

令和3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

「全原子・粗視化分子動力学による細胞内分子動態の解明」

成果報告書

令和4年5月30日

国立研究開発法人 理化学研究所

杉田 有治

目次

1.	補助事業の目的.....	1
2.	令和3年度（報告年度）の実施内容.....	2
2-1.	当該年度（令和3年度）の事業実施計画.....	2
	（1）全原子・粗視化分子動力学の最適化と並列化.....	2
	① GENESISの最適化と並列化（理研・杉田）.....	2
	② 粗視化モデル SPICA-FFの開発（岡山大・篠田）.....	2
	（2）データ科学的手法による粗視化モデルパラメタの最適化.....	2
	① データ同化によるシミュレーションと実験データの統合（埼玉大・松永）.....	2
	② 粗視化モデル CafeMolのパラメタ最適化（京大・高田）.....	2
	（3）生体分子系のモデリングと「富岳」を用いたシミュレーション.....	3
	① 細胞内分子混雑と液液相分離（理研・杉田）.....	3
	② 遺伝子転写機構の解明（京大・高田）.....	3
	③ 核内蛋白質・核酸複合体の動的モデリング（量研機構・河野）.....	3
	④ ウィルス（岡山大・篠田）.....	3
	⑤ 多剤排出トランスポータの分子動力学シミュレーション（東大・篠田）.....	3
	（4）プロジェクトの総合推進.....	3
2-2.	実施内容（成果）.....	5
	（1）全原子・粗視化 MDの最適化と並列化.....	5
	① GENESISの最適化と並列化（理研・杉田）.....	5
	② 粗視化モデル SPICA-FFの開発（岡山大・篠田）.....	6
	（2）データ科学的手法による粗視化モデルパラメタの最適化.....	12
	① データ同化によるシミュレーションと実験データの統合（埼玉大・松永）.....	12
	② 粗視化モデル CafeMolのパラメタ最適化（京大・高田）.....	18
	（3）生体分子系のモデリングと「富岳」を用いたシミュレーション.....	20
	① 細胞内分子混雑と液液相分離（理研・杉田）.....	20
	② 遺伝子転写機構の解明（京大・高田）.....	24
	③ 核内蛋白質・核酸複合体の動的モデリング（量研機構・河野）.....	27
	④ ウィルス（岡山大・篠田）.....	33
	⑤ 多剤排出トランスポータの分子動力学シミュレーション（東大・篠田）.....	38
	（4）プロジェクトの総合的推進.....	45
2-3.	活動（研究会の活動等）.....	46
2-4.	実施体制.....	47

補助事業の名称

「富岳」成果創出加速プログラム

全原子・粗視化分子動力学による細胞内分子動態の解明

1. 補助事業の目的

細胞はバクテリアから人に至るあらゆる生命における基本要素である。細胞内で働く蛋白質や核酸などの生体高分子の機能を理解することは基礎生命科学における重要な課題であるだけでなく、疾患の原因解明や薬剤開発などにおいても必要である。X線結晶構造解析やNMR（核磁気共鳴）、クライオ電子顕微鏡などの構造生物学によって生体高分子の立体構造の理解は着実に進む一方で、細胞内環境においてこれらの生体高分子が機能を発現する詳細については未解明の問題が多い。例えば、iPS細胞などの多能性幹細胞は再生医療などに重要な役割を果たすことが期待されており、その作製には数多くの遺伝子の中から4つの遺伝子を細胞に導入する必要があることが京都大学の山中教授らによって発見された。しかし、この4つの遺伝子によって発現される転写因子（蛋白質）がどのように遺伝子発現を制御しているのか、その実態は未解明である。すなわち、生命科学における多くの基本的な問題がブラックボックスとして放置されたまま、医療や創薬などへの応用が議論されている。

このような基本的な問題を解決する方法のひとつとして、計算科学・データ科学・実験科学を融合した新しい科学によって、原子・分子からボトムアップ的に生命現象を理解する試みが発展しつつある。この試みは、単純な構造を持つ原核細胞（バクテリアなど）について「細胞まるごと」の理解を深めるために、ヒト細胞などの複雑な真核細胞においては「細胞環境（細胞質・細胞膜・細胞核）を考慮した分子機能」を理解するために有効である。我々は「京」を用いることで、バクテリア細胞質に存在する非常に多くの蛋白質や核酸を含む系の全原子分子動力学(Molecular Dynamics, MD)シミュレーションを実現した。この計算は1億個以上の原子を含む2016年当時世界最大級の計算であった。また、蛋白質間に働く弱い相互作用が蛋白質の構造安定性や分子間相互作用に与える影響を予測したことで高く評価されている。また、米国ロスアラモス研究所のスーパーコンピュータTrinityを用いた国際共同研究によって、染色体の一部を含む10億原子を超える世界最大の系のMD計算も行なった。これらの計算は理研で開発しているMDソフトウェアGENESISを用いて実行されたものであり、非常に大規模な生体分子系の全原子MDシミュレーションを行う上では世界的にも優位な立場を築いている。

全原子MDシミュレーションでは各分子に含まれる分子運動を正確に解く必要があるため、スーパーコンピュータを用いたとしても計算できる時間はナノ秒 (ns: nanosecond, 10^{-9} 秒) からマイクロ秒 (μ s: microsecond, 10^{-6} 秒) 程度に限定されており、遺伝子発現など生命科学における本質的な問いに答える計算を行うことは困難であった。本研究では、「京」などで実績のある全原子MDシミュレーションに加えて、複数の原子を束ねた粗視化分子モデル（粗視化モデル）を用いた超大規模MDシミュレーションを「富岳」上でGENESISを用いて行うことにより、 μ mスケールの生体分子系に関するミリ秒 (ms: millisecond, 10^{-3} 秒) 以上の遅い生命現象の解明を目指す。大規模な生体分子系の粗視化MDシミュレーションは過去にほとんど例がなく「富岳」の優れた計算性能とGENESISを用いて可能になる超並列計算を組み合わせることで初めて実現する。このプロジェクトのメンバーである高田彰二（京都大学）らが開発しているCafeMolや篠田渉（岡山大学）らが開発しているSPICAなどの生体高分子の粗視化モデルは、世界的にも優位性の高いものであり、実際、既に多くの成果を挙げている。本プロジェクトではこれらのモデルをGENESISに

導入し、「富岳」上で大規模粗視化MDシミュレーションを行うとともに、分子構造と相互作用の詳細については、全原子MDシミュレーションの知見と組み合わせる。粗視化MDシミュレーションで用いられる分子内および分子間相互作用を記述するパラメタは、立体構造データベースを用いて経験的に決定されてきた。このパラメタの信頼性を改善するために、機械学習などのデータ科学の方法を用いて最適化する。さらに、別プロジェクトなどで連携している実験科学の結果を取り込み、計算結果と比較していくことで分野融合を行い新しい研究分野を創生する。

2. 令和3年度（報告年度）の実施内容

2-1. 当該年度（令和3年度）の事業実施計画

（1）全原子・粗視化分子動力学の最適化と並列化

① GENESISの最適化と並列化（理研・杉田）

分子動力学ソフトウェア GENESIS を「富岳」で活用するために必要な機能の追加等を行う。量研機構の河野らと協力して、蛋白質・核酸等の効率的な構造探索を実現するアルゴリズムを導入する。また、名大篠田や京大高田らと協力して粗視化モデル SPICA と CafeMol を GENESIS に導入し、「富岳」に対して最適化・並列化を行い、高速化を推進する。

② 粗視化モデル SPICA-FF の開発（岡山大・篠田）

粗視化 SPICA 力場の分子ライブラリを充実し、様々な混合脂質＋蛋白質系に対して、汎用化を行う。理研杉田と協力して、SPICA 力場の GENESIS への導入・最適化を継続して実行し、分極水バージョンの pSPICA 力場の移植も追加して行う。

（2）データ科学的手法による粗視化モデルパラメタの最適化

① データ同化によるシミュレーションと実験データの統合（埼玉大・松永）

前年度では、分子動力学シミュレーションの結果と実験データを統合するための機械学習（隠れマルコフモデル）に必要なツール群を、ライセンスが必要な MATLAB 言語から Julia 言語へ全面的に移植し広く使用できるように拡張した。また、高速原子間力顕微鏡データへも応用できるように整備し、応用範囲を拡大した。本年度では FRET や原子間力顕微鏡の 1 分子計測データへの応用研究を推進し、「富岳」を用いたデータ同化シミュレーションを実施することでターゲット分子の構造揺らぎの詳細な動態を調べる。

② 粗視化モデル CafeMol のパラメタ最適化（京大・高田）

粗視化モデル CafeMol は、蛋白質・核酸の計算を行うために必要なアミノ酸・塩基・それらの相互作用などについてのパラメタが求められている。前年度のテスト計算の結果を踏まえて、分子・細胞レベルのシミュレーションに必要な細胞膜を構成する脂質分子および非特異的蛋白質・蛋白質相互作用の粗視化パラメタのさらなる改良を行い、より精度の高い分子動力学シミュレーションの実現に向けて研究を推進する。名大篠田や埼玉大松永と連携して、全原子シミュレーションの結果を参照とした機械学習による分子内・分子間相互作用の検証と最適化を試みる。

(3) 生体分子系のモデリングと「富岳」を用いたシミュレーション

① 細胞内分子混雑と液液相分離 (理研・杉田)

細胞質は多くの生体分子で混み合った環境であり、この環境における蛋白質の構造・ダイナミクス・機能の関係を調べる。特に、蛋白質と基質の相互作用がターゲット以外の蛋白質が存在する場合にどのように変化するか、また、ATPなどの基質が存在することで凝集を防ぐ分子機構を解明する。また、蛋白質と核酸が含まれる液液相分離の問題に取り組む第一歩として、粗視化分子を用いた三次元構造モデルを構築し、CafeMolを用いた粗視化分子動力学によって構造と安定性を調べる。また、連携研究者である東工大・北尾と協力して細菌べん毛のモデリングを行う。

② 遺伝子転写機構の解明 (京大・高田)

染色体の高スループット実験から得られるHi-C/Micro-C/MNase-seq等の実験情報を用いて構築した三次元クロマチン構造粗視化モデルに、ChIP-seq等の実験情報を用いて結合蛋白質(リンカーヒストン、転写因子等)を付加し、細胞核内状況に近いモデルを構築する。このモデルについて、粗視化モデルCafeMolを用いて詳細な動態シミュレーションを実施し、遺伝子座における転写活性と染色体高次構造に関する知見を得る。

③ 核内蛋白質・核酸複合体の動的モデリング (量研機構・河野)

理研杉田らと協力して、蛋白質・核酸等の効率的な構造探索を分子動力学計算プログラム GENESIS に実装し、前年度に構築したヌクレオソーム複合体の原子モデルについて分子動力学計算を実行し、その構造体の構造多様性や構造遷移の仕組みを調べ、クロマチン動態の知見を得る。

④ ウィルス (岡山大・篠田)

前年度に引き続き、核酸を含有するB型肝炎ウイルス(HBV)ビリオンのモデリング及び分子シミュレーションにより、ビリオンの動態と相互作用の解明を目指す。本年度は、エンベロープを模倣するため、前年度にモデリングで作成したエンベロープ蛋白質を多数取り込んだ脂質膜を構築し、その平面膜及びベシクルについて構造安定性を解析し、さらにカプシドを内包するエンベロープ全体構造の分子シミュレーションを実行する。

⑤ 多剤排出トランスポータの分子動力学シミュレーション (東大・篠田)

前年度に構築した大腸菌モデル膜に埋め込まれた多剤排出トランスポータAcrABZ-To1C複合体の分子動力学シミュレーションを実行し、大腸菌モデル膜と多剤排出トランスポータ合体の相互作用を解析する。また、本年度は新たに、大腸菌膜全体の5%程度しか存在しないが分子量の大きい脂質のカルジオリピンの力場を作成し、さらにリアルな大腸菌モデル膜を構築する。

(4) プロジェクトの総合推進

プロジェクト全体の連携を密としつつ円滑に運営していくための研究成果報告会・シンポジウムを適宜、オンラインまたはオンサイトで開催する。プロジェクト内の研究の進捗状況および成果の発表のためのワークショップ・セミナー等を開催する。プロジェクトで得られた成果については学会発表等によ

り、積極的に公表する。また、他のプロジェクトとの連携などにより、効率的・効果的な研究の推進を行う。

2-2. 実施内容（成果）

（1）全原子・粗視化 MD の最適化と並列化

① GENESIS の最適化と並列化（理研・杉田）

分子動力学 (MD) シミュレーションは様々な生命現象を理解するのに強力なツールである。水分子を含む全原子モデルは自由エネルギー地形が険しくかつ時間刻みが短いため、大規模の系を長時間計算することは困難である。蛋白質残基ベースの粗視化 (CG) モデルは比較的正確性が低い、全原子モデルが抱える問題点を解決できると期待されている。しかし、クロマチン等の巨大生体分子は不均一な粒子密度分布を持つためドメイン分割による並列化が難しく、CG モデルで未だに応用されている例は少ない。本研究では CG モデルに対して効率的な動的負荷分散を持つ “Cell-Based kd-Tree “と呼ばれる新しい並列化スキームを開発した。

既にGENESISに導入されている全原子の並列化スキームと同じく、系をセルと呼ぶドメインに分ける。しかし全原子モデルでは各CPUに対して同じセル数が与えられるが (図1-1-1左)、本モデルでは各CPUに対しセル数ではなく、粒子数が同じになるようにした。まずは、系を同じ粒子数を持つ2つのドメインに分ける (図1-1-1右)。次に各ドメインを更に2つに分け、ドメイン数がCPU数と同じになるまで分割を繰り返す。従来の全原子の配列の形のままであるメモリの問題が発生する恐れがあるため、新しいドメイン分割スキームに基づいて新しい配列と相互作用スキームを設計した。

約760,000のCG粒子を含むクロマチン系に対して、本研究で開発した並列化スキームのベンチマーク計算を行った。その結果、負荷不平衡が的確に減少されることにより、演算性能が3~4倍向上したことを確認した。加えて、メモリを効率的に利用できるため、使用するCPU数を更に減らすことが可能になり、従来のCGMDソフトウェアに比べて10倍以上の性能向上が期待できる。

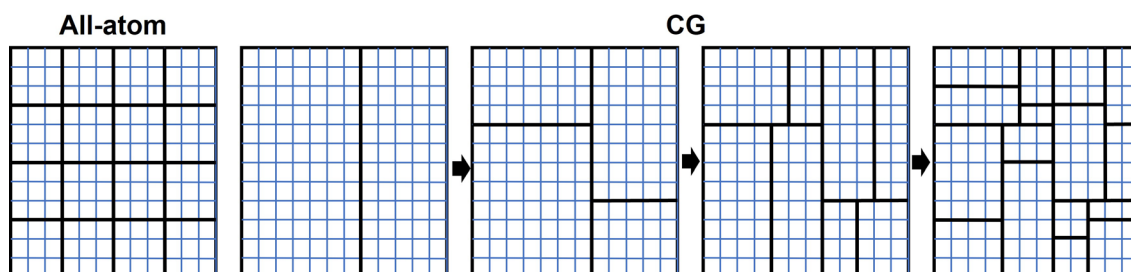


図1-1-1：全原子モデル（左）と粗視化モデル（右）のドメイン分割スキームを2次元で表現した。

16MPIを用いた計算例。

令和3年度には蛋白質残基ベースの粗視化モデルを複数、GENESISに導入した。導入されたモデルを用いて生体分子のMDシミュレーションを行う際には、初心者にとっていくつかの困難点がある。一つの主要な問題点は不均一生体系のMD入力ファイルを作成する作業が複雑になり得ることである。この作業が確立されている全原子力場と異なり、各CGモデルがそれぞれ固有の特徴を持ち、多種類の生体分子が存在する系に対して、MD入力ファイルを一貫したやりかたで作成するツールはこれまで存在しなかった。そこで、残基レベルのCGモデルを用いたMDシミュレーションの入力ファイルを簡単に作成するツールボックスを開発した。このツールボックス”GENESIS-cg-tool”は、蛋白質・天然変性蛋白質・DNA・RNA等様々な生体分子に対応し、すでにフリーソフトとして公開した (図1-1-2, Tan et al. *PLoS. Comp.*

Biol. 2022)。本ツールボックスは効率的な完全なオープンソースかつ最先端である Julia プログラミング言語で書かれている。GENESIS-cg-tool のエントリーポイントは“aa_2_cg.jl”というスクリプトであり、UNIX 系シェルのコマンドラインから実行できる。このスクリプトは PDB (protein data bank) を、主な入力ファイルとして扱い、モデルに関係するいくつかの引数を受け取る。最初に PDB ファイルから原子座標が読み込まれる。次に PDB 構造から短距離と長距離相互作用パラメタ (基準値も含む) が決定され、最後に各モデルに基づく力定数等のパラメタが決定される。

GENESIS-cg-tool を用いて RNA ポリメラーゼやヌクレオソーム等、蛋白質と核酸双方を持つ様々な不均一体系が容易に作成できる。1,024 個のヌクレオソームを含む人工的なクロマチン構造をこのツールを用いて作成した。本系では折り畳まれたヒストンに対し AICG2+モデルを、天然変性ヒストンテールに対し HPS モデルを、DNA に対し 3SPN.2C を、さらに蛋白質-DNA 間に対しては配列非特異的水素結合モデルを利用した。更に、長距離相互作用の決定部分では Julia のスレッド並列ライブラリを用いたことにより MD ファイル作成が高速化された。

