

課題名 (タイトル) :

タンパク質・核酸など生体高分子の分子シミュレーション

利用者氏名 : ○木寺 韶紀、寺田 透、福田 育夫、藤崎 弘士、今井 隆志、
古田 忠臣、森次 圭、松永 康佑

所属 : 和光研究所 次世代計算科学研究開発プログラム
次世代生命体統合シミュレーション研究推進グループ
分子スケール研究開発チーム

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

すべての生命活動は、タンパク質や核酸など生体分子が担っている。したがって、これら生体分子の運動や反応を、物理法則に基づく分子シミュレーションによって再現することは、生命科学の究極の目標である。理化学研究所をはじめとした国内外の研究機関が強力に推進した構造ゲノミクスの成果として、多数の生体高分子の立体構造情報が入手可能となっている。これらの立体構造情報を利用して、生体分子の機能に対する理解を進め、さらに改変や薬剤設計に発展させていくために、分子シミュレーションに対する重要性が高まっている。

本申請の代表者がチームリーダーを務める次世代計算科学研究開発プログラム分子スケール研究開発チームでは、次世代スーパーコンピュータの計算能力を最大限に活用して、生体高分子の立体構造形成・機能発現メカニズムの解明を高精度かつ効率的に行うことのできる、新規アルゴリズムの開発と、これを実装するプラットフォームとなるマルチコピー・マルチスケール分子動力学シミュレーションソフトウェアの開発を行っている。また、並行してシミュレーションによってどのようなことが明らかになるか具体的に示すための応用研究も行っている。

2. 具体的な利用内容、計算方法

網羅的シミュレーションと応用研究においては、分子動力学シミュレーションソフトウェア AMBER および CHARMM、NAMD、GROMACS を用いた。計算対象に含まれる低分子化合物 (リガンド) の力場パラメータの計算には、量子化学計算ソフトウェア Gaussian および GAMESS を適用した。計算対象のタンパク質系

の周囲に十分な数の水分子を配置したのち、定温・定圧で~100 ns の分子動力学シミュレーションを行った。

藤崎、森次、松永が開発した新規アルゴリズムについては、寺田が開発を統括しているマルチコピー・マルチスケール分子動力学シミュレーションソフトウェアに実装した。これらの方法では、異なるパラメータを与えた数百から数万の系のコピー (レプリカ) を発生させ、それらの間の相互作用を考慮しながら並行してシミュレーションを行う。Lennard-Jones クラスタやアラニンダイマー、シニユリンなどの簡単なテスト蛋白質について実証実験によりアルゴリズムの正しさを検証したのち、adenylate kinase などの生体分子の系に適用し、これらの新規手法の精度と効率の評価を行った。

今井が実施した 3D RISM 計算では、自然科学研究機構分子科学研究所で開発されたプログラムを用いた。本計算では、生体高分子を含む空間を細かいメッシュに切ることによって溶媒分子の分布を表現するため、大規模容量メモリ計算機により計算を実行した。

3. 結果

(1) 網羅的シミュレーション

木寺、寺田、古田は、生体高分子の運動に関する統一的な理解につながる知識を抽出することを目指して、多数の生体高分子に対して、網羅的な分子動力学シミュレーションを行った。具体的には、Protein Data Bank に立体構造データが登録されているタンパク質のうち、配列が似ている (配列相同性が 95%以上) にもかかわらず、ドメイン運動によって立体構造が異なっており ($C\alpha$ RMSD が 1 Å 以上)、かつその違いが結晶場に起因すると予想されるタンパク質ペアを対象とした。今年度は、これらタンパク質ペアのうち、膜タンパク質

ではなく、単量体で機能していると考えられる 11 ペアについて、リガンドを取り除いた後、水溶液中の環境で 10 ns の分子動力学シミュレーションを行った。ここでは、PDB ID を入力としてシミュレーションのセットアップを行うシミュレーション管理システムと、結果を解析するツールの開発を併せて行った。

