 新領域創成科学研究科 (〒 277-8562 千葉県柏市柏の葉 5-1-5 E-mail: kutsuna@kt.rim.or.jp)

ABSTRACT. Based on the idea that a cell is just a cluster of chemical reactions, we are trying to implement whole cell simulation of a unicellular organism by describing every single reaction inside the cell in terms of physical chemistry. As you imagine, this attempt is highly dependent on the products of Genome Projects and the recent progress in computational chemistry. We developed a model of *Escherichia coli*, the most convenient species of bacteria to model we think, because of its historical background in experiments and many biological data ever disclosed and stocked. Though our project is still in progress and not perfect, our simulator now enables you to simulate the behavior of the entire cell, visualize and manipulate it interactively. We believe it is not until this kind of system is realized that biology reaches comprehensions and descriptions of phenomena with high quantitative accuracy and high time resolution which “wet” molecular biology has never achieved.

1 背景

今日、様々な生物種で次々にゲノムプロジェクトが完了しており、ヒトゲノムプロジェクトも 2000 年にドラフトシークエンスが公開され、その後データベースも徐々に精度を増している。もはや遺伝子の塩基配列はインターネットを通じて誰もが手にできる時代となっている。そしてポストゲノム時代と呼ばれる生物学の次の段階として、現在はプロテオミクスと呼ばれる分野が注目されている。プロテオミクスの中心課題となっているのが、proteome 解析や transcriptome 解析と言われる、生物を構成する全タンパクの相互作用情報と発現情報の解析である。これらは、生物の体の中で、注目するタンパクが、いつ、どこで、どのくらいの量発現し、それがどのような分子と相互作用するか、ということをも明らかにすることを究極的な目標としている。つまりゲノムプロジェクトで得られた遺伝子配列の意味するタンパク質の機能を知というのがプロテオミクスの目的となっている。現在、バイオ関連ベンチャーも含めた様々な研究機関が、遺伝子チップや様々な分子生物学的手法を駆使して、これらのデータの蓄積にしのぎを削っている。

プロテオミクス研究の結果として、個々のタンパクの役割は明らかになると予想されるが、生命現象の包括的な理解ということ考えた場合、実はまだ知識が不十分であると考えられている。生物の設計図としてのゲノムが読み解かれ、一つ一つの部品であるタンパク質の性質が分かっても、それらがどのような反応ネットワークを構成し、その各々の反応がどのような時間的スケールで起こっているかという情報が次に必要になる。プロテオミクス研究が終わった後に必要となるであろう、こうした反応のネットワークの解析が進むことによって本来の意味で生物の理解が行われ、さらにはその完全な制御というものも実現すると信じる。

実は、既にバスクウェイ解析と呼ばれる類似の研究分野は存在するが、これは単にタンパクの相互作用情報を定性的な反応経路のデータとして蓄積することに重点を置いたものである。し

かし真に生命の記述を行うには、分子のネットワークを定量的に記述することが必要であると考えられる。また反応によって増減する分子の数の時間変化を追跡することも重要と考えられる。一部の発現情報データベースプロジェクトでも、定量的な時間変化の重要性は言われているが、反応ネットワークの解析ではおそらくこれよりずっと精度の高い情報が必要となる。すなわち分子の数が議論できる定量的な化学反応の起こる時間スケールでの時間分解能である。

また以上の流れとは別に、最近ではバイオインフォマティクスと呼ばれる学問分野がさかんである。現在の生物学では、ゲノムプロジェクトによる DNA の塩基配列データを始めとした大量のデータが生み出され、蓄積され続けており、これらの膨大なデータを処理するために計算機を利用するというのは自然な成り行きであった。既に DNA の塩基配列・タンパク質の立体構造といった情報のデータベース等のインフラが整い、DNA 配列解析のアルゴリズムなどが開発され、生物学を研究する上で無くてはならないツールとなっている。さらにタンパクの立体構造予測・機能予測や、タンパク質とリガンドのドッキング解析についても、有効なアルゴリズムが考案されつつあり、これらを利用することで生体分子の振る舞いを計算機の上だけで記述できる基盤が徐々に整いつつある。これらバイオインフォマティクス分野の背景には、計算機化学の理論が進歩し、生体分子のような高分子の振る舞いも量子化学的に計算できるようになったことがある。分子軌道法、ab initio 法をはじめ、様々な理論をベースにした量子化学計算パッケージが既に配布されており、これらを使うことで理論化学者なくとも、分子の物性値を容易に計算できる環境はかなり整備されてきているといえる。