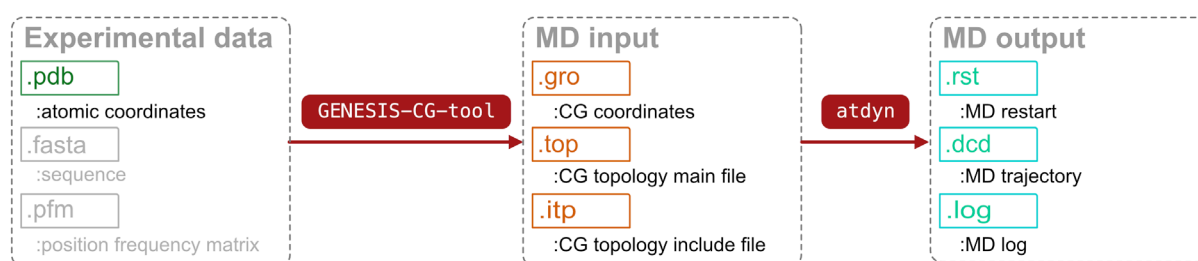


図 1-1-2 : GENESIS-cg-tool を用いた GENESIS/ATDYN で CGMD シミュレーションのフローチャート。

導入された CG モデルを用いて、転写因子の相分離挙動および DNA への結合のシミュレーションを行った。最近公開された異なる HPS モデルのパラメタを GENESIS で試した。転写因子の相分離系の相互作用強度を実験データに基づいて構成し、チューニングを行った。

② 粗視化モデル SPICA-FF の開発 (岡山大・篠田)

SPICA 力場は重原子 3 つ程度を含むグループを一つの粗視化粒子 (相互作用点) と見なす解像度を持つ粗視化力場であり (図 1-2-1)、そのパラメタ決定は、熱力学量の実験データ、全原子 MD から得られる分子分布関数や自由エネルギーデータを参照して行う。また、全原子力場のようにクロス相互作用パラメタをコンビネーションルールによって決めることができないため、分子ライブラリの拡充には多大な労力を必要とするが、完成した力場モデルは、例えば脂質膜では、界面張力、密度、溶媒和自由エネルギーを再現し、粒子の分布関数 (構造) もほぼ全原子モデルの結果を再現する精度を持つ。さらに、脂質膜系では膜弾性係数 (膜面積圧縮係数、曲げ弾性係数、ガウス係数) も実験をよく再現すると同時に、膜の線張力もよい一致を示すため、メゾスコピックな脂質膜系の集合構造 (モルフォロジー) を予測的に計算することができる大きなメリットである。今年度は、蛋白質モデル (図 1-2-1) の改良として、ホスファチジルコリン (PC) 脂質膜との相互作用パラメタの改善を行い (Kawamoto et al. *J. Chem. Theory Comput.* 2022)、さらに新たな脂質膜 (ホスファチジルエタノールアミン (PE)、ホスファチジルセリン (PS)、ホスファチジルイノシトール (PI)、ホスファチジルグリセロール (PG)) との相互作用パラメタを決定し、多様な

蛋白質・脂質膜複合系で使用可能とすべく、力場の汎用化を行った。

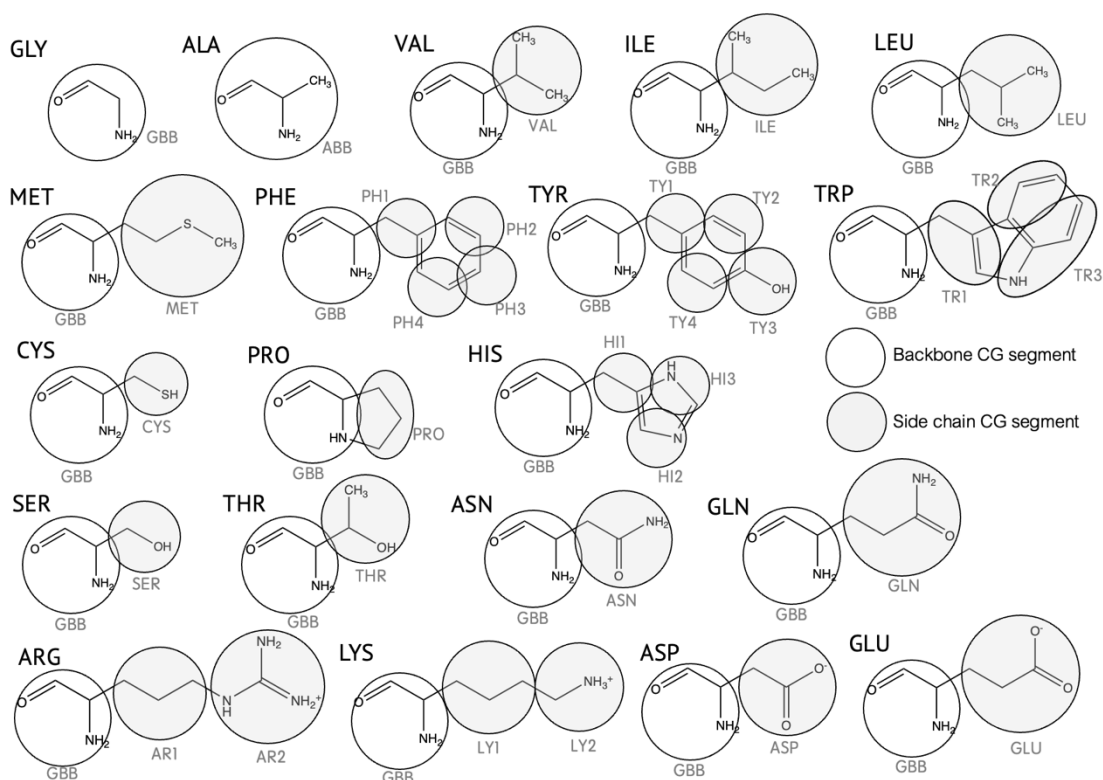


図 1-2-1: 各アミノ酸残基の粗視化粒子の定義。アミノ酸骨格サイト(GBB, ABB)と側鎖サイトからなる。ベンゼン環を持つ粒子は、その平面構造を表現するため、粗視化粒度を細かく設定している。灰色の文字は各粗視化サイトの名前を示す。

新しい蛋白質モデルの側鎖粒子と脂質膜の相互作用の修正として、CHARMM36 力場を用いた全原子モデルでのアミノ酸側鎖アナログの膜を横切る自由エネルギープロファイルを参照データとして用いた。以前は OPLS-AA 力場を用いていたが、他の脂質種との整合性を保つために、CHARMM36 力場を採用している。計算は Adaptive Biasing Force (ABF) 法を用いて、膜中心からの距離 $Z = 0-30 \text{ \AA}$ の領域の自由エネルギーを計算した。計算を効率化し、収束を早めるために 5 \AA のウィンドウに分け、それぞれ $400-500 \text{ ns}$ の ABF-MD を行った。粗視化力場では各ウィンドウ、 300 ns の計算でプロットの線幅以内で収束した。17 種のアミノ酸側鎖アナログに対しての自由エネルギー計算結果を図 1-2-2 に与える。

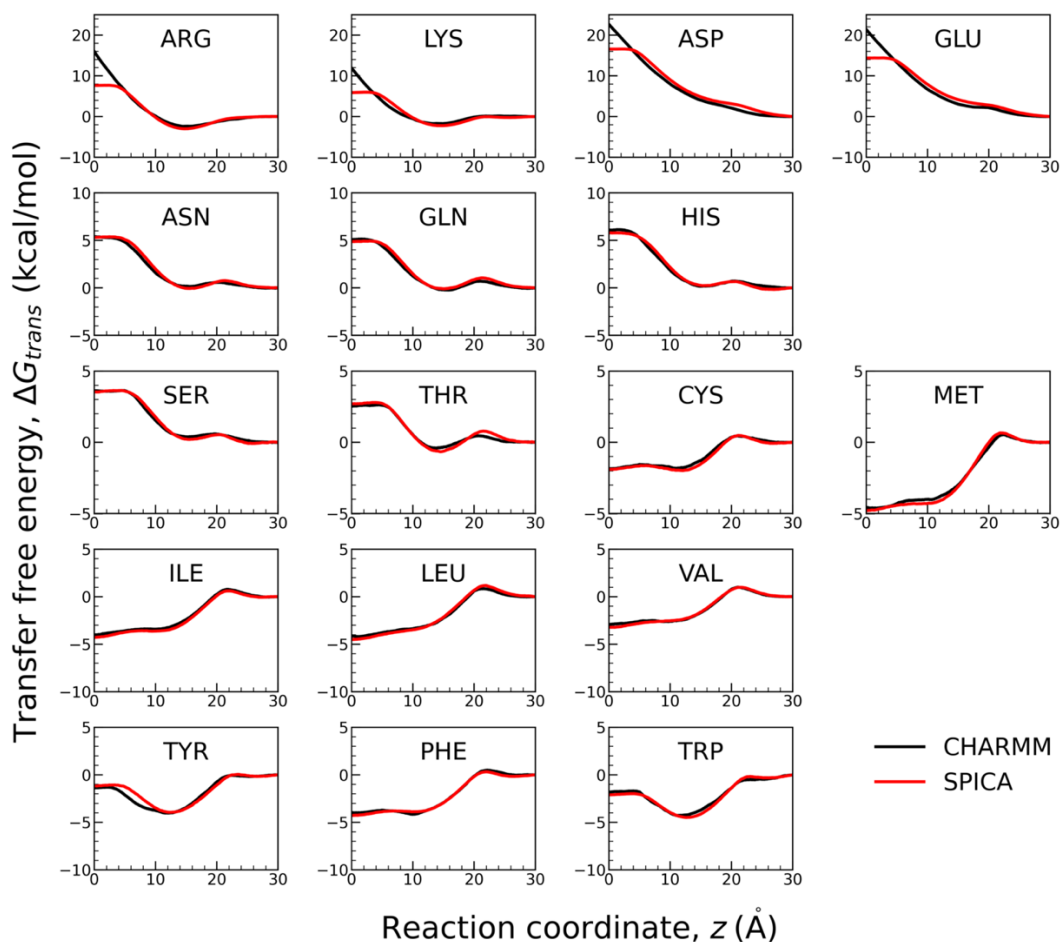


図 1-2-2: アミノ酸側鎖アナログ分子の DOPC 脂質膜を横切る自由エネルギープロファイルの比較。全原子力場 CHARMM36(黒)と SPICA(赤)の比較。反応座標は膜垂直軸に沿っての脂質膜重心からの距離。

図に示したように、SPICA 力場は全原子モデルとほぼ同等の自由エネルギープロファイルを与える。ただ、top panel にあるイオン性の 4 残基については、膜内部での自由エネルギー障壁を多少、過小評価する傾向が見られた。これは、イオン性分子は水分子を伴って膜内部に侵入し、細い水の糸 (water string) を形成するが、SPICA 力場では water string の再現が難しいためである。この問題は極性を持つ水モデルを使用する pSPICA 力場で完全に解決される (Miyazaki et al. *J. Chem. Theory Comput.* 2020)。イオン性分子が膜を透過するようなイベントを伴う現象の解析には、pSPICA 力場の使用が望まれる (Miyazaki & Shinoda, *BBA-Biomembranes*, 2022)。ここでは、DOPC 膜に対する自由エネルギープロファイルの結果を載せたが、同様に POPE、POPS、POPI、POPG 膜についてもほぼ同等精度で全原子モデルの結果を再現する粗視化パラメタの生成を達成している。

この他、脂質膜-蛋白質ペプチド骨格相互作用パラメタは、Orientation of Proteins in Membranes (OPM) database に登録されている 100 種以上の膜周辺タンパク、ペプチドの脂質膜へのアンカリングの深さやペプチドの tilt 角を再現するように調整した。また、膜貫通型ペプチドの脂質膜中での二量体化自由エネルギーを再現するようにペプチド骨格と側鎖間の相互作用を決めている。これらの相互作用パラメタ調整は、ペプチド主鎖骨格に対して Elastic Network Model (ENM) を採用し、ペプチド・タンパクの 2 次構造に対する拘束をかけた計算を行っている。ENM は非常に簡便で扱いやすいが、蛋白質の大きな

構造変化を扱えないという問題もあるため、ENM の代替として Go-like モデルを SPICA においても使用可能とする拡張を行っている。

SPICA 力場の MD ソフトウェア GENESIS への組み込みを、前年度から継続して行っている。特に本年度は、分極水モデルを元に粗視化力場を再構築した pSPICA 力場の GENESIS への組み込みを行った。pSPICA 力場の構築の際に使用した LAMMPS を用いた MD 計算の結果(物性量や構造)を再現することを確認した。pSPICA 力場における SPICA 力場との大きな違いは、水分子(とイオン分子)の取り扱いである。SPICA 力場では、水モデルとして、3つの水分子を1つの電荷を持たない粗視化(CG)粒子として取り扱い、水と他粒子との相互作用は Lennard-Jones (LJ) 12-4 型ポテンシャルを使用して記述する。しかし、pSPICA 力場においては、水分子はダイポールを持ち、正負の異なる符号を持った同じ大きさの電荷をもつ 2 サイトにより 1つの CG 水分子が表現されている。このような分極水の表現は、先に少し述べたように、イオン種が膜中を拡散する場合や膜に細孔を生成するような現象を精度良く再現するために必要である。また、pSPICA の分極水間のポテンシャルの関数形も異なり、水-水の vdW 相互作用は次のような LJ12-5 型のポテンシャル関数を用いている。

$$U_{LJ}(r_{ij}) = \frac{12}{7} \left(\frac{12}{5}\right)^{\frac{5}{7}} \epsilon_w \left[\left(\frac{\sigma_w}{r_{ij}}\right)^{12} - \left(\frac{\sigma_w}{r_{ij}}\right)^5 \right]$$

また、水粒子内のダイポールは固定で、すなわち水内サイト間距離を SHAKE アルゴリズムなどのホロノミック拘束のアルゴリズムを用いて、時間発展させる必要がある。電荷間の相互作用については、比誘電率 3.2 を用いて、particle mesh Ewald (PME) 法のようなカットオフ無しの静電相互作用計算により評価される。他の脂質分子や蛋白質分子について、相互作用パラメタは SPICA と異なり、分極水に合わせて最適化されているが、関数形も含めて相互作用計算の扱いは同様である。

表 1-2-1: pSPICA 力場のテスト系として、POPC 平面膜の MD シミュレーションの計算条件。

Number of POPC molecules	1, 152
Number of Water molecules	64, 742
Total Number of CG particles	147, 916
Box size	190 x 190 x 200 [Å]
LJ Cutoff	15 [Å]
Integrator	velocity Verlet
Times step size	10 [fs]
Temperature Control	Velocity rescaling
Temperature	310 [K]
Pressure Control	Parrinello-Rahman Semi-isotropic coupling
Pressure	1 [atm]

GENESIS/SPDYN に pSPICA 力場が正しく移植されたかを判定するため、表 1-2-1 に示すような計算条件で POPC 脂質二重膜での MD シミュレーションの結果を比較した。まず、GENESIS と LAMMPS を用いて、同

じ初期構造に対してポテンシャルエネルギー、各粒子にかかる力の値が数値精度の範囲内で一致することを確認した。また、それぞれのプログラムで MD 計算を行い、膜物性量が統計精度内で一致することを確認した。例えば、脂質分子あたりの膜面積の値は $62.5 \pm 0.5 \text{ \AA}^2$ (GENESIS), $62.4 \pm 0.4 \text{ \AA}^2$ (LAMMPS) と両者はよく一致した。この他、膜厚、疎水鎖のオーダーパラメータ、各粒子の分布関数などを比較したところ、すべてが非常によい一致を示した。これにより、pSPICA 力場が GENESIS に正しく移植できていることを確認した。

さらに、SPICA 力場を用いた「富岳」での本格計算(長時間 MD)を行う必要があるため、GENESIS-SPICA の「富岳」での並列計算の最適化と高速化に、理研杉田グループと共に取り組んだ。前年度までに、「富岳」において、MPI 並列化と OpenMP スレッド並列のハイブリッド並列での計算速度の測定を行ってきた。それにより、コア数に比例して、弱い線形のスケール並列が確認された。今年度は、これらの測定結果をもとに、さらなる「富岳」での並列計算の最適化と高速化に取り組んだ。特に、SPICA による MD での実効性能を追求するため、アルゴリズムによる高速化を目的として、マルチタイムステップ(MTS)法を用いた MD を GENESIS-SPICA で適用した。MTS 法では、短周期で変化する比較的強い力と長周期で変化する弱い力を異なる時間刻み(タイムステップ)を用いて数値的に運動方程式を解く。特に後者を長い時間刻みに取ることで、エネルギーの保存性を保ちつつ、その演算回数を減らすことができる。この弱い力は PME の逆空間からの寄与に取ることで、並列計算において通信負荷の高い高速フーリエ変換(FFT)の演算回数を減らすことができるため、並列計算効率が向上する。MTS 法では、時間刻みをより多段階で調整することができるが、本検討では、分子内相互作用+分子間相互作用の実空間部分の相互作用(短距離力)を短時間刻みで更新し、PME の逆空間の相互作用(長距離力)を長時間刻みで更新する 2 つの時間刻みでベンチマークを実行した。なお、熱浴変数の運動の更新も長時間刻みで行っている。時間刻みの取り方は、一般に、全エネルギーなどの保存量の確認によって決められるが、ここでは、より厳しい基準として、系の粒子のダイナミクス、特に並進の時間自己相関関数として、平均二乗変位(MSD)を使い、自己拡散係数が影響を受けない範囲で、時間刻みを延ばし、MTS を採用することとした。表 1-2-1 に示す POPC 平面膜系でシングルタイムステップ(STS)と MTS の MD 計算を行い、水粒子と脂質分子の MSD を用いて評価した。ただし、この検討では、温度を 300K として計算している。図 1-2-3 が示すように、水分子の MSD は時間刻み 5fs と 10 fs の STS で全く同一である。そのため、これらの量を短距離力の時間刻みとして採用し、長距離力の時間刻みを長くしていった。MTS 法を用い、短い時間刻みを 5 fs と 10 fs のどちらを用いても、長い時間刻み 50 fs までは、MSD が STS のものから変化しないことがわかった。脂質分子の MSD を同じ MD から計算した結果を図 1-2-4 に示す。脂質分子は 1, 152 分子であり、8 万以上含まれている水分子よりも統計量が少ない。MD 計算時間が 25 ns であり、統計誤差を考慮すると、脂質の MSD も、これらすべての条件において、統計誤差内で一致している。MSD の傾きとして求められる POPC 脂質の側方自己拡散係数は $3.5 \pm 0.1 \times 10^{-6} \text{ cm}^2/\text{s}$ で一致している。粗視化 MD では、相互作用の平均化により自由エネルギー面が平滑化されて、自己拡散係数が実験(や全原子モデル)より 1 桁近く高くなることに注意が必要である。以上より、採用した時間刻みの組み合わせによる MTS 法の計算条件は系のダイナミクスを変えていないことが確認された。現在の GENESIS プログラムの通信の制約上、長時間刻みを 50 fs より長くすることはできない。そのため、最大効率の計算条件としては、MTS 法では短距離力を 10 fs 毎に更新し、長距離力と熱浴パラメータを 50 fs 毎に更新するのが、計算物性量を変えない範囲で、最も効率がよいことがわかった。

