

PaCS-MD と異常検知を援用したタンパク質構造遷移経路の探索

原田 隆平

筑波大学計算科学研究センター

1. はじめに

2021 年度（前期）スーパーコンピューター若手・女性利用者推薦制度の採択課題「PaCS-MD と異常検知を援用したタンパク質構造遷移経路の探索」の研究成果を報告する。報告内容は、採択期間において開発した生体機能に関するタンパク質の構造変化を効率的に探索する計算手法（構造サンプリング法）の開発である。

2. イントロダクション

生体内ではタンパク質をはじめとする生体分子が機能を担い、生命を維持している。タンパク質は構造揺らぎを巧みに利用することで機能を発揮している。例えば、ドメイン間の大規模な構造変化により基質（リガンド）を認識したり、他のタンパク質と会合して複合体構造を形成したり、様々な構造変化が生体機能に関係している。これらの構造変化を追跡することは、生体機能の解析において非常に重要である。構造変化を直接追跡する計算科学的手法として分子動力学シミュレーション（MD）は強力であり、理論研究者のみならず実験研究者にも利用されている。MD は、運動方程式を逐次的に解くことで時々刻々変化するタンパク質ダイナミクスを追跡できる。しかしながら、生体機能に関するタンパク質の大規模構造変化は通常の MD が追跡可能な時間スケールよりも遥かに長時間で観測される「生物学的レアイベント」であるため、抽出が困難である。MD 専用計算機 [1, 2] などを利用すれば、タンパク質のサイズは小さいもののミリ秒に及ぶ長時間 MD が可能になってきてはいるものの、これらの極めて高価な計算機を利用できる研究者は限られており、全ての研究者に解放されているわけではない。故に、現実的かつ合理的な計算コストによりタンパク質の長時間ダイナミクスを抽出する「レアイベントサンプリング手法」の開発が望まれている。

本研究では、タンパク質の構造遷移探索法として開発したカスケード選択型分子動力学計算（PaCS-MD: Parallel Cascade Selection MD）[3] と機械学習の 1 つである異常検知を援用し、生体機能に重要なタンパク質の長時間ダイナミクス（レアイベント）を効率的に抽出する「異常検知型 PaCS-MD (ad-PaCS-MD)」を開発した。具体的には、「構造変化のし易さ」と「構造の異常さ」を異常検知により対応付け異常度が高い構造を特定し、PaCS-MD の初期構造に採用する。PaCS-MD の各サイクルにおいて「異常度が高い構造 = 遷移確率が高い構造」から短時間 MD を繰り返すことで構造遷移を誘起し、構造変化（遷移経路）を自動的に予測する ad-PaCS-MD を開発する。ここで ad-PaCS-MD の開発に先立ち、PaCS-MD の構造サンプリング戦略の概要を記載する。

PaCS-MD はタンパク質の構造変化（構造遷移）を効率的に探索可能なレアイベントサンプリング手法である。図 1 に示すように PaCS-MD の具体的な計算手順は、(1) 複数の短時間 MD から遷移確率の高い分子構造を選択し、(2) 選択構造を初期構造として再び短時間 MD を実行し繰り返すことで効率的な構造探索を可能にする。非常に単純な計算スキームであるが、(1)と(2)からなるサイクルを繰り返すだけで効率的に生体分子の構造遷移を誘起させることができる。常に遷移確率の高い分子構造を選択して選択した分子構造から短時間 MD を繰り返すサイクルにより、生体機能に重要な構造変化を効率的に誘起する。PaCS-MD は様々なターゲットに適用され、生体機能に重要な構造遷移の抽出に成功してきた。

[4-6]

PaCS-MD の詳細に関してはレビュー論文

[7, 8] を参照にされたい。PaCS-MD は、バイアスをかけることなく構造間の遷移経路を探索する。従来法では、構造遷移を促進するため系の温度を高温にして MD を実行するか、系の特定の部分に外力や外部ポテンシャルといった「バイアス」を付加することがほとんどである [9, 10]。従来法に対して本手法は、反応座標を設定してしまえば、遷移確率が高い初期構造を選択し、短時間 MD により構造リサンプリングを繰り返していただくだけで、バイアスは使用しない。更に、パラメーターを最適化するための準備計算も必要としない。本手法により効率的に構造探索が可能であるのは、構造リサンプリングを通して「稀にしか発生しない分子構造」の出現確率を選択的に上昇させているからである。初期構造の選択にあたり様々な反応座標が考えられる。あらかじめ遷移先の終構造が分かっている場合に関しては、終構造に対する平均自乗距離 (root-mean-square deviation: RMSD) が考えられる。例えば、構造リサンプリングにより得られるトラジェクトリを RMSD の値でランキングし、値の小さい順にスナップショットを複数個選択し、短時間 MD の初期構造として構造探索していくことにより、徐々に終構造へ近づいていき、レアイベントを抽出することができる。ここで、反応座標は任意に指定可能であり、RMSD に限らず指定できる。実際の適用では、抽出したいレアイベントを的確に記述する合理的な反応座標を指定することが重要になる。また、遷移先の構造が未知である場合、PaCS-MD の拡張も必要になる。PaCS-MD の拡張にあたり、遷移先の終構造が未知である場合、かつ、反応座標の指定ができるだけシステムに依存しないようする、という 2 点が重要になってくる。