以上述べてきたように、ゲノムサイエンスの次のステップとして、従来の分子生物学の実験手法が苦手としていた精度の高い定量的な時間分解能で生命活動を記述することが強く求められている。そしてそのためには、膨大なデータを計算機で処理しつつ、生体の全体像を見失うことなく生命を記述できるシミュ

レーション系を作るのが最もストレートな解決法であると考えている。

このような背景のもと、本プロジェクトでは、計算機による量子化学計算を駆使し、出来る限り近似をせずに求めた生体分子の物性データを用い、生物を構成する細胞の全反応系をシミュレーションできるシステムの構築を目指している。

2 目的

本プロジェクトの最終目的は、生物をその周囲の環境の物理現象も含めて記述することのできるシミュレーション空間を定義することにある。このような長期的な目標のもと、現在は単純な単細胞生物を対象にして、一つの細胞内での生命現象、および外部との相互作用を記述できる系の開発に集中している。本年度は、

- 分子の立体構造から現実的な時間内に、シミュレーションに使える精度で反応速度定数を計算する方法の模索
- 細胞内の拡散を考慮した濃度系のモデル化とその実装
- 細胞内の膜系の運動と周囲の溶液との相互作用のモデル化とその実装

の3点を目標とした。

3 期待される効果と活用のシナリオ

計算機による細胞シミュレーションは、細胞内の反応ネットワークを、実験生物学では想定しなかったような精度の高い定量性と時間分解能で記述することにより、ゲノムプロジェクトが終わり、プロテオミクス研究によるデータ蓄積もある程度充実した段階の将来の生物学に、細胞とその内部における生命現象を包括的に理解するための強力な手段を提供する。これにより、初めて生命現象が物理化学的な理論基盤と矛盾しない形で記述され、現代の工学におけるのと同様の *quality* で、細胞の、ひいては生命の制御が実現することになると考えられる。

全細胞シミュレーションの意義と効果として、以下のような事柄を考えている。

(1) 創薬分野

現在、創薬分野では、薬の候補となる物質全てに対して膨大な時間と手間をかけて、実験的に薬効をチェックしている。本プロジェクトのような細胞シミュレーションを使えば、計算機上でこれらのチェックを行うことが可能になると考えられる。もちろん、完全に実験的手法を置換しうる訳ではないが、例えば1万種の候補物質の振る舞いを計算機上の仮想細胞でシミュレーションし、見込みのありそうなものだけを抽出することで、100種に絞ることができたとしたら、薬の開発における費用と時間を大幅に節約できることになる。また計算機上の仮想細胞に対する複数の薬剤の投与により、副作用の予測も効果的に行えるようになると期待される。そもそも、薬剤というのは、(1) 薬剤が体内の作用部位に到達すること、(2) 薬剤が作用部位で作用を発揮すること、の二つの要素が共に達成されて初めて効き目を顕わす。従来の創薬においては、(2) の薬剤の化学的性質が比較的重視されてきた。しかし、昨今のゲノムプロジェクトによる SNPs (single nucleotide polymorphisms) 解析などにより、患者側の遺伝的要因により薬剤が作用部位に到達できなかったりする場合があることが知られてきて、そのせいで大量の薬剤が体内の意図しない部分に到達し、副作用の発現に寄与したりすることが明らかになってきている。ここで、(1) の視点、すなわち薬の吸収・体内分布・代謝・排泄の4要素による体内での薬物動態を把握することが副作用を減らし、患者によりやさしい治療につながるとして再認識され始めているのである。

