

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 1 日現在

機関番号：82401

研究種目：若手研究(A)

研究期間：2013～2016

課題番号：25711011

研究課題名(和文) 真核細胞の信号伝達系まるごとシミュレーションのためのグリーン関数反応動力学法開発

研究課題名(英文) The parallel Green's Function Reaction Dynamics method enables whole signal transduction pathway simulations

研究代表者

高橋 恒一 (Kouichi, Takahashi)

国立研究開発法人理化学研究所・生命システム研究センター・チームリーダー

研究者番号：20514508

交付決定額(研究期間全体)：(直接経費) 20,900,000円

研究成果の概要(和文)：真核生物のシグナル伝達経路は、発生、分化やがんなど人の営みに深くかかわる様々な細胞現象の根幹をなしている。その原理を解明しようとする計算生物学の分野でも研究が続けられてきたが、細胞膜や核、染色体などの複雑な構造を含む細胞内でほんの数個の分子のふるまいが決定的な意味を持つシグナル伝達経路を分子一つ一つのふるまいから計算することはこれまで非常に困難であった。我々はこの困難を解決すべくグリーン関数動力学法と呼ばれる新規技法を開発し、厳密かつ高速な反応拡散シミュレーションを実現した。本研究ではこれをさらに並列化することでひとつの細胞をまるごと含めたシグナル伝達経路の1分子粒度での計算を可能にした。

研究成果の概要(英文)：Signal transduction pathways in eukaryotic cells are significant to understand cellular phenomena highly related to human life, e.g. cancer and cellular development and differentiation. In the field of computational biology, even though many algorithms have been developed, it is still difficult to simulate the signal transduction pathway at the single-molecule resolution, where only a few of molecules can affect the cellular fate. To overcome this difficulty, we have been developed a novel method named the enhanced Green's Function Reaction Dynamics (GFRD) method, which realizes fast and exact reaction-diffusion simulations in the single-particle level. In this study, we newly developed the parallelization algorithm of the GFRD method. This algorithm scales up simulations in time and space, and enables single-particle simulations of whole signal transduction pathways containing comprehensive structures such as cellular membrane, cell nucleus, and DNA.

研究分野：計算生物学

キーワード：並列化 シミュレーション 細胞生物学

1. 研究開始当初の背景

地球上の全ての生命活動は細胞内の生体高分子間の反応の連続、すなわち生化学反応経路の働きにより立ち現れる。細胞内生化学反応経路の挙動をモデル化、シミュレーションする技術は、細胞内現象を定量的に理解、予測、さらに操作する上で不可欠である。これまでの生化学反応経路シミュレーションは、細胞内空間をよくかき混ぜられた均一な溶液として単純化して考え、濃度平均場近似を用いて反応経路を電気回路のようなネットワークモデルとして扱う事が多かった。

しかし、実際には細胞内のタンパク質は細胞膜上や細胞質、核といった空間的に隔離した場に配置されており、さらに細かく見れば細胞膜上の脂質ラフトや細胞骨格等それぞれの場で複合体形成や足場タンパクによる係留の作用を受け不均一に分布している。このような特異な微小環境における分子の運動と反応の様態は、ミカエリス=メンテンの酵素反応速度論やスモルコフスキーの反応拡散理論といったこれまで標準的に用いられてきた古典的枠組みに拡張を迫る。「細胞環境」における分子の運動や反応の動態を *in vivo* で直接観察する事はいまだ困難であり、計算機を用いたモデル化が力を発揮する。

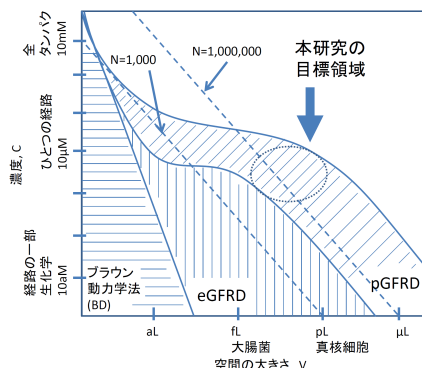


図1 本研究で開発する新手法の性能目標

2. 研究の目的

分子一つ一つの運動を解像する1分子粒度計算手法には大きく分けて連続空間内に粒子を配置する粒子法と離散空間内に粒子を配置する格子の二つに分類されるが、粒子法のほうがより柔軟性が高く精度の高い計算が可能である。粒子法としてはブラウン動力学(BD)が長く使われてきたが、細胞規模の問題に適用するには膨大過ぎる計算コストという難点があった。

