

研 究 報 告 書

「生体分子動態解析のためのデータ同化基盤の開発と応用」

研究タイプ: 通常型

研究期間: 2016 年 10 月～2020 年 3 月

研 究 者: 松永 康佑

1. 研究のねらい

近年、生体分子の構造の「動態」(分子構造の動き)が細胞内化学反応の制御に重要な役割を担っていることがわかってきた。そこで、その原理を解明するために、高解像度な動態を計測する手法が求められている。動態を原子解像度で観測する手法として、計算機を用いた分子動力学シミュレーションは強力であるが、力場パラメータが完全ではないため、得られる動態情報が正確とは限らないのが欠点である(Piana et al., *Curr. Opin. Struct. Biol.* 2014)。一方で、1 分子計測などの実験計測では、ナマの分子の動態を観測することができるが、生体分子にラベルした蛍光分子間距離などの不完全な構造情報しか得ることができない。そこで我々は、統計数理の手法を適用することで、シミュレーションと計測データを統合させるデータ同化手法の開発に取り組んできた。これまでに我々は、データ同化の代表手法である粒子フィルタを粗視化ペプチドモデルのシミュレーションとエミュレートした 1 分子計測データへ初めて導入し、両データを同化させることで従来では見えなかった隠れた構造情報を推定できることを示した(Matsunaga et al., *J. Chem. Phys.* 2015)。

しかしながら生体分子特有の事情として、「1 分子計測データの時間スケールと分子シミュレーションに大きなギャップがある」という問題があり、粒子フィルタなどのスタンダードなデータ同化手法を実データへ応用することは困難である。例えば典型的な分子では、1 分子計測データの 1 ステップが 10^{-6} から 10^{-3} 秒である一方で、シミュレーションを実時間で一週間実行して到達できる時間スケールがやっと 10^{-6} 秒というのが現状であり、両データを同化するのは難しい。

この問題を打ち破るため、本研究ではシミュレーションと計測の時間スケールの間にマルコフ状態モデルという統計モデルを介したデータ同化手法を提案する。マルコフ状態モデルは、分子の空間・時間スケールを粗視化し、代表構造間の遷移確率パラメータで動態を表現する。そこで、シミュレーションで構築した事前モデルに対して計測データを使って遷移確率パラメータを補正することで同化する。正確なモデルが構築できれば、解析的に中間状態や構造変化のパスウェイを求めて動態の詳細を得ることができる。本研究のねらいは、このデータ同化手法を様々な計測データに対して汎用化するとともにタンパク質フォールディングやシグナル伝達系などの重要な現象へ応用することにある。

2. 研究成果

(1) 概要

本研究のねらいは、提案するシミュレーションと計測データのデータ同化手法を開発し応用することを通じて様々な計測データに対して汎用化するとともにタンパク質フォールディングやシグナル伝達系などの重要な現象を理解することにある。そのために、以下の 3 つのテーマ

について研究を行った。

●研究テーマ A「小タンパク質・解像度の高い計測データへの応用」

WWドメインタンパク質のフォールディングをターゲットとして、分子動力学シミュレーションと1分子 FRET 計測時系列データを同化して解析した。その結果シミュレーションのみでは捉えられなかった、他の変異実験と整合性のある中間状態、フォールディングパスウェイを特定することに成功した(論文発表 1)

●研究テーマ B「中タンパク質・解像度の低い計測データへの応用」

Protein G タンパク質のフォールディングをターゲットとして、シミュレーションと1分子 FRET 計測時系列データを同化して解析した。Protein G タンパク質の構造を普通のシミュレーションでサンプリングすることは困難であるため、効率的なサンプリング法を提案し、マルコフ状態モデルを構築した。計測データとの同化の結果、アンフォールドで広がった状態が安定化され、そのフォールディングパスウェイを解析中である。

●研究テーマ C「データ同化手法の汎用化」

データ同化手法の汎用化として、味覚レセプターの不活性化・活性化を制御する構造変化をターゲットとしてシミュレーションと高速 AFM 像を同化して解析した。シミュレーションデータから AFM 像のエミュレートを行いそれと計測データを使って同化中である。

