

プロジェクト名(タイトル):

分子動力学シミュレーションを用いた生体分子の構造および反応機構の解明

利用者氏名:

杉田 有治(1)、○伊東 真吾(1)、大出 真央(1)、大滝 大樹(1)、種村 潔人(1)、新津 藍(1)、本橋 昌大(1)、森 貴治(1)、八木 清(1)、Haeri Im(1)、Hisham Dokainish(1)、Weitong Ren(1)、Yaokun Lei(1)

理研における所属研究室名:

(1)理化学研究所 開拓研究本部 杉田理論分子科学研究室

2次元自由エネルギー地形の解析によるトリプトファン合成酵素におけるアロステリーの解明

(担当:伊東)

1. 本プロジェクトの研究の背景、目的、関係するプロジェクトとの関係

トリプトファン合成酵素 (TRPS) は、 α および β サブユニットからなる酵素複合体である(図1)。 α サブユニットにおいてインドール-3-グリセロールリン酸(IGP)よりL-Trp合成の反応物であるインドールを生成し、 β サブユニットでは生成されたインドールとL-セリン(L-Ser)を用いてL-トリプトファン(Trp)を生成する。 α サブユニットの活性中心にIGPが結合することにより、 β サブユニットの構造および酵素活性が変化する“アロステリー”を有することが知られている。このようなアロステリーを持つ酵素の反応メカニズムは酵素反応の基礎的な理解を深めるためだけでなく、創薬の面からも長年研究されてきたが未だによく解明されていない。

本研究は、TRPSの酵素反応における複雑なアロステリーの仕組みをUmbrella Sampling(US)法を用いた分子動力学計算(MD)による2次元自由エネルギー面の解析から、原子レベルでの解明を目指すものである。TRPSの α および β サブユニットは、各サブユニットへのリガンドの結合状態でOpenおよびClosedの2状態の構造を持つことが知られている[M. F. Dunn, Arch. Biochem. Biophys., 2013]。我々はリガンドの結合状態の異なる3つの状態(apo, IGP結合, and IGPおよび β リガンド結合)においてUS-MDで各サブユニットの開き具合をRMSDで定義したものを反応座標として平均力ポテンシャル(PMF)を計算し、各状態でどのような構造が安定な構造であるかを解析した。またサブユニット間の水素結合ネットワークの変化を解析し、片側のサブユニットの構造変化がどのように反対側のサブユニットの構造変化に及ぼすメカニズムについて解明に努めた。

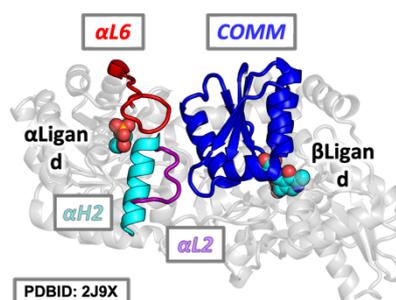


図1 TRPSの構造(PDBID:2J9X)

α サブユニットと β サブユニットから構成される。

2. 具体的な利用内容、計算方法

Apo (PDBID: 1K8X)、IGP結合 (PDBID: 1WBJ) および IGP/ β リガンド結合 (PDBID: 2J9X) 状態を初期条件とし、AMBERFF14SB/GAFF2 力場を用いて US-MD 計算を行った。また US の反応座標として、塩橋を形成する β R141 と β D305 の N-O 間原子距離および β R141 と β S299 の N-O 間原子距離の線型結合を利用した。本計算後、Multi-state Bennett Acceptance Ratio (MBAR) 法による Reweight を用いることで、上記の反応座標と結晶構造を Reference とする RMSD を反応座標として加えた2次元の自由エネルギー地形を解析した。

3. 結果

本計算により得られた2次元 PMF を図2に示す。得られた PMF より、IGP結合状態において一般に Open 構造と呼ばれる構造より大きく β サブユニットが開いている構造が安定な状態(図2中金丸)として存在することが判明した。また Open 構造と大きく開いた構造に対して水素結合解析を行い比較することで両者においてサブユニット間水素結合ネットワークが異なることを発見した(図3)。

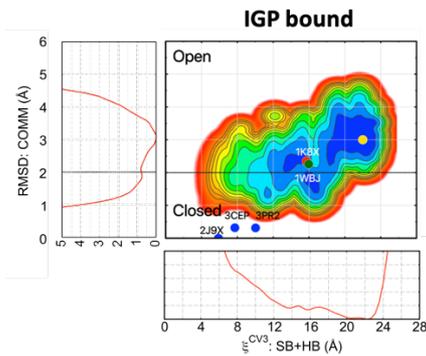


図2 塩橋の原子間距離および Closed 構造を reference とした RMSD を軸とする2次元自由エネルギー地形
金色の丸印が今回発見した大きくβサブユニットが開いた構造となる。

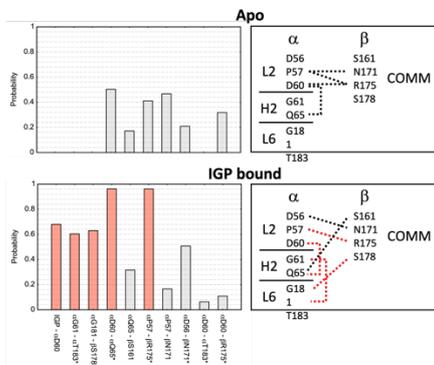


図3 Apo および IGP 結合状態に対するサブユニット間水素結合ネットワーク

4. まとめ

我々の研究によって、IGP 結合体に結晶構造解析では確認されていないβサブユニットが大きく開いた構造が安定に存在することが示唆された。このような構造はβリガンドを活性中心に導く経路を作ることに役立つことから妥当な構造であると考えられる。また IGP がαサブユニットに結合することでα/βサブユニット間の水素結合ネットワークが変化し、この変化が TRPS のアロステリーに深く関わっているのではないかということ突き止めた。