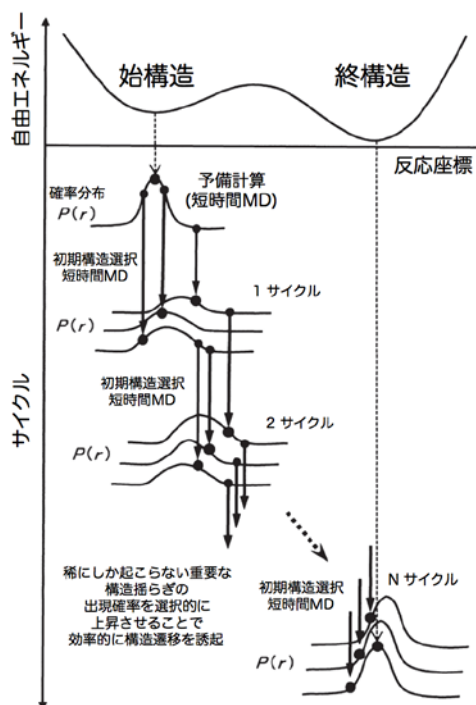


図 1: PaCS-MD の概念図

3. 手法開発 (ad-PaCS-MD)

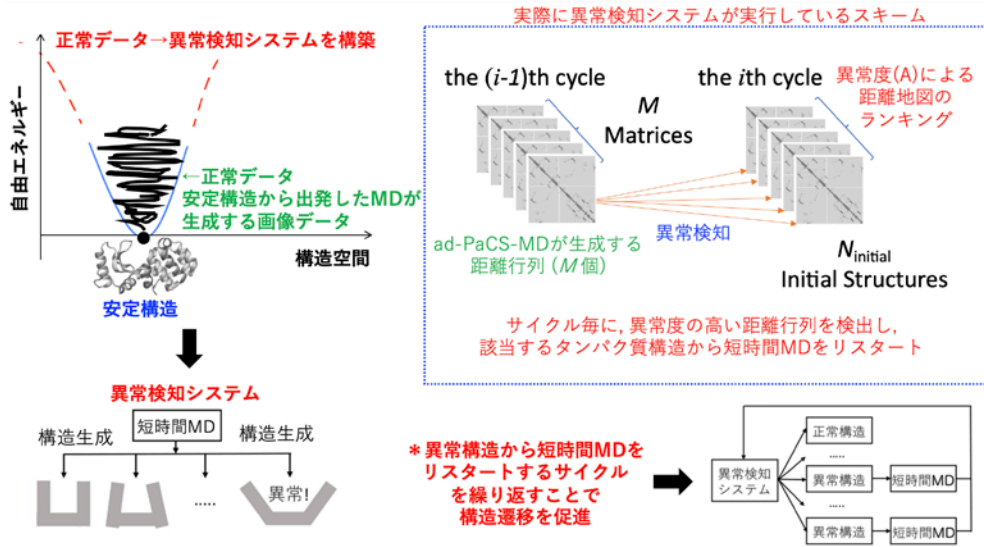


図2: ad-PaCS-MD の設計思想

本研究では PaCS-MD の拡張として、タンパク質の構造遷移を効率的に抽出する PaCS-MD と機械学習 (異常検知) を援用することで生体機能に重要なタンパク質の長時間ダイナミクス (レアイベント) を効率的に抽出する「異常検知型 PaCS-MD (ad-PaCS-MD)」を開発した (図2). PaCS-MD は、構造遷移を誘起する確率が高い初期構造を特定し、短時間の分子動力学計算 (MD) を繰り返すことで効率的にレアイベントを抽出する. 本研究では PaCS-MD を拡張し、より効率的にレアイベントを抽出可能な方法へ発展させた. 具体的には、「構造変化のしやすさ」と「構造の異常さ」を異常検知により対応付けて異常度が高い構造を特定し、PaCS-MD の初期構造に採用する. 対応付けは、タンパク質の残基間距離を要素とする距離行列を画像として蓄積して機械学習し、異常度が高い距離行列を他の準安定状態に構造遷移しやすいタンパク質構造とみなした. ad-PaCS-MD の各サイクルにおいて「異常度が高い構造 = 遷移確率が高い構造」から短時間 MD を繰り返すことで構造遷移を誘起し、構造変化 (遷移経路) を自動的に予測可能な ad-PaCS-MD を開発した. 以上の開発思想に基づき Reedbush-H 上で ad-PaCS-MD を実装した.