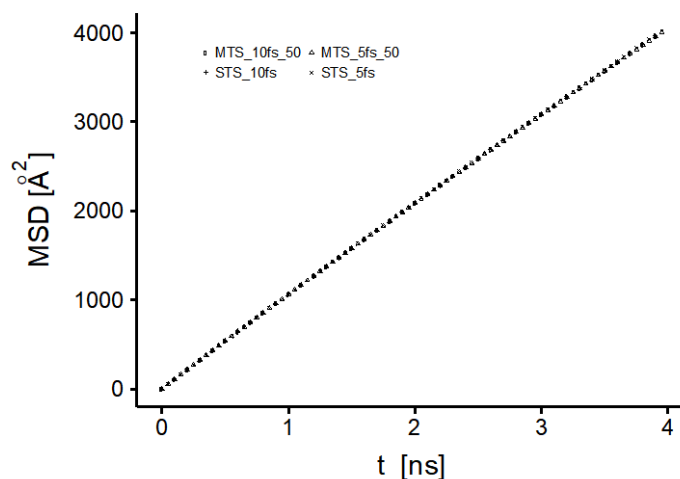


図 1-2-3: GENESIS で SPICA 力場を用いて平面膜系で計算した水分子の平均二乗偏差。計算方法は MTS と STS を用いた。MTS については、短距離相互作用は 5fs もしくは 10fs の頻度で計算を行い、長距離相互作用については、50 fs の頻度で計算を行なった。この計算条件において、統計誤差はシンボルの大きさよりも小さく、MTS は STS とプロット精度内で完全に一致した。

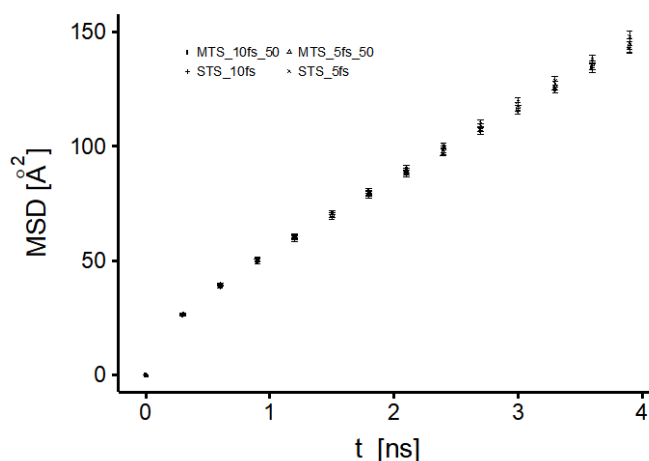


図 1-2-4: POPC 平面膜系で計算した POPC 分子の 2 次元平均二乗変位 (MSD)。MTS と STS を用いた。これらのすべての計算条件で、MSD は統計誤差内でほぼ一致した。

表 1-2-2: ST と MTS での計算速度比較。計算は「富岳」8 ノード 384core (64MPI, 6openmp) で行った。

時間刻み	計算速度 (ns/day)
ST 5 fs	7.0
ST 10 fs	13.9
MTS 5 fs - 50 fs	11.8
MTS 10 fs - 50 fs	21.1

次に、応用計算として長時間の CGMD を行うウィルス系での計算速度を「富岳」において実測した。本計測においては、使用する総コア数を変更して 1 日当たりの MD 計算時間を推定しており、その結果を図

1-2-5 に示す。計算に使用した系は、課題(3)-④で実際に使用する約 300 万粒子からなる B 型肝炎ウイルスエンベロープ系である。倍精度計算を用いて、比較のために、STS 法で 10fs 刻みを用いた場合の測定も行った。MTS 法における相互作用のアップデートは前出の膜系と同じであり、短距離力を 10fs の時間刻み幅、長距離力を 50fs の時間刻みとした。並列計算は MPI と OpenMP のハイブリッド計算により行い、最大のプロセス数はプログラムと計算系の制約を考慮し、計算効率の向上が見られる範囲まで増加させたところ、最大 1728 MPI と 12 OpenMP スレッドとわかった。その結果最大のコア使用数は 20,736 コアとなった。このときのパフォーマンスが 10fs-50fs の MTS では 234 ns/day、10fs 刻みの STS では 139 ns/day である。MTS の使用により、STS よりも 168 % への計算性能の増加が確認された。また、MTS による逆空間の力の演算回数の減少は、高速フーリエ変換(FFT)の演算を避けることができるため、並列性能の向上にもつながる。そのため、並列性能の線形スケーリング性を確認することを目的として、MPI の個数を減らした条件でも計測を行った。計測結果が図 1-2-5 である。図により、総コア数とパフォーマンスの間には弱い線形性能が確認された。MTS の使用は逆空間計算の評価回数の減少につながるため、STS よりも MTS の方がプロットの傾きが大きくなっていることが確認された。したがって、時間刻み幅による計算効率の向上だけでなく、並列性能の向上も MTS によりもたらされ、168 % への計算速度の増加につながったものと考えられる。

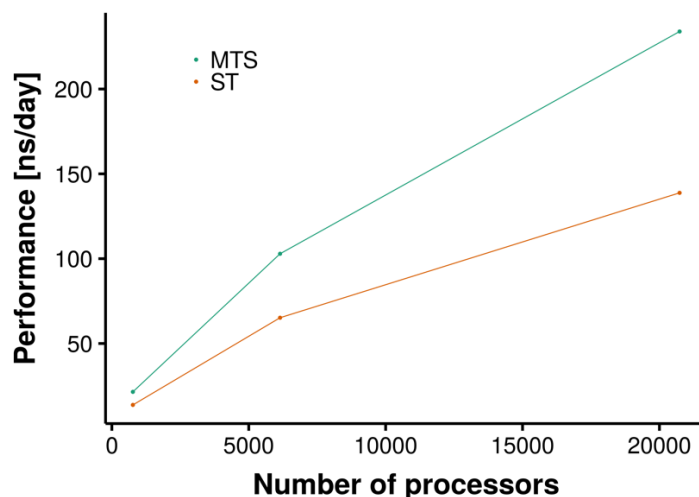


図 1-2-5: 「富岳」上で GENESIS を用いて、B 型肝炎ウイルスのエンベロープ粒子系、約 300 万原子系において、倍精度計算を用いて、10fs 刻みの STS 法と 10fs と 50fs 刻みの MTS 法を用いて、使用するプロセス数と 1 日当たりの MD 時間の計測値。

(2) データ科学的手法による粗視化モデルパラメタの最適化

① データ同化によるシミュレーションと実験データの統合 (埼玉大・松永)

生体分子のアンサンブルやダイナミクスを原子解像度で観測する手法として MD シミュレーションは強力である。しかし、用いる力場パラメタの精度が不十分な場合に、MD シミュレーションを行って得られるアンサンブル/ダイナミクス情報が不正確で実験とも整合的でなくなってしまうという弱点がある。それに対して、1 分子計測などの実験計測は、分子のアンサンブル/ダイナミクスを直接観測できるという強み

がある一方で、計測のために用いるプローブなどを経由した不完全な構造情報しか得ることができないという弱点がある。そこで近年盛んに研究されているのが、MDデータと実験計測データを統合してより高精度・高解像度の構造アンサンブル/ダイナミクスをモデリングするためのデータ解析手法である。これまでに我々は、統計数理のデータ同化手法に対して、MDのデータ解析で用いられるマルコフ状態モデルを導入して、生体分子向けのデータ同化手法を開発・応用してきたが(Matsunaga and Sugita, *JCP, eLife* 2018)、未だ一種類の計測データについてのみの対応で、小蛋白質のフォールディングに応用が留まっている。そこでデータ同化手法を他の実験計測手法でも取り扱えるように範囲を拡大すること、またより大きなサイズの細胞内現象へ関わる蛋白質へ応用可能にすることが求められている。

前年度において、データ同化に必要な機械学習(隠れマルコフモデル)に必要なツール群を整備し、さらに高速原子間力顕微鏡(HS-AFM)データへ適用できるように手法開発を行った。今年度は応用研究として、開発したツールと手法を用いて、味覚受容体のHS-AFMデータに対する応用研究を行い蛋白質機能にとって重要な構造ダイナミクスやアンサンブルを解析した。

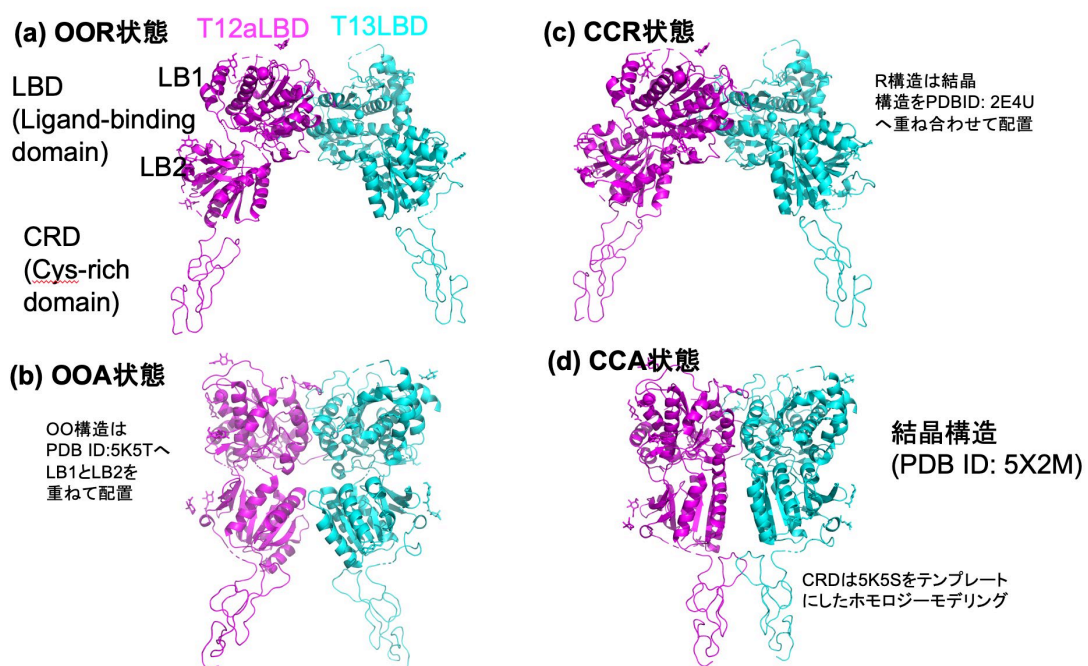


図2-1-1: 味覚受容体の結晶構造とモデリングで作成した他の構造。(a)両方のモノマーのLigand-binding domain (LBD)がオープンであり(OO型)、ヘテロ二量体がオープン型(R型)であるOOR状態。(b)OO型であり、ヘテロ二量体がクローズ型であるOOA状態。(c) CCR状態。(d)結晶構造がとられているCCA状態。

味覚受容体はClass-C G蛋白質共役型受容体(GPCR)ファミリーに属する蛋白質であり、脊椎動物において様々な味覚の感知を担っている。味覚受容体の構造は、ヘテロ二量体から成り、細胞外に500アミノ酸残基ほどのリガンド結合ドメインを持つ。最近、Nuemketらによって解かれたメダカ由来の味覚受容体の結晶構造を図2-1-1に示す(Nuemket et al. *Nat. Commun.* 2017)。Ligand-binding domainは味覚物質を結合し、結合に伴って細胞膜側のドメイン(Cysteine-richドメイン)が構造変化を起こし、下流へとシグナルが伝達されると考えられている。しかしながら、どのように構造変化を引き起こして活性化状態へと至るのか?途中の中間状態はあるか?なぜヘテロなのか、非対称性がどのようにリガンド認識に使わ

れるのか?などの具体的な問題が分かっていない。そこで本研究では、味覚レセプタのHS-AFMデータとMDシミュレーションとのデータ同化を行い、構造ダイナミクスや中間状態構造を詳しく解析した。

最初に、データ同化を行うための準備として、味覚レセプタの粗視化モデルの分子動力学シミュレーションを行ないとりうる構造を網羅的にサンプリングし、マルコフ状態モデルを構築した。まず、味覚レセプタの細胞外ドメインを作成するために、活性化状態である結晶構造(PDB ID:5X2M)に対して、そのホモログであるPDB ID:5K5Sをテンプレート構造としてCysteine-rich domainをホモロジーモデリングにより付加した。次に、2つのLigand-binding domainsがそれぞれ開いた構造を作成するために、他のホモログであるPDB ID:5K5Tへ両ドメインを重ね合わせた。また一方で、Cysteine-rich domainが開いた構造を作成するために、PDB ID:2E4U構造へ重ね合わせて構造を作成した。ただしこの場合はLigand-binding domainsのインタフェース部分がクラッシュしてしまったために、2E4Uへ重ね合わせた構造をターゲットとしてPDB ID:5X2M構造から全原子モデルのTargeted MDを行い、構造を作成した。同様の計算をLigand-binding domainsが開いた構造でも行った。得られた構造は全て、その後でC α 原子へpositional restraintを課した全原子モデルMDを行なって細かな箇所をrelaxさせた。次に、粗視化モデルとしてC α 原子ベースのKaranicolas-Brooks Go-modelを作成し、全原子のモデリングで得られた4つの構造を安定化するようにmacro-mixingさせてquad-basinのポテンシャルエネルギーを作成した。作成したモデルでMDシミュレーションを行い様々な状態間を遷移する網羅的なトラジェクトリを得た。全てのMDシミュレーションはGENESISを用いて行った。

得られた粗視化モデルのトラジェクトリのデカルト座標に対して主成分解析を行った。その結果、第1主成分はCysteine-rich domain間の距離、第2主成分は各Ligand-binding domainのサブドメイン(LB1 and LB2)間の距離に相当することがわかった。我々はこの第1主成分と第2主成分の空間でk-means clusteringを行い、マルコフ状態モデルの状態を定義した。今の粗視化モデルの場合、状態の数(クラスタの数)として50を用いると、マルコフ性の近似のもとで元のダイナミクスをよく捉えることがimplied timescale解析から示唆された。そこで50個の状態を用いてMDトラジェクトリから遷移確率を推定し、マルコフ状態モデルを構築した(図2-1-2)。

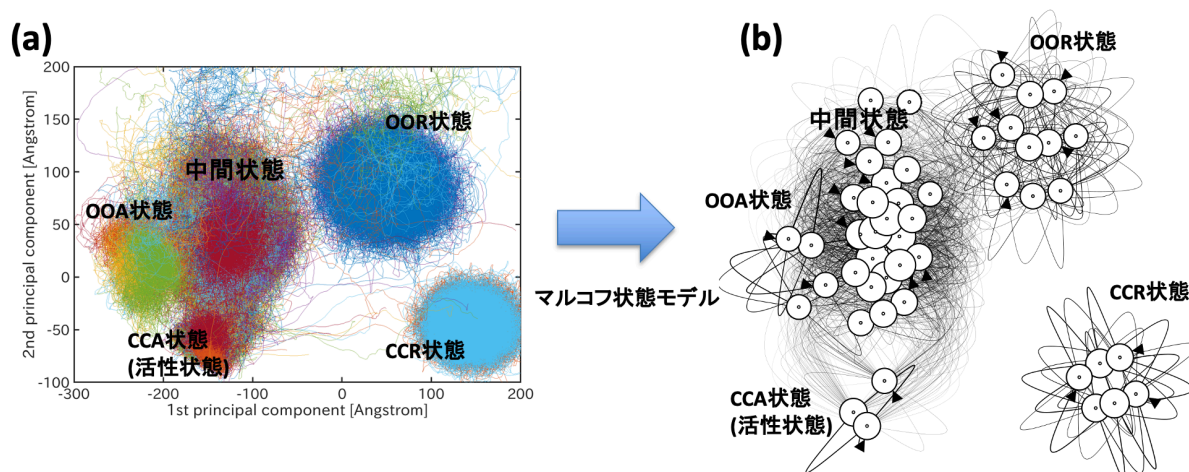


図2-1-2: 粗視化モデルシミュレーションの結果とそこから構築したマルコフ状態モデル。(a)粗視化モデルの分子動力学シミュレーションのトラジェクトリを第1主成分軸と第2主成分軸へ射影したもの。(b)トラジェクトリをクラスタリングして遷移行列を推定することで構築したマルコフ状態モデル。ノードの

