

令和6年度高性能汎用計算機高度利用事業

「富岳」成果創出加速プログラム

「生体分子シミュレータを基にした  
大規模推論システムの開発と応用」

成果報告書

令和7年5月30日  
国立大学法人 埼玉大学

松永 康佑

## 目次

目次.....	- 1 -
補助事業の名称.....	- 2 -
1. 補助事業の目的 .....	- 2 -
2. 令和6年度（報告年度）の実施内容 .....	- 2 -
2-1. 当該年度（令和6年度）の事業実施計画 .....	- 2 -
2-2. 実施内容（成果） .....	- 3 -
2-3. 活動（研究会の活動等） .....	- 14 -
2-4. 実施体制.....	- 14 -
別添1 学会等発表実績	

## 補助事業の名称

「富岳」成果創出加速プログラム

生体分子シミュレータを基にした大規模推論システムの開発と応用

体系的番号： JPMXP1020230119

### 1. 補助事業の目的

生体分子シミュレータを活用して、実験データから逆問題を解き、分子構造やダイナミクスに関する大規模な統計的推論を実現する研究開発を行う。ここでは、富岳向けに最適化されている GENESIS に注目し、若手を中心とするメンバーで統計的推論に応用できるようにする。具体的には、GENESIS を推論システムにおける尤度計算部品とみなし、ベイズ推論・機械学習/AI と組み合わせて構造空間の探索範囲をうまく絞り込むことで、これまで不可能であった高次元サンプリングを実現し生命科学の新たな知見を得る。

### 2. 令和6年度（報告年度）の実施内容

#### 2-1. 当該年度（令和6年度）の事業実施計画

(サブ課題1) マルコフ状態モデルを介した構造ダイナミクスの推論（埼玉大・松永）

本課題は、大規模分子シミュレーションと高速原子間力顕微鏡(AFM)データを融合させるデータ同化を行い、天然変性タンパク質の構造遷移ダイナミクスの推論を行うことを目的とする。一般に MD シミュレーションで到達できる時間スケールと高速 AFM データの時間解像度にはギャップがあり、両者を同化することは困難であるが、これまでに松永らは両者の間に中間的なスケールのモデル(マルコフ状態モデル)を介することでデータ同化を行い構造ダイナミクスを推論することに成功してきた。しかし一方で、マルコフ状態モデルを構築するには MD シミュレーションで当該タンパク質がとり得る構造を網羅的にサンプリングする必要があり、適用できる分子系が限られてきた。令和6年度は、天然変性タンパク質の構造サンプリングとマルコフ状態モデルの構築を行う。

(サブ課題2) 遺伝子発現と共役するゲノム3次元構造動態の推論（京大・高田）

本課題は、大規模生体分子シミュレーションにより、ゲノム3次元構造に依存した遺伝子発現制御の分子機構の推論を行うことを目的とする。Hi-C等のゲノムデータに基づいてヘテロなゲノム構造アンサンブルを推論するために、ベイズ推論ベースの *metainference* 法を発展させるとともに、マルチスケールシミュレーションを実施する。令和6年度は、ES細胞 Nanog 遺伝子座の複数の構造モデルについての長時間シミュレーションとその解析を行う。ES細胞 Nanog 遺伝子座の複数の構造動態モデルについての超解像イメージングによる評価を行う。

(サブ課題3) クライオ電顕データからの構造アンサンブル推論（名大・Tama）

本課題は、クライオ電顕の画像データから生体分子の構造アンサンブルやサンプル不均一性についての情報を推論するための手法を開発し、実験データに応用することを目的とする。具体的には、クライオ電顕個々の画像を解析し、それぞれが捉えている生体分子構造を推論することによりサンプル内の構造の

多様性に関する情報を得ることを目的とする。令和 6 年度は、クライオ電顕データを用いた構造アンサンブルの推論として、実験データへの応用を行う。

(サブ課題 4)エラー・ノイズを含む実験データからの統合的立体構造推論 (東理大・森)

本課題は、クライオ電顕実験データとクロスリンク質量分析実験データを統合的に組み合わせてタンパク質複合体の立体構造を予測するための技術基盤の構築と応用を行うことを目的とする。電顕像の解像度が低く、電顕像のみからは構成タンパク質の配向がはっきりしない場合、アミノ酸間距離情報を含むクロスリンク質量分析実験データがしばしば参照されるが、クロスリンク質量分析データは実験条件によっては 30%近くエラーが含まれることもあり手動モデリングによる主観が入る。この問題をベイズ推論をベースにした MELD 法の応用により解決することが目的である。令和 6 年度は、ヌクレオソームおよび転写因子に対する電顕フィッティングおよび MD+MELD 計算を行う。また、GENESIS において MELD 計算+電顕フレキシブルフィッティング+レプリカ交換法が実行できるように機能を改良する。

理研・小林らは、引き続き上記の課題と協力して、課題計算を富岳で有効にスケールさせるためのプログラムの高度化を行う。必要な場合には GENESIS への機能追加を行う。

明大・中村らは、引き続き上記の課題と協力して、推論を効率的に行うための機械学習/AI モデリングを行う。具体的には、課題 1 の松永らのサロゲートモデル構築を協力する。

(プロジェクトの総合推進)

課題内の連携を推進するとともに、分野全体の底上げを行うために、毎月オンラインでの勉強会(Fugaku monthly seminar)を開催する。また、4月に Wuhan で国際ワークショップ(ICT-HPCC24)を松永が Chair として共催し、6月には神戸で国際ワークショップ(CHARMM-GUI/GENESIS Tutorial Workshop)を共催し、本課題で得られた成果発表を行う。その他、前年度に引き続き引き続き学術変革等の他のプロジェクトと連携し効果的な研究の推進を行う。

## 2-2. 実施内容 (成果)

(サブ課題 1) マルコフ状態モデルを介した構造ダイナミクスの推論 (埼玉大・松永)

本課題では、高速 AFM による実験計測データ (AFM 画像) と MD シミュレーションから得られる構造アンサンブルのデータ同化を通じて、天然変性タンパク質 (Intrinsically Disordered Protein: IDP) の構造遷移を推論する手法の開発と応用を目的とする。従来、高速 AFM と MD シミュレーションは時間スケールの違いから、直接的な統合が困難であり、両者の間を橋渡しする中間的なスケールのモデルが必要とされてきた。我々はこれまでに、構造状態を有限個の離散状態として表現し、それらの遷移確率からキネティクスを記述するマルコフ状態モデルを導入し、実験と理論のデータ同化を進めてきた (Ishizone, et al., *J. Chem. Theory Comput.* **20**, 436 (2024))。しかしマルコフ状態モデルの構築には、対象分子が取り得る構造状態の網羅的サンプリングが必要であり、柔軟性に富んだ IDP に対しては特に困難な課題であった。令和 6 年度は、この課題を克服するため、物理ベースのモデルと機械学習を組み合わせた構造サンプリング手法の開発と、その適用に取り組んだ。

