

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 3 日現在

機関番号：82401

研究種目：若手研究(B)

研究期間：2012～2013

課題番号：24770159

研究課題名(和文) タンパク質構造変化データの統計解析

研究課題名(英文) Statistical analysis of conformational changes of proteins

研究代表者

松永 康佑 (MATSUNAGA, YASUHIRO)

独立行政法人理化学研究所・計算科学研究機構・基礎科学特別研究員

研究者番号：60464525

交付決定額(研究期間全体)：(直接経費) 2,200,000円、(間接経費) 660,000円

研究成果の概要(和文)：本研究では、近年の大規模な生体分子分子動力学シミュレーションから得られるトラジェクトリデータを統計処理し、科学的知見を得るための統計解析フレームワークを開発した。特に、近年シミュレート可能となってきたタンパク質機能に関わる構造変化のデータを使って以下の2つの実証研究を実施し、その有効性を示した。(1)アデニル酸キナーゼの数マイクロ秒トラジェクトリを統計解析し、マイクロ秒オーダーで状態遷移するローカルな部位を特定した。(2)多剤排出トランスポーターAcrBの構造変化トラジェクトリから自由エネルギー推定計算を行い、プロトン化状態の変化によって自由エネルギー地形が変化することを示した。

研究成果の概要(英文)：With the help of recent advances in computer hardware and software, it is becoming possible to simulate functionally relevant conformational changes of biomolecules. On the other hand, however, the amount of simulation data is rapidly increasing and it is becoming difficult to treat such large amount of data without statistical analysis. In this study, we have developed a framework for the statistical analysis of molecular dynamics simulation data. The developed framework was successfully applied to conformational change data of two cases: (i) The most slowest degrees of freedom (involving the P-loop and a hinge region) in Adenylate Kinase was identified from 9 microseconds simulation data. (ii) Free energy differences along the conformational change pathways of multidrug transporter AcrB was successfully evaluated.

研究分野：生物学

科研費の分科・細目：生物科学・生物物理学

キーワード：統計数理 タンパク質 分子動力学シミュレーション アデニル酸キナーゼ 多剤排出トランスポーター

1. 研究開始当初の背景

近年のソフトウェアと計算機ハードウェアの急速な進歩により、生体分子の分子動力学シミュレーションは、生体分子研究の重要な手法の一つとなっている。しかし一方で、シミュレーションから得られる大量のデータを如何にして有効に処理・解釈し、科学的知見を得るかという問題も生じている。特に、分子動力学シミュレーション専用計算機(Anton, MDGRAPE-4)、GPGPU 計算、巨大なスーパーコンピュータ(京コンピュータ等)の登場により、分子動力学シミュレーションから得られるデータ量は近年飛躍的に増大しており、事前の統計処理なしにはトラジェクトリデータを“見る”こともままならない時代に突入している。

生体分子シミュレーション研究では、トラジェクトリデータを統計処理するための共通フレームワークが存在せず、個々の研究者がそれぞれで統計処理コードを書くことが多い。これは研究結果の再現性を担保できない可能性を生むということの他に、(1)開発した統計処理法を気軽にシェアできない、(2)複数の手法を組み合わせることが困難、(3)オーバーフロー(logsumexp 等)処理の不注意で生じた結果が見られる、(4)他分野の研究者や学生がトラジェクトリ I/O を含めて一からコーディングすることが困難、などといった問題を生んでいる。既存の汎用的な統計処理フレームワーク(例えば R 等で整備されている統計処理パッケージ)を使用するという可能性も考えられるが、生体分子研究には特有の処理(構造フィッティング、自由エネルギー推定、周期境界条件等)が多いため現実的ではない。

以上の背景を鑑みて、本研究では近年の分子動力学シミュレーションから得られるビッグデータを統計処理するためのフレームワークを開発する。特に近年シミュレート可能になってきたタンパク質機能に関わる構造変化データ、を応用ターゲットとして統計解析ツールを開発し、実証研究を通してその有効性を示す。

2. 研究の目的

本研究では、近年の大規模な生体分子分子動力学シミュレーションから得られるトラジェクトリデータを統計処理し、科学的知見を得るための統計解析ツールを開発・整備する。特に近年シミュレート可能となってきたタンパク質機能に関わる構造変化データに対して実証研究を実施し、その有効性を示すのが目的である。そのために、