こうした背景のもとで、薬剤の化学的性質を理論計算で推測する *in silico drug design* の分野のみならず、計算機シミュレーションとしての仮想細胞を使った薬物動態の把握が、動物実験などを積み重ねる従来の実験的手法に比べて短期間に、圧倒的に低コストで薬物を開発できる鍵になるとして注目され、創薬分野で新たな投資対象となってきている。

(2) オーダーメイド医療

医療の現場において、SNPsなどのゲノムの個人差が治療に対する *response* に変化を生じることが注目されつつある。患

者の遺伝的バックグラウンドの違いによって、一つの薬がある患者には著効し、ある患者には副作用をもたらす、ということが起こり得る。細胞シミュレーションは、このような患者による治療効果のばらつきを予測することに威力を発揮することが期待される。特定の患者の細胞をオーダーメイドでシミュレーションし、薬の効果などをあらかじめ判定するといったことをルーチン化すれば、副作用の発生も予測することが可能となる。こうした情報を治療にフィードバックすることで、治療の進行と並行してその精度の高い評価ができ、より質の高い医療を提供できるようになると考えられている。

(3) 人間の仮説創出能の補助

現在の分子生物学的な実験手法においては、例えばある重要な機能を持つタンパク質の候補として複数のアミノ酸配列が分かっているとき、それらを網羅的に調べるのが通例である。もし運が良ければ、最初に調べたタンパクが目的のものということもあるが、そういうことは必ずしもいつも望めない。それでも実験者はある仮説を立てて、可能性の高い候補から調べていく訳であるが、あらかじめ仮想細胞を用いて、ある程度の精度で候補となる分子の振る舞いに「あたり」を付けることができるれば非常に便利である。このように、実験者の仮説形成の補助手段として細胞シミュレーションを用いれば、目的のものである可能性が最も高いと予測された分子から実験を始めることで、実験に必要な時間・コスト・手間などの *resource* を大幅に減らし、生物学分野での実験を飛躍的に加速できると期待される。

(4) 学問横断的な生命現象の理解

全細胞レベルのシミュレーションは従来の実験手法では想定しえなかった精度の高い定量性と時間分解能で生命現象を記述できる。こうしたシステムは、今まで莫として知れなかった生命の全体像を、物理的な適応複雑系として把えて議論する際に強力なツールとなると予想している。

生命をシステム論的にうまく再定義することは、物理学と生物学の間の垣根を取り払い、かつてない精度での生物の制御を実現する上でとても重要だと考えられる。具体的には、微生物を計算機上で自由にデザインして工業利用したり、多細胞生物を工学的にコントロールする道が開けるものと考えている。

(5) 進化の理解

全細胞シミュレーションが実現すれば、実際に仮想生物を世代を越えて生かして試すことが可能となる。これにより、従来は現象論的に解釈することでしか議論できなかった生物の進化についても新たな理解が得られ、生命の起源といったトピックについても新たな知見がもたらされるものと信じている。

4 システムの概要

本システムは以下のモジュールから構成される。

- (1) 化学反応速度自動計算システム “Ainur”
- (2) 3次元反応拡散系モデル “ReDif”
- (3) シミュレータ汎用インターフェース “libea”
- (4) 細胞内巨大構造物記述モデル “GS”
- (5) 可視化系 “Palantir2”

以下では、個々のモジュールの機能について説明する。

(1) 化学反応速度自動計算システム “Ainur”

Ainur は、

- a. Substance mode
- b. Docking mode
- c. Reaction mode の3つのモードを持ち、それぞれ、

- | | |
|-------------------|------------------------------------|
| a. 分子の立体構造 | → 拡散係数 |
| b. タンパクとリガンドの立体構造 | → { 酵素複合体の立体構造
阻害定数 (K_i) } |
| c. 反応式 | → 反応速度定数 |