ターゲットの IDP として、高速 AFM による計測データが得られているオートファジー関連タンパク質 Atg13 を選定した。このタンパク質は末端の構造ドメインと長い天然変性領域 (IDR) を含み、構造的多様性が極めて高いことが知られている。まず、前年度に味覚受容体 t1r に対して評価を行った能動学習に基づく構造サンプリング法を適用するため、Atg13 の全原子 MD シミュレーションを試みた。Atg13 は結晶構造が知られていないため、構造ドメインは AlphaFold2 を用い、Disordered 領域や Modeller を用いて初期構造を作成した。生体分子シミュレーターである GENESIS (Jung, et al., *WIREs Comput. Mol. Sci.* **5**, 310 (2015)) により、IDP に対応している CHARMM36m 力場を用いて、富岳上でレプリカ交換並列 MD シミュレーションを行って広範な構造空間の探索を実施した。その結果として得られた構造群は、AFM 観測値に対して著しくコンパクトであり、計測データから推定される慣性半径のおよそ半分に留まった。これは、使用された全原子力場が IDP の広がり十分に表現できていないことが原因であると考えられる。さらに、IDP は数多くの構造状態を取り得るため、全原子モデルにおけるサンプリングの計算コストは膨大であり、現実的な時間スケールでは網羅性を確保することが困難であることが確認された。この問題を回避するため分子モデルの粒度を落とし、粗視化モデルによるモデリングを行った。粗視化モデルでは、1 アミノ酸残基を 1 粒子として扱うモデルを採用し、計算効率を大幅に向上させた。特に、液液相分離挙動の再現に適した HPS モデルを採用し、各残基の疎水性スケールを明示的に取り入れることで、IDP の空間的な広がりを適切に表現可能とした。AFM 画像の空間分解能は数ナノメートルオーダーであるため、大域的な構造推論では原子分解能は必須ではなく、今回採用した粗視化スキームは計測分解能との整合性から見ても合理的な選択である。実際に得られた構造は、AFM による慣性半径とよく一致し、物理的妥当性を備えた構造アンサンブルを構築することが可能となった。

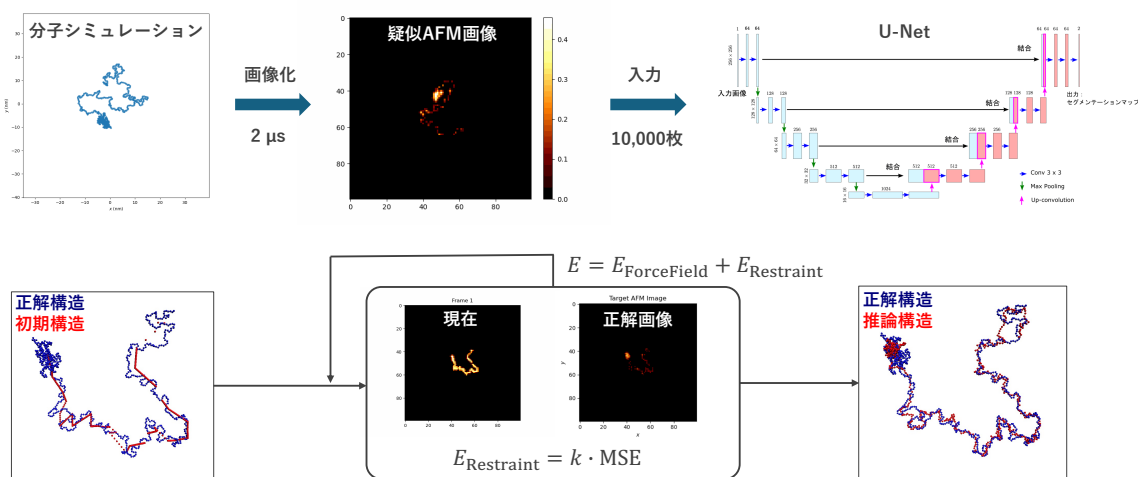


図 1-1: 開発した構造アンサンブル推論システムのプロトコル。上段は分子シミュレーションから作成した疑似 AFM 画像により初期構造を生成する U-Net モデルの概念図。下段は生成された初期構造から構造を推論するフレキシブルフィッティングの概念図。

サンプリングした構造を計測データと同化するために、我々は統合モデリングの一環としてフレキシブルフィッティング法を導入した。これは、対象構造に外部から束縛ポテンシャルを付加し、実測画像との整合性を保ちながら構造を最適化する手法である。GENESIS にフレキシブルフィッティング機能を組み込む前のテスト実装として、OpenMM のカスタム力場を利用し、フレキシブルフィッティングを試行

的に実装・適用した。しかし、IDP のように構造が不安定かつ揺らぎが大きい系に対しては、通常のフレキシブルフィッティングでは十分なフィッティング精度が得られなかった。具体的には、正解構造から離れた初期構造に対して束縛を加えると、フィッティング中にすぐに局所的な構造にトラップされてしまい、正解の構造との大域的な一致を得ることが難しいことがわかった。また、構造全体を単純に並進・回転して整列させた場合でも、柔軟な IDP の構造には対応しきれず、大域的な一致が困難であった。

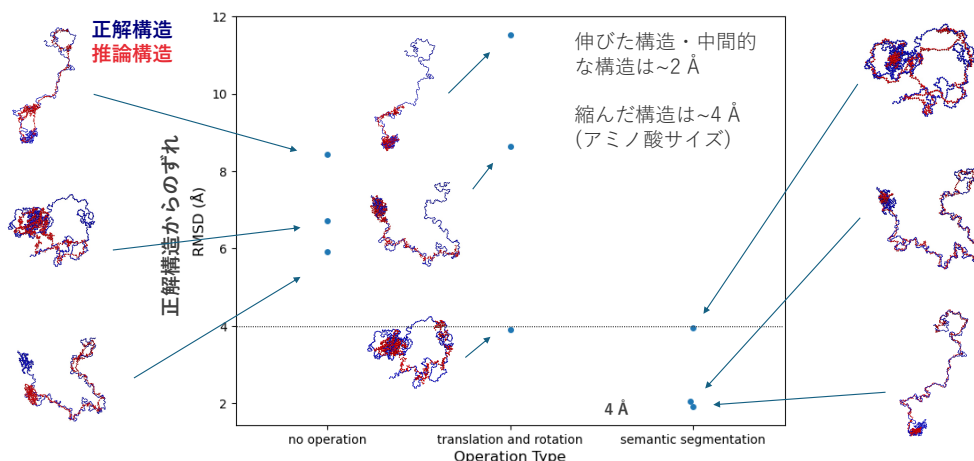


図 1-2: テストデータでの構造推論精度。左・中・右がそれぞれ操作なし、並進・回転、セマンティックセグメンテーションによる結果。セグメンテーションでは正解構造(青)に推論構造(赤)がよく合致する。