近年のグリーン関数反応動力学法(Green's Function Reaction Dynamics: GFRD)と呼ばれる一群の新規計算手法が出現により、粒子法の適用範囲が飛躍的に広がった。本提案者らが開発した、近似を用いない正確な粒子反応拡散法として現在最速の手法であるThe enhanced Green's Function Reaction Dynamics: eGFRD法を用いると、典型的なバクテリアである大腸菌程度の空間規模(1fL)で、1つの信号伝達経路に関与するタンパク分子

の総濃度である数uM程度、分子数にして 10^3 程度の問題を解く事が可能である。

本研究では、現行のeGFRD法よりも1000倍高性能な新規手法を開発し、1pL規模の空間内で 10^6 分子以上の問題を解けるようにする(図1)。結果としてより重要な生物学的対象である酵母やほ乳類などの真核細胞の1本の信号伝達経路をまるごとシミュレーション可能にする。

3. 研究の方法

(1) グリーン関数反応動力学法(GFRD)

従来、1分子粒度計算手法として用いられてきた離散時間技法であるBD法は、局所的に起こる分子同士の衝突・反応を正確に捉えるため、系全体を非常に小さな時間幅で同期させる必要があった。GFRD法はこの問題を解決するため、保護球と呼ばれる境界条件によって系を粒子の周囲ごとに局所的に分割し、離散事象技法を用いることで典型的な濃度でBD法の 10^5 倍という高速かつ精密な1分子粒度計算を実現した。GFRD法ではまず、各粒子の周辺に互いに重ならない仮想的な保護球を考える。粒子の拡散運動はこの保護球内に留まる限りは他の粒子の影響を受けない1体問題として解ける。その解であるグリーン関数 $p_1(r, t|r_0, t_0)$ によって、保護球内の粒子の位置と保護球から外にでる時刻の確率分布が得られる。

$$\partial_t p_1(r, t|r_0, t_0) = D \nabla^2 p_1(r, t|r_0, t_0),$$

$$p_1(r, t_0|r_0, t_0) = \delta(r - r_0), \quad p_1(a, t|r_0, t_0) = 0$$

GFRD法は各粒子の計算をこのグリーン関数に従って非同期的に進めることで高速化を実現した。一方で、並列計算を行う場合、非同期的であるため頻繁にプロセス間通信が必要になるため、性能を発揮することができないという欠点があった。

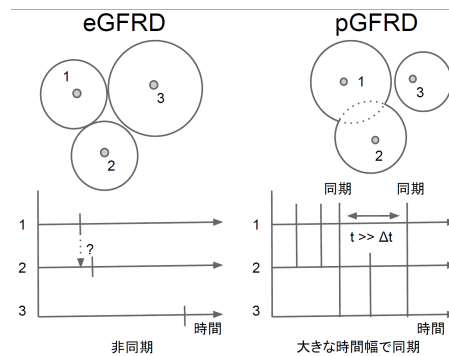


図2 従来法と並列化手法の比較

(2) 並列グリーン関数反応動力学法(pGFRD)

本研究では、申請者らが新たに考案したグリーン関数動力学法の大規模並列化計算手法である並列グリーン関数動力学法(parallel Green's Function Reaction Dynamics: pGFRD)を開発する。pGFRD法では階層的ステップングによる部分的な同期を採用することで、GFRD法の高速度性能を維持したまま並列化を実現することができる(図2)。

まず、pGFRD法の同期時間幅としてGFRD法の平均的な時間幅を採用し、この時間幅を元に各粒子の保護球を作成する。このとき、近傍に粒子が存在せず、保護球が重ならない場合にはGFRD法同様グリーン関数を用いることができる。また保護球が重なった場合にも、関わる分子の数は局所的であり、BD法と比べて大きな時間幅で計算できる。従ってpGFRD法ではGFRD法と同等の時間幅で同期し、それに合わせてプロセス間通信を行うことで、GFRD法の性能を維持しながら 10^3 スケールの高効率な大規模並列化が可能となる。