5. 今後の計画・展望

本研究における成果に関して現在論文を投稿中である。また、本計算において古典計算から TRPS のアロステリーのメカニズムについて迫ったが、αおよびβサブユニットで起きる反応に関してはよくわかっていない。このような化学反応を理解するためには量子化学(QM)計算が必要であり、次のステップとして密度汎関数法レベルでの MD 計算を用いてα/βサブユニットにおける化学反応の解明を目指す。

非線型な手法を用いた蛋白質動態の次元削減

(担当:大出)

1. 本プロジェクトの研究の背景、目的、関係するプロジェクトとの関係

N 粒子から構成される蛋白質の自由度は $3N-6$ であり、数万原子で構成される蛋白質分子の動態は超高次元空間上に分布する関数として表現される。しかしながら、大抵の場合蛋白質の構造変化を記述する本質的な自由度は $3N-6$ よりもずっと少ないため、蛋白質構造変化の分布関数を、2,3次元の低次元空間へ適切に射影することでその動態を直観的かつ定量的に理解することができる。

蛋白質動態の次元削減手法には主成分分析(PCA)や time-lagged Independent Component Analysis (tICA)がよく用いられる。これらは計算コストが低い線形な手法でありながら多くの蛋白質の動態をよく記述できることが知られているものの、例えばフォールディング/アンフォールディングのような複雑な動態に対してはあまり有効でない。一方で、t-distributed Stochastic Neighborhood Embedding (t-SNE) に代表される非線型な手法は複雑な動態にも適用できると期待されるものの、非線型な手法では一般にデータ量に対して指数関数的に計算量が増加するため、数万から数十万のサンプリングされた構造を扱う MD 計算に対しては応用があまり現実的ではないという問題点があった。

2018 年に発表された非線型次元削減手法 UMAP [McInnes and Healy, *arXiv*, 2018]は t-SNE と同程度以上の性能を有しながら極めて高速に動作するという利点を持つ。そこで本課題では UMAP の蛋白質動態次元削減への応用可能性を検討するべく大規模な構造変化を示す複数の系について分子動力学(Molecular Dynamics: MD)シミュレーションを行い、UMAP およびそのほかの手法による動態の解析を行った。

2. 具体的な利用内容、計算方法

本課題については前年度に protein G B1 ドメインおよび Src キナーゼ Sh3ドメインの折れ畳み MD シミュレーションを行っており、今年度は新たにアデニル酸キナーゼ (Adenylate Kinase: ADK)ドメイン運動のシミュレーションと MD トラジェクトリーの解析を実施した。ADK については主鎖 Cα 原子レベルでの粗視化モデルを用い、MD 計算ソフトウェア GENESIS によってシミュレーションを行った。

3. 結果

ADK は図4に示す通り 3 つのドメイン (Core, LID, NMP) からなる蛋白質である。酵素としての機能発現時には、まず LID が Core ドメインについて閉構造(Semi-closed)を取り、次に NMP が閉構造を取ることで完全な閉構造(Closed)に至る、という逐次的な構造変化を示すことが知られている。ADK MD トラジェクトリーを UMAP によって二次元平面へ射影したところ、主に 3 つの安定な状態からなる動態分布が得られた(図5)。これらはそれぞれが Open, Semi-closed, Closed に対応しており、従って UMAP は ADK の動態を極めてよく記述できたと言える。

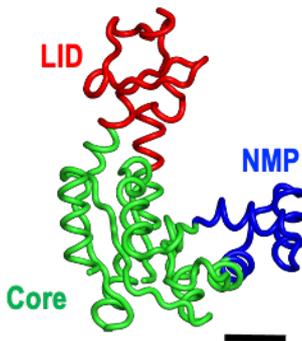


図4 ADK の各ドメイン

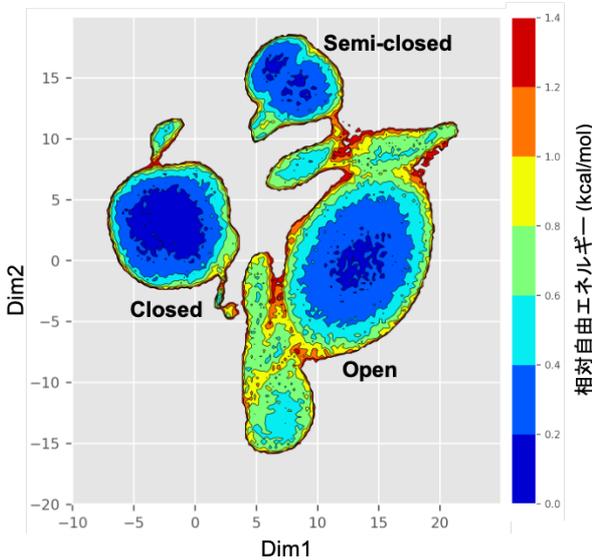


図5 ADK 動態分布

4. まとめ

粗視化モデルを用いて蛋白質ドメイン運動の MD シミュレーションを行い、得られたトラジェクトリーを UMAP によって解析した。UMAP での次元削減により得られた自由エネルギー地形は ADK の機能発現に伴う構造変化過程をよく記述できており、UMAP は MD 計算で得られた蛋白質動態の解析に効果を発揮する手法であると期待される。本研究成

果は第 21 回日本蛋白質科学会年会、第 59 回日本生物物理学会年会において発表を行い、また学術論文への投稿を予定している。

5. 今後の計画・展望

今後は UMAP による次元削減の実際的な応用可能性を検討するべく、その他の系、特に粗視化でない全原子 MD シミュレーションについて解析を試みる予定である。

De novo 設計膜貫通 α ヘリックスペプチドポアの MD 計算による構造サンプリング

(担当:新津)