これらの問題を解決するため、AFM 画像から精度の良い初期構造を推定する手法として、中村グループと協力しセマンティックセグメンテーションを導入した。セマンティックセグメンテーションとは、画像を複数の領域に分割し、各領域に意味的なラベルを与える画像処理手法である。本研究では、MD シミュレーションから生成した疑似 AFM 画像と対応する構造ラベルデータを教師データとし、U-Net ベースのモデルを PyTorch 上に構築した。学習済みのネットワークは、画像から各残基の位置に対応する領域を予測可能であり、この予測を用いて、AFM 画像に大域的に整合した初期構造を生成した。この初期構造を用いてフレキシブルフィッティングを適用することで、局所的最適化の問題を回避しつつ、AFM 画像との整合性を高精度に確保することが可能となった。得られた構造精度はアミノ酸残基レベルに達し、空間分解能として約 4 Å 程度の位置推定が可能であることを確認した。

また、開発した手法を、1,000 枚以上の大量の高速 AFM 実測画像へ適用するために、小林グループと協力し GENESIS と PyTorch 間のインターフェースを実装し、フレキシブルフィッティングを大規模の富岳上で実行できるコードを開発した。これらの開発技術を用いて、現在実データへの応用研究を実施中である。

## (サブ課題 2) 遺伝子発現と共役するゲノム 3 次元構造動態の推論 (京大・高田)

本課題は、大規模生体分子シミュレーションにより、ゲノム 3 次元構造に依存した遺伝子発現制御の分子機構の推論を行うことを目的とする。Hi-C 等のゲノムデータに基づいてヘテロなゲノム構造アンサンブルを推論するために、バイズ推論ベースの *metainference* 法 (Bonomi, et al., *Sci. Adv.* **2**, e1501177 (2016)) を発展させるとともに、マルチスケールシミュレーションを実施する。

多細胞生物は多様な細胞種をもち、多様性は細胞種ごとに異なる遺伝子発現に起因する。これが高次生体現象の基礎となっている。細胞種依存的な転写制御において、エピゲノム制御およびそれと密接に関連したクロマチンの3次元折りたたみ構造の変化が主要な役割を果たす。本課題では哺乳類の胚性幹(ES)細胞に着目し、幹細胞とその分化で主要な機能を果たす遺伝子群の転写制御とクロマチン構造の関係を、細胞スケールのMDシミュレーションによって推論する。哺乳類ES細胞のコア遺伝子ネットワークは山中因子を含む3つの遺伝子、*Oct4*, *Sox2*, *Nanog* で構成され、ES細胞の多能性と自己複製能の維持に必須である。本課題ではマウス *Nanog* 遺伝子座(図2-1)を本研究の標的とし、Micro-CおよびChIP-seqなどの実験データをもとに、その遺伝子座の折りたたみ構造と発現活性化機構を推論する。令和5年度までに、我々が開発したHi-C Metainference法を、1kb分解能およびそれより詳細なヌクレオソーム分解能のクロマチンモデルを用いてマウス *Nanog* 遺伝子座に適用し、詳細な構造状態を特定することに成功している (Brandani et al. *PNAS Nexus* 3, pga226 (2024))。

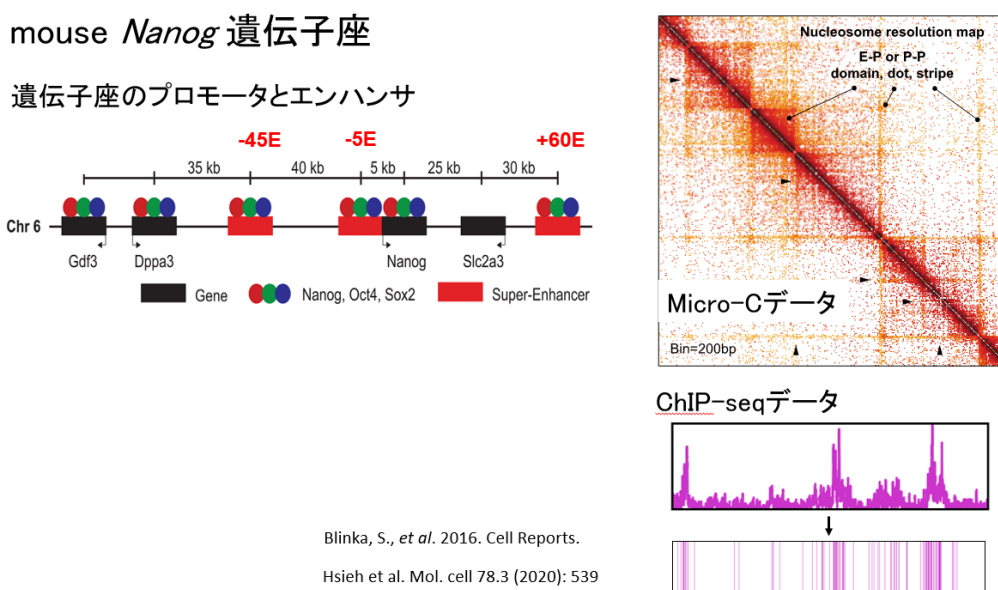


図2-1: 対象としたマウス *Nanog* 遺伝子座の特徴と主な実験データ。(左上) *Nanog* 遺伝子座200kb領域は、そのプロモータ、3つのスーパーエンハンサ (-45SE, -5SE, +60SE)、および関連する他の3つの遺伝子座 (プロモータ) を含む。(右上) Micro-Cデータ。(右下) ChIP-seqデータの一部。

令和6年度は、これまでに得られたES細胞 *Nanog* 遺伝子座の複数の構造モデルについて、長時間シミュレーションとその解析を行った。富岳上で、GENESISにより残基分解能の構造モデルを用いて、長時間分子動力学シミュレーションを実施し、得られた構造サンプルの特徴を分析した。構造モデルは、約1000個のヌクレオソーム、約400個のリンカーヒストン、7個のRNA合成酵素(Pol II)、4種類の転写因子(*Oct4*, *Sox2*, *Klf4*, *Nanog*)各20個、コアティベータ(*Brd4* 20個、*p300* 3個)を含む。得られた構造からまず、micro-Cデータを再現するためには、ヘテロな構造アンサンブルが必要であったことから、クロマチンはヘテロな構造の集合であることがわかる。代表的な構造を見ると(図2-2左)、かなりコンパクトな構造からずっと広がった構造まで大きな多様性をもっていた。どのクロマチン構造においても、RNA合成酵素(Pol II)部分は、ドメインの表面に位置しておりアクセス可能となっている。また、エンハンサを中心に、その近傍にコアクティベータ(*BRD4*, *p300*)、その周辺に転写因子群(TFs)が分布しており、エンハン