さらに、GFRD法がこれまで不得手であった細胞膜や細胞骨格、細胞内小器官などの複雑な一次元、二次元、三次元表面と粒子の相互作用および表面上での分子間相互作用のためにポリゴンを用いた高効率な境界条件処理を開発する。

4. 研究成果

(1) 分散メモリ並列計算による拡散の実装

eGFR法をはじめとする粒子反応拡散手法においてもっとも計算コストを要するのは粒子の拡散過程である。pGFR法では、OpenMPIを用いた分散メモリ並列計算によって拡散過程を並列化した。前述の通り、eGFR法は確率論的かつ離散イベント駆動型のアルゴリズムである。離散イベント駆動型のアルゴリズムでは各ステップが順次的に計算される必要があり、並列化には向かない。そこでpGFR法では大規模並列化のため、離散時間型のアルゴリズムを採用することでこの問題を解決した。従来のeGFRD法ではまず周囲の粒子の状況に合わせて保護球と呼ばれる領域を設定し、その保護球から外へとび出す時刻をグリーン関数によって与えられる確率に従って決定する。保護球は他の粒子と可能な限り干渉しないよう粒子ごとに最適な様々な大きさになる。一方pGFRD法では、まず並列計算の同期時間幅 t が与えられ、これに従って保護球の大きさ $r_i = 4(4D_i t)^{1/2}$ が粒子の状況によらず決定される。 D_i は粒子の拡散係数である。 r_i は時間幅 t の間に粒子がほとんど外へ出ないだけの十分な大きさを持つ。ただし、厳密には非常に小さな確率で保護球外に出るため、あらかじめその確率を計算し、必要に応じてロールバックを行う。次に保護球同士の重なり合いを確認する。これにより保護球が重なり合った粒子はまとめてドメインと呼ばれるグループに分割される。粒子は同じドメイン外の粒子に同期時間内では干渉することがないため、ドメイン毎の計算を並列化できる。以上の処理は、計算する容積を立方格子に分割し、各計算ノードで並列的に行われる。つづいて各ノードで決定されたドメインのうち、隣のノードの領域と重なり合うドメインを通信によって受け渡し、隣のノードから受け取ったドメインと自身のドメインの重なりを確認する。隣のノードのドメインを介してさらに

大きなドメインを形成する可能性があるため通信は二度行われる。この通信ののち、再び並列的に各粒子の t 後の位置を決定する。ドメインが1つの粒子(保護球)しか含まない場合はグリーン関数を、複数の粒子を含む場合はBD法を用いる。複数粒子を含むドメインでも、粒子間の距離は十分遠い場合が多いため、単純なBD法ではなく、粒子間の最短距離から動的にステップ幅を決定する階層的ステップングを採用することで高い性能が得られる。最後に粒子の拡散後、各ノードの計算領域外に移動した粒子を隣のノードへ通信し、ステップを終える。以上、1ステップあたり3回×三次元=9回の通信により、計算の厳密さを失わず離散時間型の分散メモリ並列化が可能となった。

さらに、各ノードは複数のドメインを有することが一般的である。これらの計算は完全に非同期的であるため、OpenMPを用いた共有メモリ並列化との併用が可能である。ただし、ドメイン毎のロードバランスが均等でないと計算速度は改善されるものの並列化性能そのものは悪くなるため、同期時間幅を最適化することが必要となる。

(2) 反応とグリーン関数の実装

つづいて、反応過程の実装とグリーン関数ライブラリの整備を行った。まず、一次反応はBD法においては従来通り、単一粒子のドメインでは一次反応の時刻を確率的に引き、同期時間まで逐次的に処理していく。一次反応後に拡散速度が変化する場合や二つの粒子にかい離する場合など保護球の大きさが同期時間内に可能性があるため、保護球を確定する前に最初の一次反応時刻を引き、それが同期時間内であればその反応を加味して保護球の大きさを決める。次に、二次反応については標準的にはBD法で扱い、グリーン関数による解析解が得られる場合は後から分岐を追加できるようにした。eGFRD法が対応していた三次元的に自由拡散する2粒子の場合はpGFRD法においてもグリーン関数を用いて反応拡散方程式を直接解く。