1. 本プロジェクトの研究の背景、目的、関係するプロジェクトとの関係

膜タンパク質の de novo 設計は、天然タンパク質やその改変体を持つ機能・安定性を越えた新しい分子を得る手段となると共に、膜タンパク質のアミノ酸配列-立体構造相関を知ることができる。本研究では膜タンパク質の最も基礎的な構造である α ヘリックスペプチドポアを対象とし、実験結果と合わせて動的構造と機能評価を実現することを目的とする。これまでに de novo 設計した膜貫通 α ヘリックス CCTM ペプチドは、分析超遠心の結果から疎水環境で 5 量体を形成することが明らかとなっている。また一分子チャンネル電流測定と構造モデリングの結果から、脂質二重膜中では 5 量体 Single Channel state と 10 量体以上のいくつかのサイズのポアを形成する Multiple conductance state の 2 状態を電位依存的にとることが示唆されている(図6)。しかしながらこのようなポアの大きい構造変化がどのようなメカニズムで起こるかはわかっていない。そこで本プロジェクトでは、実験の時間・空間分解能では確認できないポアの構造と会合数変化を全原子分子動力学計算によりサンプリングする。これにより CCTM ペプチドのアミノ酸配列が脂質二重膜中でとりうる構造とそこに存在するペプチド間、ペプチド-脂質間相互作用を明らかにし、今後の α ヘリックスを基盤とした膜タンパク質の de novo 設計に資することを目指す。

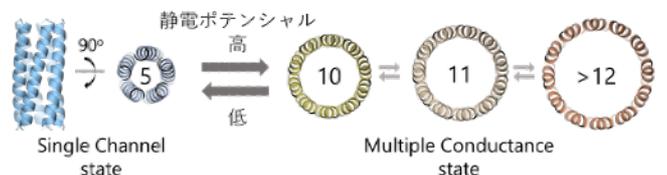


図6 CCTM ペプチド 5 量体のモデル構造と脂質二重膜中での構造変化の概要

2. 具体的な利用内容、計算方法

CCTM ペプチド 5 量体、7 量体、13 量体について conventional MD 計算を 1 マイクロ秒ずつ実施した。脂質とペプチドのダイナミクスを向上させるため、系の温度は 353K とした。ペプチドポアの初期構造は昨年度のプロジェクトで最適化した構造モデルを用いた。力場は CHARMM36 を使い、実験条件とそろえるために脂質は DPhPC、イオンは 1M KCl、ペプチドポアを含む計算ボックスを CHARMM-GUI を利用して作成した。

3. 結果

13 量体の計算結果の分析から、13 量体ポアは 6 量体および 7 量体に分割することがわかった。13 量体ポア構造モデルの理論コンダクタンス、および分割後の理論コンダクタンスは一分子チャンネル電流測定の実験コンダクタンスと矛盾しない結果となった。また分割する際には分割面から遠いペプチドが脂質の垂直方向に対して大きく傾くことから、脂質の疎水領域の厚みと CCTM ペプチドの疎水領域の長さが構造変化の鍵となることが示唆された。

4. まとめ

CCTM ペプチドポアの構造と会合数の変化について、高温での全原子分子動力学計算を実施した。特に 13 量体ポアの計算結果から、構造変化のメカニズムは、ポアの分割とヘリックスの大きな傾きによるものであると考えられた。

5. 今後の計画・展望

CCTM ペプチドポアの構造変化をアミノ酸配列の最適化により制御することができれば、これまで成功例のない電位依存性ペプチドチャンネルの開発に貢献できると期待される。今後は系に外部電場を導入した際の効果や CCTM ペプチド改変体の MD 計算を実施し、新たなペプチドチャンネルを作成していく。

分子動力学計算による F₁-ATPase のリン酸放出経路の探索

(担当:本橋)

1. 本プロジェクトの研究の背景、目的、関係するプロジェクトとの関係

F₀F₁-ATP 合成酵素の可溶性部分である F₁-ATPase は、ATP 合成・加水分解と共役し回転運動を行う回転分子モーターである。六量体リング $\alpha_3\beta_3$ に回転子 γ が突き刺さった

構造をしており (図7)、 $\alpha_3\beta_3$ に対して γ が 120° 回転するとともに 1 つの ATP を加水分解する。120° 回転はさらに 80° と 40° のサブステップに分けることができ、80° と 40° サブステップ前のポーズは、それぞれ ATP-binding dwell state、catalytic dwell state と呼ばれている。F₁-ATPase の反応スキームのほとんどは 1 分子計測により解明されたが、加水分解の生成物である無機リン酸 (Pi) の放出経路は未だ明らかではない。近年、クライオ電子顕微鏡法 (Cryo-EM) により明らかとなった好熱菌 F₁-ATPase (TF₁) の構造 [Sobti et al., Nature Communications, 12, 4690 (2021)] では、ADP・Pi 状態の触媒部位から Pi がミオシンに見られるような back door を介して放出される可能性が指摘された。本研究では、この TF₁ の構造を用いた分子動力学 (MD) 計算により Pi の放出経路とそのタイミングの特定することを目指す。

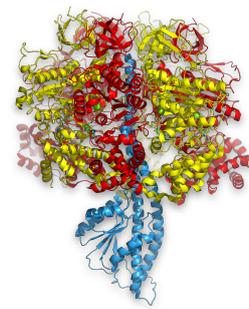


図7 TF₁の構造

α サブユニット(赤色)、 β サブユニット(黄色)、 γ サブユニット(水色)の $\alpha_3\beta_3\gamma$ 複合体で形成される。

2. 具体的な利用内容、計算方法

TF₁ の post-binding dwell state (PDBID: 7L1Q) と catalytic dwell state (PDBID: 7L1R) を初期構造とし、分子動力学ソフトウェア GENESIS を用いた MD 計算を行なった。タンパク質の力場には CHARMM36m を使用し、ATP, ADP, Pi にはパラメータを修正した CHARMM 力場[Y. Komuro et al., Journal of Chemical Theory and Computation 10, 4133-4142 (2014)] を使用した。TF₁ の Cryo-EM 構造から、Pi は 80° サブステップ回転に伴い放出されると予想される。そこで、まず初めに、post-binding dwell state (start) から catalytic dwell state (end) への Targeted MD を実行し、80° サブステップ回転の中間状態構造 (images) を取得した。その後、この中間状態に対して、Pi を solute region とした Generalized Replica Exchange with solute Tempering (gREST) を行い、Pi のトラジェクトリーを得た。