サを起点とした転写活性化に寄与する構造が見いだされた。

次に、クロマチン上の連続した12ヌクレオソーム断片の折れたたみ構造と転写活性（遺伝子発現）やゲノム物性との相関を分析し、それらに顕著な相関があることを見出した（図2-2右）。局所的に広がった構造をとっているのは、プロモータ、エンハンサなどシス制御領域と呼ばれる領域の近くであり、転写因子およびコアクティベータの結合頻度が顕著に高くなっている。また、ヌクレオソーム間のつなぎ目にあたるリンカーDNAが長いほど、リンカーDNAの長さ非対称性が強いほど、クロマチンは広がった構造になっていることが分かった。

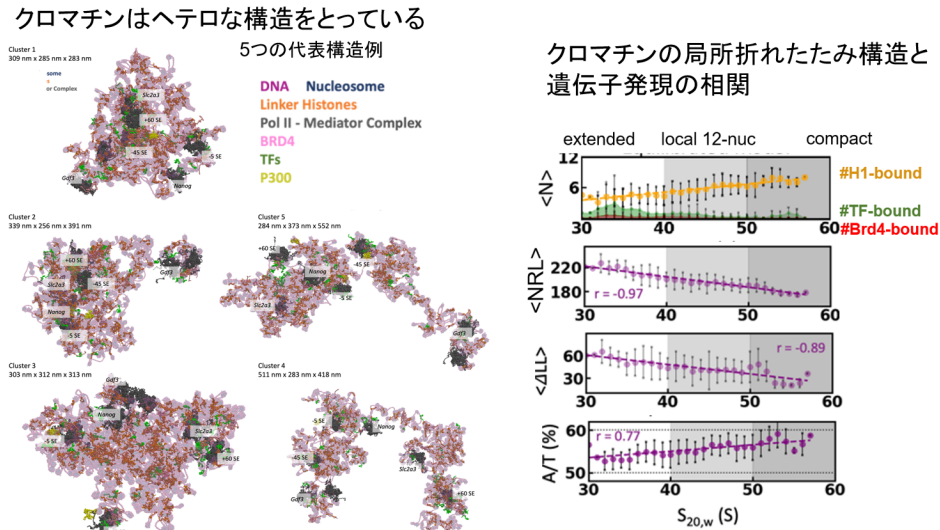


図2-2: 大規模構造サンプリングによって得られたヘテロなマウス *Nanog* 遺伝子座のクロマチン構造（左）と、その局所構造が転写活性（遺伝子発現）因子と関係する様子（右）。（左）大規模構造サンプルからクラスタリングにより得られた5つの代表構造。もっともコンパクトな上の構造から右下に向けて、より広がった構造を描いた。（右）局所的な12ヌクレオソームの構造のコンパクトさ（横軸  $S_{20,w}$  が大きいほどコンパクト）と転写因子の結合（TF-bound）やコアクティベータ(Brd4-bound)の確率等の関係を示す。NRL: nucleosome repeat length,  $\Delta LL$ : asymmetry in linker length, A/T: ゲノムDNA上のA/T含有量。

また、ES細胞クロマチンの複数の構造動態モデルについての超解像イメージングによる評価を行った。我々が推論に用いたデータは、micro-CデータやChIP-seqデータなどであり、これらは結合の有無を教えてくれるものの、直接的に距離の情報ではない。そのため、得られた構造モデルの空間的広がりが、実際のクロマチンの広がりと一致するかどうかは自明ではない。そこで、推論（モデリング）に用いたデータとは独立な実験データで距離情報を与えるものとして、超解像イメージングデータを用いて、クロマチン構造モデルの広がりについての検証を行った。超解像イメージングデータが利用できる場合について、推論によるクロマチン構造モデルが実験と整合した広がりをもつことを確認した（図2-3）。

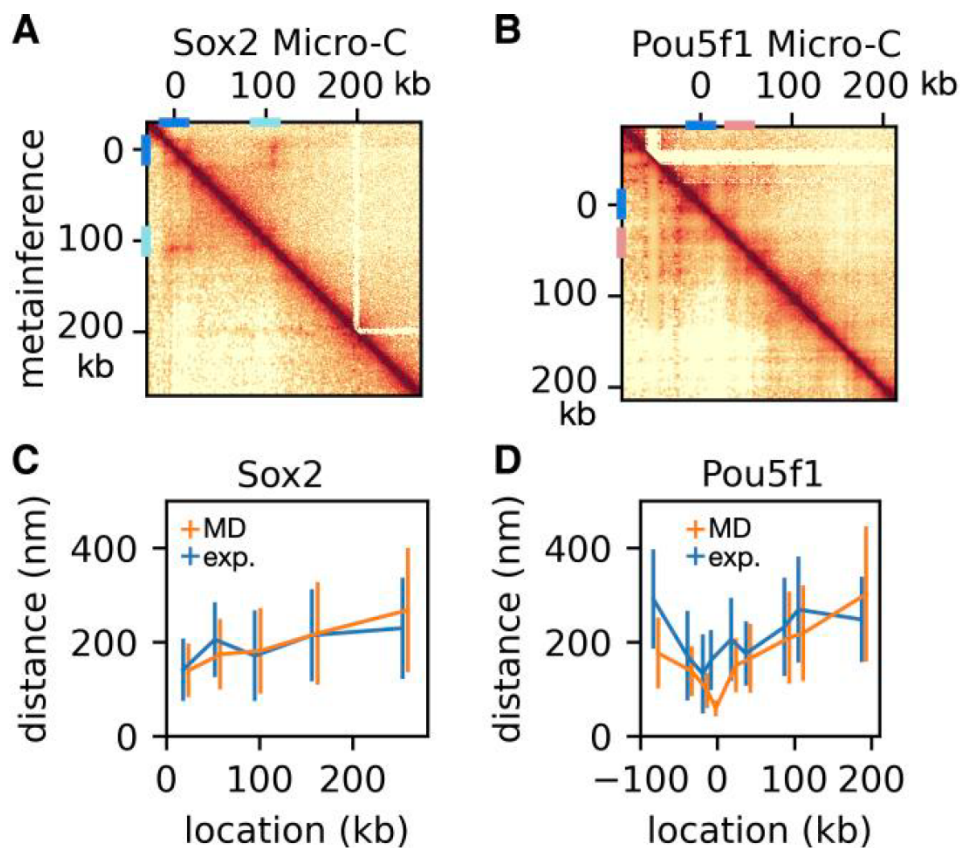


図 2-3: クロマチン構造モデルの超解像イメージングによる評価。(A,B)構造モデルから得られた Hi-C マップと、実験による micro-C コンタクトマップの比較。(C,D) 構築したクロマチン構造モデル(MD)と超解像イメージングデータ(exp)によるクロマチン上の 2 点間距離の比較。

マウス胚性幹細胞 Nanog 遺伝子座全体のクロマチン構造モデルは、多くの実験データに基づいて推論され、前例のないほど詳細な構造モデルとなった。このクロマチン構造モデルについての富岳による長時間分子動力学シミュレーションデータは、極めて膨大なデータであり、まだ明らかにできていない豊富な現象を内包していると考えられる。今後、より多くの観点から、データの定量的な解析を続ける必要がある。また、転写因子やコアクティベータを削除、あるいは過剰に配置した状況でのクロマチン動態を分析し、それらを比較することが重要である。