の計算を行う。さらに Reaction mode は Docking mode を、Docking mode は Substance mode をそれぞれ再帰的に起動し、これにより物性値の自動計算が達成される。また、全ての計算結果は可読的なテキストファイルに自動的に save され、もし同じ計算が指定された場合、データディレクトリに save データが残っていれば、再計算は行われず前の計算結果が出力される。これは将来、後述の libea 経由で Ainur が起動されることを想定し、シミュレーション中にリアルタイムに反応速度・拡散係数などの物性値の query が行われる場合に対応するための設計である。(但し、現在反応速度導出の過程で、人間による GUI からの手動操作を必要とするステップが二箇所存在している。これは今後のバージョンで自動化される予定である。)

シミュレーション途中の物性値計算を実現することの意義としては、次のように考えている。

既存のプロジェクトとして、計算機上での細胞シミュレーションを目指すいくつかの計画が存在する。それらはいずれも、本能的には反応速度論の微分方程式を数値積分することで細胞内の濃度の変化を時系列で追跡するというスタイルを採用している。しかし、これら従来のシステムではシミュレーションに必要なとなる反応速度定数や分子の初期濃度などのパラメータとして、文献値を用いたり、GA などの最適化によって推測した値を用いている。

文献値とはすなわち実験によって決定されたパラメータである。全細胞シミュレーションをしようと考えた時には、必要なパラメータを文献を漁って拾い集めてくる事になるが、まずシミュレーションに必要な全てのパラメータが実験によって過去に決定されているという保証はない。さらに文献検索という作業には当然ある程度の労力と時間を要し、学術の実験の補助手段として、あるいは創薬・バイオ産業分野における一要素技術として、細胞シミュレーションを様々な系に対して適用したいと考えたとき、汎用性は非常に低く使い勝手は良いとはいえなくなってしまう。

こうした事態を打開するために、一部のプロジェクトでは GA や NN によるパラメータの最適化による問題の回避を試みているものがあるが、このような手法は本来理論的根拠に薄く、こうした計算されたパラメータを使ったシミュレーション結果が、たとえ一見妥当そうに見えたとしても、信頼性はあまり高くないと考えられる。さらにシミュレーションをする系が変わるたびに、GA や NN によるパラメータ最適化をやり直さなければならず、汎用性という意味ではやはり望ましい解決法ではない。

今回、Ainur で実現を試みている反応速度の自動計算は、分子起動法パッケージ MOPAC2000[3] を援用し、量子化学や物理化学を駆使して、完全に理論計算のみにより反応速度定数の導出を行っている。これは、実験値が存在しない速度定数についてもパラメータ導出の可能性を与え、またシミュレーションに用いるパラメータに理論的な裏付けと妥当性を与える。それと同時に、過去に全く想定されていなかったような物質と物質がシミュレーションの途中で出会い、反応する事になった場合も、その反応が起き得るかどうか、また反応が起こる場合その効率はどうのくらいかということとその場で判定する手段を提供する。

このようにシミュレーションの中途での分子物性・反応速度の導出が出来る、というコンセプトを持つことのメリットは非常に大きい。なぜならシミュレーション開始前に文献検索やパラメータ最適化を終わらせていなければいけない従来の方法の場合では、事前に想定された化学反応しかシミュレートすることができないが、Ainur のような予測スタイルを確立すれば、シミュレーション開始前には予測できなかった未知のシミュレーションに対応できる可能性が飛躍的に高まるからである。

(2) 3 次元反応拡散系モデル "ReDif"

濃度表現・立方格子モデルを採用した 3 次元反応拡散モデルで、細胞内外の生体分子の挙動を細胞内局在を考慮しつつ追跡することを実現した。

細胞内外で起こっている複数の化学反応と、複数の溶質の拡散をシミュレートし、濃度分布の時間軸上の変化を計算する。シミュレーションの対象となる空間は立方格子によって均等に分割される。格子内においては反応が、隣接する格子と格子の間(「境界」)では濃度勾配にしたがった拡散が起こる。

なお任意の格子において、反応速度定数および溶質濃度を変更することができる。また任意の境界において、拡散係数を変更することができる。このことからシミュレーション空間で「反応のおこりやすさ」「物質移動のしやすさ」は時間的・空間的に非均質であり、自由に設定できる。これによって、生体膜の発揮する半透性、オルガネラのような細胞内外に展開する微環境の違いによる酵素活性の変化といった現象を、反応拡散モデルの外につくりこむことが可能となっている。また、濃度分布の