3. 結果

Solute region の Pi を 2790K で加熱し、Pi と ADP の β -リン酸の距離の時間変化を解析した(図8)。その結果、end 構造に近い image は容易に Pi が放出されることがわかった。さらに image3, image4 では end 構造の場合とは異なる経路による放出が確認された。

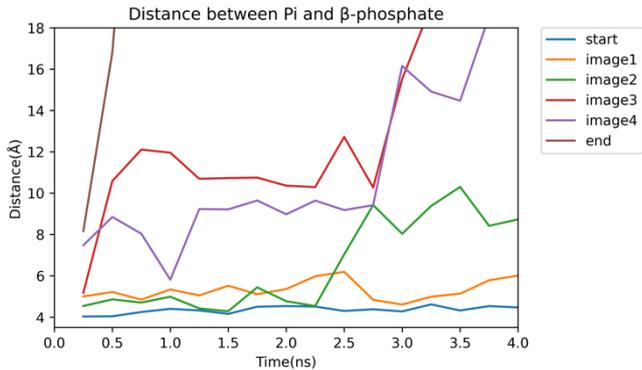


図8 Pi と ADP の β -リン酸の距離の時間変化

4. まとめ

TF1 の Pi 放出経路とタイミングを特定するために、Targeted MD と gREST 法を用いた MD 計算を行なった。Pi を加熱し、タンパク質や MgADP との相互作用を小さくすることで、計算可能な時間スケールで Pi の放出を確認できた。さらに、得られた経路は image3,4 と end では異なるものだった。

5. 今後の計画・展望

Target MD は人工的な力によって始状態から終状態へ構造変化させるため、得られる中間状態の構造にはバイアスが存在する。今後は、String 法による正確な中間状態の取得や別の手法を用いたシミュレーションにより、詳細な Pi の放出経路の探索を行う予定である。

MD 計算による *De Novo* モデリング構造の最適化

(担当: 森)

1. 本プロジェクトの研究の背景、目的、関係するプロジェクトとの関係

近年、構造生物学において、クライオ電子顕微鏡を用いた単粒子解析が盛んに行われており、実験により得られたタンパク質の密度マップから、分子構造を精密にモデリングすることが重要となっている。対象とするタンパク質が新規構造の場合、*De Novo* モデリングが必要となる。*De Novo* モデリングでは計算科学に基づく手法が広く用いられ、これまでに ROSETTA や MAINMAST などのさまざまな方

法が開発されてきた。MAINMAST では主鎖構造を予測するため、全体構造をモデリングするためには、そこから側鎖を生成し、さらに構造を最適化する必要がある。一方、このような過程では、キラリティーエラーや芳香環貫通、シスペプチドなどの構造エラーが発生しやすく、*De Novo* モデリングでは多数の候補構造を生成するため、効率良く構造エラーを解消して高速に構造を最適化し、最も確からしいデコイを選ぶプロトコルが必要となる。本研究では、MAINMAST により得られた *De Novo* モデリング構造に対して、MD 計算に基づくフレキシブル・フィッティングを行い、最適な構造最適化プロトコルを確立することを目指した。

2. 具体的な利用内容、計算方法

本研究では、*De novo* 構造モデリングにより得られた $C\alpha$ モデルを効率よく最適化する方法として「SAUA-FFR 法」を提案した。これは $C\alpha$ モデルから側鎖を生やした後、融合原子モデルを用いてシミュレーテッド・アニーリング MD 計算を繰り返す方法である。本研究では、異なる8つの系 (F420-reducing hydrogenase α subunit (FrhA), 20S proteasome core subunit (PCS), Sputnik virophage (SV), Bordetella phage (BPP-1), transient receptor potential vanilloid 1 (TRPV1), CARD domain of mitochondria anti-viral signaling protein (MAVS), bombyx mori cytopovirus 1 (BmCPV-1), and porcine circovirus 2 (PCV2)) を対象として、電顕フレキシブル・フィッティングを行い、全原子モデルと融合原子モデルの比較を行った。

3. 結果

CHARMM C19 力場の融合原子モデルと implicit solvent EEF1 モデルを用い、各タンパク質 500 個のデコイに対して 100 ps \times 5 サイクルの最適化計算を行なった。その結果、CHARMM C36m 力場の全原子モデルと GB/SA モデルを用いた方法 (SAAA-FFR) と比べて二次構造の形成が促進され、マップとの相関係数が増大するとともにデコイは正解構造により近づくことが分かった(図9)。これは、融合原子モデルでは水素が省略されているため、全原子モデルと比べてタンパク質内部の構造に隙間があり、最適化の際に原子が動きやすいためであると考えられる。また、フレキシブル・フィッティングでは分子構造がマップに一致するような強い拘束をかけるため、フィッティングの際

に構造エラーを自動的に解消することが難しく、フィッティングの前にあらかじめエラーを取り除いておくことが重要であることがわかった[T. Mori, G. Terashi, D. Matsuoka, D. Kihara, and Y. Sugita, *J. Chem. Inf. Model.*, **61**, 3516–3528 (2021)].

