

スパース表現に基づく重要因子同定法の開発と機械学習技術の生体分子シミュレーションへの応用

Development of Important Feature Identification Methods Based on Sparse Structure Representations and Applications of Machine Learning Techniques to Molecular Dynamics Simulations

山守 優

高度情報科学技術研究機構

1. 研究目的

本研究では、分子動力学 (MD) シミュレーションから得られる大規模時系列データに対して、スパース構造学習に基づく異常検知手法を応用することで、生体分子の構造変化や機能発現に関わる重要因子を自動的に抽出する解析手法の開発を目的とした。近年、低温電子顕微鏡法や AlphaFold に代表される構造解析・構造予測技術の発展により、高精度なタンパク質構造を取得することが容易となった。一方で、生体機能の理解には静的構造のみならず、構造変化や協奏的運動などの動的情報が重要である。しかし、多くの機能発現は MD シミュレーションで直接追跡可能な時間スケールを超えており、膨大なトラジェクトリデータから重要な自由度を抽出する一般的手法は十分に確立されていない。

そこで本研究では、残基間距離や内部座標を対象としたスパース構造学習ベースの異常検知法を開発し、構造変化に関与する重要因子を低コストで抽出することを目指した。さらに、得られた重要因子を主成分解析などの方法と組み合わせることで、生体分子の自由エネルギー地形や機能発現機構の解析を高度化することを目的とした。

2. 研究成果の内容

2025 年度は、スパース構造学習に基づく異常検知手法を MD シミュレーション解析へ適用するためのプログラム実装・高速化を進めるとともに、複数の生体分子系への応用を実施した。

まず、酵素タンパク質 T4 lysozyme や PDZ3 domain を対象として、二つの状態に対する 1 マイクロ秒規模の MD シミュレーションのトラジェクトリに対して開発手法を適用した結果、約 30 程度の重要残基ペアが抽出され、それらを用いた主成分解析により、状態転移に対応する主要な協奏運動を再現することに成功していた。これらの結果は、先行研究で実施された数十マイクロ秒規模の長時間 MD シミュレーションで得られていた知見を、より低い計算コストで抽出可能であることを示している。2024 年度中に投稿していたこの結果について、開発手法について論文投稿およびコード整備を進め、2025 年度中に ACS Omega 誌への論文掲載および GitHub 上でのコード公開に至った。

また、本年度には解析手法の適用範囲を拡張し、近年構造解析が進展したヒト味覚受容体 hTAS1R ヘテロダイマーへの応用を開始した。甘味受容体 hTAS1R2-hTAS1R3 および旨味受容体 hTAS1R1-hTAS1R3 を対象に、apo 状態および holo 状態の MD シミュレーションを実施した。

特に、AlphaFold2 を用いて複数の二量体モデルを構築し、モノマー状態および二量体状態における構造揺らぎを比較解析した結果、二量体形成によって結合部位がより開いた構造をとることが示唆された。さらに、異常検知解析によって、機能的構造変化に關与する重要残基間距離の候補を抽出した。これらの成果は、生物物理学会 2025 年度年会において発表を行った。

これにより、本研究で開発した解析法が、MD シミュレーションに対する重要因子発見法として有効であることを示した。

3. 学際共同利用プログラムが果たした役割と意義

本研究では、Pegasus および Miyabi-G といったスーパーコンピュータを活用し、大規模 MD シミュレーションおよび解析計算を実施した。

対象とした生体分子系は数十万から百万原子規模に達し、particle mesh Ewald 法を用いた長距離静電相互作用計算を含む大規模並列計算が必要であった。また、複数状態 (apo/holo 状態など) および複数条件に対してマイクロ秒スケールの MD シミュレーションを実施する必要があるため、研究室レベルの計算機環境では対応が困難であった。学際共同利用プログラムによって提供された計算資源により、長時間 MD シミュレーションを多数実行するとともに、大規模トラジェクトリ解析を並列的に進めることが可能となった。その結果、構造変化に關わる重要因子を高精度かつ低コストで抽出する解析技術の有効性を検証できた。

さらに、本プログラムを通じて、MD シミュレーション、機械学習、統計解析を横断する学際的研究を推進できた点も大きな意義である。特に、異常検知やスパース構造学習といったデータ科学的手法を生体分子シミュレーションへ応用することで、新しい MD データ解析基盤の構築へ繋がった。

4. 今後の展望

今後は、2026 年度の継続課題として、本研究で開発した重要因子同定法をさらに高度化し、より大規模かつ複雑な生体分子系への適用を進める予定である。

特に、味覚受容体複合体について、apo/holo 状態間の比較解析を継続し、特定のヘテロダイマーのみが機能を示す機構について、構造ダイナミクスの観点から説明することを目指す。また、異常検知によって抽出された重要因子を Collective Variable として Metadynamics などの効率的サンプリング法へ組み込むことで、自由エネルギー計算の高精度化と高速化を進める。

5. 成果発表

(1) 学術論文

(2) 学会発表

Yamamori, Y., Morita, R., Shigeta, Y., Harada, R., “Molecular dynamics study on the structural basis and mechanism of recognition of human taste receptor heterodimers,” 第 63 回生物物理学会年会, ポスター発表, 2025 年 9 月 24 日-26 日.

(3) その他

使用計算機	使用計算機に○	配分リソース※		
		当初配分	移行*	一般利用による追加
Pegasus	○	17,600		-
Miyabi-G	○	49,500	-	-
Miyabi-C	-	-	-	-
※配分リソースについてはノード時間積をご記入ください。 *バジェット移行を行った場合、「+2000」「-1000」のように記入				