面積が状態の平衡確率、エッジの太さが遷移確率を表す。

マルコフ状態モデルを可視化した結果、quad-basinの粗視化モデルを用いているにも関わらず、00R状態とCCA状態の間に準安定な中間状態が存在することが示唆された。実際、トラジェクトリのプロットのレベルでも多くのサンプル点が00R状態とCCA状態の間にあることが見てとれる。この中間状態領域の代表構造を解析した結果、ヘテロ二量体のうちの味覚物質の認識を担っている側であるT1r2aLDBのLigand-binding domainがクローズド構造をとっている一方で、もう一方側であるT1r3LBDはオープン構造をとっている非対称構造であることがわかった。このことから00R状態やCCR状態といった不活性化状態から、非対称な中間状態を通して活性化状態へ至る構造変化のパスウェイが粗視化モデルのシミュレーションレベルでも示唆された。

次に、構築したマルコフ状態モデルをHS-AFMデータを用いて補正してより高精度なモデルを得るために、隠れマルコフモデルを用いたデータ同化を行った。隠れマルコフモデリングでは、HS-AFM計測で得られる一連の画像を時系列データと見なして、モデルからその時系列が得られる確率(尤度)が最大化するように遷移確率を補正する。ただし、マルコフ状態モデルの各状態は3次元構造である一方で、HS-AFMデータは2次元画像であり、両者を比較して尤度を計算するプロトコルは自明ではない。ここでは、前年度開発したHS-AFM用の隠れマルコフモデリング手法を適用してデータ同化を行った。前年度開発した手法では、HS-AFM画像の時系列データ(フレーム1~ T)に対する尤度は以下のように定義される:

$$L(m_1, \phi_1, dx_1, dy_1, \dots, m_T, \phi_T, dx_T, dy_T) \\ = \sum_{s_1=1}^M \dots \sum_{s_T=1}^M p(s_1) L(I_t | m_1, \phi_1, dx_1, dy_1) \prod_{t=2}^T T_{s_{t-1}s_t} L(I_t | m_t, \phi_t, dx_t, dy_t)$$

ここで、 I_t は t フレーム目のAFM像、 m_t, ϕ_t, dx_t, dy_t はそれぞれ t フレーム目のマルコフ状態モデルの状態の構造、向き、 x 座標と y 座標の並進である。更に、 $T_{s_{t-1}s_t}$ は構造と向きとで定義された状態 s_{t-1} から s_t への遷移確率である。 $L(I_t | m_t, \phi_t, dx_t, dy_t)$ は、 t フレーム目のAFM像 I_t に対する尤度である。HS-AFM画像には構造の向きの情報があるので、マルコフ状態モデルの状態が構造だけでなく向きとの直積で拡張されている。ただしHS-AFM計測では、一般に計測対象の分子は基盤と相互作用しているため、回転の緩和は遅く、隣接したフレーム間では似た回転方向をとっていることが多い。そこで前年度は、あらかじめ回転角度が離れた状態間の遷移確率をゼロとして、大きな回転が生じないような拘束を遷移確率へ課したBaum-WelchアルゴリズムによりHS-AFMデータから精度良く遷移確率を推定する手法を開発した。この手法を実際の味覚受容体のHS-AFMデータへ適用した。

データ同化することで補正したマルコフ状態モデルを図2-1-3に示す。HS-AFM計測は、味覚物質であるリガンド(L-Glu)がある条件(5mM)と存在しない条件で行っており、双方の条件でそれぞれデータ同化を行った。まずリガンドありの条件では、期待通りにリガンド結合状態の結晶構造であるCCA状態が安定化しており、活性化状態として選ばれていることがわかる。他の非活性化状態と予想される状態(CCR状態、00R状態)はシミュレーションデータのみから作成したマルコフ状態モデルに比べて不安定化している。一方で、リガンドなしの条件では、不活性化状態の一つである00R状態が安定化することがわかった。特にもう一つの非活性化状態の候補であるCCR状態は安定化しないことがわかった。以上のことより、味覚受容体の活性化へ至る構造変化パスウェイとして、不活性化状態である00R状態においてリガンドが結合

し、途中の非対称な中間状態を経由して、活性化状態であるCCA状態へ至るという描像を得ることができた。ただし、リガンドなしの条件ではCCA状態も安定化することがわかった。これは味覚受容体が基盤と強く相互作用しているoutlierデータが影響していると思われる(後述の解析を参照)。

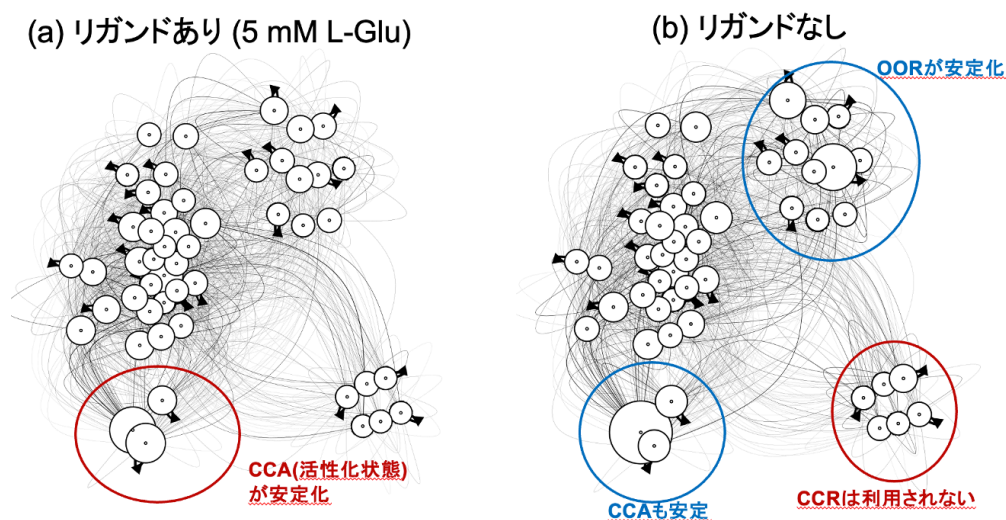


図 2-1-3: 高速原子間力顕微鏡データと同化することで補正したマルコフ状態モデル。(a) リガンドあり (5 mM L-Glu) の条件における計測データを同化した結果。(b) リガンドなしの条件における計測データを同化した結果。

データ同化をして得られたマルコフ状態が適切か否かを調べるために、2通りの検証を行った。1つ目の検証では、隠れマルコフモデルで学習した遷移確率が適切に時系列データの揺らぎ(HS-AFM 動画)を表現できているかどうかを調べるために、データ後のリガンドあり・なしのマルコフ状態モデルを用いて、学習には用いなかったテストデータであるリガンドあり・なしの時系列データ(HS-AFM 動画)の尤度を計算した(図 2-1-4)。もし各モデルの遷移確率がリガンドあり・なしの時系列データの揺らぎを上手く捉えているならば、リガンドあり・なしの条件をよく識別できるはずである。両者の尤度を比較した結果、例外はいくつかあるが、概ねリガンドあり・なしのHS-AFM 動画を識別できていることがわかった。

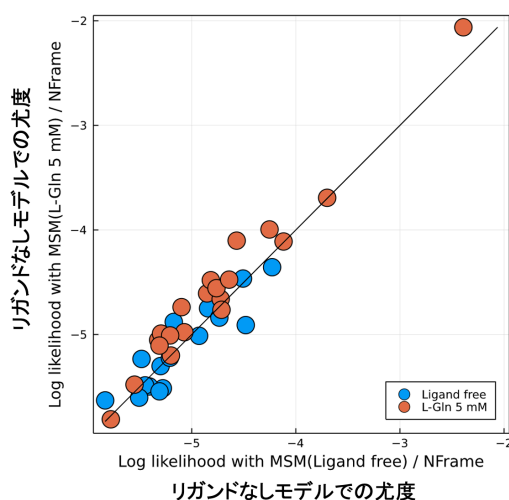


図 2-1-4: リガンドあり・なしの条件のマルコフ状態モデルを用いて、リガンドあり・なしの条件における高速原子間力顕微鏡データの尤度を計算し識別した結果。

2 つ目の検証として、他の独立に行なわれた実験データと整合的か否かを検証した。最近、味覚受容体の FRET 計測が行われており、リガンド濃度に依存して、Cysteine-rich domain 間の距離分布のピークがシフトすることが分かっている (Liauw et al. *Nat. Chem. Biol.* 2021)。そこで、マルコフ状態モデルから Cysteine-rich domain 間の距離分布を計算し両者を比較した。まず、リガンドありの条件では Cysteine-rich domain 間の距離が小さい箇所にピークがあり、定性的には一致することがわかった。一方で、リガンドなしにおける FRET 計測の結果は一つのピークが距離が大きい箇所に存在するが、マルコフ状態モデルでは2つのピークが生じることがわかった。定性的な比較では、FRET のピークは OOR 状態に帰属できることがわかった。一方で、マルコフ状態モデルの CCA 状態に対応するピークは FRET では見られないことがわかった。

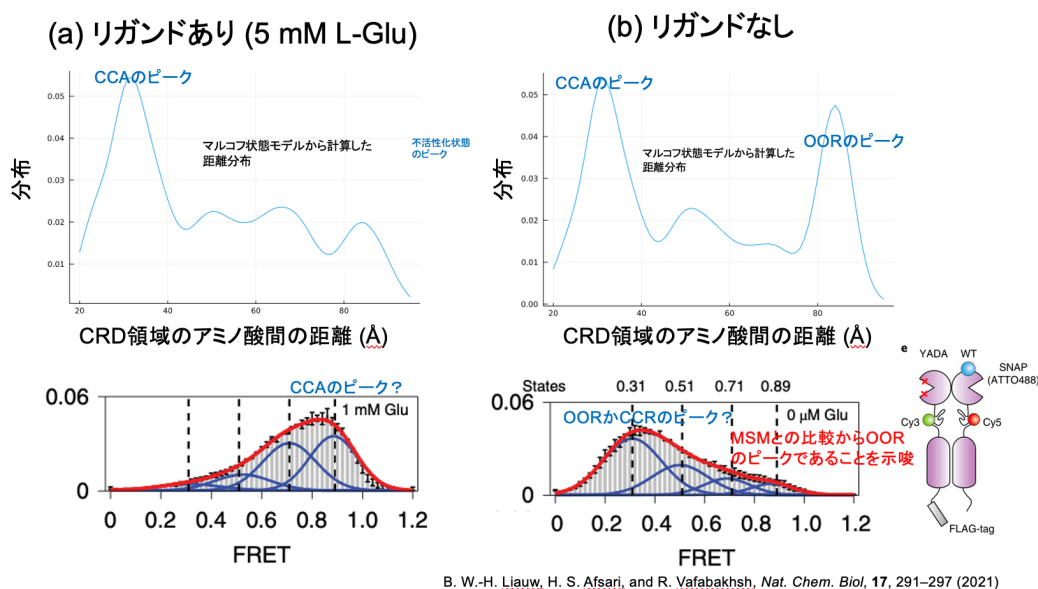
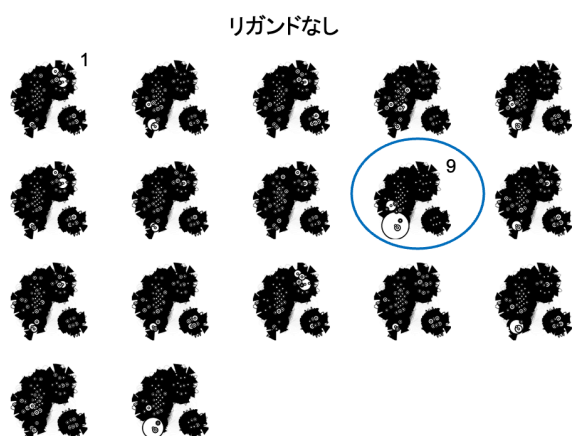


図 2-1-5: マルコフ状態モデルを FRET 計測と比較した結果。(a) リガンドあり (5 mM L-Glu) 条件で比較した結果。(b) リガンドなし条件で比較した結果。

リガンドなしの条件において、CCA 状態が安定化することの起源を調べるため、HS-AFM データの動画セット (1,000 フレーム程の動画が 17 本から成る) を一つ一つ調べて outlier なデータが存在するか調べた。17 本の動画に対して個別にデータ同化を行い、マルコフ状態モデルを補正し、動画毎のモデルを可視化した (図 2-1-6)。その結果、およそ 2 つの動画だけで突出して CCA 状態が安定化することがわかった。更に、これらが outlier となる原因を探るために、動画の各フレームで分子のどの回転方向において尤度が高くなっているかを調べた。その結果、CCA 状態が安定化する動画ではヘテロ二量体が Cysteine-rich domain を「足」として基盤に直立している回転方向が有意に選ばれており、基盤との相互作用が偶々強いことで、二量体がオープン・クローズ変化しにくくなっていることが示唆された。以上のように、このデータ同化を通して、マルコフ状態モデルの補正を行うだけでなく、実験データの outlier とその原因まで探り、実験へのフィードバックができるようになったことは意義深い。

(a) 各AFM動画からマルコフ状態モデルを計算した結果



(b) 各動画から推定したタンパク質の配向

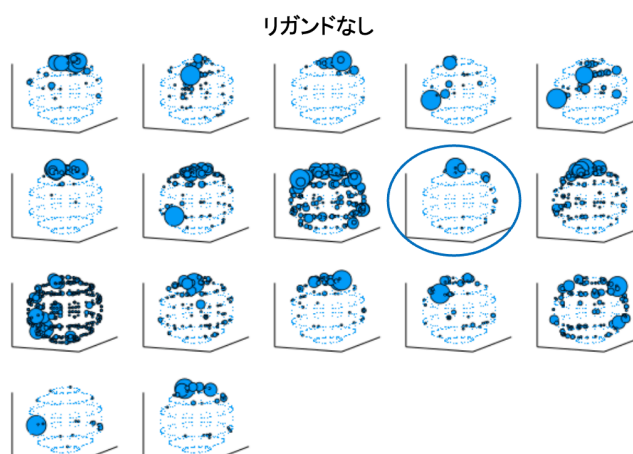


図 2-1-6: 高速原子間力顕微鏡の各動画でデータ同化を行った結果。(a) リガンドなし条件の各動画を同化することで得られたマルコフ状態モデル。(b) 同条件の各フレームで推定された分子の回転方向。

② 粗視化モデル CafeMol のパラメタ最適化 (京大・高田)

粗視化モデル CafeMol は、蛋白質・核酸の計算を行うために必要なアミノ酸・塩基・それらの相互作用などについてのパラメタが求められている。前年度のテスト計算の結果を踏まえて、分子・細胞レベルのシミュレーションに必要な細胞膜を構成する脂質分子および非特異的蛋白質・蛋白質相互作用の粗視化パラメタのさらなる改良を行い、より精度の高い分子動力学シミュレーションの実現に向けて研究を推進した。具体的には、以下の段落で記載の通り、液液相分離(LLPS)を定量的に再現すべく、アミノ酸間に働く非特異的な相互作用の精細チューニングを行うとともに、アミノ酸・ヌクレオチド間の非特異的な相互作用についても、標的となるクロマチン系の実験データと比較しながら精細チューニングを行った。

粗視化モデルにおけるアミノ酸間に働く非特異的な相互作用のチューニングは、過去数年の間に、LLPSシミュレーションに必須のものとして世界各地で急激に進められている。本研究ではまず、理研・杉田グループとの共同研究により、Best グループ、Mittal グループ、Lindorff-Larsen グループが提案した互いに類似している 3 つのモデルに対して、標的転写因子 Nanog 等の LLPS シミュレーションを実施し、実験結果と比較することでモデルの良否を検討した。つまり、転写因子の LLPS が実現できるか、また再現される臨界濃度は実験値と近いかどうか、を比較した。結果として、どのモデルも定性的には転写因子の LLPS を実現するが、その臨界濃度は実験値と大きく異なっていた。また、Best らのモデルは望ましくなくない振舞いを示したため採用しないことに決定した。次に、残る 2 つのモデルそれぞれのエネルギー強度を決めるパラメタ ϵ を変化させて、LLPS 臨界濃度を実験値と近い値にするようにチューニングを実施した。結果として、共通のパラメタ ϵ 値で 3 種類の転写因子の LLPS 臨界濃度を再現できたのは、Lindorff-Larsen グループが提案したモデルであった。また、先行研究で使われていた ϵ 値が 0.2 であるのに対して、チューニングした最適 ϵ 値は 0.22 であった。図 2-2-1 に、最適 ϵ 値を用いた転写因子 Nanog の LLPS 形成シミュレーションで得られた典型的な構造(A)、それに基づく臨界濃度評価(B)を示す。

また、Nanog は LLPS 臨界濃度より低い濃度でホモ 2 量体を形成することが実験的に示唆されている。その相互作用様式を明らかにするために、GENESIS および GROMACS を用いて Nanog の Trp 繰返し領域の全原子 MD シミュレーションを行った。Trp の相互作用状態の一例を図 2-2-1 (C)に示す。