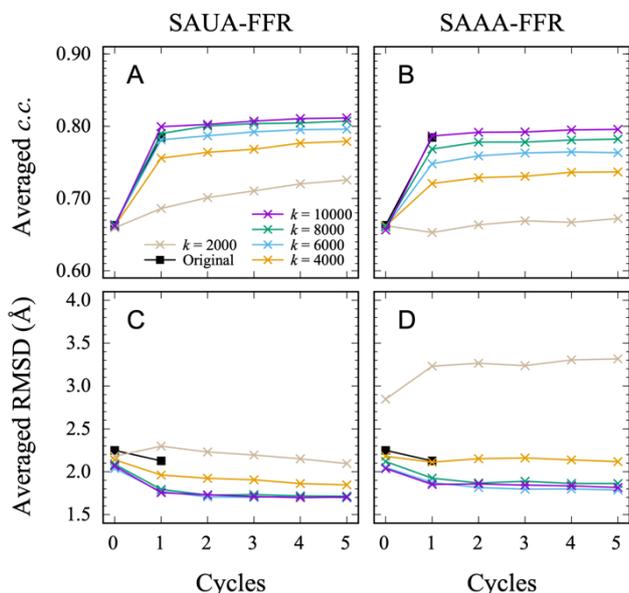


図9 PCV2 に対する SAUA-FFR と SAAA-FFR による構造最適化計算の比較 (A, B):最終構造において相関係数上位5つのモデルの相関係数の変化, (C, D) 相関係数上位5つのモデルの RMSD の変化

4. まとめ

フレキシブル・フィッティングに基づく構造最適化法として、融合原子モデル、implicit solvent モデル、シミュレーション・アニーリング MD を用いる SAUA-FFR 法を提案した。Do novo モデリング構造に対して手法を適用したところ、全原子モデルと比べて2次構造の形成が促進され、正解構造により近づくことがわかった。SAUA-FFR 法により、精密な構造モデリングが可能になると期待している。

5. 今後の計画・展望

フレキシブル・フィッティングでは、バイアス項の力の定数 k の選び方がしばしば問題になる。実験で得られる電顕密度マップには、ノイズや局所低解像度領域が含まれるため、フレキシブル・フィッティングで強い k を用いると、overfitting によって分子構造が歪む可能性がある。今後は、このような overfitting を軽減するようなプロトコルの確立を目指す。

分子動力学計算プログラム GENESIS における QM/MM 法の開発と応用

(担当:八木)

1. 本プロジェクトの研究の背景、目的、関係するプロジェクトとの関係

QM/MM 法は、興味ある空間領域を高精度な量子化学 (QM) 計算で扱い、周囲環境を古典力場 (MM) で扱うハイブリッド法である。杉田グループでは、分子動力学 (MD) 計算プログラム GENESIS (<https://www.rccs.riken.jp/labs/cbrt>)を開発している。最近、我々は GENESIS に QM/MM 法を導入した [K. Yagi, K. Yamada, C. Kobayashi, and Y. Sugita, *J. Chem. Theory Comput.* **15**, 1924 (2019)]. 本課題では、GENESIS における QM/MM 計算の機能を拡大し、酵素反応に対する応用計算を実施する。

2. 具体的な利用内容、計算方法

【QSimulate との連携】最近、塩崎らは並列性に優れた量子化学計算プログラム、QSimulate (<https://qsimulate.com>)を開発した。我々は、GENESIS と QSimulate をライブラリー接続し、並列性を損なわないインターフェースを開発した。これにより、DFTB 法で数 ns/day、DFT 法で数十 ps/day の高速な QM/MM-MD 計算が可能になった(図10)。

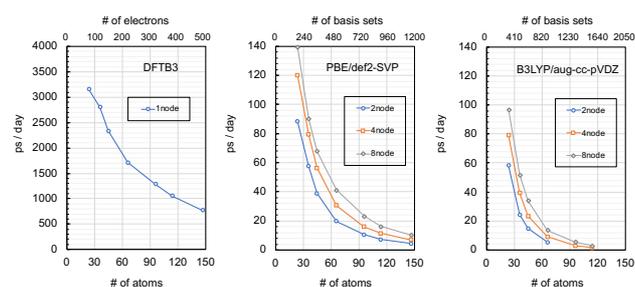


図10. HOKUSAI-BWで測定したGENESISとQSimulateによるQM/MM-MD計算のパフォーマンス。

【反応経路探索法の実装】String 法による最小エネルギー経路 (MEP) 探索を GENESIS に実装した。String 法は既知の反応物と生成物の 2 点をつなぐ経路をイメージの数珠つなぎで表現し、イメージを発展させることで MEP を求める chain-of-replica 法の 1 つである。String 法では、各イメージを①gradient の方向へ動かす、②等間隔に整列しなおす、という操作を繰り返すことで MEP が得られる。

【酵素反応への応用】 開発した方法を triosephosphate isomerase (TIM)による酵素反応へ応用した。TIM は、dihydroxyacetone phosphate (DHAP) から glyceraldehyde 3-phosphate (GAP)を生成する4つのプロトン移動反応を触媒する酵素である。これらの MEP を求め、さらにアンブレラサンプリング法により反応自由エネルギーを計算した。

3. 結果

図11に、本研究により得られた TIM の酵素反応の反応物(I)、中間体(II~IV)および生成物(V)を示す。String 法により、I~VをつなぐMEPを求めた。それぞれの経路に対し16イメージを用い、5thread x 128 MPI の計算資源により、全経路を2.6時間で求められた。その後、アンブレラサンプリング計算と自由エネルギー(PMF)計算では、64 ウィンドウに対して、それぞれ12 ps (時間刻み0.5 fs × 24,000ステップ)の MD 計算を実行した。このような計算は非常に大きな資源を必要とし、従来の方法では計算できなかったが、開発した GENESIS/ Q-Simulate プログラムと HOKUSAI-BW を用いることで、約1,500 ノード時間(6万コア時間)で実行することができた。

図10に MEP と PMF のエネルギープロフィールを比較する。各中間体の相対的な順位には変化がないが、PMF ではエネルギーの絶対値が下がる。実験的には、この反応の見た目の障壁は15 kcal/molであり、PMFで得られた障壁とよく一致する。

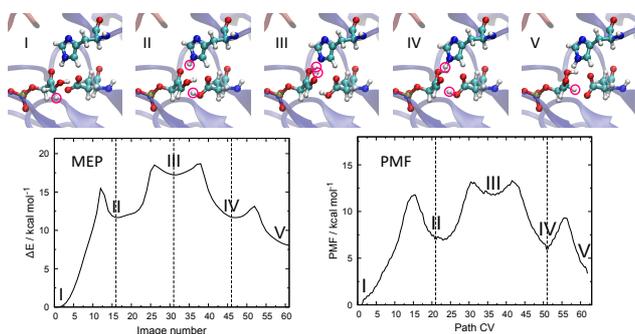


図11. QM/MM 計算により得られた TIM による4つのプロトン移動反応の中間体の構造(上)と反応経路に沿ったポテンシャルエネルギー(MEP)と自由エネルギー(PMF)の変化。量子化学計算のレベルは B3LYP-D3/aug-cc-pVDZ。