以下の3つを実施する。

(1)**基本フレームワークの整備**:トラジェクトリデータを統計解析にのせるための前処理として、分子シミュレーション特有のファイルフォーマットに対応した I/O 機能、特定の原子団を選択するためのテキスト・インデックス処理、水素結合・コンタクト判定のための効率的なカットオフ距離内ペア原子探索ルーチン(Grid-cell アルゴリズム)等を整備する。また、構造フィッティング・原子間距離・角度・二面角・慣性半径・距離行列・コンタクトマップ等の分子構造計算ルーチンを整備する。他に、分子特有の構造間距離(Root mean square displacement; RMSD)に対応したクラスタリングルーチン(K-means クラスタリング等)を整備する。

(2)**構造変化に関わる遅い自由度を同定する統計処理法の応用**:応用・実証研究として、GPGPU を用いて酵素タンパク質であるアデニル酸キナーゼの長時間(数マイクロ秒)シミュレーションを実施し、得られた長時間トラジェクトリデータから構造変化に関わる遅い自由度を抽出する統計解析を行う。このために、高野・光武ら(JCP, 2011)と成富・淵上(JCP, 2011)によって独立に開発された緩和モード解析/ time-structure based Independent Component Analysis (tICA)を応用する。

(3)**構造変化における自由エネルギー変化を推定する統計処理法の応用**:応用・実証研究として、代表者が別プロジェクトで実施する多剤排出トランスポーター AcrB の薬剤排出を伴う構造変化(機能的回転とよばれる)シミュレーションのデータを使って、自由エネルギー推定を行う。そのために、ボロノイ分割と自由エネルギー推定法 (Multistate Bennett Acceptance Ratio; MBAR)を組み合わせた独自の手法を整備する。

3. 研究の方法

(1)の基本フレームワークを整備するにあたって、MATLAB 上に Toolbox の形で実装していくことにした。MATLAB は商用ソフトであるが、高度な関数を使用しない限りはフリーソフトである Octave で実行できるものがほとんどである。念の為に、基本フレームワーク実装中は、Octave 互換で動作するかのチェックを行った。

基本フレームワークとして、ファイル I/O(AMBER/CHARMM のパラメータ・トポロジー・トラジェクトリファイルの入出力)、原子団選択

(文字列選択・インデックス選択・距離内選択)、構造計算(最小二乗フィッティング・原子間距離・角度・二面角・慣性半径・距離行列・コンタクトマップ等)、クラスタリング(K-means・K-center・Information-based clustering 等)を整備した。また高度手法として、次元削減法(主成分分析・tICA)、自由エネルギー推定法(WHAM・MBAR法)、キネティック解析(遷移行列推定・マルコフクラスタリング等)を実装した。

テストデータをクラスタリング計算している際に、構造フィッティングによる RMSD 計算が律速になることが判明し、OpenMP 並列化した C 言語ルーチン(MEX ファイル)を呼び出す形とした。また他に、(3)で用いる自由エネルギー推定計算(MBAR)にも律速が見つかり、同じく OpenMP 並列化した C 言語ルーチン(MEX ファイル)とした。

(2)のアデニル酸キナーゼの長時間シミュレーションでは、本研究の予算で購入した GPU マシンと分子動力学シミュレーションパッケージ Amber を使って GPGPU 計算を SPFP 精度で実施し、9 マイクロ秒のトラジェクトリデータを得ることができた。シミュレーションには、Amber FF12SB 全原子モデル力場と、水(TIP3P モデル)を露わに含んだ系を用いた。総原子数は 40,353 個であり、このサイズとしては相当に長いシミュレーションができた。

(3)の多剤排出トランスポーター AcrB のシミュレーションでは、京コンピュータ(一般利用枠、課題番号 hp120027)を使用して、通常のシミュレーションとは異なる手法である最小自由エネルギー経路探索法(istring法)を使って AcrB の構造変化経路を求めた。求めた構造変化経路周りで、トータルとして 1.6 マイクロ秒のアンブレラサンプリング(束縛付きのシミュレーション)を行い、自由エネルギー推定のためのデータを得た。AcrB の総原子数は 480,074 個であり、前例にない規模の自由エネルギー推定といえる。