(サブ課題 3) クライオ電顕データからの構造アンサンブル推論 (名大・Tama)

This project aims to investigate the continuous conformational variability and dynamics of biological macromolecules from large-scale cryo-EM single-particle 2D images using MD simulations employing the MDSPACE algorithm (Vuillemot et al., *J. Mol. Biol.* **435**, 167951 (2023)). This year, the research focuses on applying the MDSPACE algorithm to experimental cryo-EM data to generate continuous structural ensembles through large-scale MD simulations guided by single-particle images. It is structured in three stages: 1) RELION analysis of cryo-EM micrographs (Scheres, *J. Struct. Biol.* **180**, 519 (2012)) to obtain high-quality single-particle images, followed by the reconstruction of an average

3D volume to allow accurate estimation of the pose Euler angles; 2) Large-scale MD-based flexible fitting using MDSPACE implemented in GENESIS using the selected high-quality single-particle images (~1.2 million); 3) Structural ensemble analysis and interpretation using both linear (principal component analysis, PCA) and non-linear (uniform manifold approximation and projection, UMAP) dimension reduction methods.

### 3.1 A total of 1.2 million high-quality particle images processed by RELION analysis

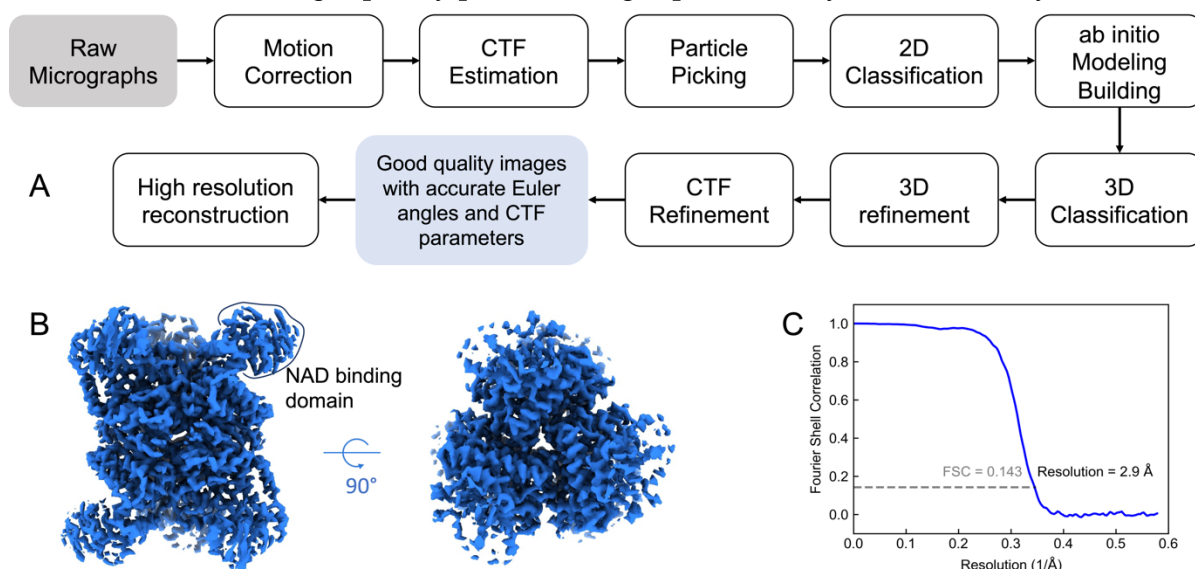


Figure 3-1: (A) A simplified cryo-EM single-particle analysis workflow. (B) Final 3D reconstruction of the GDH hexamer complex using ~1.2 million good-quality images, shown in two orientations, rotated by 90° relative to each other. The flexible region of the NAD binding domain has a lower resolution than the core domain. (C) Fourier shell correlation (FSC) curve indicates an average resolution of 2.9 Å for the reconstruction from 1.2 million good-quality particle images at the 0.143 cutoff criterion.

Cryo-EM micrographs of glutamate dehydrogenase (GDH) hexamer were obtained by the researchers at RIKEN and JASRI. MOTIONCOR2 software was used for dose-fractionating motion correction to compensate for beam-induced motion and drift, and the CTFFIND4 program was used to estimate the contrast transfer function (CTF) parameters. All image processing steps were carried out using RELION5.0. Approximately 6.4 million particle images (each 360 × 360 pixels in size) were selected using the auto-pick LoG (Laplacian-of-Gaussian) method. First, we selected the first 20% particle images for reference-free 2D classification to obtain an initial 3D volume of GDH. Low-quality particles, such as those with ice contamination, dust, or incomplete/clustered GDH particles, were discarded. Resulting good-quality particle images were used to construct an initial 3D volume, which serves as a reference map for subsequent 3D classification and 3D refinement calculations. To reduce the computational cost, all picked particle images were initially down-sampled to 90 pixels × 90 pixels and then rescaled back to 120 pixels × 120 pixels, 180 pixels × 180 pixels, and 360 pixels × 360 pixels for separate rounds of 3D classification and 3D refinement under C1 symmetry. This approach helped

further refine good-quality particle images and improve the final resolution. Next, CTF refinement (including asymmetrical and symmetrical aberrations and per-particle defocus values estimates) and another round of 3D refinement were performed to determine accurate Euler angles and CTF parameters for each high-quality particle image (Figure 3-1). In the end, the selected high-quality particle images were used to achieve a final high-resolution reconstruction.

### **3.2 Large-scale MD-based flexible fitting in GENESIS on Fugaku**

To capture the dynamic structural variability of the GDH hexamer, we performed large-scale molecular dynamics (MD)-based flexible fitting using the MDSPACE algorithm implemented in ATDYN engine of the GENESIS program. The flexible fitting is achieved by iteratively adjusting the atomic coordinates to increase the correlation coefficient (CC) between the simulated projections of the molecular coordinates in the MD simulation and the experimental cryo-EM particle images. The ~1.2 million high-quality particle images obtained in section 3.1 were used as the experimental restraint to drive the flexible fitting process. Each particle image, associated with refined Euler angles for the pose, served as a target for MD-based flexible fitting. The initial structure (PDB code: 1EUZ) was used to build a coarse-grained Ca Go-model. The fitting simulations were parallelized across thousands of nodes on Fugaku by using an automation tool to handle the fitting of large-scale image datasets. As the optimal force constant for fitting biasing varies depending on the force field and target image, a suitable parameter for the coarse-grained Ca Go-model of GDH was identified by testing a range of values and selecting the one that yielded the most rapid increase in the CC between simulated and experimental images, while avoiding structural distortions. Fitted conformations were collected from the last frame of the simulations. The MD-based flexible fitting of all datasets resulted in a diverse ensemble of fitted conformations, each relates to the particle conformations in individual cryo-EM particle images.