(2) 詳細

●研究テーマ A「小タンパク質・解像度の高い計測データへの応用」

研究期間前半では、WWドメインタンパク質のフォールディングをターゲットとして、分子動力学シミュレーションと1分子 FRET 計測時系列データをデータ同化した。WWドメインは37残基からなる小タンパク質であり、マルコフ状態モデルを構築するための構造サンプリングも比較的やりやすい。また計測データも時間解像度がよく、最初の実データ応用のターゲットとして適当であった。まず、WWドメインの全原子分子動力学シミュレーションを長時間行い、タンパク質構造を網羅的にサンプルした。得られたシミュレーショントラジェクトリを統計解析し、ダイナミクスをマルコフ遷移として近似するマルコフ状態モデルを構築した。次に、構築したマルコフ状態モデルを隠れマルコフモデルとして、1分子計測時系列データを使った機械学習を行い、遷移確率パラメータを補正した。その結果、1分子計測データをよく再現するマルコフ状態モデルを構築することができた。データ同化後のモデルを解析することで、フォールディングにおける中間構造(遷移状態)が他の変異実験との整合性があることを確かめ、データ同化によるモデリングの妥当性が支持された。更に、パスウェイ解析をすることでフラックスの大きなフォールディングパスウェイを求め、WWドメインのフォールディングが Wako-Saito-Munoz-Eaton モデルで説明できることを示し、論文発表した(論文発表 1、図 1)。

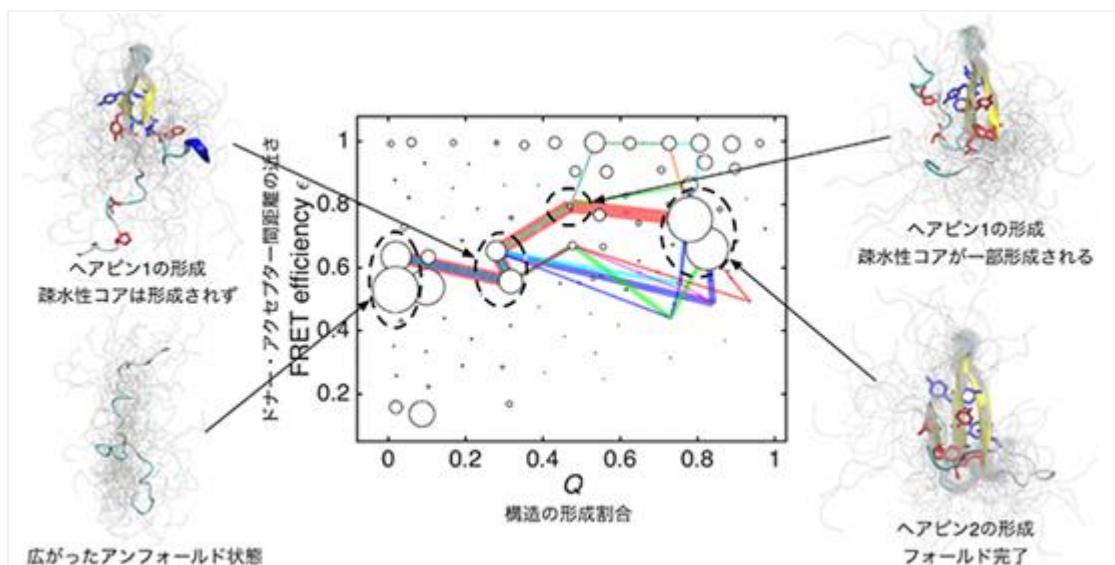


図 1. データ同化により解明された
WWドメインタンパク質のフォールディングパスウェイ

●研究テーマ B「中タンパク質・解像度の低い計測データへの応用」

研究期間中盤では、Protein G タンパク質のフォールディングをターゲットとして、分子動力学シミュレーションと 1 分子 FRET 計測時系列データをデータ同化した。Protein G は 56 残基からなる中タンパク質であり、WWドメインよりも 100～1000 倍遅いフォールディングレートを持つ。したがって、構造空間をシミュレーションでサンプリングできるかという課題があった。そこで、同領域の CREST 小松崎グループと共同でタンパク質の Q 座標(天然構造におけるコンタクト形成割合)上で多くの独立なシミュレーションを同時にはしらせ、途中で消したり分裂させながらレアな箇所を効率よくサンプルする手法(Adaptive sampling)を行った。その結果、普通のシミュレーションで想定するよりも 5 倍の短さで構造空間をサンプルすることができた。

またこの時期、領域会議のコメントで「簡単なモデルで手法の有効性を示したほうがよい」とあり、それを受けて粗視化ペプチドモデルで本研究の提案するデータ同化手法を再検討した。その結果、エミュレートした 1 分子 FRET 計測時系列データから状態遷移をよく推定できることを示すことができた。また副産物として、計測データを時系列データではなくて分布データにしたときにも、各状態の平衡確率を推定できることを提案し論文発表した(論文発表 2)。そこで得られた知見も以下の Protein G の解析へ活かした。