変更や新たな反応の追加など、任意の時点で恣意的摂動を加え、その影響を見ることが出来る。

ReDif は、基本的にライブラリであるが、様々な使い方を想定して多くのインターフェースを提供し、汎用性を追求した作りになっている。ユーザは C ライブラリ、サーバ、Python モジュール、コマンドのいずれについても、じかに利用できる。また、後述の libea を介して ReDif サーバを利用することができる。

既存の細胞シミュレーションプロジェクトには、細胞内の濃度勾配を記述できるモデルは少なく、あっても 2 次元の簡単で非現実的なモデルである。しかし細胞内の物質の局在や濃度勾配が、生命現象の理解の上で非常に大切なことは既に生物学の大前提として明らかに示されており、全細胞シミュレーションをするなら、細胞内の物質局在も扱いたいと考えるのは、生物学者ならばごく自然の成り行きであろう。細胞内の物質の濃度勾配が、特定の生命現象に重要であるという例は枚挙に暇がない。

そうした背景のもと、今回 ReDif で実現された 3 次元の反応拡散モデルは、細胞内の物質の局在を記述するための必要不可欠な第一ステップであるといえる。後述の GS による膜系のモデル化と合わせて、細胞分裂や genomic DNA の構造変化など、いわゆる細胞の目に見える「動き」は、物質の細胞内局在を記述することによって初めて記述される。これは計算機上での生命現象の完全再構築を目指す立場からは、避けては通れない過程となっている。

(3) シミュレータ汎用インターフェース "libea"

libea は、制御/描画系 (Palantir2) と数値計算 engine (ReDif・GS 等) の間を繋ぐ human interface を構築するためのライブラリである。

- a. 制御/描画系 (Palantir2) と計算 engine 相互の data 構造の変換

- b. 制御/描画系への network 機能の提供

の 2 点を実現する API 群で、TCP/IP protocol 上に実装されている。

プロジェクトでは全ての計算エンジンへのアクセスを libea 経由で行うことを想定しており、libea を用いることで user program は統一的な interface で個々のエンジンを利用できる。現在、GUI を持った Palantir2 が libea 上に構築されているが、libea 自体はそれだけで細胞シミュレータの interface となりうる汎用性の高いライブラリとなっている。

libea はまた、シミュレーションに必要な物質や反応の簡易データベースとしての機能も提供する。

(4) 細胞内巨大構造物記述モデル "GS"

GS は質点ばねモデルを採用した膜モデルであり、libea から使用することを想定したライブラリになっている。

膜の力学モデルとして、膜を細胞表面上に一様に分布するいくつかのスパーズな質点で代表させている。これらの質点の間をばねで連結させた質点ばねモデルを定義してやることにより、理論的な側面を度外視すれば、ある程度の膜の現実感が表現できるようになる。それぞれの質点の質量、あるいは質点間のばねのばね定数やダンパ値等は可変である。

ここで、技術的な問題であるが、数値計算の分野ではよく知られている、流体の抵抗に相当するダンパを導入することが重要で、このエネルギーの減少係数がないと、膜は次第に発散してしまう。

GS の役割としては、外部と内部の圧力・浸透圧等を考慮した変形、また内部流体モデルとの相互作用を行うこと、あるいは細胞内膜構造を扱うことなどが求められるが、そのためには流体力学を考慮して膜系と拡散現象との相互作用を記述することが不可欠と考えられた。当初かなりの時間を割いてこの膜の流体力学モデルのモデル化が検討された。ReDif の格子モデルとの相性の問題で実装は先送りされたが、この過程で流体シミュレーションについて一定の知識を蓄積できた。

(5) 可視化系 "Palantir2"

まだ未完成であるが、汎用の細胞シミュレータの GUI として必要と思われる機能を模索し、実装目標とした。その過程においては、

- 実験生物学者・実験生化学者の立場から実験を行う上で必要な仮説創出過程での補助手段となりうること

- 学生の立場から多くの因子が絡み合う複雑な生命活動を直感的に理解できるインターフェースとなりうること
- 創薬・バイオ産業などの産業分野で実用した際になるべく使い勝手が良いこと