図 1 は、PDB ID が 1e4f の T 鎖と 1e4g の T 鎖のペアについてシミュレーションを行った結果を示す。左は結晶構造の重ね合わせを、右はシミュレーション中の 0.1 ns ごとのスナップショットについて計算したペアワイズ RMSD が最も小さい構造のペアの重ね合わせを示している。ここから、結晶構造では結晶場の影響で 2.7 Å の構造の違いがあるが、水溶液中の環境で、分子動力学シミュレーションを行うことによって、差が 1.7 Å まで小さくなることが明らかとなった。熱揺らぎを考慮すると、これらは実質的に同じ構造に収束したと言える。11 ペアのうち 7 ペアについて、同様な傾向が見られた。一方、11 ペア中 3 ペアについては、シミュレーション中のペアワイズ RMSD が結晶構造よりも顕著に大きくなった。RMSD が小さくなったタンパク質ではドメインが複数のリンカーでつながっているのに対し、RMSD が大きくなるタンパク質では、ドメインが 1 本のリンカーによってつながれているという特徴があった。このように、タンパク質の立体構造は、実験上のアーティファクトである結晶場によってゆがめられており、分子動力学シミュレーションはこのゆがみを取り除く上で有用であるといえる。また、シミュレーションを多数の対象について行うことで、タンパク質が共通に持つ運動の法則性を見出すことができることが示された。

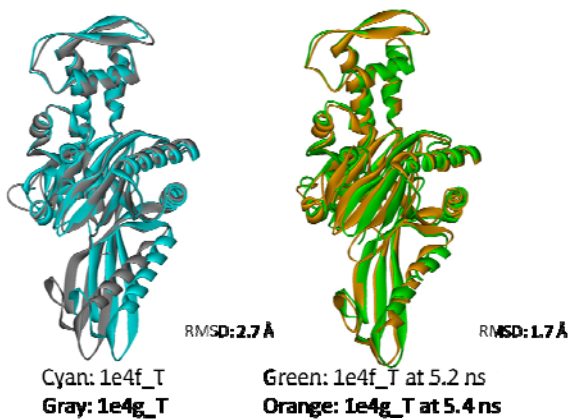


図 1 : 結晶構造 (左) とシミュレーション中のスナップショット (右) の重ね合わせ

(2) マルチコピー・マルチスケール分子動力学シミュレーションソフトウェアの開発

次世代計算科学研究開発プログラム分子スケール研究開発チームでは、次世代スーパーコンピュータの計算能力を最大限に活用して、生体高分子の立体構造形成・機能発現メカニズムの解明を高精度かつ効率的に行うことのできる、新規アルゴリズムの開発を行っている。寺田、森次、古田、松永は、これら新規アルゴリズムを実装するためのプラットフォームとなる、マルチコピー・マルチスケール分子動力学シミュレーションソフトウェアの開発を行っている。今年度は、generalized Born 法や、particle mesh Ewald 法など、基盤となる計算ルーチンの高機能化と並列効率を高めるチューニングを行った。また、マルチコピーシミュレーションのテストを行い、図 2 に示すように、156 コア (1 コピーあたり 2 コア) から 2,028 コア (1 コピーあたり 26 コア) まで、ほぼリニアにスケールすることを確認した。

Ribose binding protein(271残基、4071原子)で10psのストリング法計算のベンチマーク結果

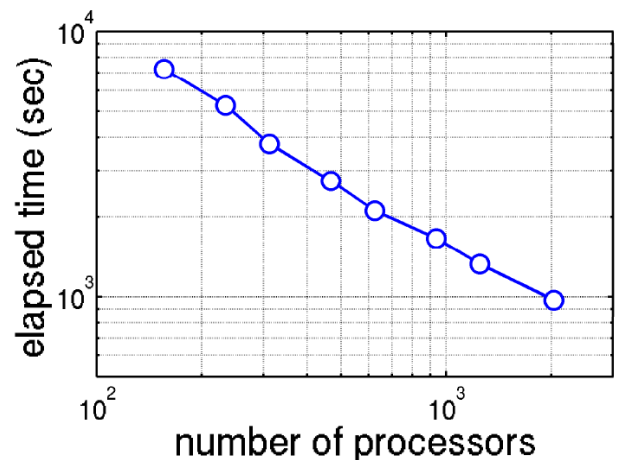


図 2 : マルチコピーシミュレーションのベンチマーク結果

(3) 拡張アンサンブル Onsager-Machlup 法の開発・実装

藤崎は、パスサンプリング法の一種である、Onsager-Machlup 法と、配位空間の強力なサンプリング手法であるレプリカ交換法を結合し、2次元の簡単なモデル系におけるパスサンプリングの問題に適用して、その有効性を調べた。その結果、初期のパスにトラップされる問題を克服し、非平衡の分布を正しく得ることができることが分かった (図 3)。