4. まとめ

本課題では、GENESIS の新しい QM/MM 機能として、Q-Simulate とのインターフェースを作成し、これを TIM の酵素反応に対し応用した。MEP とそれに沿った自由エネルギーを計算し、実験との良い一致を得ることができた。本研究の成果は論文として発表した[K. Yagi, S. Ito, and Y. Sugita, *J. Chem. Phys. B* **125**, 4701 – 4713 (2021)]。

5. 今後の計画・展望

今後は TIM 以外の様々な酵素反応に対する応用計算を計画している。また、同時に自由エネルギー計算の計算負荷を減らす新しい近似法の開発にも取り組む。

Investigation of structural characteristics of Hero proteins using MD simulations

(担当: Im, Niitsu)

1. Background and purpose of the project, relationship of the project with other projects

Most proteins denature and aggregate under stress conditions, such as near-boiling temperatures, drying and freezing, and high salinity. In cells, molecular chaperones regulate protein folding and prevent aggregation or misfolding of client proteins, protecting proteins after such stress shocks. Recently, some of the intrinsically disordered proteins (IDPs) are also found to function as “molecular shields”. Heat-resistant obscure (Hero) proteins are identified in Human and Drosophila, predicted to be IDPs, and remain soluble even after boiling at 95°C. In vivo experiments have demonstrated that Hero proteins protect proteins from denaturation under stress conditions. Hero proteins can also block the aggregation of several types of pathological proteins in cells and Drosophila strains modeling neurodegenerative diseases. These functions of the Hero protein as a “molecular shield” are unique and likely to be biologically significant. However, the molecular mechanism of their function is unknown.

In this project, we perform MD simulations on all atoms to understand the properties of the Hero proteins and investigate and compare the structural and thermodynamic properties of the Hero proteins.

Furthermore, we intend to demonstrate how the features of the amino acid sequence of Hero proteins relate to the properties of the disorderly structure, and the function of preventing denaturation and aggregation of client proteins.

2. Specific usage status of the system and calculation method

We performed all-atom MD simulations at 300K and 0.15M NaCl for Hero proteins.

We employed the CHARMM36m force field for Hero proteins and TIP3P model for water. We used GENESIS to perform 100 ns MD simulations for Hero proteins, starting from the AlphaFold2 predicted structures.

3. Result

The predicted atomistic structure of Hero9 holds an α -helix around the C-terminal region (Fig 12). During the 100ns MD simulation, the helicity was maintained high. On the other hand, the N, C-terminal regions of the Hero9 peptide were disordered. Compared to the C-terminal region, the N-terminal region appears to form interactions with other amino acids.



Fig 12. AlphaFold2 predicted structure for Hero9.

4. Conclusion

In this project, we investigate the structural and thermodynamic properties of Hero proteins by using all-atom MD simulation. We performed 100 ns MD simulations for the Hero proteins using GENESIS. During 100 ns MD simulations, the helix of the Hero9 protein was stable, but the N-terminal region of the protein was disordered and exhibited interactions with other amino acids in the structure.

5. Schedule and prospect for the future

We will further perform MD simulations of Hero9 to understand the relationship between dynamic structural properties and functions. We will also conduct simulations of other Hero proteins, and compare structural features and differences among Hero protein family.

Exploring Conformational Space of Phosphoglycerate kinase in the presence different substrate

(担当:Ren)

1 Background and purpose of the project, relationship of the project with other projects

Phosphoglycerate kinase is a transferase presenting in all living organisms which catalyzes the reversible transfer of a phosphate group from 1,3-bisphosphoglycerate (1,3-bPG) to ADP producing 3-phosphoglycerate (3PG) and ATP. The enzyme is a two-domain protein. 1,3-bPG and 3PG bind to the N-terminal domain, while the nucleotide substrates, MgADP or MgATP bind to the C-terminal domain. Experiments have shown that the substrate binding to PGK will trigger a large structural transition from open conformation to fully closed state. Recently, our collaborators found that the ligand-binding cooperativities between adenine-nucleotides and 3PG is dependent on the type of adenine-nucleotides by NMR spectroscopy. MgADP and 3PG bind to hPGK with negative cooperativity, whereas MgAMPPNP, a nonhydrolyzable ATP analogue, and 3PG bind to it with positive cooperativity.

Previously, we have applied replica path sampling to study the structural transition of 3PG and adenine-nucleotide bound PGK. In this project, we aim to investigate the conformational dynamics of PGK in the presence and 3PG and different adenine-nucleotides (MgADP and MgATP respectively) by conventional atomistic MD (cMD) simulations and decipher the mechanisms of the negative binding cooperativity between 3PG and MgADP and positive binding cooperativity between

3PG and MgATP.

2 Specific usage status of the system and calculation method

We performed multiple independent cMD simulations of wild type PGK in the presence of 3PG and different nucleotide substrates (MgADP or MgATP). For each system, we ran 10 trajectories starting from the open and closed state respectively. Each cMD simulation lasted 50ns. In addition, to examine the effect the D374N mutation on the binding cooperativities between 3PG and nucleotide substrates, we also conducted similar simulations of D374N mutant. The total simulations time was about 2 μ s.

3 Result

Our simulations revealed the intrinsic flexible dynamics of PGK. Starting from the open state, 3PG and nucleotide substrate bound PGK remains in the open form within 50ns. Large conformational fluctuation of the enzyme can be observed which can induce more open conformations as shown in the bottom panel of Figure 13. On the other hand, the fully closed state is unstable for substrates-bound PGK. A domain motion occurred very quickly during the cMD simulations to induce the conformational transition from the fully closed state to a half-closed state (top panel of Figure 13). Moreover, MgATP induced a more compact half-closed state compare to MgADP, which might be related to the different binding cooperativity between 3PG and adenine-nucleotide. In addition, we found that the charge complementary among 3PG and nucleotide in the PGK active site play an important role in the cooperative binding of substrates.