シミュレーショントラジェクトリを統計解析し、ダイナミクスをマルコフ遷移として近似するマルコフ状態モデルを構築した。その結果、Adaptive sampling で得られた多数の短いシミュレーションデータからもマルコフ状態モデルの遅い運動(遷移確率の固有値・固有ベクトル)がきれいにフォールディング/アンフォールディング運動と相関を持つことがわかった。これはマルコフ性の近似が妥当であることを示している。その後、構築したマルコフ状態モデルを機械学習(隠れマルコフモデリング)を使って実験計測データに合うように遷移確率パラメータを補正した。この段階で、状態の定義など様々な試行錯誤をしたが、機械学習部分のプログラムを高速化させることで従来の 1/3 の実行時間へ短縮することができた。その結果、WWドメインの結果と同じく、アンフォールド状態においては、シミュレーションだけよりも伸びた状態が安定

化されることがわかった(図 2)。Protein G のシミュレーションでは、わざわざアンフォールド状態で空間的に拡がるように設計された力場パラメータを使っただけに、それでもまだアンフォールド状態が拡がるのは予想外であった。この状態からのフォールディングパスウェイを解析し、論文として発表する。

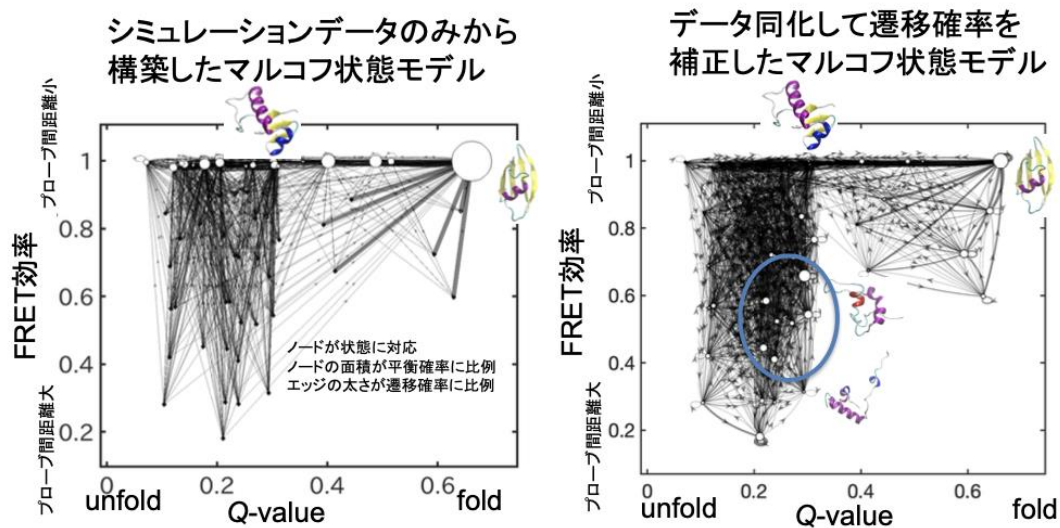


図 2: Protein G のデータ同化による拡がったアンフォールド状態の安定化

●研究テーマ C「データ同化手法の汎用化」

これまでは 1 分子 FRET 計測という特定の種類の計測データを扱っていたが、研究期間後半では他の種類の計測データへの汎用化を行った。岡山大学の山下教授、理研の杉田主任研究員と共同で味覚レセプターの不活性化・活性化を制御する構造変化の動態について、分子動力学シミュレーションと高速 AFM 像を用いて解析した。まずターゲットである味覚レセプターのモデリングと分子シミュレーションを行ってマルコフ状態モデルを構築した。味覚レセプターは、生理条件下では 4 つのこととなる構造(CA 状態、CR 状態、OA 状態、OR 状態)をとると考えられているが、X 線結晶解析では CA 状態が解けているのみである。そこで山下教授らが、他の 3 つの状態について、CA 状態構造や他のタンパク質の構造を雛形として推定するホモロジーモデリングを行った。ただし、ホモロジーモデリングだけでは、原子間の衝突や主鎖同士の間からまりが発生したために、その付近の構造は我々が物理的な全原子モデルシミュレーションを行ってリファインメントした。その後、得られた 4 状態周りを安定とする粗視化モデルを構築し、4 状態間を変遷するシミュレーションを行った。これまでの仮説では、味覚レセプターが活性化してシグナルが伝達する際には 4 状態間を決まったパスウェイで遷移するとされており、それを検証するのが目的となる。まずシミュレーションデータのみからマルコフ状態モデルを構築し、その後各状態の代表構造を使って高速 AFM 像のエミュレーションを行った(図 3)。高速 AFM 像エミュレーションでは、同領域 CREST 高田グループと連携して、同グループで開発したプログラム(afmize)を使用した。現在エミュレーション像と計測像の比較から遷移確率パラメータの補正を行っており、得られた知見を論文として発表する。