転写因子Nanog 液液相分離シミュレーション

チューニングした $\epsilon = 0.22$ による計算結果（スナップショット）

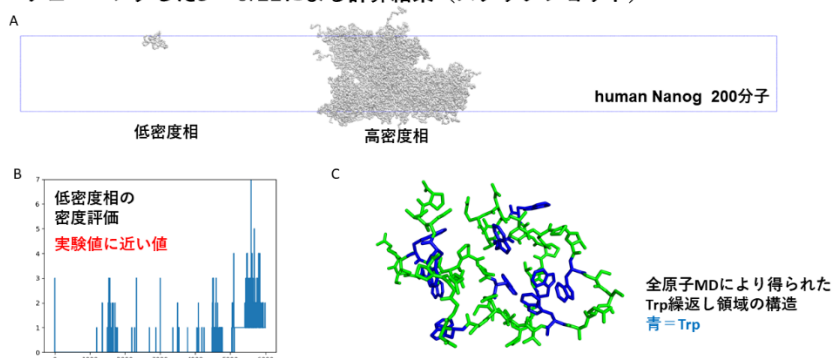


図2-2-1: 液液相分離シミュレーションによるパラメータチューニング。 A) チューニングされたパラメータによる転写因子NanogのLLPS構造。 B) LLPSを起こした状態における低密度相の密度からLLPS臨界濃度を見積り、実験値と比較したところ概ねよい一致を示した。 C) 全原子MDシミュレーションによるNanog Trp繰返し領域の構造。粗視化モデル作成の参考として利用。

本課題の(3)②の標的となるクロマチン系はヌクレオソームを1,000個程度含む巨大分子系である。これは、従来、粗視化モデルCafeMolで主に行われてきた1個から3個程度のヌクレオソームを含む系と全く異なる規模である。従来規模では問題にならなかった蛋白質・DNA間の相互作用、特に長距離で支配的な静電相互作用、の小さな誤差が巨大系では致命的な問題を生じる。そこで本課題では、カギを握ると考えられるヌクレオソーム-ヌクレオソーム間のスタッキング強度を実験値と比較し、それをもとに粗視化モデルの精細チューニングを行った。ヌクレオソーム-ヌクレオソーム間のスタッキング自由エネルギーについて、シミュレーションの結果と実験による測定値を比較した(図2-2-2)。その結果、蛋白質アミノ酸とDNAリン酸の引力相互作用を計算する際のリン酸の部分電荷の値を、従来用いていた $-1.0q$ から対イオン凝集を反映した値である $-0.6q$ に変更することで実験値とよく一致する結果となることが分かった。さらに検証実験として、 $-0.6q$ に変更した場合に、ヌクレオソーム12個を含むモデルクロマチンを用いて、その沈降係数の新しいパラメータによるシミュレーションからの推定値が、実験値とおおむね一致することを確認した(図2-2-3)。

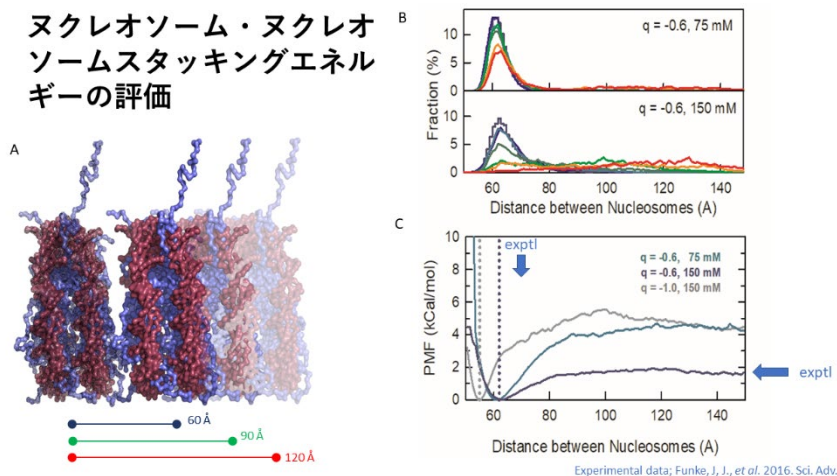


図 2-2-2: ヌクレオソーム-ヌクレオソーム間のスタッキング強度比較。(A) ヌクレオソーム間の距離を変化させたアンボレラサンプリングの模式図 (B) 電荷 $q=-0.6$ の場合の計算結果 (C) そこから求めた自由エネルギー曲線と実験値との比較。

12ヌクレオソーム繊維の沈降係数

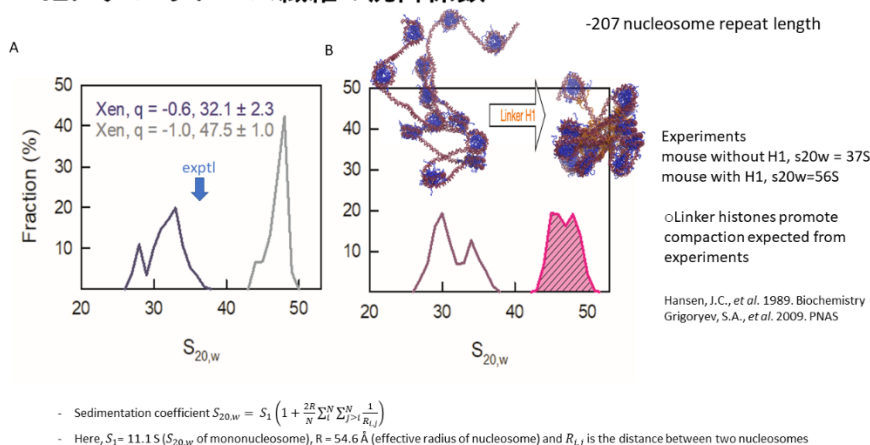


図 2-2-3: ヌクレオソーム 12 個を含むモデルクロマチンの沈降係数比較。(A) リン酸電荷 $q=-0.6$ が実験値と近い結果を与えるのに対して、 $q=-1$ は過度にコンパクトになる。(B) リンカーヒストン H1 が結合すると、H1 がいない時と比べてコンパクト化する。その変化は、実験値とおおむね合致した。

(3) 生体分子系のモデリングと「富岳」を用いたシミュレーション

① 細胞内分子混雑と液液相分離 (理研・杉田)

これまでに我々が行ってきた細胞内分子混雑環境の計算において、従来の分子力場を用いた場合には、蛋白質・水間に働く相互作用に比べて蛋白質間の相互作用が強いため、複数の蛋白質が凝集し、拡散定数が低下するなどの問題が生じていることがわかってきた。この問題を解決するため、CHARMM36m力場に対してNBFIXと呼ばれるパラメタを調整することで、蛋白質・水および蛋白質・蛋白質の相互作用のバランスをとることができる。NBFIXは天然変性蛋白質の構造が、全原子MDシミュレーションでは実験と比較してコンパクトな構造になってしまう問題を解決するためにも調整されてきた。ここで問題となるのは、異なる分子力場によって最適なバランスを示すNBFIXパラメタが違うことである。我々が過去に行った蛋白

質Villin headpiece subdomain (Villin)の混雑系に関するMDシミュレーションでは、CHARMM36力場（水はTIP3Pモデル）に対して、NBFIX=1.09が最適な値であることがわかった。AMBER力場を用いたMDシミュレーションの結果を改善するNBFIXパラメタとして1.1が提案されており、それと近い値であった。しかし、同じ値を最新のCHARMM力場（CHARMM36m）に適用したところ、蛋白質が変性する（すなわち、蛋白質・水間の相互作用が強すぎる）という傾向があることがわかった。そこで、最新のCHARMM36mにおける蛋白質・水と蛋白質・蛋白質間の分子間相互作用のバランスを最適化するNBFIXパラメタを検討した。

初めに温度レプリカ交換法を用いて、小ペプチド(AAQAA)₃の構造安定性について調べた（図3-1-1）。その結果、NBFIX=1.04において構造安定性が急激に低下することがわかった。このことからNBFIX=1.03が、水中での蛋白質の立体構造安定性と凝集を回避する最適なパラメタであることが示唆される。Villinを用いた混雑系をモデリングし、長時間のMDシミュレーションを実行し、混雑系での並進と回転拡散の挙動を調べた（図3-1-2）。その結果、構造については(AAQAA)₃の結果と矛盾せず、NBFIX=1.03においても安定に保たれていた。その上で、Villinの並進と拡散を計算したところ、NBFIX=1.03において有意な増加が確認された。以上よりNBFIX=1.03が上記二つの観点からみて混雑系における適切なパラメタ設定であると結論付けた。以上を考慮すると、CHARMM36mに対しては、NBFIX=1.03が最適なパラメタであると示唆されるため、さらに多くの分子系に関して調査を継続する。

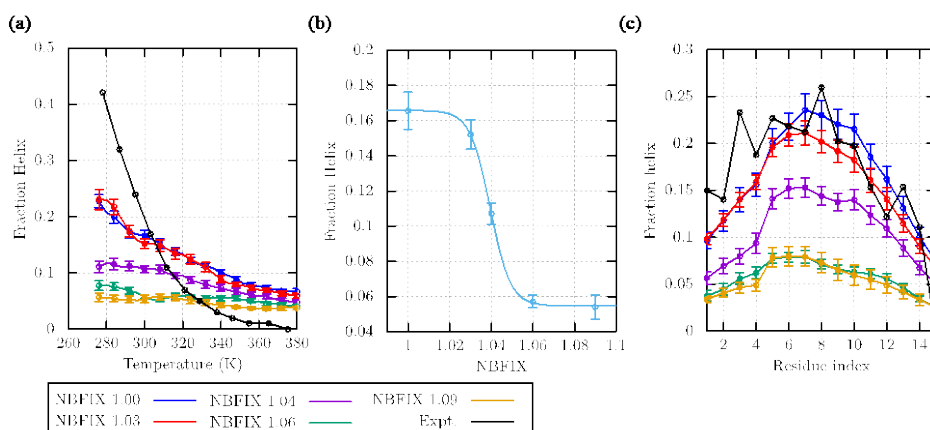


図3-1-1: 異なるNBFIXパラメタを用いた(AAQAA)₃水溶液中の温度レプリカ交換シミュレーション。

(a) ヘリックス形成率の温度依存性。(b) 300Kにおけるヘリックス形成率のNBFIX依存性。(c) 300Kにおける残基毎のヘリックス形成率。

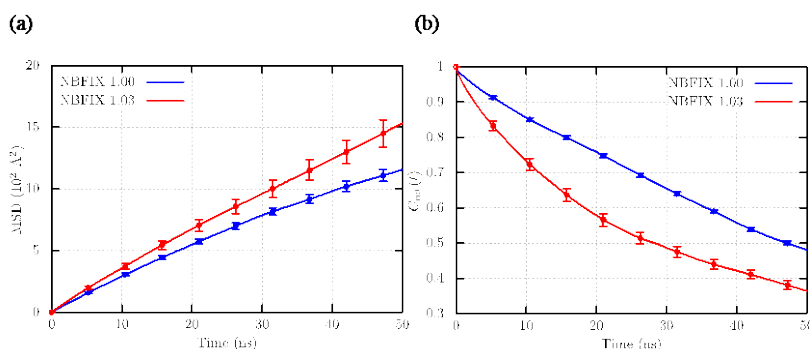


図3-1-2: Villin混雑系における異なるNBFIXパラメタを用いたMDシミュレーション。(a) 平均二乗変位。(b) 回転相関関数。

細胞内には、蛋白質や核酸、その複合体などの生体高分子に加え、膨大な数の代謝物やイオンが共存している。中でも ATP などのヌクレオチド三リン酸が分子量の観点から大部分を占めている。近年、ATP が様々な蛋白質に非特異的に結合し、凝集や液-液相分離構造 (LLPS) の形成を制御している可能性が示された (Patel and Malinovska, *Science* 2017)。これは蛋白質水溶液の実験によって示されているが、分子レベルの作用機構については不明である。一方で我々は、過去に実行されたバクテリア (マイコプラズマジェニタリウム: MG) の細胞質の大規模分子動力学シミュレーション (総原子数約 1,000 万~1 億個: 蛋白質総数約 200~1,500 個) の結果から、ATP をはじめとするヌクレオチド三リン酸類が様々な蛋白質表面に非特異的に結合している事を明らかにした (Yu et al. *eLife* 2016)。

このような背景の下、代謝物の影響をより詳細に評価するために、我々は新たに小型の細胞質モデル (原子数 100 万個: 蛋白質総数 32 個) を作成し、代謝物 (イオンを含む) を含む系 (MG_{cyt}) (図 3-1-3 左) と、代謝物は全て取り除き、溶媒として水分子と系を中和するための最小限のイオンのみを含む系 (MG_{wat}) それぞれについて約 1.3 μs の MD シミュレーションを実行した。その結果、MG_{wat} では時間経過とともに蛋白質の凝集が進行してゆくのにに対し、MG_{cyt} では凝集が大きく抑制されている事が示された。凝集の度合いは、異なる蛋白質間で一定の距離以内にある原子ペアの数 (コンタクト数) を計上することで評価した (図 3-1-3 右)。これについては実験結果 (Patel and Malinovska, *Science* 2017) と定性的な一致を得られたが、なぜ、代謝物、特にヌクレオチド三リン酸の存在が蛋白質凝集を抑制しているかについての分子論的詳細については現在解析中である。

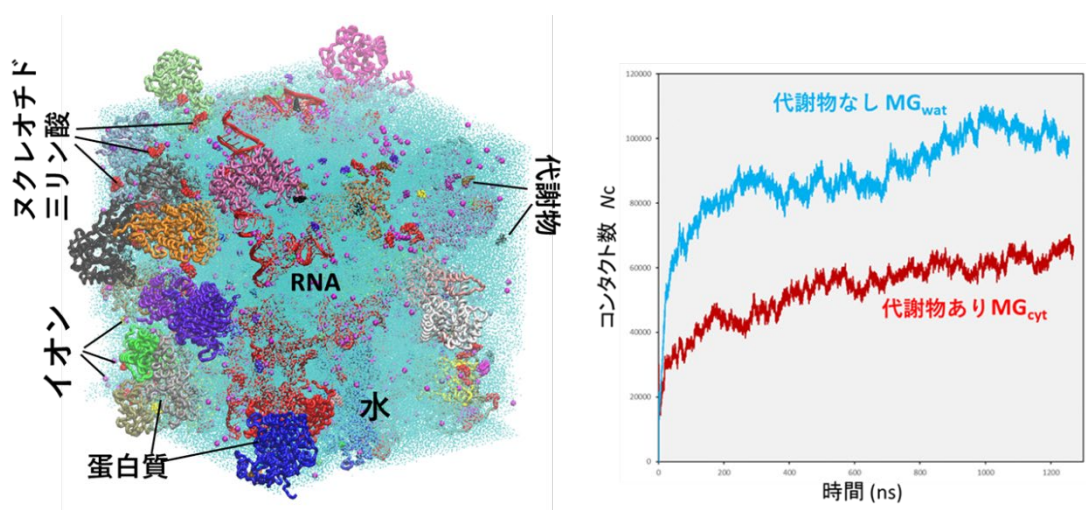


図 3-1-3: (左) 細胞質モデル MG_{cyt}. (右) コンタクト数の時間変化

連携研究者である東工大・北尾は、細菌べん毛の研究をおこなった。細菌べん毛の回転モーターは細胞内外のプロトンまたはナトリウムイオン勾配を用いて駆動され、巨大な回転子の周りに 7~8 個程度の固定子が配置された複雑な超分子構造体である。細菌べん毛の回転はシグナル蛋白質によって反転可能であるがその分子機構はまだ明らかになっていない。固定子 MotAB には MotA の 5 分子が MotB2 分子を取り巻くようにして固定子の 1 ユニットが形成されることが明らかになってきた。共同研究者が低温電子顕微鏡解析で得た未発表の *Aquifex aeolicus* (*A. aeolicus*) の MotA5 量体の立体構造は、完全な 5 回対

称の 5 量体を示している (図 3-1-4 左の赤色部)。一方、最近決定された他の種の MotA/B 複合体の結晶構造では、5 量体の対称性に歪みが生じている。この非対称性に関連して、MotA と MotB の間にイオンを透過するチャンネルが形成され、イオン透過が固定子に何らかの動きを生じさせ、それがさらに MotA の回転などの回転子の動きを生じさせるのではないかという仮説が提唱されている。そこで前年度より、MotA/B 複合体における MotA 非対称化のメカニズムを MD シミュレーションで研究してきた。具体的には、*A. aeolicus* の MotA の 5 量体モデルに、*Campylobacter jejuni* (*C. jejuni*, PDB ID: 6YKM) の立体構造を参照して MotB の 2 量体をモデリングで挿入した立体構造モデルを構築し、MD シミュレーションを行ってきた。前年度、1.5 μs 程度の計算をすでに行ったが、実験によって前年度用いたものよりもより精密な立体構造モデルが得られたので、改めて 1 μs のシミュレーションを 10 回行い、*A. aeolicus* の MotA 5 量体の MotB 2 量体存在下での対称性の変化を検証した。このシステムは約 22 万原子からなる。