の3点を念頭に置いた。
具体的な機能としては、以下のようなものを実装中である。

- ReDif/GS のシミュレーション条件の setup と、濃度などの変数の初期値の入力
- ReDif/GS のシミュレーション条件・変数のリアルタイムな変更 (add solute/change solute)
- ReDif/GS のシミュレーション結果のリアルタイムな可視化 (等濃度面・volume rendering・cutting plane)
- シミュレーション空間に対するマウスによる直感的な操作 (rotate/pan/zoom/pick)
- ReDif/GS のシミュレーション結果のファイル書き出し (録画機能) とその再生・解析 (record)
- Ainur へのインターフェースとして、シミュレーションと seamless な物性計算を実現する

(但し、項目 f. については未実装となっている。)
開発中の Palantir2 の snapshot を以下に示す。

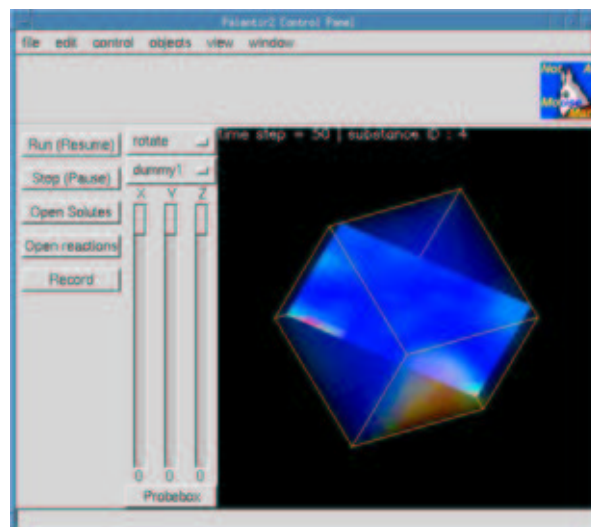


図 2: 濃度の cutting plane 表示

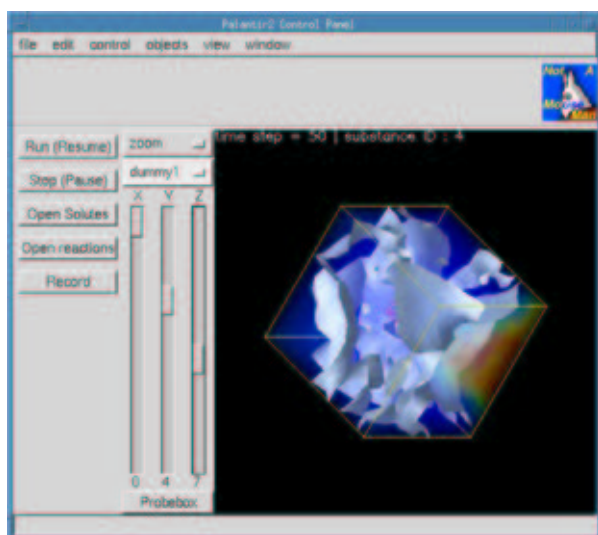


図 1: 等濃度面/volume rendering 表示

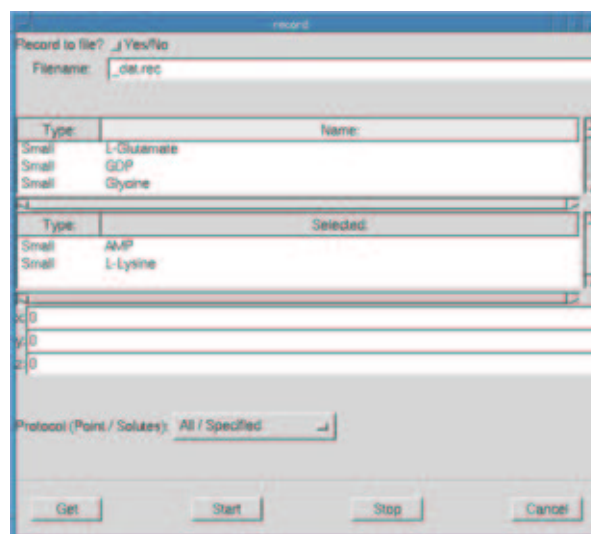


図 3: 録画機能設定画面 (解糖系周辺反応データを読み込んだところ)