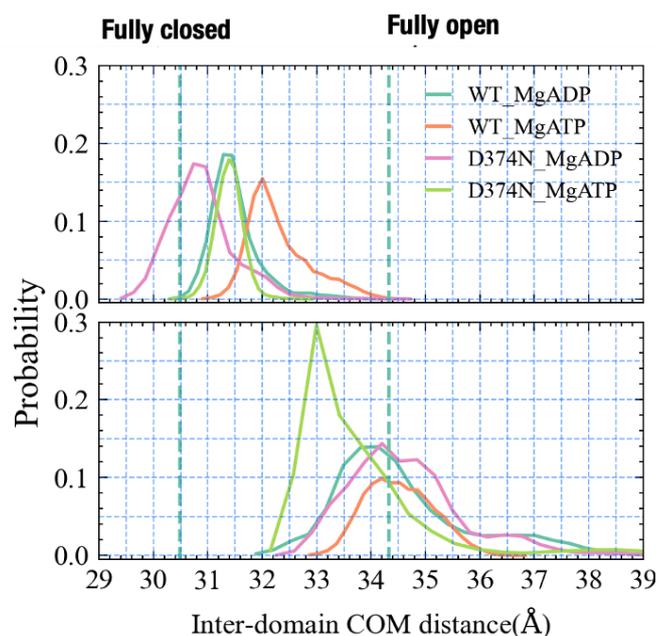


Figure13: Distribution of the inter-domain distance of wild type and D374N mutated PGK in the presence 3PG and nucleotide substrates at open (top) and semi-closed (bottom) state respectively.

4 Conclusion

- Our simulations indicate that substrates-bound PGK shows large conformational fluctuation in the open state.
- We found that the fully closed state is unstable even in the presence 3PG and adenine-nucleotides. Instead, a stable half-closed state is firstly revealed in our simulations.
- We show that charge complementary between 3PG and nucleotides at the active site of the enzymes determines the structures of the half-closed state. The D374N mutation which alter the electrochemical environment at the active site dramatically affect the communication between 3PG and adenine-nucleotide.

Exploring Catalytic Mechanism of Enzymatic Reactions by Multiscale Machine Learning Model

(YaoKun Lei)

1. Background and purpose of the project, relationship of the project with other projects:

Chemical reaction involves the motion of electrons so that quantum calculation is necessary for simulation. But it is well known that quantum mechanics (QM) calculation is very time-consuming. One solution is to fit the quantum results by parametric model. The classic force field follows this route but cannot be applied to chemical reaction due to the simplicity of energy terms. To keep the balance between accuracy and time cost, machine learning (ML) model known for its powerful potential to fit any continuous function is used to fit energy and force generated by QM calculation. But most existing models are single-scale model in which the coordinates of all atoms are used as input to model to predict the total energy and force of the system. But for enzymatic reaction system, the input dimension is extremely large due to its system size, resulting in huge memory cost and high computation complexity. In addition, the common reference method applied to enzymatic reactions is QM/MM method where energy is composed of potential of QM and MM region and interaction between two regions. The last two terms are calculated by classic force field which is a simpler parametric model than ML model. Therefore, it's a waste of resources to fit the results of a simpler parametric model by another complex one. Therefore, to apply machine learning model to enzymatic reaction. Here we adopt a multi-scale model called FieldSchNet in which only force, energy and charge of QM region is predicted by ML model and the rest two terms are processed by classic force field. Therefore, the input to ML model is just the coordinates of QM region and the external electric field applied on each QM atom so that the input's dimension is much less than single scale model. We integrate machine learning model with GENESIS to simulate chemical systems. This method is firstly

applied to a text book reaction, the substitution between Cl⁻ and CH₃F for test.

2. Specific usage status of the system and calculation method

Most of the calculations were performed by the end of the previous fiscal year. Wherein, 100 picoseconds GaREUS simulation was performed to sample the configurations used as training dataset using 32 replicas.

3. Result

We train the model with the training dataset collected by GaREUS sampling. The resulting mean absolute value of prediction error of energy and force is respectively 0.879 kcal/mol and 0.318 kcal/mol/Å. In addition, during the test we found that introducing the electrostatic potential applied on each QM atom by solvents to ML model can improve the prediction accuracy (Figure 14(a)). With an optimized model, we carry out GaREUS simulation again but force generated by ML model. Unfortunately, the system is gradually pushed to the region with larger prediction error and then the simulation will break down due to large force (Figure 14(b)).

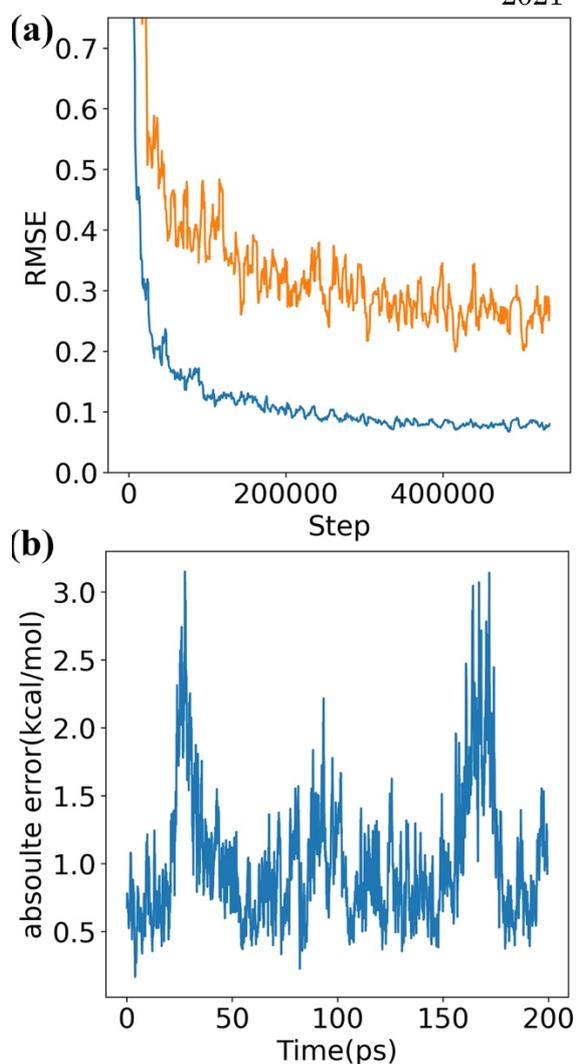


Figure 14 (a). Root mean square error (RMSE) of predicted energy of QM region with training step. Blue Line: introducing electrostatic potential applied on each QM atom into ML model, Orange Line: original model. (b) the evolution of absolute value of prediction error of energy with simulation.