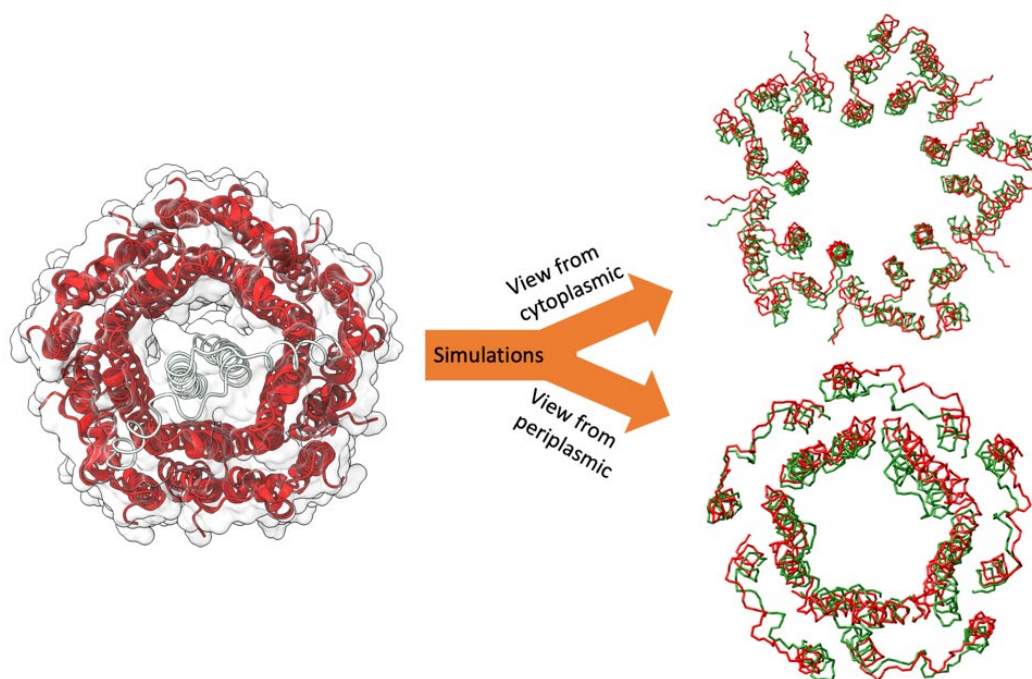


図 3-1-4: MotB の 2 量体が挿入されることで *A. aeolicus* MotA の 5 量体に生じる歪み。左図は MotA/MotB 複合体の MD シミュレーションを開始する前の初期立体構造モデル。左側で赤の NewCartoon モデルで示した部分は MotA を、白の NewCartoon モデルで示した部分は MotB を表す。図右の緑で示したのは緩和後の MotA の構造で、比較のためにもとの構造を赤で示してある。MotB の構造は見やすさのために省略している。

図 3-1-4 右に示すように、MotB の二量体が挿入されると、細胞質側 (右上) とペリプラズム側 (右下) の両方で MotA の対称性が崩れることがわかった。ペリプラズム領域のヘリックスは、MotA の 5 量体単独の場合よりもコアに対して内側に向かっていることが見て取れる。対称性の変化をより定量的に調べるために、共同研究者が注目している 4 つの残基、Asn31 (ペリプラズム領域の外側残基)、Pro176 (ペリプラズム領域の内側残基)、Glu96 (細胞質領域の外側残基)、Asn123 (細胞質領域の内側残基) を選

び、隣接するサブユニットの残基間距離の時間変化を調べた (図 3-1-5)。Asn31 はどの組の MotA サブユニットでもほぼ同じ距離を示しているが、残りの残基は大きくずれており、特に細胞質領域の Glu96 と Asn123 は対称性が崩れており、MotA サブユニット D-E と E-A の間の距離は、他のモノマーよりも著しく大きくなっている。その結果、MotB の存在によってサブユニット E は、5 量体のコアから離れていることが分かった。つまり、MotA のサブユニット E は他のサブユニットと比較して最も結合が緩く、MotB との間に最もスペースがあることが分かった。この状態から、MotA が MotB の周りで回転する可能性があるかどうか Metadynamics 等のシミュレーションで検証を続けている。



図 3-1-5: MotA サブユニット間の距離の時間変化を調べたもの。左のカラムから隣接するサブユニットの Asn31、Pro176、Glu96、Asn123 間の距離を示している。また上から、サブユニット A と B、B と C、C と D、D と E、E と A のペアについてそれぞれ示している。

② 遺伝子転写機構の解明 (京大・高田)

染色体の高スループット実験から得られる Hi-C/Micro-C/MNase-seq 等の実験情報を用いて構築した三次元クロマチン構造粗視化モデルに、ChIP-seq 等の実験情報を用いて結合蛋白質 (リンカーヒストン、転写因子等) を付加し、細胞核内状況に近いモデルを構築した。このモデルは低解像度のポリマーモデルであるため、粗視化モデル CafeMol を用いて詳細な動態シミュレーションを実施した。クロマチンと転写因子等を含む遺伝子座における染色体高次構造に関する立体構造モデルを構築することに成功し、「富岳」での MD シミュレーションを行った。

多細胞生物は多様な細胞種をもち、細胞種ごとの多様性は細胞種ごとに異なる遺伝子発現に起因する。これが高次生命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。異常クロマチン3次元構造は発現異常を引き起こし致死あるいは疾患の原因となる。ここでは哺乳類の胚性幹(ES)細胞に着目し、発生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明らかにする。哺乳類ES細胞のコア遺伝子ネットワークは3つの遺伝子、*Oct4*, *Sox2*, *Nanog*で構成される。*Oct4*, *Sox2*は、いわゆる山中因子に含まれる。3つの遺伝子はお互いに協同的に転写を活性化しあう関係にあり、それがES細胞の多能性と自己複製能の維持に必須である。

そのなかで、複雑で興味深い転写制御が実現されていると考えられているマウス*Nanog*遺伝子座を本研究の標的とする。*Nanog*遺伝子座の制御領域には遺伝子本体のすぐ上流にあるプロモータと、そこから45kb上流、5kb上流および60kb下流に合計3つのスーパーエンハンサ(SE)が存在する。プロモータおよび3つのSEには、*Oct4*, *Sox2*, および*Nanog*蛋白質の結合モチーフが複数存在し、これら蛋白質の結合が*Nanog*遺伝子の転写を活性化する。プロモータと3つのSEが物理的に近距離に存在する可能性をもつことは、近年のHi-C法、micro-C法などの実験から明らかになっている。また、近距離になった状態で、コアクティベータと呼ばれるMediatorやBrd4などが、プロモータとSEとの相互作用を介助することが想定されている。RNA合成酵素および基本転写因子は、プロモータ近傍に存在すると考えられる。しかし、これら、転写制御因子、コアクティベータ、転写マシナリーなどの分子群が実際にどのような3次元構造で、どのような相互作用によって、転写制御を実現するのか、ほとんど分かっていないのが現状である。本研究では、メソスコピックモデル、粗視化モデル、全原子モデルの3階層にわたるマルチスケールのモデリング、MDシミュレーションによってこの分子機構を明らかにする。

前年度の本課題研究において、*Nanog*遺伝子座70kb領域について、micro-C法に基づくメソスコピック構造モデルを得たが、そこでは構造多様性を考慮しなかった。実際には、染色体構造は細胞ごとに大きく異なり、そのアンサンブルは非常にヘテロであることが知られており、micro-C法等で得られるデータはそのアンサンブル平均である。本年度、その多様性を顕に考慮するために、あらたにメタ推定法をクロマチン構造モデリングに適用した。メタ推定法では、ヘテロなクロマチン構造をポリマーモデルのn個のレプリカとして表現する。本研究では、前年度の70kbから拡張して*Nanog*遺伝子座200kb領域について、1kbを1粒子で表現しレプリカ数n=128を用いて、多様なクロマチン構造をモデリングすることに成功した(図3-2-1)。

次に、メソスコピックポリマーモデルから残基粒度の粗視化モデルCafeMolへの変換、いわゆるバックマッピング法について、前年度開発したプロトコルの改良を行った(図3-2-2)。まず、MNase-seq、化学マッピングによるヌクレオソームのゲノム配列上の位置情報、ChIP-seq、GRO-seq等によるDNA結合蛋白質のゲノム配列上の結合位置情報をもとに、ゲノム配列上に約1,000個のヌクレオソーム、1個のRNAポリメラーゼMediator複合体、転写因子*Oct4*, *Sox2*, *Nanog*が結合する位置を決定した。超解像イメージングデータを参考に、転写因子*Oct4*, *Sox2*, *Nanog*を各30分子含めることとした。次に、これら蛋白質が“装飾”した状態のDNAについて、前年度と類似の方法で3次元構造を構築した。すなわち、一方の端から順次、粗視化モデルのヌクレオソームサンプルとリンカーDNAサンプルを交互に伸長させてクロマチン構造を構築していく。ここで、あらかじめ各ヌクレオソームあるいはリンカーDNA断片にRNAポリメラーゼ等の蛋白質を結合させておく。伸長させて構築した3次元上に配置した各ヌクレオソームの位置と、メソスコピッ

クモデル3次元構造の対応する位置との一致度を表すスコア関数を設定する。多様なヌクレオソームおよびリンカーDNAの構造要素を組み合わせて、このスコア関数を最適化するものを探索する。プロトコル中のパラメタを調整し、次の粗視化MDシミュレーションが問題なく行えるバックマッピング法を確立することができた。得られたモデルに後で、コアクティベータであるBrd4を加えて、粒子間の軽微な重なりを除去して、粗視化モデルシミュレーションの初期構造とした。

メタ推定法によるmicro-Cデータからのポリマーモデリング

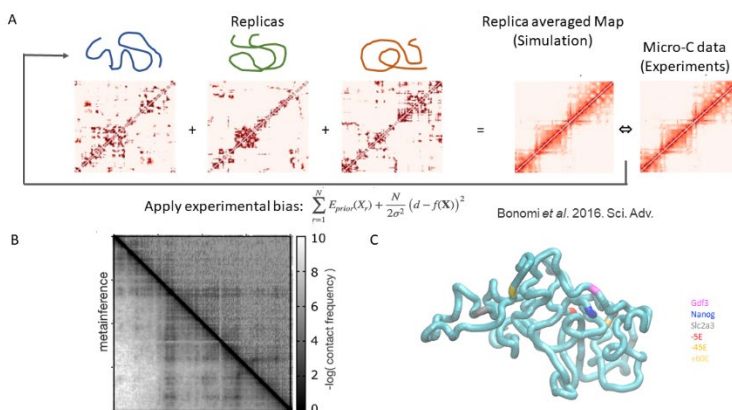
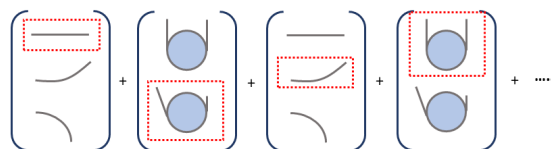


図 3-2-1: メタ推定法による micro-C データからのメゾスコピック・ポリマーモデルシミュレーション。
 (A) メタ推定法の概念図。N 個のレプリカを用意し、その平均が実験データと一致するようにモデリングする。(B) *Nanog* 遺伝子座 200kb 領域の micro-C データ (右上三角) とメタ推定法のレプリカ平均 (左下三角) は、良い精度で一致している。(C) メタ推定法で得られたメゾスコピック・ポリマー構造。

バックマッピング法

- Build and grow a CG model end-to-end
- Sample from a conformational ensemble of Nucleosome and Linker DNA



- Cost function: Minimize clashes and distance from mesoscale model
- Appropriate ensembles of PIC, TFs are included to the process

図 3-2-2: メゾスコピックポリマーモデルから、残基粒度の粗視化モデルへのバックマッピング法の概念図。

以上により得られた *Nanog* 遺伝子座 200kb 領域の残基粒度の粗視化モデルを用いて、「富岳」で GENESIS を用いて MD シミュレーションを実施した。これまでに、「富岳」を用いて 10^7 MD ステップのシミュレーションに成功した (図 3-2-3)。まず、前年度の初期テスト計算と同様に、得られたクロマチン構造は一樣ではなく、配列に沿って連続した数個~10個程度のヌクレオソームがクラッチを形成する傾向が見て取れた。また、*Nanog* 遺伝子プロモータに結合している RNA ポリメラーゼ Mediator 複合体は、-5SE の向きに伸びているが、-5SE 領域と物理的に相互作用するほどには接近していない。系内に各 30 分子含まれる転写因

子Oct4, Sox2, Nanogは、全体のなかでスーパーエンハンサ(SE)周辺に比較的高密度で存在する。またこれらの約30%の分子はクロマチンから解離し3次元的な拡散運動をしている。一方で、現段階では、いまの転写因子数では、クロマチン内にLLPSが形成される様子は観察されていない。

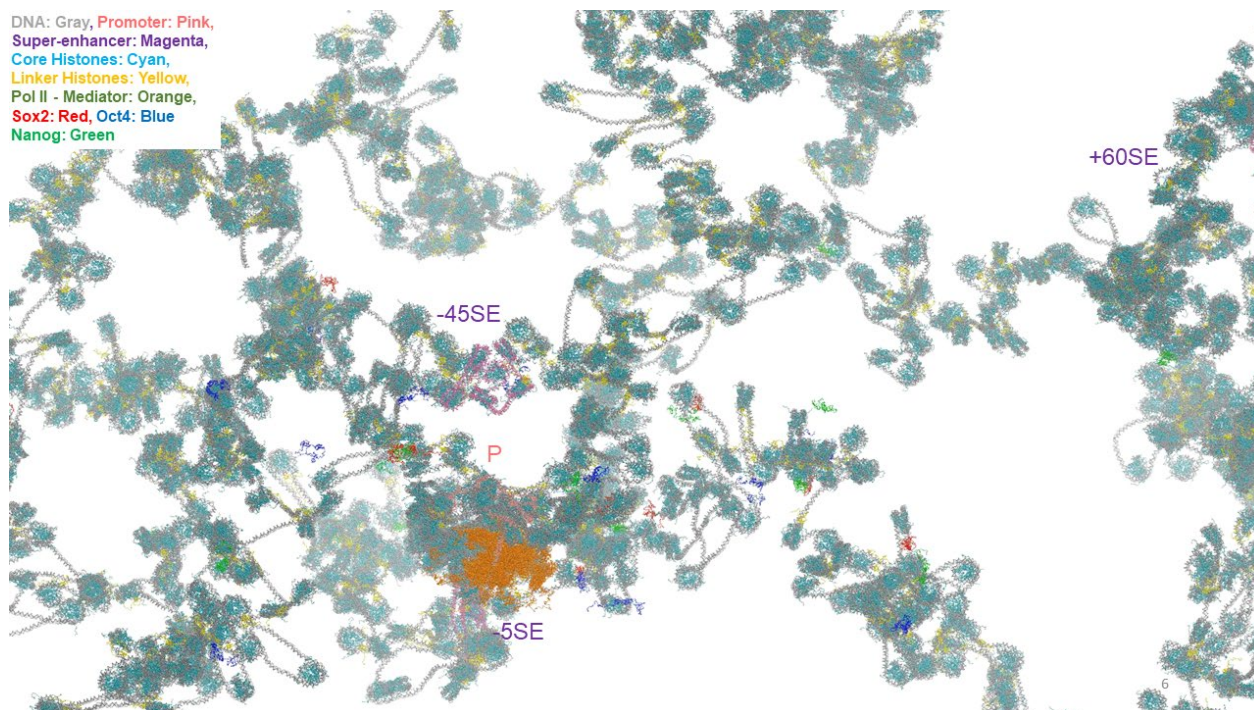


図3-2-3: *Nanog*遺伝子座200kb領域の粗視化MDシミュレーション構造。約1,000個のヌクレオソーム(シアン)、リンカーヒストン(黄)、1個のRNAポリメラーゼMediator複合体(オレンジ)、30個のOct4(青)、Sox2(赤)、Nanog(緑)を含む。P: *Nanog*プロモータ、SE: スーパーエンハンサ。

これらは注目すべき結果ではあるが、信頼できる結果になるためにはさらなるモデルの精密化が必要である。具体的に、標的遺伝子座200kb領域中にRNAポリメラーゼ、転写因子Oct4, Sox2, Nanog、コアアクティベータBrd4を何分子ずつ含めるのが適切か、またそのDNA上の配置はどうするのがよいか、再検討が必要である。

③ 核内蛋白質・核酸複合体の動的モデリング(量研機構・河野)

遺伝子発現の制御機構を明らかにするためには、細胞核内のゲノム収納状態を理解する必要がある。真核生物のゲノムDNAは、ヌクレオソームというDNAと蛋白質からなる構造体を形成し、直径数マイクロメートル程度の細胞核内に収納されている。ヌクレオソームは、ヒストン蛋白質8量体の周りに約150塩基対のDNAが2回巻き付いた構造体であり、細胞核内にはこのような構造体が数珠つなぎになって、コンパクトに折れたたまっている。DNAに刻まれた遺伝子が働くためには、個々のヌクレオソームに巻き付いたDNAが他の分子に読まれなければならないが、ヌクレオソーム構造のままの状態では大部分のDNAには分子が直接アクセスすることができない。本年度は、1) 遺伝子活性が不活発な領域のヌクレオソーム付近に多く存在することが知られているHP1蛋白質の構造多様性とヌクレオソームの相互作用様式、2) ヌクレオソームに結合して転写を活性化するGATA3蛋白質について調べた。

HP1 蛋白質の構造多様性とヌクレオソーム相互作用様式

遺伝子発現が不活発な領域では、ヌクレオソームどうしがコンパクトな構造を取っている。その領域に多く存在する蛋白質の一つが HP1 蛋白質である。HP1 蛋白質は、ホモダイマー蛋白質で、ヌクレオソームを構成するヒストン H3 の N 末領域にある 9 番目のリジンがトリメチル化されると、それをターゲットにして結合する。HP1 は、**図 3-3-1** に示すように、クロモドメイン (CD) とクロモシャドウドメイン (CSD) の 2 つの安定なドメインを持つが、それ以外の N 末領域 (NTE)、ヒンジ領域 (HR)、C 末領域 (CTE) は特定の安定な構造を持たない天然変性領域であり、その領域がどのように HP1 ダイマー内で相互作用しているか、また、ヌクレオソームとどのように相互作用するかは、構造学的な見地からはほとんど分かっていない。