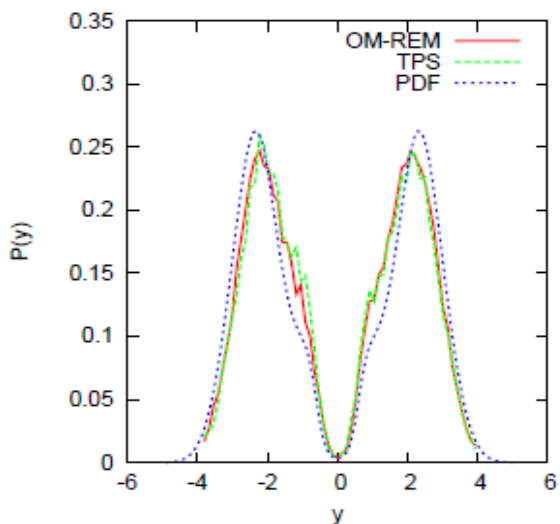


図 3 : Bolhuis ポテンシャル上でのパスの分布密度

(4) String 法の開発・実装

分子動力学計算の並列計算では、一般に長距離相互作用項を評価するための通信がボトルネックとなり、並列度を上げることは難しい。並列度を上げて、有効に超並列計算(>1024 コア)を実現するための方法として、松永は、経路空間サンプリングの一種である string 法を開発・実装した。string 法とは、単一の分子動力学計算の代わりに、多数のレプリカをタンパク質の構造変化パス上に配置し、それらレプリカを並列に計算することで構造変化パス空間をサンプルする手法である。レプリカ間の相互作用は弱く、通信は疎であるために並列度の高いアルゴリズムである。まず、Ribose binding protein を用いたベンチマークテストを行い(超並列 PC クラスタ使用)、およそ 2,000 コアまでよくスケールすることを確かめた(図 2)。次に、Adenylate kinase に適用し(超並列 PC クラスタ 1024 コア使用)、より自由エネルギーの低い構造変化パスを得ることに成功した(図 4)。

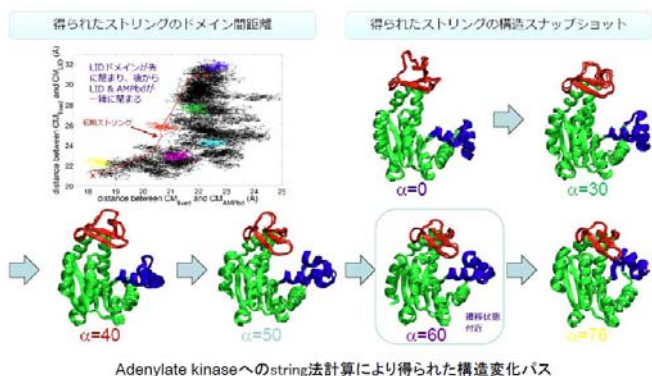


図 4 : Adenylate kinase の構造変化パス

(5) 逐次モンテカルロ法を用いた 1 分子 FRET 実験データの時系列解析

分子動力学計算の妥当性を評価するには、最終的には実験データと組み合わせた検証を行うことが重要である。特に、1 分子 FRET 実験では、タンパク質の部位間距離の時系列情報を得ることが可能であり、この情報と分子動力学計算を組み合わせることで、タンパク質ダイナミクスに関するより詳細な検証が可能となる。松永は、統計学で開発された逐次モンテカルロ法を 1 分子 FRET 時系列へ応用する実証研究を行った。逐次モンテカルロとは、多数のレプリカを用いたシミュレーションから、実験データを説明するレプリカをベイズの定理に基づいて逐次取捨選択していく技術である。実証研究では、米国プリンストン大の H. Yang 氏と共同で、Adenylate kinase の残基間距離の 1 分子 FRET データに対して、逐次モンテカルロ法を適用した。現在までの研究では、実験データではノイズのためにおよそ単峰的に見える分布から、Adenylate kinase で安定に存在すると言われているオープン型とクローズド型の分布を取り出すことに成功した(図 5)。