ad-PaCS-MD のサンプリング効率を調べるため、加水分解酵素である T4 リゾチーム (T4L) に適用して Open-Closed 構造遷移をレアイベントとして抽出できるか検証した. 具体的には T4L の Open 構造から MD を実行し、得られたトラジェクトリを機械学習したのち、異常検知システムを構築した. この異常検知システムを用いてトラジェクトリに含まれる異常度の高い (出現確率の低い) タンパク質構造を抽出し、短時間 MD を繰り返した. 結果として、T4L の Open 構造からスタートして Closed 構造へ至る遷移経路を抽出することに成功した. ad-PaCS-MD から得られたトラジェクトリをもとにマルコフ状態モデルを構築し、自由エネルギーを計算した (図3左). 自由エネ

T4L (164残基)の自由エネルギー地形

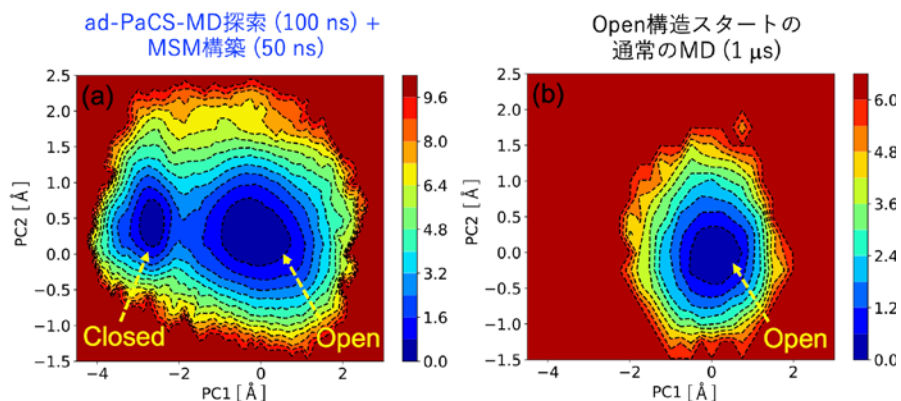


図3: ad-PaCS-MD (左) と 通常の MD (右) の探索効率の比較

ルギー地形が示すように, Open 構造からスタートして Closed に至る 2 構造間の遷移経路を抽出することができる. 通常の MD ではこのようなレアイベントを抽出することが難しいため, Open 構造からスタートしても Closed 構造に至る遷移経路を抽出することはできない (図 3 右). 以上より, 生体分子に ad-PaCS-MD を適用することで, 現実的な計算コストで効率的にレアイベントを抽出でき, 生体機能の解析に貢献することが期待される. 現在, ad-PaCS-MD を複合体の解離プロセス抽出に適用した応用研究を実施している. 複合体の場合は, タンパク質間相互作用を距離行列で的確に記述できるため, 詳細な解離課程を抽出できる.

4. まとめと今後の展望

PaCS-MD を ad-PaCS-MD として拡張することにより, 遷移先の構造が未知である場合にも適用可能になり, 出発構造の周辺に存在する順安定状態への遷移経路が自動的に抽出できるようになる. また, タンパク質の残期間距離に基づく距離行列を初期構造に採用しているため, システム依存で反応座標を定義する必要がなく, 汎用性が高い. 今後は ad-PaCS-MD を適用し, タンパク質とタンパク質, あるいは, タンパク質-基質といった複数の分子や化合物が結合して複合体を形成する秩序形成過程や解離過程といった複雑な構造変化を抽出していきたい. 拡張にあたり, 距離行列は複合体に関しても容易に定義できる.

5. 謝辞

本研究は, 東京大学情報基盤センター「若手・女性利用者推薦」2021 年度 (前期) の採択課題として, Reedbush-H を利用して実施された.

参考文献

- [1] D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y.B. Shan, W. Wrighers, *Science* 330 (2010) 341-346.
- [2] K. Lindorff-Larsen, S. Piana, R.O. Dror, D.E. Shaw, *Science* 334 (2011) 517-520.
- [3] R. Harada, A. Kitao, *J. Chem. Phys.* 139 (2013) 035103.
- [4] T. Baba, R. Harada, M. Nakano, Y. Shigeta, *J. Comput. Chem.* 35 (2014) 1240-1247.
- [5] R. Sato, R. Harada, Y. Shigeta, *Biophysics and Physicobiology* 13 (2016) 311-319.
- [6] J. Fujita, R. Harada, Y. Maeda, Y. Saito, E. Mizohata, T. Inoue, Y. Shigeta, H. Matsumura, *J. Struct. Biol.* 198 (2017) 65-73.
- [7] R. Harada, Y. Takano, T. Baba, Y. Shigeta, *Phys. Chem. Chem. Phys.* 17 (2015) 6155-6173.
- [8] R. Harada, *Bull. Chem. Soc. Jpn.* 91 (2018) 1436-1450.
- [9] G.M. Torrie, J.P. Valleau, *J. Comput. Phys.* 23 (1977) 187-199.
- [10] G.M. Torrie, J.P. Valleau, *J. Chem. Phys.* 66 (1977) 1402-1408.