後述する二次元、一次元の構造体上の反応拡散ではさらに複数のグリーン関数を利用する。GFRD法の開発は、オランダ原子分子物理研究所(AMOLF)のten Woldeグループと共同で推進しており、互いに様々な状況におけるグリーン関数を導出・実装してきた。そこで、pGFRD法のメインコードとは別にグリーン関数ライブラリをまとめて公開し(https://github.com/ecell/greens_functions/)、pGFRD法を含む複数のGFRDアルゴリズムの実装から共通の資源として利用できるようにした。

(3) 構造体と構造体上の反応拡散の実装

eGFRD法は離散イベント駆動型であるため、実装が複雑であり、構造体を含む反応拡散の実装もまた非常に困難であった。他方、pGFRD

法は離散時間型の採用で並列化が可能になった以外に実装が平易になるという利点があり、構造体の実装も行えるようになった。具体的には三次元空間中にポリゴンもしくは真球で表現された二次元構造体と折れ線で表現された一次元構造体をそれぞれ利用できるようにした(図3、4)。シグナル伝達系において、前者は細胞膜・核膜など、後者はゲノムや微小管などを表すものとして用いられる。ポリゴンは Standard Triangulated Language (STL)形式のファイルを読み込み、折れ線については点列を直接与える。

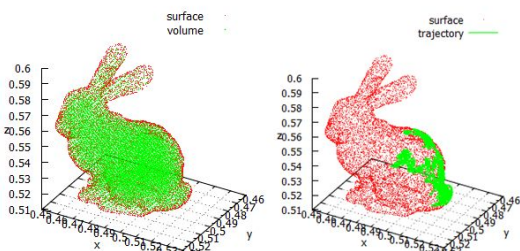


図3 ポリゴン内とポリゴン上の分子の拡散の例

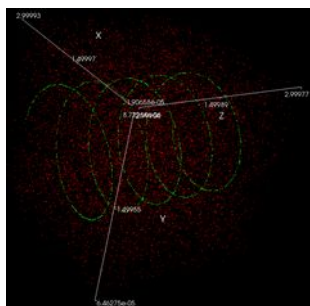


図4 一次元構造体との反応拡散の例

構造体上にある粒子の拡散は三次元空間中の拡散と比べて制限される(例えば線分上の拡散では円筒形の保護領域が必要となる)が、実装を容易にするため pGFRD 法では全ての場合について球形の保護領域を用いる。これにより、効率は幾分落ちるが、構造体から離れる反応が含まれている場合についても同じ保護球によって守られるため厳密さを維持できる。構造体上にある分子はその次元に応じたグリーン関数を用いて移動距離を確率的に決定し、構造体に沿って位置を更新する。ポリゴン上の拡散の場合、厳密さを保つためには保護球が頂点を含む場合には特殊な境界条件を持つグリーン関数を用いる必要がある。これは実際に本研究によりグリーン関数ライブラリに追加された。ただし、pGFRD 法では並列化が主目的であるため、簡単のため省略し、単に繰り返し折り返して位置を決定している。厳密さを保つ実装については開発を継続中である。

(4) 手法の性能評価

作成された手法について並列化の性能評価を行った。研究室が有する並列計算機を用いて 27 から 343 プロセスまでについて弱スケーリングを確認した。pGFRD 法は空間を立

方格子に分割するため、並列数に制限がある。今回の性能評価では各軸の分割数を同じにし、 3^3 、 4^3 、 5^3 、 6^3 、 7^3 プロセスを利用することとした。また、並列計算機の各ノード当たりのプロセス数に偏りがあった場合、並列度が落ちることがわかったため、性能評価では各ノード当たりのプロセス数が均等になるように指定している。pGFRD 法ではメモリアクセスを行う関数が全体の 17%程度 of 計算時間を占めているため、特定のノードに負荷が偏るとメモリアクセスのキャッシュ待ちが発生しているのではないと思われる。