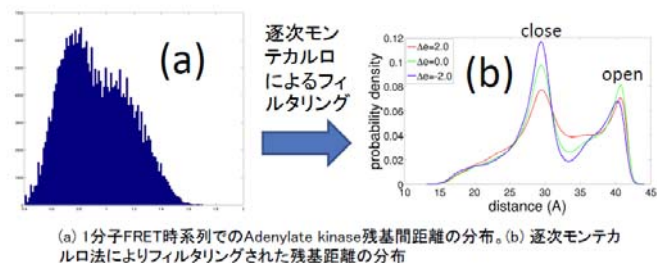


図 5 : 1 分子 FRET 実験と逐次モンテカルロ法を用いたの解析から得られる Adenylate kinase 残基間距離分布

(6) MM-CG シミュレーション法の開発・実装

粗視化 (CG) モデルにより蛋白質の構造変化に関わる運動を高速にシミュレートしつつ全原子力場 (MM) モデルの高精度な構造アンサンブルを得ることを目的として、互いに独立な MM と CG の自由度を連成した新たなシミュレーション法の構築を試みた。MM のなかの粗視化自由度と CG 自由度との拘束により MM と CG 自由度を連成させる。レプリカ交換法と組み合わせることにより、MM-CG 拘束のない、MM のみの高精度なアンサンブルを得ることが可能になる。森次は、マルチコピー・マルチスケール分子動力学シミュレー

シミュレーションソフトウェアにこのレプリカ交換 MM-CG シミュレーション法を実装した。10 残基の人工蛋白質であるシニョリンに適應した結果、マルチカノニカル MD で得られる結果と同様の自由エネルギー地形を効率よく計算できることを示した (図 6)。

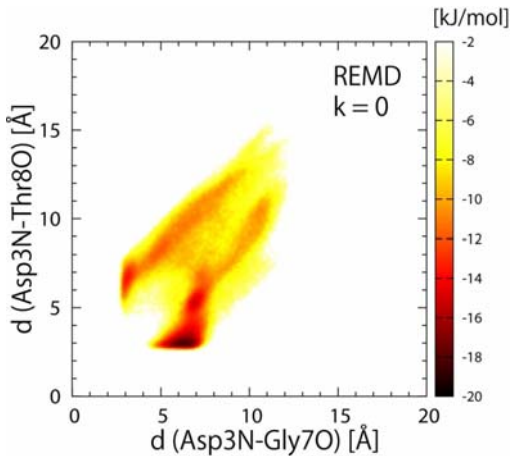


図 6 : MM-CG 法によるシニョリンの自由エネルギー地形

(7) 3D-RISM-based ligand mapping 法を用いた応用研究

今井は、3D-RISM-based ligand mapping法を用いて多剤排出トランスポーターAcrBに対する一連の薬物フラグメント分子の親和性のマッピングを行うことにより、AcrBの輸送ドメイン内空洞における化学官能性を調べた。その結果、AcrBの輸送ドメイン内空洞に少数の“多官能性”のリガンド結合サイトが存在することを発見した。そして、それらの“多官能性”サイトを結ぶことにより、AcrBの多剤排出経路を明らかにした。

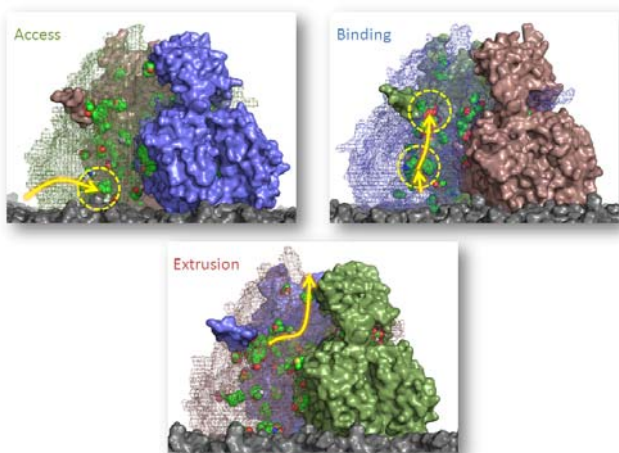


図 7 : AcrB の多剤排出経路

(8) Force-switching Wolf 法の実装

Wolf らが提案した簡便で効率の良い静電相互作用エ

ネルギーの計算法を利用して、正確な全エネルギー計算を行い、かつ、正しく導出された力の関数を用いてコンシステントな MD 計算を実現する手法を開発している。福田は、本手法をあらわに水を扱ったアラニンダイマー系に適用して計算精度を検証し、実用的な数値になることを確認した。また、自由エネルギーの計算を行い、標準的な手法である Particle Mesh Ewald 法の計算と比較して同様な結果を得た。さらに、精度を上げるための新手法の開発も行い、予備計算にて良好な結果を得た。