### **3.3 Structural ensemble analysis and interpretation**

The structural conformations obtained in MD simulations were subjected to comprehensive ensemble analyses to extract meaningful conformational motions and identify dominant dynamic modes. Principal Component Analysis (PCA) was applied to all subunits of fitted conformations to extract collective motions underlying the structural variability (Figure 3-2A). The fitted structural models of the GDH subunits occupy a broad region in the first and second principal component (PC) space, reflecting large conformational variability of GDH within the dataset. The first PC accounted for a large portion of the variance, which captured a hinge-bending motion transitioning between closed and open conformations, while the second PC represented a shear motion orthogonal to the primary transition. The ensemble of the structural models obtained by MDSPACE fitting densely covered the conformational space between the closed and open states, consistent with previously-reported results, while the distribution suggests that sampling in pure MD simulations, even performed for a very long timescale, may be insufficient to capture the full range of GDH variability. To evaluate the

conformational dynamics of individual subunits, clustering analysis on all subunits of the fitted GDH hexamers was performed. Six major structural clusters revealed distinct regions of conformational states (Figure 3-2B), illustrating diverse intermediate states bridging closed and open conformations.

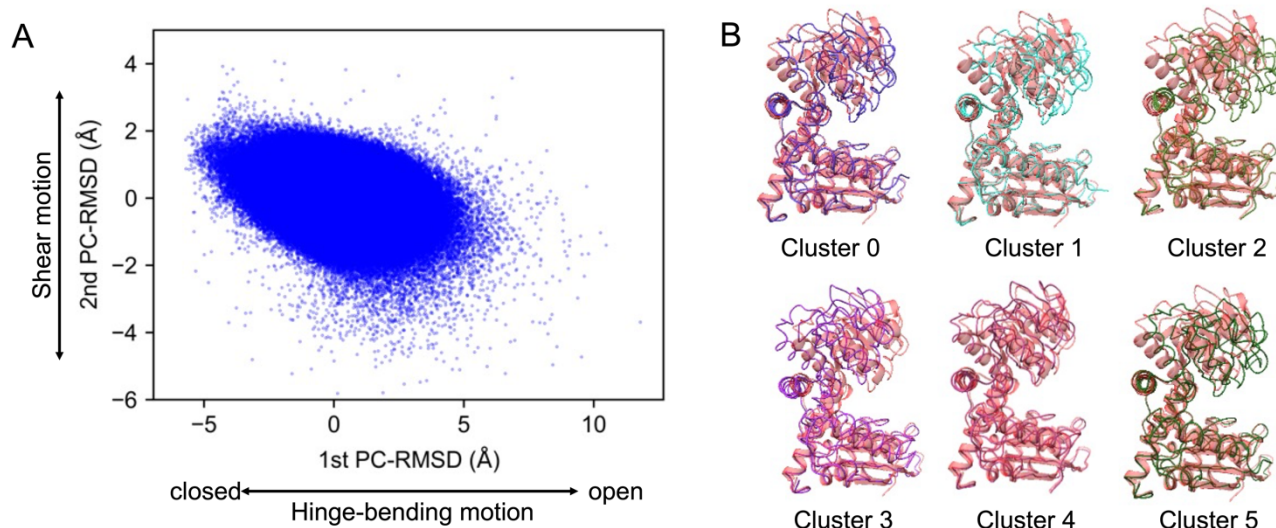


Figure 3-2: (A) PCA of subunits of 1.2 million MDSPACE-fitted structures were projected onto the first two principal components in RMSD representing hinge-bending (PC1) and shear motions (PC2) between core domain and NAD binding domain. (B) Representative structures of six classified clusters aligning with subunit A of 1EUZ (open, in red) by core domain (residue 24-180). In each panel, the half-transparent red cartoon represents the reference conformation, overlaid with the cluster-representative structure (shown in solid colors).

(サブ課題 4)エラー・ノイズを含む実験データからの統合的立体構造推論 (東理大・森)

昨年度、クライオ電顕データおよびクロスリンク質量分析データを組み合わせた MELD 法 (MacCallum, et al., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6985 (2015)) による立体構造予測が、比較的小さなタンパク質であるカルシウムイオンポンプに適用可能であることがわかった。そこで本年度は、より大きく複雑なヌクレオソーム-DNA 結合タンパク質複合体への応用可能性を検討した。本年度も擬似実験データを用いて、GENESIS (Jung, et al., *WIREs Comput. Mol. Sci.* **5**, 310 (2015))に導入された MD+MELD 計算の性能評価を行った。

まず、本研究で用いたエラーおよびノイズを含む擬似実験データの作成方法について述べる。単粒子解析では、タンパク質の揺らぎが大きい部位で解像度が低下し、側鎖だけでなく主鎖や全体の密度もほとんど見えなくなり、タンパク質の全体像や配向が不明になることがある。そのような欠損データを模倣するため、構造が既知のヌクレオソーム-CENPA 複合体を用いて、既知立体構造に基づく擬似クライオ電顕データを作成した (図 4-1)。ここでは CENPA の揺らぎが大きく、ヌクレオソームと接する部分のみ密度が見える状況を再現し、CENPA の全体の 45%の原子に対して密度を生成した。さらに、エラーを含む擬似クロスリンク質量分析データも作成した。クロスリンク質量分析は、通常、構造未知のタンパク質を含む溶液にクロスリンカー分子を添加し、Lys などの反応性に富むアミノ酸に架橋を導入、タンパク質を断片

化・質量分析することで近接するアミノ酸配列を同定する手法である。このデータには、タンパク質の会合や凝集に由来し、クロスリンカーの長さを超える距離の残基間に架橋が検出されるといったエラーが含まれる。本研究では、このようなエラーを含む、overlength ペアが 88%を占める擬似データを作成した。

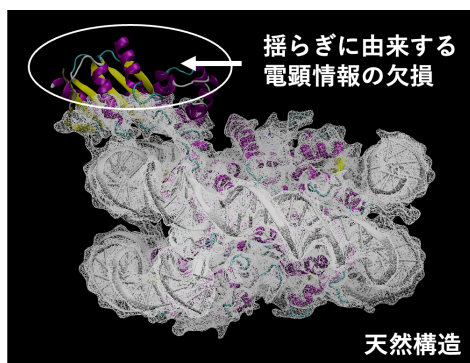


図 4-1: ヌクレオソーム-CENPA 複合体の擬似的クライオ電顕実験データ

MELD 法は、ベイズ確率に基づく構造推定法である。実験データを束縛条件とする MD 計算法の一種であり、予測構造と大きく乖離する実験データを自動的に束縛条件から除外する特徴を持つ。本研究では、粗視化モデルの一種である AICG2 モデルを用い、クライオ電顕データとクロスリンク質量分析データを組み合わせて MELD 法による構造予測を行った。MELD 計算を行う際、クライオ電顕に関する束縛エネルギーについては、全粒子にガウス関数を配置して simulated map を作成し、欠損を含む擬似実験データとの相関係数 (cc) を算出した。クロスリンク質量分析データについては、全束縛データのうち 88% をエラーとみなし、自動的に除外する設定とした。CENPA をヌクレオソームから離れた場所に置いて初期構造とし (天然構造からの RMSD=12.2Å)、初期乱数を変えて 100 回の計算を実行した。