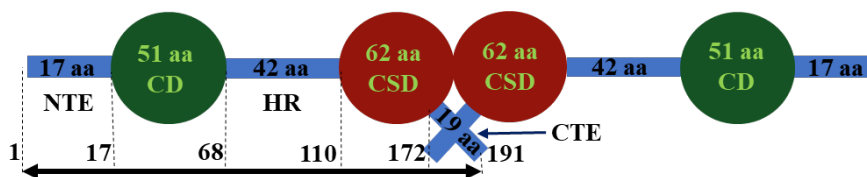


図3-3-1: HP1のトポロジー。ダイマーの1次構造。N末領域 (NTE)、クロモドメイン (CD)、ヒンジ領域 (HR)、クロモシャドードメイン (CSD)。

そこで、まず、分子モデリングにより、全原子モデルを構築した。大きな構造変化が期待されるため、高田チームで開発している粗視化分子シミュレーションソフトウェア CafeMolを用いて、HP1ダイマーの構造多様性を調べた。**図3-3-2** に示す初期構造からシミュレーションを繰り返して、非常にコンパクトな構造から大きく広がった多様な構造を得ることができた。得られた構造群に対して分子内の残基間のコンタクトに基づいてクラスター解析を行ったところ、20個のクラスターに分類された。代表的な構造を**図3-3-3**に示す。この図に示された構造から分かるように、HP1ダイマーのコンパクトな構造はHP1モノマー自体がコンパクトになることで実現され、2つのHP1どうしのクロスコンタクトはほとんど見られなかった。このことは、HP1ダイマーが広い空間を探索し、ターゲットとなる結合の相手、つまり、ヒストンH3のN末領域を見つけることが示唆される。

次に、HP1とヌクレオソームの相互作用様式について、全原子分子動力学計算によって調べた。ヒストンH3のN末領域とHP1のクロモドメインが結合した結晶構造とヌクレオソームの結晶構造をもとに、分子モデリングにより、HP1-ヌクレオソーム複合体の原子構造をモデリングした。そして、HP1をヌクレオソームから40 Å離れた位置に置いて、シミュレーションを行った (**図3-3-4**)。シミュレーションは、HP1のN末領域、クロモドメイン、ヒンジ領域を持つ分子とHP1から天然変性領域 (N末領域とヒンジ領域を除く、つまり、クロモドメイン領域のみ) を削除した分子に対して行った。これにより、天然変性領域の働きを調べることができると期待される。

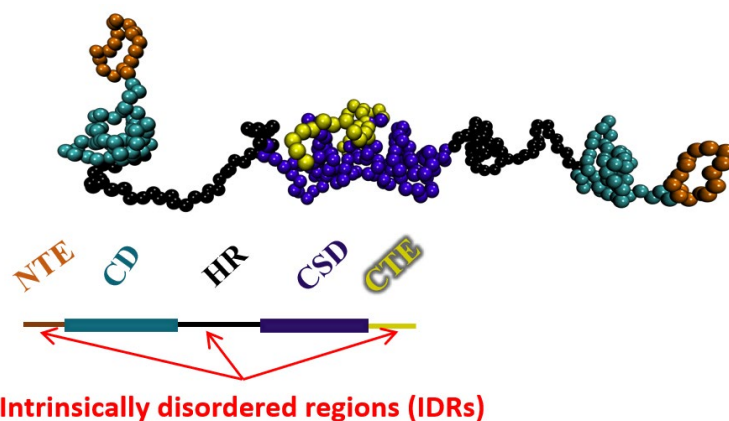


図3-3-2: HP1ダイマーの初期構造。ドメインごとに色分けをして表示。

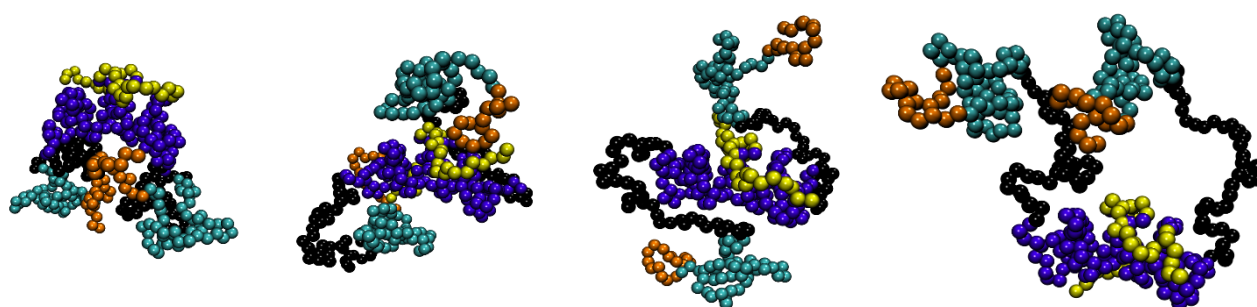


図3-3-3: HP1ダイマーの多様な構造の代表例。コンパクトな構造（左）から広がった構造（右）。

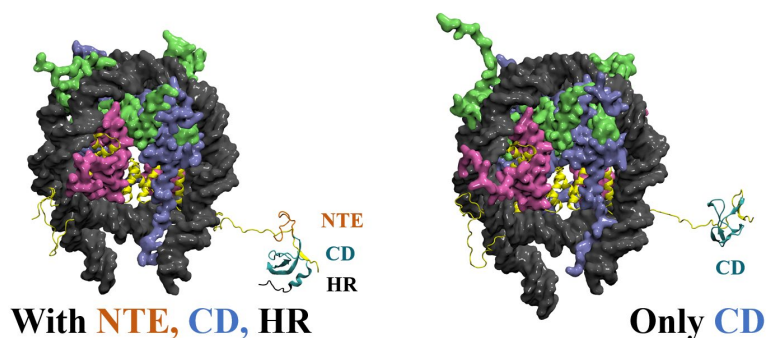


図3-3-4: HP1の初期配置。N末領域（NTR、オレンジ）、クロモドメイン（CD、シアン）、ヒンジ領域（HR、黒）を持つHP1（左）、クロモドメインのみのHP1（右）。

全原子MDシミュレーションの結果を、図3-3-5に示す。N末領域とヒンジ領域を持つHP1は、それらがDNAと結合し、その部分が足場となってクロモドメインはDNAに沿って広い範囲に分布することが分かった。一方、クロモドメインのみのHP1は、ヌクレオソームのDNAの特定の主溝（赤○のところ）に結合すると、それ以上の動きは見られなかった。これらの結果は、N末領域とヒンジ領域がHP1の相互作用の多様性を生み出す要因になっていることを示した。

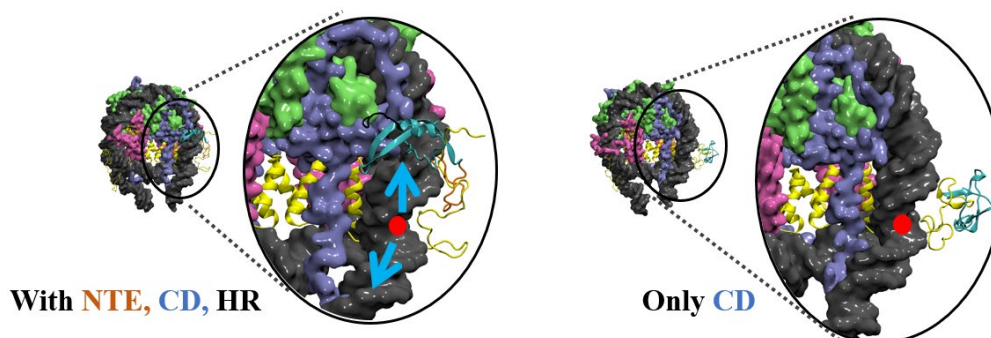


図3-3-5: HP1の初期配置。N末領域 (NTR、オレンジ)、クロモドメイン (CD、シアン)、ヒンジ領域 (HR、黒) を持つHP1 (左)、クロモドメインのみのHP1 (右)。

GATA3転写因子とヌクレオソーム複合体

GATA3 転写因子は、特定の DNA 配列に結合して遺伝子の発現を制御する転写因子である。通常の転写因子は、ヌクレオソーム構造状態の DNA には結合できないが、この転写因子はヌクレオソーム構造状態の DNA に結合することが知られている。しかし、結合によって、ヌクレオソームの構造、運動、安定性といった物性がどのように変化するかはよくわかっていない。そこで、これらを調べるために、GATA3 とヌクレオソーム複合体の全原子 MD シミュレーションを行っている。

前年度は、既知の GATA3 とオリゴ DNA 複合体とヌクレオソーム単体それぞれの X線結晶構造と、GATA3 とヌクレオソーム複合体の単粒子クライオ電子顕微鏡解析によって得られた電子密度とを用いて、原子モデルを構築することに成功した。そして、「富岳」にて全原子 MD シミュレーションを実施することで、GATA3 がヒストンテールやヒストン DNA と安定した相互作用をすることが分かった。今年度は、理研杉田グループと協力して、蛋白質・核酸等の効率的な構造探索を GENESIS に実装し、前年度に構築した GATA3-ヌクレオソーム複合体の原子モデルについて MD シミュレーションを実行し、X 線溶液散乱データとの比較することで溶液での構造を明らかにするとともに、GATA3 が結合することによってヌクレオソームに起こる運動や構造の変化を調べた。

前年度に構築したヌクレオソーム複合体の原子モデルに対して、研究代表者の理研杉田グループと協力して、GENESIS 2.0beta に新しく実装した蛋白質・核酸等の効率的な構造探索方法を用いた MD シミュレーションを実施した。実装した探索法は Adaptively Biased Molecular Dynamics (ABMD) 法と呼ばれるメタダイナミクス法のひとつで、シミュレーション中の反応座標値に応じてガウス型の局所的エネルギーを累積して生成されるバイアスポテンシャルを、通常の原子間相互作用ポテンシャルに追加する。これにより、広い範囲のサンプリングを効率的に実現することができる。

GATA3 には DNA 配列を認識する 2 つのジンクフィンガー (N 末ジンクフィンガーと C 末ジンクフィンガー) がある。前年度に構築した、電子顕微鏡像に基づいた GATA3 とヌクレオソームの複合体では、N 末ジンクフィンガーがヌクレオソーム DNA のスーパーヘリックスロケーション (SHL) 6.5 (ヒストン H3 α N および H3N 末テールの DNA 結合箇所付近) に結合し、C 末ジンクフィンガーがヌクレオソーム DNA の SHL 5.5 (ヒストン H2A の L2 ループの DNA 結合箇所付近) に結合した構造を得た (図 3-3-6 左)。

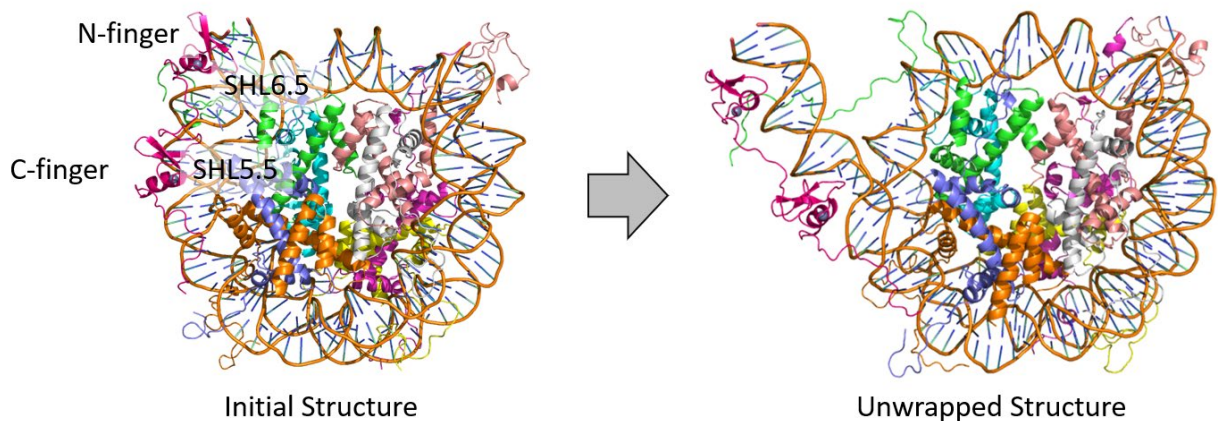


図 3-3-6: モデリングした GATA3 とヌクレオソーム複合体 (左)。ABMD で得た DNA を解離した構造 (右)。

ところが、X 線溶液散乱の散乱曲線とモデリングした複合体構造から計算した散乱曲線は、広角側で一致しなかった。そこで、ABMD 法アルゴリズムを用いた全原子 MD シミュレーション (系のサイズは $164\text{\AA} \times 164\text{\AA} \times 164\text{\AA}$ 、原子数は約 38 万原子) により、DNA が一部解離した様々な構造をサンプリングした (図 3-3-6 右)。しかし、DNA の解離の程度が異なる様々な構造から計算される散乱曲線は、どれも溶液散乱データに一致することはなかった。つまり、溶液中では DNA が解離した構造ではない、ということが分かった。

次に、GATA3 の 2 つのフィンガーのうち、片方が DNA から解離した構造を考えた。電頭による単粒子構造解析から、GATA3 に対する電子密度はヌクレオソームの電子密度に比べて小さいことから、このような発想に至った。フィンガーが解離した構造をサンプリングするため、GENESIS 2.0beta に実装した ABMD 法を用いた。ABMD 法で用いる反応座標は、ヒストンコアの重心 2 つのジンクフィンガーの重心との間の距離とした。バイアスポテンシャルは、シミュレーション時間内 (数十から 200ns 程度) で反応座標上を約 55\AA の初期状態から 100\AA までのジンクフィンガー離脱状態が生成されるように設定した。

18 通りの初期構造に対して、ABMD シミュレーションを実施した結果、N 末ジンクフィンガーと C 末ジンクフィンガーの多様なヌクレオソーム DNA 結合状態を得ることに成功した。N 末ジンクフィンガーが離脱する (C 末ジンクフィンガーは DNA と結合する) パターンが 5 通り、C 末ジンクフィンガーが離脱する (N 末ジンクフィンガーは DNA と結合する) パターンが 9 通り、2 つのジンクフィンガーが DNA と結合したまま DNA がヒストンコアから少しだけ解離するが大きな構造変化は見られなかったパターンが 4 通りであった。このことは、C 末ジンクフィンガーと DNA の結合、N 末ジンクフィンガーと DNA の結合、DNA とヒストンコアの結合、の順番に結合が強くなることを示唆している。N 末ジンクフィンガーは DNA のみならず H3 ヒストンの N 末テールとも相互作用していることが多いことが観測された。このことから、ヒストンテールとも安定した相互作用をもつ N 末ジンクフィンガーのほうが、ヒストンテールと相互作用があまりない C 末ジンクフィンガーよりも、ヌクレオソームに対する結合が強いと考えられる。

更に、ABMD シミュレーションで得られた GATA3 とヌクレオソームの様々な結合状態について計算した X 線小角散乱データを実際の実験データと比較したところ、C 末または N 末ジンクフィンガーが離脱した構造の中に、実験データとよく合致する構造が見つかった (図 3-3-7)。通常の ABMD シミュレ

ーションではバイアスポテンシャルをゆっくりと増大させ、系ができるだけ平衡状態に近いように構造変化を促す。それに対し、今回はバイアスポテンシャルを早く増大させることで、GATA3 の離脱反応を促進させる工夫を行った。

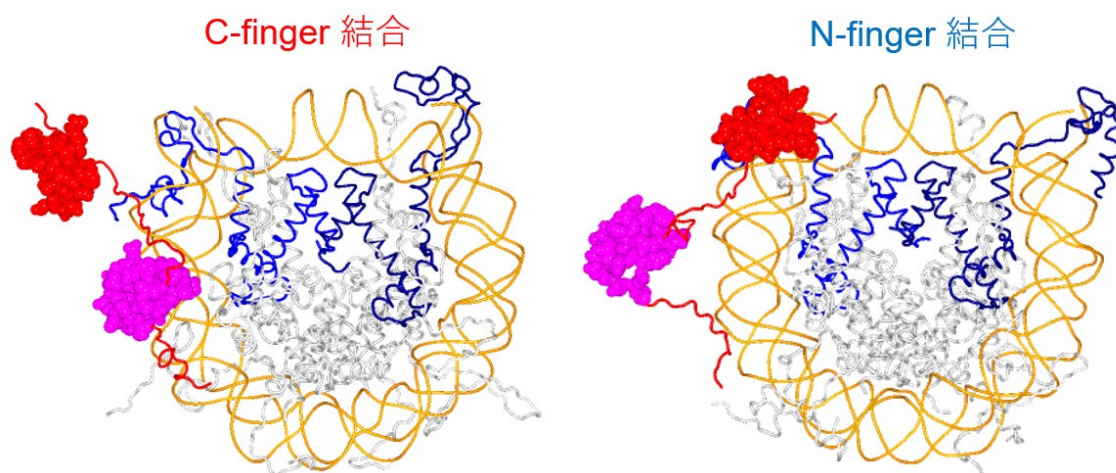


図 3-3-7: 全原子 MD シミュレーションにより得られた X 線小角散乱データに合う GATA3 とヌクレオソームの結合状態。 C 末ジンクフィンガー (マゼンタ) が結合し、N 末ジンクフィンガー (赤) が解離した構造 (左)。2 つあるヒストン H3 をそれぞれ青、紺色で示す。N 末ジンクフィンガーが結合し C 末ジンクフィンガーが解離した構造 (右)。

次に、X 線小角散乱データをよく再現する、どちらかのフィンガーが解離した構造について、通常の MD を実施し、構造や動きを調べた。フィンガーが離脱した状態は、GATA3 の 2 つのフィンガーがヌクレオソームに結合した複合体の計算に用いた $164 \times 164 \times 164 (\text{\AA}^3)$ サイズの系では正確に分子の動きを扱えない可能性がある。そこで、GATA3 が外れた状態の構造に対しては、系のサイズを大きくし (系のサイズは $190 \times 190 \times 190 (\text{\AA}^3)$ 、原子数は約 68 万原子)、通常的全原子 MD シミュレーションを $500\text{ns} \times 20$ 本実施した。