5 評価

(1) 化学反応速度自動計算システム "Ainur"

現在、Ainur では、リガンドドッキング過程でのリガンドの結合部位の指定と反応経路計算に含める仮想反応分子系の指定に人間の手が必要となっている。

計算時間は、計算に考慮するアミノ酸残基の数と計算開始時の初期構造の歪み具合 (ground state からの離れ具合) に依存するが、2~3 個程度を選択した場合で、1~6 時間とやや幅がある。

計算に考慮するアミノ酸残基の数が増えれば計算時間はどんどん大きくなる。この場合はしかし、無視する範囲が減るため当然それだけ精度は良くなると思われるが、大きな系での計算は、時間的な制約から今回の未踏期間中には未検証となっている。

なお、反応経路予測ルーチンの成功の可否は、初期構造の良し悪しに非常に sensitive であり、今のシステムでは良い初期構造を setup してやるのがなかなか難しい。今後のアルゴリズム改善を課題としている。

2nd reaction rate constant [sec^{-1}]	Ainur	Experimental[1]
Aldolase	3.981136	1010
Enolase	2988	1880
Phosphoglucose Isomerase	405.126	1100
Fumarase	1277.977	2000
Phosphofructokinase	66.127	134
Glyceraldehyde 3-phosphate Dehydrogenase	688.7661	103
Phosphoglycerate Kinase	0.135054	56000
Phosphoglycerate Mutase	528.3476	910
Hexokinase	5.4×10^{-26}	690

表 1: 反応速度

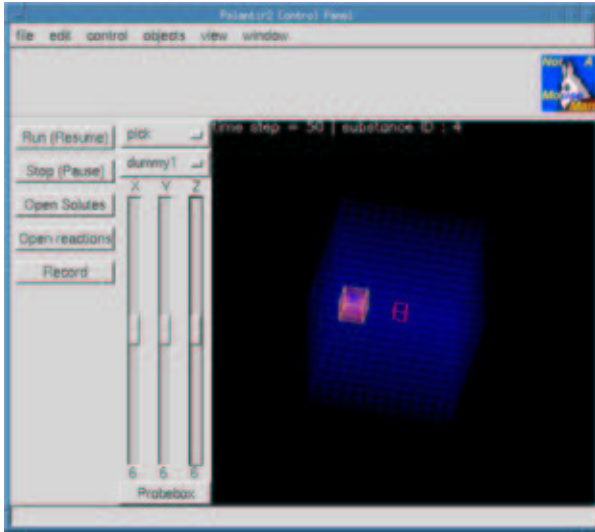


図 4: シミュレーション途中で系の格子を選択して物質を添加

反応速度計算で、今までに終了したのは、表 1 の 9 個の酵素である。(なお、開発中の暫定性能評価であり、有効数字と誤差の評価はしていない。)

現在、あまり満足すべき精度は出ていない。今年は今まであまり明らかでなかった反応速度定数の計算可能性を示すことが第一の成果となった。しかし、現行の Ainur は、未踏期間という制約があったため、全ての過程で一番簡単な、かつ計算機コストの低い手法を選択して動いているため、精度改善の余地はたくさんある。今後、理論面・実装面で精度の改善作業を進めて行く上で、煩雑な計算過程が自動化されていることは、大きな強みになると考えている。

拡散係数計算で、今までに文献値と比較できたのは、表 2 の 6 分子である。

Diffusion coefficient [m^2/sec]	Ainur	Experimental[2]
Sucrose	5.30×10^{-10}	5.22×10^{-10}
Glucose	6.43×10^{-10}	6.73×10^{-10}
Glycine	9.42×10^{-10}	1.06×10^{-9}
Water	1.83×10^{-9}	2.26×10^{-9}
Methanol	1.20×10^{-9}	1.58×10^{-9}
Ethanol	9.39×10^{-10}	1.24×10^{-9}