図 4-2(a)はクライオ電顕データとクロスリンク質量分析データの両方を束縛条件とした MELD 計算、図 4-2(b)はクライオ電顕データのみを束縛条件としたフレキシブルフィッティング、図 4-2(c)はクロスリンク質量分析データのみを束縛条件とした MELD 計算における、天然構造からの RMSD の時間変化を示す。最も予測精度が高かったのは両方の実験データを用いた場合であり、正解構造に到達した試行の数は単純な電顕フレキシブルフィッティングより多く、cc が 0.85 以上の構造も得られた (図 4-2d)。一方、クロスリンク質量分析データのみを束縛条件とした場合は、すべての計算で正解構造に到達しなかった (図 4-2c)。これらの結果は合理的であり、GENESIS に導入された MELD 法による束縛が有効に機能していることを示唆している。

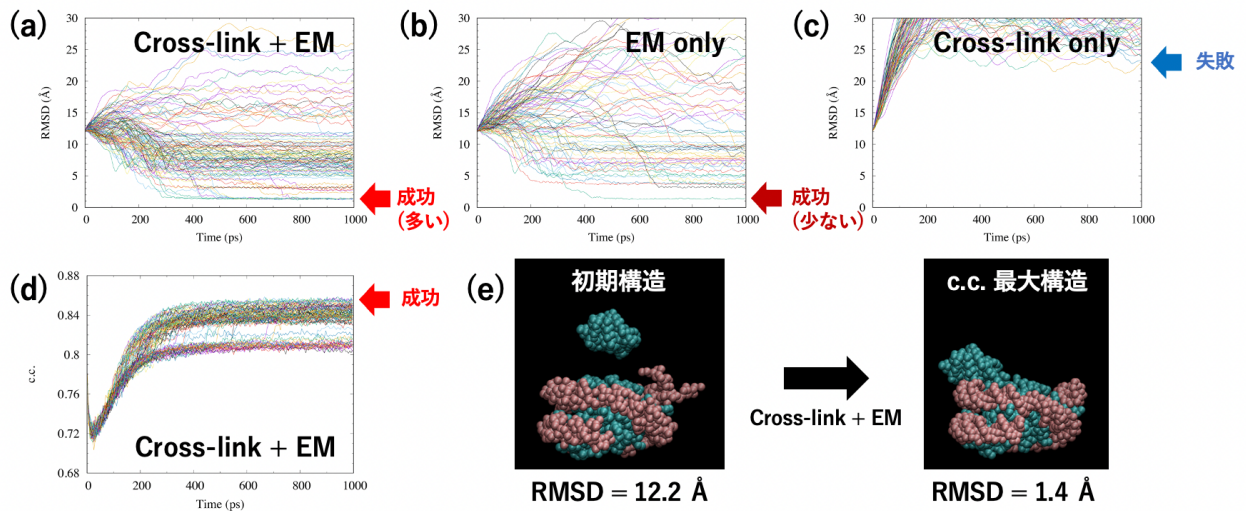


図 4-2. クライオ電顕実験データとクロスリンク質量分析データを組み合わせたヌクレオソーム-CENPA 複合体に対する立体構造予測

郷モデルに基づく粗視化モデルの大きな懸念点の一つは、天然構造で接触していないアミノ酸間に反発力が働くことである。本研究では、ヌクレオソームと CENPA を離れた状態でパラメーターを作成しているため、MELD 法によるフレキシブルフィッティングの過程でヌクレオソームと CENPA 間に大きな反発力が生じることになる。しかしながら、実際には、たとえ反発力が存在しても、ヌクレオソームおよび CENPA は安定した構造を保ち、最も cc が高い予測構造は天然構造に極めて近い (RMSD = 1.4 Å) ことがわかった (図 4-2e)。以上の結果から、AICG2 モデルを用いてもドメイン内で構造が大きく崩れることはほとんどなく、クライオ電顕データとクロスリンク質量分析データを組み合わせた MELD 法による統合的構造予測が、ヌクレオソーム複合体にも十分適用可能であることが示された。

### 2-3. 活動（研究会の活動等）

#### 国際会議 “International Conference on Theoretical and High Performance Computational Chemistry (ICT-HPC)2024” の運営と課題成果発表

2024年4月19日-22日に中国の武漢において対面で開催された国際会議 “International Conference on Theoretical and High Performance Computational Chemistry (ICT-HPC)2024” にて課題代表の松永が会議座長を務めた。会議では、最先端の理論計算科学について活発に議論した。アジアを中心として37名の招待講演があり、本課題から松永を始め Giovanni Brandani 講師(京大)、宮下治博士(理研)が招待講演を行い、本課題を宣伝するとともに、これまでに得られた成果について報告した。その他に大学院生を中心としたポスターセッションが開催された。

#### 国際ワークショップの開催 “IUPAB2024 Hands-on Training Program CHARMM-GUI/GENESIS MD Tutorial”

2024年6月30日-7月2日に国際ワークショップ “IUPAB2024 Hands-on Training Program CHARMM-GUI/GENESIS MD Tutorial” を理化学研究所神戸キャンパスにおいて、理化学研究所計算科学研究センターの杉田有治チームリーダー、Florence Tama ユニットリーダー、Lehigh 大学の Wonpil Im 教授、大阪公立大学の森次圭教授、と本課題が共催して、対面と Zoom のハイブリッドで開催した。ワークショップでは、海外(主に欧米の学術機関)からの著名な分子シミュレーション研究者を招待し、最先端の生物物理に関するシミュレーション研究について議論した。招待講演者15名のうち本課題から課題代表の松永と Florence Tama ユニットリーダー(理研、名古屋大)が招待講演を行い、本課題を宣伝するとともに、これまでに得られた成果について報告した。現地では91名が参加し、Zoom では25名が参加した。

#### マンスリーセミナーの開催

昨年度に引き続き課題内外から若手研究者を迎えて分子シミュレーションと機械学習/AIに関する Zoom セミナーをシリーズで開催した。令和6年度は、石曾根毅博士(明治大学)、唐澤直之博士(埼玉大学)、Jaewoon Jung 博士(理研)、林周斗准教授(東京科学大)、寺田透教授(東大)、川合毅氏(埼玉大学)、林慶浩助教(統数研)の方々にセミナーを行っていただき、最先端の技術について学ぶとともに、次世代のシミュレーション研究について議論した。