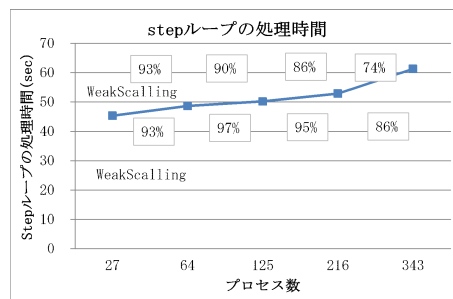


図5 反応を含まないコードの弱スケーリング性能評価

まず、三次元中かつ拡散のみを含む場合について性能評価を行った(図5)。1fLあたり6000個(およそ10uM)の直径5nm粒子が $5\mu\text{m}^2/\text{sec}$ の拡散速度で拡散するものとした。これらの値は典型的な真核細胞におけるシグナル伝達系の特性から見積もられた。弱スケーリングであるため、1プロセスあたり1fL(=1 μm^3)を計算する。全体空間の境界では周期境界条件を用いている。同期時間幅は1usecとした。1ノードあたりのプロセス数は27に固定されている。図に見られる通り、27プロセスを基準として343プロセスで74%、216プロセスで86%の性能を得た。216から343プロセスの比較では86%と他と比べてスケーリングが落ちていることからさらに大規模な並列計算機であればスケーリングも改善されるものと思われる。

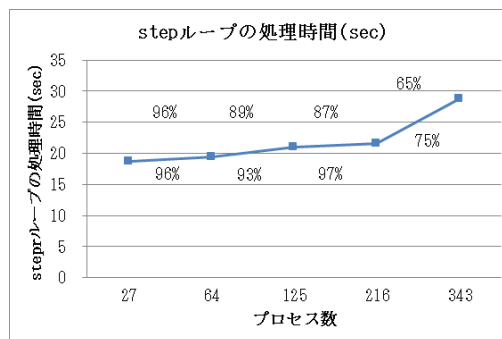


図6 反応を含んだコードの弱スケーリング性能評価

つづいて、反応を含む実装についても、同じ条件で同期時間幅のみを0.1usecに変更した場合で性能評価を行った(図6)。同じく27プロセスを基準として、343プロセスで65%、

216 プロセスで 87%の性能を得た。上の場合と同じく、216 から 343 プロセスの比較では 75%と他と比べてスケールリングが落ちているため、343 プロセスにおける性能は下がっているがそれ以外では拡散のみの場合とおおよそ同じ弱スケールリング性能を得られていることが分かった。

(5) pGFRD 法実装の公開

先に述べた通り、pGFRD 法のグリーン関数に関わるコードは独立したライブラリとして公開済みである。また、pGFRD 法のコードも <https://github.com/ece11/pgfrd/>において論文の準備ができ次第公開する予定である。これらのコードはいずれも自由なオープンソースソフトウェアとして GNU General Public License version 2 (GPLv2)ライセンスで世界に向けて公開され、誰でも自由に利用・改変することができる。当研究室では、今後も引き続き pGFRD 法の改良と応用を続けていく。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 件)

〔学会発表〕(計 3件)

Kazunari Kaizu, Kozo Nishida, Arjunan Satya, Koichi Takahashi, A composite approach of a single-particle-level simulation with multiple algorithms and spatial representations., International Conference on Systems Biology (ICSB) 2015, 2015年11月24日、シンガポール(シンガポール)
Kazunari Kaizu, Kozo Nishida, Masaki Watabe, Arjunan Satya, Kazunari Iwamoto, Koichi Takahashi, E-Cell System Version 4.0: an Integrated Platform for Single-particle-level Simulations, 2015 SIAM Conference on Computational Science and Engineering, 2015年3月15日, Salt Lake City (U.S.A.)

Kazunari Kaizu, Kozo Nishida, Masaki Watabe, Arjunan Satya, Kazunari Iwamoto, Koichi Takahashi, E-Cell System Version 4.0: An Integrated Platform for Single-particle-level Simulations., 15th International Conference on Systems Biology, 2014年9月14日, Melbourne (Australia)

〔図書〕(計 件)

〔産業財産権〕

出願状況(計 件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

取得状況(計 件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等
https://github.com/ece11/greens_functions
<https://github.com/ece11/pgfrd/> (論文と同時に公開予定)

6. 研究組織

(1) 研究代表者

高橋 恒一 (TAKAHASHI, Kouichi)
国立研究開発法人理化学研究所・生命システム研究センター・チームリーダー
研究者番号：20514508

(2) 研究分担者

()

研究者番号：

(3) 連携研究者

()

研究者番号：

(4) 研究協力者

()