4. 研究成果

開発した統計解析フレームワークを、オープンソースとして BSD ライセンスで公開した(<https://github.com/ymatsunaga/mdtoolbox>)。他の研究者も使用できるように、今後ドキュメント・チュートリアル等を更に整備する予定である。

開発中に、整備中のフレームワークを利用して粗視化モデルフォールディングシミュレーションデータのダイナミクス解析、タンパク質複合体の振動データ解析、酵素タンパク質の構造変化

データ解析へ援用し、それぞれ学術論文として出版した。

(2)のアデニル酸キナーゼのシミュレーションでは、GPGPU 計算を使ったシミュレーションにより 9 マイクロ秒という長時間におよぶトラジェクトリデータを得ることができた。ただし、RMSD 値の時系列をみると活性状態であるクローズド構造までは構造変化していないことがわかった。ただし、オープン状態の中にもいくつかの準安定状態があり、マイクロ秒オーダーにわたってその間を遷移していることが分かった。

開発した統計解析フレームワークを応用した結果を図 1 に示す。図 1(a)左は、生体分子シミュレーションデータの解析で一般的に用いられている主成分解析の結果である。9 マイクロ秒のトラジェクトリを最も揺らぎの大きい方向である第一主成分へ射影している。図 1(a)右は、第一主成分モードの方向を、構造上に白矢印で示している。第一主成分の特徴として、タンパク質の硬い部位(ドメイン)がまとまって揺らぐ運動を捉えており、幾つかの状態遷移はみてとれるが、“ノイズ”的な運動も多く含まれていることがわかる。一方で、図 1(b)は緩和モード解析/tICA の第一モードへ射影した結果である。主成分に比べて、マイクロ秒に渡るシャープな状態遷移運動を捉えていることがわかる。この第一モードを構造上に矢印で可視化した結果、これはドメイン運動ではなく、ローカルな部位(P ループとドメイン間ヒンジ)の運動であることがわかった。結果として、ローカルな部位であってもマイクロ秒のオーダーで状態遷移し得るという事実が判明し、主成分との比較からこのローカルな運動がドメイン運動と関連している可能性があることが示唆された。

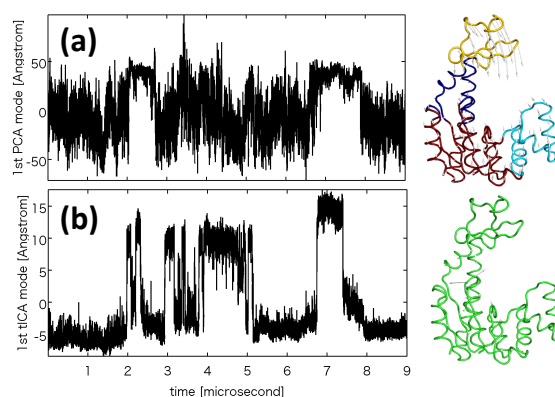


図 1: アデニル酸キナーゼの 9 マイクロ秒トラジェクトリデータの統計解析の結果。(a)主成分分析の第一主成分の時系列とその主成分モードを矢印で表現したもの。(b)tICA の第一成分の時系列とそのモードを矢印で表現したもの。

(3)の多剤排出トランスポーターAcrBのシミュレーションでは、京コンピュータ(一般利用枠、課題番号 hp120027)を使用して、様々な構造変化経路まわりのアンブレラサンプリングデータを得ることができた。開発した統計解析フレームワークを応用した結果を図2に示す。図2は、最小自由エネルギー経路探索法(istring法)の各サイクルで得られた構造変化経路周りの自由エネルギー変化を、経路に沿ったインデックスの関数として示している。自由エネルギー推定には、ボロノイ分割と自由エネルギー推定法(MBAR法)を組み合わせる独自の方法を適用した。図2から、istring法のサイクルが進むに従って自由エネルギーの低い経路を探索できていること、最終経路付近で自由エネルギーが収束していることがわかる。AcrBのサイズ(およそ50原子系)で自由エネルギー推定計算ができた例はほとんどなく、インパクトのある結果といえる。また、AcrBの膜貫通領域のプロトン化状態を変えたシステムで同様の計算を行った結果、構造変化経路に沿った自由エネルギー地形が変化することがわかった。これは、AcrBがプロトン濃度勾配を駆動力として構造変化するメカニズムの解明に繋がる結果である。