### 2-4. 実施体制

業務項目	担当機関	担当責任者
(1) (サブ課題1) マルコフ状態モデルを介した構造ダイナミクスの推論	〒338-8570 埼玉県さいたま市桜区下大久保 255 国立大学法人埼玉大学	松永 康佑
(2) (サブ課題2) 遺伝子発現と共役するゲノム3次元構造動態の推論	〒606-8501 京都府京都市左京区吉田本町36 番地1 国立大学法人京都大学	高田 彰二

<p>(3) (サブ課題3) クライオ 電顕データからの構造アンサ ンプル推論</p>	<p>〒464-8601 名古屋市千種区不老町 国立大 学法人東海国立大学機構 名古 屋大学</p>	<p>Florence Tama</p>
<p>(4) エラー・ノイズを含む 実験データからの統合的立体 構造推論</p>	<p>〒162-8601 東京都新宿区神楽坂1-3 学校法 人 東京理科大学</p>	<p>森 貴治</p>
<p>(プロジェクトの総合推進)</p>	<p>〒338-8570 埼玉県さいたま市桜区下大久保 255 国立大学法人埼玉大学</p>	<p>松永 康佑</p>

## 別添 1 学会等発表実績

### 1. 学会誌・雑誌等における論文掲載

No.	掲載した論文（発表題目）	発表者氏名	発表した場所（学会誌・雑誌名等）	発表した時期
1	YeeD is an essential partner for YeeE-mediated thiosulfate uptake in bacteria and regulates thiosulfate ion decomposition	Mai Ikei(Nara Institute of Science and Technology) Ryoji Miyazaki(Nara Institute of Science and Technology) Keigo Monden(Nara Institute of Science and Technology) Yusuke Naito(Nara Institute of Science and Technology) Azusa Takeuchi(Nara Institute of Science and Technology) Yutaro S. Takahashi(Nara Institute of Science and Technology) Yoshiki Tanaka(Nara Institute of Science and Technology) Keina Murata(Tokyo University of Science) Takaharu Mori(Tokyo University of Science) Muneyoshi Ichikawa(Fudan University) Tomoya Tsukazaki(Nara Institute of Science and Technology)"	PLOS Biology, 22, e3002601 (2024)	2024/04
2	Multiscale Bayesian simulations reveal functional chromatin condensation of gene loci	Giovanni B Brandani(Kyoto University), Chenyang Gu(Kyoto University), Soundhararajan Gopi(Kyoto University), Shoji Takada(Kyoto University)	PNAS Nexus, Vol. 3, Issue 6, pgae226 (2024)	2024/06
3	Explicit description of viral capsid subunit shapes by unfolding dihedrons	Ryuya Toyooka(The University of Tokyo) Seri Nishimoto(The University of Tokyo) Tomoya Tendo(The University of Tokyo)	Communications Biology, Vol. 7, 1509 (2024)	2024/11

		Takashi Horiyama(Hokkaido University) Tomohiro Tachi(The University of Tokyo) Yasuhiro Matsunaga(Saitama University)		
4	AFM observation of protein translocation mediated by one unit of SecYEG-SecA complex	Yui Kanaoka(Nagoya University) Takaharu Mori(Tokyo University of Science) Wataru Nagaike(Nagoya University) Seira Itaya(NAIST) Yuto Nonaka(Nagoya University) Hidetaka Kohga(NAIST) Takamitsu Haruyama(NAIST) Yasunori Sugano(NAIST) Ryoji Miyazaki(NAIST) Muneyoshi Ichikawa(Fudan University) Takayuki Uchihashi(Nagoya University) Tomoya Tsukazaki(NAIST)	Nature Communications, Vol. 16, 225 (2025)	2025/01

## 2. 国際会議・シンポジウムにおける口頭・ポスター発表

No.	発表した成果（発表題目、 口頭・ポスター発表の別）	発表者氏名（所属機関）	発表した場所（学会名 等）	発表した 時期
1	Integrative modeling of biomolecular dynamics from simulations and single-molecule experiments (口頭, 招待講 演)	Yasuhiro Matsunaga(Saitama University)	International Conference on Theoretical and High Performance Computational Chemistry 2024	2024/04
2	An integrative approach to the molecular modeling of genes (口頭, 招待講演)	Giovanni Brandani (Kyoto University)	International Conference on Theoretical and High Performance Computational Chemistry 2024	2024/04
3	Integrative modeling approach combining	Osamu Miyashita (RIKEN)	International Conference on	

	molecular dynamics simulations and experimental data to study dynamic structures of biological molecules		Theoretical and High Performance Computational Chemistry 2024	
4	Flexible Fitting of Coarse-Grained Models to AFM Images of Intrinsically Disordered Proteins(ポスター発表)	Sakura Homma(Saitama University) Yasuhiro Matsunaga(Saitama University)	IUPAB2024	2024/06
5	Learning force field parameters from ensemble-averaged data with a differentiable approach(ポスター発表)	Yohei Sako(Saitama University) Yasuhiro Matsunaga(Saitama University)	IUPAB2024	2024/06
6	Optimal transport maps for targeted free energy estimation(ポスター発表)	Tsuyoshi Kawai(Saitama University) Yasuhiro Matsunaga(Saitama University)	IUPAB2024	2024/06
7	Exploring the Continuous Conformational Variability of Glutamate Dehydrogenase Using Cryo-EM Single-particle Images and MD Simulations(ポスター発表)	Tinging Wang(RIKEN Center for Computational Science) Osamu Miyashita(RIKEN Center for Computational Science) Hideki Shigematsu(Japan Synchrotron Radiation Research Institute (JASRI)) Masaki Yamamoto(RIKEN SPring-8 Center) Florence Tama(Nagoya University; RIKEN Center for Computational Science; Institute of Transformative Bio-Molecules, Nagoya University)	IUPAB Congress 2024	2024/06
8	Unraveling the Structure and Dynamics of Biomolecules via Computational Modeling (口頭発表)	Florence Tama(Nagoya University; Institute of Transformative Bio-Molecules, Nagoya University; RIKEN Center for Computational Science)	European South Atlantic Biophysics Congress	2024/06
9	Unraveling the Structure and Dynamics of	Florence Tama(Nagoya University; Institute of Transformative Bio-	IUPAB2024 Hands-on Training Program CHARMM-	2024/07

	Biomolecules via Integrative Modeling	Molecules, Nagoya University; RIKEN Center for Computational Science)	GUI/GENESIS MD Tutorial	
10	Integrative Modeling of Protein Dynamics from High-speed Atomic Force Microscopy and Molecular Dynamics Simulations (口頭, 招待講演)	Yasuhiro Matsunaga(Saitama University)	IUPAB2024 Hands-on Training Program CHARMM-GUI/GENESIS MD Tutorial	2024/07
11	Exploring Continuous Conformational Variability via Cryo-EM Single-Particle Imaging and MD Simulations(口頭, 招待講演)	Florence Tama(Nagoya University; Institute of Transformative Bio-Molecules, Nagoya University; RIKEN Center for Computational Science)	Emerging Theoretical Approaches to Complement Single-Particle Cryo-Electron Microscopy	2024/10