表 2: 拡散係数

巨大分子については文献値との比較がまだ出来ていないが、比較的理論的なハードルの低かった拡散係数の計算精度については楽観視している。

(2) 汎用細胞シミュレーション環境

Ainur を除く ReDif・libea・GS・Palantir2 は、現在連携して動作し、汎用細胞シミュレーション環境を実現する。

開発方針として、Ainur の統合や GS の再設計などが終了するまで生物学的各論には着手しないつもりであるが、テスト目的に使っているデータでの動作検証では、大腸菌の解糖系の代謝経路を中心とした 277 分子種・45 反応・ $50 \times 50 \times 50$ 格子の系での動作を確認している。

6 関連研究

細胞シミュレーションを試みているプロジェクトとしては、慶應大学先端生命科学研究所の E-CELL Project[4] や米 University of Connecticut Health Center の Virtual Cell Project[5] が世界的に有名である。

これらに対する本プロジェクトの advantage としては、

- 1) 反応速度定数・拡散係数を理論計算で導出できること
- 2) 物質の細胞内局在を 3 次的にシミュレートできることの 2 点と考えている。

7 今後の課題

今回の未踏期間中に達成できなかったこと、明らかになった問題点を列挙する。今年得られた最大の知見であり問題点でもあるのは、GS による膜系のモデル化において流体力学を考慮した反応拡散・巨大構造物系を実現するためには、現在の立方格子モデルによる空間の離散化を捨て去らなければならないということである。格子の自動生成や irregular grid の導入は次期 GS で考慮される。

システムの各モジュールでの今後の課題としては、以下のようなことを考えている。

(1) Ainur

- 精度の改善
→ 初期構造に強い安定した反応経路予測ルーチンの実装
- 速度の改善
→ MOPAC2002[3](localized MO) の導入
- 人間の介入が必要な過程の縮小と完全自動化
- 分子の膜透過速度の自動計算

(2) ReDif

- メモリ消費の削減
→ 濃度の極端に少ない分子種では粒子系を導入
- 膜表現への親和性を高める
→ 格子モデルの棄却、格子自動生成
- 計算機のクラスタ化、hardware 面での整備

(3) libea

- GS/Ainur への対応
- WWW 上のデータベース検索機能を持った旧版 Palantir との統合

(4) GS

- redif の内部情報(濃度・拡散速度)へのフィードバック
→ 流体力学モデルの導入
- 膜表現の格子モデルからの脱却
- トポロジ的な階層構造の論理的表現の模索

(5) Palantir2

- 新しい libea によるネットワーク上のデータベースとの連携
- Ainur との連携による物性計算

8 参加企業及び機関

有限会社 ミトンハウス (プロジェクト管理組織)

9 謝辞

本プロジェクトは、IPA(情報処理振興事業協会)の平成 13 年度未踏ソフトウェア創造事業の支援を受けました。また、PM(プロジェクトマネージャ)として貴重な助言を頂いた、電気通信大学 情報工学科の竹内郁雄教授に深く感謝致します。

最後に、本プロジェクトの企画段階から数々のミーティングを重ね、価値ある idea と inspiration を与えてくれ、貴重な時間と専門知識を提供してくれた学生グループ“白の会議”の面々に深く感謝致します。

10 参考文献

- [1] Oliver H. et al. "The Relationship between Substrates and Enzymes of Glycolysis in Brain", *J. Biol. Chem.* 1964 239:31-41.
- [2] P.W.Atkins, "アトキンス 物理化学 第4版 下巻", データ表項目 p.10, 千原秀昭, 中村亘男 訳, 東京化学同人 (1993).
- [3] "MOPAC 2000", J. J. P. Stewart, Fujitsu Limited, Tokyo, Japan (1999).
- [4] <http://www.e-cell.org/>
- [5] <http://www.nrcam.uchc.edu/>