4. 今後の計画・展望 (これまで利用した状況と継続して利用する際に行う具体的な内容)

(1) 網羅的シミュレーション

平成 21 年度は、リガンドを取り除いた状態でシミュレーションを行ったが、平成 22 年度は、リガンドの違いがドメイン運動の原因ではないことを示すために、リガンドを結合した状態で、水溶液中の分子動力学シミュレーションを行う。また、平成 21 年度は単量体で機能しているタンパク質のみをシミュレーション対象としたが、ホモダイマーなど、多量体で機能しているタンパク質も対象に加え、シミュレーションを行う。また、多量体やリガンドを含む系を扱えるよう、シミュレーション管理システムと解析ツールの改良を行う。

(2) マルチコピー・マルチスケール分子動力学シミュレーションソフトウェアの開発

平成 21 年度に引き続き、基盤となる計算ルーチンの高機能化と並列効率の向上を図る。このソフトウェアは、現時点では MPI を用いて並列化しているため、OpenMP と組み合わせたハイブリッド並列化を行うことで、並列計算効率の更なる向上を図る。また、さまざまなマルチコピー・マルチスケールシミュレーションアルゴリズムを実装し、並列計算性能および立体構造空間探索性能の評価を行う。

(3) 拡張アンサンブル Onsager-Machlup 法の開発・実装

平成 21 年度に開発した方法論を分子系に使えるようにするために、Miller-Predescu によって提唱された曲率を用いない Onsager-Machlup 法を採用し、それを小さな分子系に適用する。温度レプリカ交換法がうまく

いかない場合は、マルチカノニカル法もしくは Tsallis 統計法と OM 法を結合させる。これらの手法は、離散化された経路、各レプリカに対して、超並列計算が容易に実装できる。さらにこれらの手法をわれわれが現在、開発中の分子動力学計算プログラムに移植することで、生体分子系でのパスサンプリング計算が可能になるようにする。

(4) String 法の開発・実装

次世代スパコンを効率的に使用するためには、少なくとも RICC 超並列 PC クラスタ 8,000 コアをフルに使えるようにプログラム開発を準備しておく必要がある。平成 21 年度までには string 法で 2,000 コアまでのベンチマークテストが成功しているが、8,000 コアではまだ成功していない。来年度には継続して開発・テストを 2,000 コア以上で行い、8,000 コアで実証研究を終える。

(5) MM-CG シミュレーション法の開発・実装

平成 21 年度には方法論の構築とマルチコピー・マルチスケール分子動力学シミュレーションソフトウェアへの実装、簡単なテスト系であるシニョリンへの適用を行った。来年度には、より大きな生体分子である adenylylase kinase への適用と、そのための方法論の改良に取り組む。また、応用研究として、天然変性蛋白質の自由エネルギー地形を計算する。

平成 21 年度 RICC 利用研究成果リスト

【論文、学会報告・雑誌などの論文発表】

1. H. Fujisaki, M. Shiga, and A. Kidera, *Onsager-Machlup action-based path sampling and its combination with replica exchange for diffusive and multiple pathways*, accepted for publication in J. Chem. Phys.

【国際会議、学会などでの口頭発表】

1. 今井隆志、「分子液体論（3次元 RISM 理論）に基づく新しいリガンドマッピング法」、第 10 回創薬インフォマティクス研究会、東京、2009 年 12 月。
2. 松永康佑、「パスサンプリングによるタンパク質構造変化解析」、大阪大学蛋白質研究所セミナー「分子科学を基盤とした生命活動への理論的アプローチ」、大阪、2009 年 12 月。
3. H. Fujisaki, M. Shiga, A. Kidera, “Onsager-Machlup action-based path sampling and its combination with the replica exchange method, 3rd International Symposium on Molecular Theory for Real Systems”, 京都、2010 年 1 月。
4. 福田育夫, 米澤康滋, 中村春木 “Evaluation of Electrostatic Interaction by New Summation Principle in Molecular Dynamics” 大阪大学蛋白質研究所セミナー、横浜、2010 年 1 月。
5. 古田 忠臣、「リガンド結合タンパク質構造変化の分子動力学シミュレーション」蛋白質の基質結合や構造変化における分子揺らぎの意義を討論する会、仙台、2010 年 3 月。