4. Conclusion

- We found that introducing external potential applied on each QM atom can improve prediction accuracy.
- We applied a multi-scale model to fit the energy and force calculated by QM/MM method. The resulting prediction error of energy and force are
- We found that the fluctuation will push the system to the region with large prediction error causing the breakdown of simulation.

5. Schedule and prospect for the future

We are currently using active learning method to sample configurations with large prediction uncertainty so that the model can be further improved. After the test, we plan to apply this method to more complex reaction system: enzymatic reaction to study the catalytic mechanism.

2021 年度 利用研究成果リスト

【雑誌に受理された論文】

1. Takaharu Mori, Genki Terashi, Daisuke Matsuoka, Daisuke Kihara, and Yuji Sugita, “Efficient flexible fitting refinement with automatic error fixing for de novo structure modeling from cryo-EM density maps”, *J. Chem. Inf. Model.*, 61, 3516–3528 (2021).
2. 森貴治, 杉田有治, “分子シミュレーションによる新型コロナウイルススパイクタンパク質の糖鎖ダイナミクスの解析”, *Glycoforum*, 24, A12 (2021).
3. K. Yagi, S. Ito, and Y. Sugita, “Exploring the Minimum-Energy Pathways and Free-Energy Profiles of Enzymatic Reactions with QM/MM Calculations”, *J. Phys. Chem. B* 125, 4701–4713 (2021).
4. K. Yagi and Y. Sugita, “Anharmonic Vibrational Calculations Based on Group-Localized Coordinates: Applications to Internal Water Molecules in Bacteriorhodopsin”, *J. Chem. Theory Comput.* 17, 5007 – 5020 (2021).
5. V. A. Lorenz-Fonfria, K. Yagi, S. Ito, and H. Kandori, “Retinal Vibrations in Bacteriorhodopsin are Mechanically Harmonic but Electrically Anharmonic: Evidence From Overtone and Combination Bands”, *Front. Mol. Biosci.* 17, 749261 (2021).

【口頭発表】

1. ○Oide, M., Sugita, Y. “Capturing drastic state transitions of biological macromolecules by molecular dynamics simulation and nonlinear dimensionality reduction”, 第 59 回日本生物物理学会年会, オンライン, 2021 年 11 月 27 日.
2. ○Niitsu, A. R. Thomson, A.J. Scott, J.T. Sengel, Y. Sugita, M.I. Wallace, H. Bayley, and D.N. Woolfson. “自己会合する膜貫通ペプチドチャネルの理論設計”, 日本物理学会 2021 年秋季大会, オンライン, 2021 年 9 月 20 日.
3. ○森貴治, 寺師玄記, 松岳大輔, 木原大亮, 杉田有治, “Cryo-EM flexible fitting refinement with automatic error fixing for de novo protein structure modeling”, 第 59 回日本生物物理学会年会, オンライン, 2021 年 11 月 26 日.
4. ○八木清, 杉田有治, “分子動力学プログラム GENESIS における QM/MM 法の開発と応用”, 第 23 回理論化学討論会, オンライン, 2021 年 5 月 13 – 15 日.
5. ○八木清, “(招待講演)量子化学と分子動力学の融合が拓く高分子の分子機能解析”, 第 1 回溶液化学夏季講演会, オンライン, 2021 年 7 月 26 日.
6. ○八木清, “(招待講演)QM/MM 法の開発と生体分子の化学反応解析”, 化学反応経路探索のニューフロンティア, オンライン, 2021 年 9 月 22 日.
7. ○八木清, “(招待講演)Weight average 法による振動スペクトル計算と生体分子への応用”, 第 2 回分子集合系計算科学セミナー, オンライン, 2021 年 10 月 4 日.
8. ○K. Yagi, “(招待講演)Anharmonic vibrational calculations of hydrogen bond network in protein using vibrational quasi-degenerate perturbation theory with localized coordinates,”, *Advances in Hydrogen Bond Research*, PacifiChem, Waikiki, Hawaii (Online), 2020, 2021/12/16-21.
9. ○K. Yagi, “(招待講演)Finding the reaction path of metalloenzyme using QM/MM in GENESIS”, *XFEL & Multidisciplinary Approaches: New Challenges in Metals in Structural Biology*, Waikiki, Hawaii (Online), 2021/12/16-21.

【ポスター発表】

1. ○伊東真吾, 八木清, 杉田有治, “Mechanism of Tryptophan synthesis of TRPS”, 第 3 5 回分子シミュレーション討論会, 岡山大学, 2021 年 12 月.
2. ○M. Oide, Y. Sugita, “Nonlinear-dimensionality reduction analysis of protein dynamics”, 第 21 回日本蛋白質化学会年会, オンライン, 2021 年 6 月 18 日.
3. ○森貴治, “分子動力学シミュレーションに基づく遺伝情報転写機構の解明”, 第 8 回 HPCI システム利用研究課題 成果報告会, オンライン, 2021 年 10 月 29 日.
4. ○森貴治, 杉田有治, “新型コロナウイルス表面のタンパク質動的構造予測”, 第 8 回 HPCI システム利用研究課題 成果報告会オンライン, 2021 年 10 月 29 日.