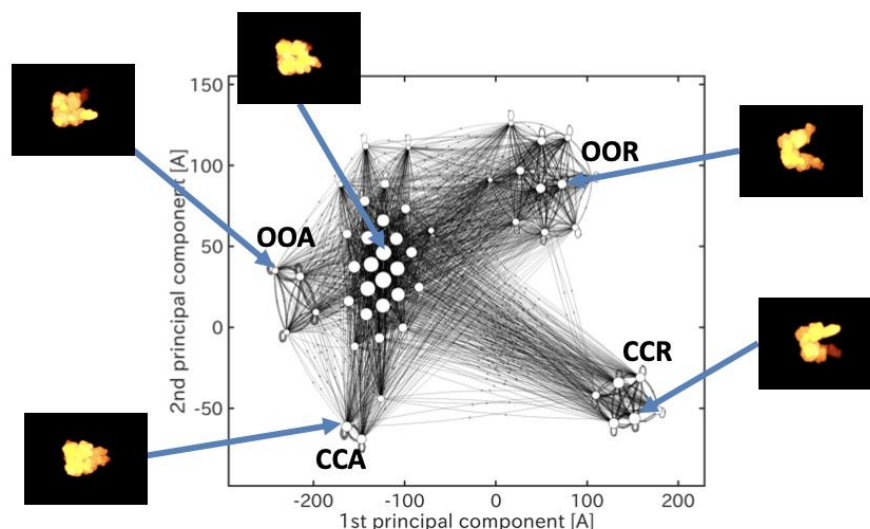


図 3. 味覚レセプターの AFM 像のエミュレーション

本研究で提案したデータ同化手法を行うためのプログラムを GitHub 上へ公開した(論文発表 2)。シミュレーションデータと計測データが既にあるとして、事前マルコフ状態モデルの構築から、計測データを用いた遷移確率パラメータの補正まで対応している。

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

最後に、本さがけ研究の成果が国際的に評価され、*Curr. Opin. Struct. Biol.*から総説を依頼され発表した(論文発表 4)。

3. 今後の展開

本研究により「個別の系のシミュレーション結果」に対して、「個別の計測データ」を用いて、従来の構造補正や分布補正ではなく「動態(構造の動き)」をその場限りで補正することは可能になってきたと考える。しかし、総括やアドバイザーからコメントがあったように、マルコフ状態のパラメータだけでなく、そもそものシミュレーションのカ場パラメータへフィードバックができるようになれば、個別の系にとどまらずシミュレーション全般に広く役に立つだろう。そのため手法として、任意のカ場パラメータに対して Onsager-Machlup action などの動的統計量を使って動態を重み付けする手法を開発していく。既にその端緒として、最尤推定を用いて粗視化モデルの適切なカ場パラメータを効率的に探索する論文を発表した(論文発表 5)。また、そのようにカ場パラメータによる動態の予測ができるようになってくると、今度はアミノ酸置換による相互作用の変化が動態へどのような影響をおよぼすのか予測することも期待できる。そのように動態を考慮したタンパク質デザインへの研究への展開していきたい。

4. 自己評価

マルコフ状態モデルを介したデータ同化手法を開発し、いくつかの生体分子が機能する際の重要な動的プロセス(中間状態・パスウェイ)を明らかにしたことで、戦略目標のひとつである「計測対象の特徴量解析技術の構築」へある程度貢献できたと考えている。

達成度に関して、基本的には研究計画に沿って研究を進めてきたが、途中領域会議でのコメントや実験研究者との連携もあり、テーマやターゲットを都度修正してきた。その上で全体の達成度を与えると70%程と考えている。研究テーマA「小タンパク質・解像度の高い計測データへの応用」に関しては、最終的な科学的知見も得られて達成できたと考えている。研究テーマB「中タンパク質・解像度の低い計測データへの応用」では、研究の進め方として最初のうちは文字通り個人で行っていたが途中からCREST 小松崎グループの助けも借りて進展させることができた。研究テーマC「データ同化手法の汎用化」についても、実験の山下教授やCREST 高田グループとの連携で進展させることができた。研究費は主に分子動力学シミュレーションを高速に行ってマルコフ状態モデルを構築するためのGPU 計算機購入にあてた。研究テーマAとBで有効に活用することができ、これまでにない量の構造を探索できたと考える。