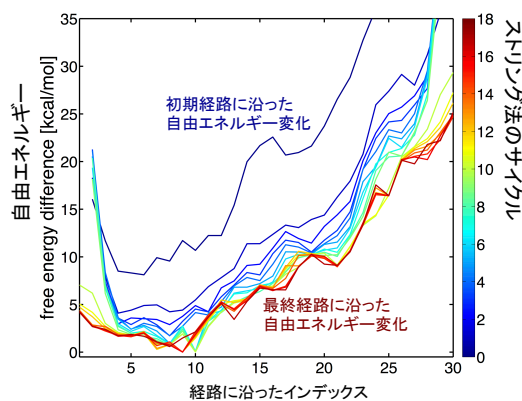


図2: 多剤排出トランスポーターAcrBの構造変化経路に沿った自由エネルギー推定の結果。横軸は経路に沿ったインデックスで縦軸は自由エネルギーを示す。線の色は、構造変化経路探索に用いたistring法のサイクル数に対応する。

以上の成果から、本研究の当初の目的は達成できたと考える。特に(3)のAcrBのシミュレーションデータ解析では、代表者が別プロジェクトで行った大規模シミュレーションと、本研究で開発した統計解析法の双方の強みを活かすことができ、国内外で見ても特徴のある成果を出せたと思われる。しかしながら、課題研究実施期間中に最終的な応用成果である(2)と(3)を学術論文

として出版できなかったのが反省すべき点である。今後、論文投稿準備をして出来る限り早く出版する予定である。開発した統計フレームワークについては、他の研究者が利用できるように今後ドキュメント等を一層整備していくとともに、他の統計解析手法(ネットワーク解析、カーネル法、因果性解析等)を追加していく予定である。

主な発表論文等

[雑誌論文](計4件)

① 松永康佑 “istring法によるタンパク質構造変化の解析” 分子シミュレーション研究会会誌 アンサンブル, 査読有, **16**, 29-35 (2014).

② Y. Matsunaga, A. Baba, C.-B. Li, J. E. Straub, M. Toda, T. Komatsuzaki, and R. S. Berry, “Spatio-temporal hierarchy in the dynamics of a minimalist protein model”, *The Journal of Chemical Physics*, 査読有, **139**, 215101 (13 pages) (2013).

<http://dx.doi.org/10.1063/1.4834415>

③ Y. Matsunaga, R. Koike, M. Ota, J. H. Tame, and A. Kidera, “Influence of Structural Symmetry on Protein Dynamics” *PLoS ONE*, 査読有, **7**, e50011 (11 pages) (2012).

DOI: 10.1371/journal.pone.0050011

④ Y. Matsunaga, H. Fujisaki, T. Terada, T. Furuta, K. Moritsugu, and A. Kidera, “Minimum Free Energy Path of Ligand-Induced Transition in Adenylate Kinase”, *PLoS Computational Biology*, 査読有, **8**, e1002555 (12 pages) (2012).

DOI: 10.1371/journal.pcbi.1002555

[学会発表](計4件)

① Y. Matsunaga, “Finding Conformational Transition Pathways in Biomolecules with the String Method and Sequential Data Assimilation”, Rare Event Sampling and Related Topics I, March 4-5 2014, ISM, Tokyo, Japan.

② Y. Matsunaga, “Sequential data assimilation of single-molecule FRET photon-counting data by using molecular dynamics simulations”, Workshop on Molecular Simulations of Biophysics and Biochemistry, November 21 2013, RIKEN AICS, Kobe, Japan.

- ③ 松永康佑, “生体分子シミュレーション研究の GPGPU による加速”, 2013 年度理研シンポジウム「理研 RICC 新搭載 GPU と科学へのアプリケーション」, 2013 年 6 月 27 日, 理化学研究所・和光.
- ④ 松永康佑, “有限温度ストリング法によるアデニル酸キナーゼの最小自由エネルギー経路探索”, シンポジウム「化学反応経路探索のニューフロンティア 2012」, 2012 年 9 月 21 日～22 日, 東京大学・山上会館.

[その他]

開発した統計解析フレームワークの公開先

<https://github.com/ymatsunaga/mdtoolbox>

6. 研究組織

(1) 研究代表者

松永 康佑 (MATSUNAGA, Yasuhiro)

独立行政法人理化学研究所・計算科学研究
機構・基礎科学特別研究員

研究者番号: 60464525