研究成果の科学技術及び社会・経済への波及効果について、本研究で開発されたデータ同化手法は今後、生体分子動態一般を調べる際に、シミュレーションと計測データが一致しなかった場合の補正法として広く応用されることが期待される。また、研究テーマC「データ同化手法の汎用化」で解析している味覚レセプターの動作原理まで踏み込むことができれば、新たな嗜好性物質の開発にも展開することができるだろう。また他にも、分子モーターやトランスポーター、例えば薬剤を排出してしまう多剤排出トランスポーターが構造変化と共役して薬剤を排出する原理を本研究のデータ同化手法を用いて従来より高解像度で観測することができれば、排出を阻害する薬剤の開発にも貢献することが期待できるだろう。

5. 主な研究成果リスト

(1) 論文(原著論文)発表

1. Y. Matsunaga, and Y. Sugita, “Linking time-series of single-molecule experiments with molecular dynamics simulations by machine learning”, *eLife*, 2018, **7**, e32668 (19 pages)
2. Y. Matsunaga, and Y. Sugita, “Refining Markov State Models for conformational dynamics using ensemble-averaged data and time-series trajectories”, *The Journal of Chemical Physics*, 2018, **148**, 241731 (7 pages)
3. Y. Matsunaga, T. Yamane, T. Terada, K. Moritsugu, H. Fujisaki, S. Murakami, M. Ikeguchi, and A. Kidera, “Energetics and conformational pathways of functional rotation in the multidrug transporter AcrB”, *eLife*, 2018, **7**, e31715 (19 pages)
4. Y. Matsunaga, and Y. Sugita, “Use of single-molecule time-series data for refining conformational dynamics in molecular simulation”, *Current Opinion in Structural Biology*, 2020, **61**, 153–159.
5. A. Shinobu, C. Kobayashi, Y. Matsunaga, and Y. Sugita, “Building a macro-mixing dual-basin Go model using the Multistate Bennett Acceptance Ratio”, *Biophysics and Physicobiology*, 2019, **16**, 310–321.

(2)特許出願

研究期間累積件数:0 件

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

1. プレスリリース・新聞報道など

- A. 1 分子計測のデータ同化による生体分子構造ダイナミクス
ー小タンパク質が折り畳まれる際の中間構造・パスウェイを特定ー
松永康佑、杉田有治
http://www.riken.jp/pr/press/2018/20180515_2/
- B. 多剤排出トランスポーターの薬剤排出機構を解明
ースーパーコンピュータ「京」で巨大分子機械の動きを計算ー
松永康佑、山根努、寺田透、森次圭、藤崎弘士、村上聡、池口満徳、木寺詔紀
http://www.riken.jp/pr/press/2018/20180312_1/
2018 年 3 月 22 日付け日経産業新聞 5 面に掲載
2018 年 5 月 29 日付け産経新聞【万象】に掲載

2. 招待講演

- A. Y. Matsunaga “Integrative modeling of protein dynamics from time-series data of single-molecule experiments and molecular dynamics simulations” 14th Asia-Pacific Physics Conference (APPC2019), Kuching, Malaysia, November 29, 2019.
- B. Y. Matsunaga “Integrative modeling of protein folding dynamics from experiments and simulations” Telluride Workshop on “The Complexity of Dynamics and Kinetics from Single Molecules to Cells”, Telluride, Colorado USA, June 20-24, 2017.
- C. Y. Matsunaga “Drug extrusion mechanism of the multidrug transporter AcrB studied by molecular dynamics simulation” 日本化学会第 97 春季年会・アジアシンポジウム 慶応日吉キャンパス 2017 年 3 月 18 日
- D. Y. Matsunaga “Markov state modeling of protein folding dynamics by combining single-molecule experiments and simulations” Simulations Encounter with Data Science, Institute of Statistical Mathematics, Tokyo, Japan, March 9-11, 2017.

他 12 件

3. The Journal of Chemical Physics, 2018 Editor's Choice

Y. Matsunaga, and Y. Sugita, “Refining Markov State Models for conformational dynamics using ensemble-averaged data and time-series trajectories”, *The Journal of Chemical Physics*, 2018, **148**, 241731 (7 pages)