この結果、N 末ジンクフィンガーのみが結合している状態と C 末ジンクフィンガーのみが結合している状態それぞれで、X 線小角散乱データに合う構造が安定に存在することが分かった。N 末ジンクフィンガーのみが DNA に結合している状態 (図 3-3-7 右) では、GATA3 が結合している DNA 末端とは反対側の DNA 末端が数塩基解離している構造が多く観測された。一方、C 末ジンクフィンガーのみが結合している状態では、GATA3 が結合している DNA 末端とは同じ側の DNA 末端がやや解離している構造が多く観測された (図 3-3-7 左)。つまり、X 線小角散乱データを再現する構造は、DNA の片側が数塩基解離した構造であることが分かった。また、ヌクレオソーム単体の MD を行い、同様に X 線溶液散乱データとの比較を行った。この場合も、DNA の片側が解離した構造のみが実験データとよく合致することが分かった。以上から、溶液中ではヌクレオソームの DNA は片側が少し解離した状態であること、GATA3 の N 末ジンクフィンガーが結合すると結合した側の反対側の DNA が少し解離することが示唆された。

以上、「富岳」を用いたシミュレーションと協力研究者からの実験データを統合的に解析することにより、ヌクレオソームに結合する 2 つの因子、HP1 蛋白質と転写因子 GATA3 について、ヌクレオソームとの相互作用様式と結合による構造や運動の変化について新たな知見を得ることができ、本年度の計画を達成することができた。

④ ウィルス（岡山大・篠田）

B型肝炎ウイルス (HBV) のエンベロープ構造のモデリングを行い、分子シミュレーションによってエンベロープの物性を解析し、HBVの感染初期過程の分子機構の解明を目指している。エンベロープ単体の構造モデリングを進め、エンベロープ全体構造を作成し、SPICA力場を用いた粗視化分子動力学(CGMD)シミュレーションにより、その構造と動的な性質を調べた。HBVエンベロープの原子レベルの構造は実験的に明らかにされておらず、その構造モデリングと分子シミュレーションによる解析は初めての試みとなる。前年度は、エンベロープ膜蛋白質であるS蛋白質二量体の構造モデリングを行い、得られた膜蛋白質をwet実験から分かっている濃度で脂質ベシクルに埋め込むことで、HBVエンベロープの初期構造を作成した。今年度は、この初期構造を元にCGMDシミュレーションを実行したが、比較検討のため、(1)カプシド含有のエンベロープ、(2)カプシドを含まないエンベロープ、(3)脂質ベシクルの3種類の構造を用意し、SPICA力場を適用しCGMDを実行した。系の平衡化計算と温度圧力一定条件下でのプロダクションランを行い、エンベロープの構造と脂質の拡散などのダイナミクスを解析を行った。カプシド含有のエンベロープ構造の分子レベルの解析は、これまでにない初めての試みであり、カプシドとエンベロープの相互作用、複合体構造の詳細を得た。

ウィルスエンベロープは宿主細胞膜由来の脂質分子とウィルス由来の膜蛋白質から構成される膜であり、HBVカプシドの外縁を覆う袋状の構造体である。エンベロープはウィルスの核酸分子を外界から保護するとともに、細胞外の受容体蛋白質との相互作用による細胞内外輸送を実現するための機能を持つ、ウィルス粒子の重要な機能を担う膜構造体といえる。したがって、ウィルスエンベロープの構造を解明することは、エンベロープ型ウィルスの研究において重要な課題である。しかしながら、これまでに、エンベロープの構成要素である膜蛋白質の構造や役割については、十分な理解が進んでおらず、エンベロープ構造自体も高解像度の情報が不足している。エンベロープはウィルス種毎に特定の粒子サイズや形状をもつが、それらはどのような相互作用や分子機構で決まっているのかなどは全く不明瞭である。特に、ウィルス由来の膜蛋白質が持つ、ウィルス粒子の構造の安定化における役割や相互作用による構造形成時における役割などを調べるには、エンベロープの分子構造モデリングが非常に重要となる。本年度は、カプシド含有のエンベロープの3 μ sのCGMDを実行することに成功した。本計算系は粒子数が300万を超えた大規模系であり、この規模の計算を3 μ sという長時間のCGMDを実行できたのは、(1)-(2)で述べたGENESIS-SPICAのチューニングを行い、また、マルチタイムステップで計算できたことによる。

図3-4-1は、カプシド含有のエンベロープ系のCGMDの最終座標のスナップショット(3 μ s後)である。水分子やイオン分子は表示していない。(a)はエンベロープを外側から見た構造で、S蛋白質はエンベロープ膜上に数多く存在し混雑環境にあり、その配置には秩序構造は観察されていない。また、S蛋白質ダイマー間の凝集も見られず、すべてのS蛋白質ダイマー間に脂質膜を保持していることが分かる。(b)は内部のカプシド分子とのエンベロープの相対配置を見るために、手前のエンベロープ構造を非表示にしたスナップショットである。カプシド蛋白質はエンベロープ内部に隙間少なく詰まっており、カプシドのスパイク部分とエンベロープ内の膜蛋白質間に相互作用が確認される。図3-4-2は、電子密度プロファイルをプロットしており、CGMDの結果と実験で求められた相対電子密度を比較している。本検討での構造モデリングで、粒子サイズを半径 約23 nm に一致するようにモデルを作成したが、エンベロープとカプシドの相互作用により、エンベロープ半径は想定よりも縮むことが分かり、本計算は数回の試行的な構造モデリン

グの結果、得られた構造である。最終的に得られたCGMDシミュレーション後の構造は、実験で得られるカプシド、エンベロープ膜内側、外側の3つのピーク位置をよく再現しており、本モデリング構造がHBVビリオン粒子のカプシドエンベロープ複合体構造と同様な電子密度を持つことが示された。カプシドエンベロープ系において、内部に配置したカプシド分子はT=4の準対称性をもった正二十面体構造をもっている。エンベロープ蛋白質は内部のカプシド分子と相互作用していることが知られているため、いくつかの実験論文においては、膜蛋白質もエンベロープ膜上で結晶構造のような秩序配置をしていることが提案されていた。しかしながら、我々の分子モデルによって、そのような完全な秩序構造はエンベロープカプシドの複合体構造を安定化させるのに不要であることが明らかになった。一方で、カプシド分子とのコンタクトのため、S蛋白質の配置において部分的に五角形構造や六角形構造配置が確認された。このような部分的な秩序構造ができることは、カプシド分子とエンベロープ膜蛋白質の相互作用が強いことに由来する合理的な結果である。ビリオン粒子において、カプシド分子と膜蛋白質の結合は強いものの、構造や機能に必要な流動性を残していると考察される。

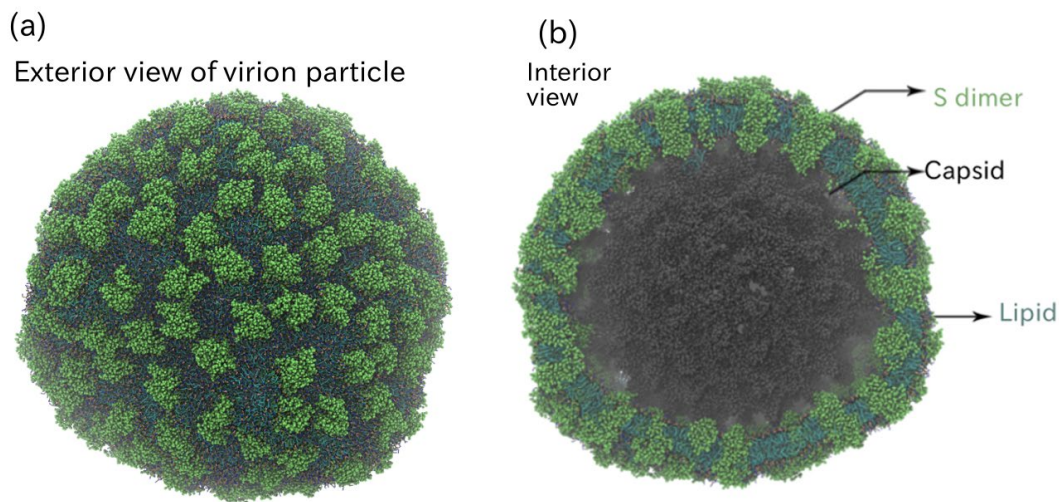


図 3-4-1: 3 μ s CGMD の最終構造。カプシド含有の HBV エンベロープの全体系の構造。(a)外側からみたエンベロープ構造、(b)手前のエンベロープを非表示にして、内部のカプシド構造を表示したもの。水分子やイオンは表示していない。粒子半径は約 22 nm で、カプシド(内側の灰色)とエンベロープ内の膜蛋白質(S dimer;黄緑のドット)は密接していることがわかる。

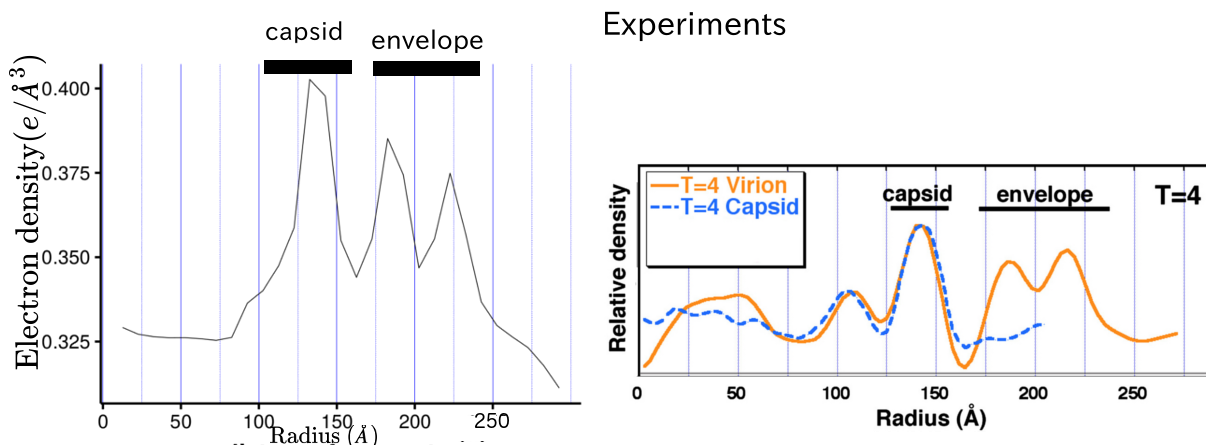


図 3-4-2: (左) カプシド含有エンベロープ系の CGMD の平衡構造から計算されたエンベロープの重心

からの距離の関数として計算された平均電子密度。(右) HBV カプシド含有エンベロープ構造 (ビリオン粒子) の実験より得られた相対電子密度。(Dryden et al. *Mol. Cell* 2006) 両者の形状はよく一致しており、カプシドエンベロープのモデリング複合構造はビリオン構造に類似していることが検証できた。

次にエンベロープ膜中の脂質分子の性質を調べるために、カプシド含有エンベロープ系、カプシド非含有エンベロープ系と脂質ベシクル系、それぞれで脂質分子の疎水鎖のオーダーパラメタを各 MD 計算の最後の 500ns の平均量として計算した。ここで計算する粗視化分子のオーダーパラメタは $^2\text{H-NMR}$ から得られるものと直接関係せず、疎水鎖中の粗視化セグメント間の結合ベクトルが膜垂直軸となす角 θ を用いて、

$$S_{zz} = \frac{1}{2} \langle 3 \cos^2 \theta - 1 \rangle$$

で定義される量であり、疎水鎖の配向分布の広さを表す指標として用いられる。POPC 平面膜では、SPICA 力場によるこのオーダーパラメタはよく検証されており、 $^2\text{H-NMR}$ の結果をよく再現する原子モデルを粗視化粒子にマッピングし、ここでの定義と同じ方法で算出したオーダーパラメタは、SPICA 力場によっても非常に良く再現できることが確認されている。図 3-4-3 に、カプシド含有エンベロープ系、カプシド非含有エンベロープ系、脂質ベシクル系中の POPC 脂質のオーダーパラメタをプロットした。横軸のセグメント番号はエステル粒子から末端セグメントにかけて疎水鎖に沿って順に振った番号である。純粋な POPC 脂質ベシクルに対して、エンベロープ膜タンパクの入った系、さらにカプシドを含有する系は、この順でオーダーパラメタが上がる傾向が見られる。すなわち、脂質分子の膜中で配向秩序は、エンベロープ膜蛋白質が膜中に存在することで高くなり、さらにカプシド蛋白質がエンベロープ内側に存在することによっても、さらに高くなるという影響を受けることが示された。すなわち、脂質分子は膜蛋白質への配位によって構造化され、さらにカプシドの存在によっても構造化が進んでおり、これはエンベロープの安定化に寄与するものと考えられる。

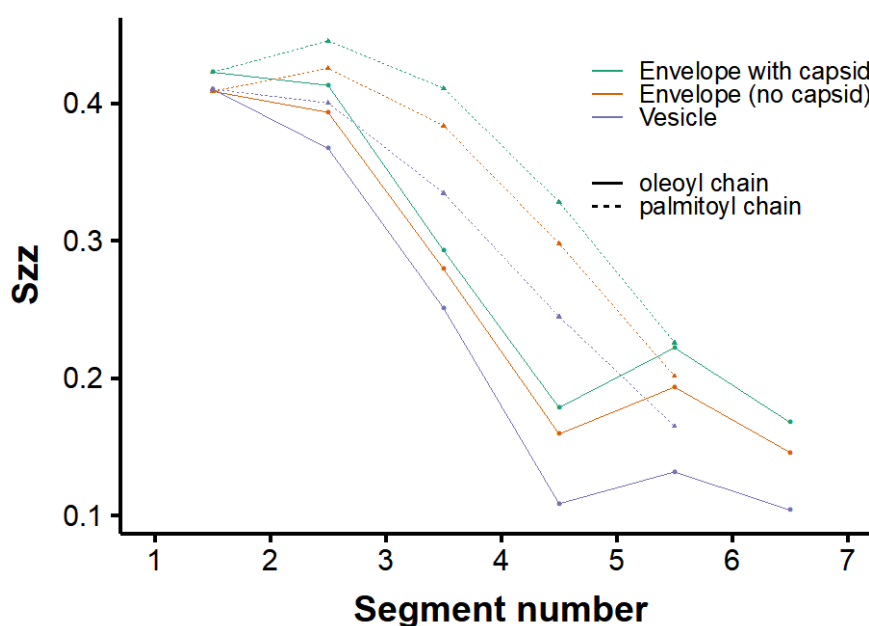


図 3-4-3: ベシクル系、カプシド非含有、カプシド含有エンベロップ系における POPC 分子の疎水鎖のオーダーパラメータ。MD 計算の最後の 500ns の平均から算出。脂質分子のオーダーパラメータは、ベシクルに比べてエンベロップ中の方が高くなり、より配向秩序の高い構造を取る。さらに、エンベロップ内部に存在するカプシドによって、オーダーパラメータがより高くなる。

表 3-4-1: カプシド含有、カプシド非含有エンベロップ系と脂質ベシクル系での内葉・外葉における脂質分子の自己拡散係数。

系	内葉脂質の自己拡散係数 D [$\times 10^{-7}$ cm ² /s]	外葉脂質の自己拡散係数 D [$\times 10^{-7}$ cm ² /s]
カプシド含有エンベロップ	5.66	11.3
空のエンベロップ	9.84	12.5
POPC 脂質ベシクル	50.4	68.9

次に、蛋白質による脂質ダイナミクスへの影響を調べるために、脂質分子の自己拡散係数をカプシド含有エンベロップ系、カプシド非含有エンベロップ系と脂質ベシクル系、それぞれで脂質二重層の外葉と内葉に分けて計算を行った。その結果を表 3-4-1 に示す。粗視化分子系では、粒度分の平均化により、一般にエネルギー地形が平滑になるため拡散係数が高くなる。SPICA 力場では、平面膜において、脂質拡散係数が実験に比べて約 1 オーダー高くなる。そのため、ここでの比較は定性的に留める。リン脂質のフリップフロップは非常に長時間で起こる現象であり、本 MD 計算中でも起きていないことを確認した。そのため、外葉、内葉の脂質はきれいに分離できる。また、球面上の側方拡散であれば、半径×回転角で球面上を移動した距離を算出できるが、カプシド含有エンベロップなどは、球状からはずれが大きいため、ここでの脂質分子の拡散は単純に 3 次元の移動距離(平均二乗変位)から算出した。時刻 $t = 10\text{--}20\text{ns}$ は平均二乗変位のスロープに乗るため、その傾きから見かけの自己拡散係数を評価している。表 3-4-1 からは、脂質の拡散係数が、膜蛋白質の存在により 5 倍以上低くなることがわかる。さらに、カプシド含有系においては、カプシド近傍の内側の層が強く脂質分子のダイナミクスに影響を与えていることがわかる。すなわち、脂質分子の拡散は蛋白質の存在により遅くなり、エンベロップ膜全体の動きはベシクルよりも遅くなっていることが示唆される。オーダーパラメータの結果と合わせると脂質分子は蛋白質の存在により構造化され、その拡散も制限されることでより堅い膜を形成し、エンベロップ膜の構造が安定化されていると考えられる。

次に、蛋白質間の構造を調べるために、カプシド蛋白質と S 蛋白質の複合体部分について調べた。図 3-4-4 は (a)CGMD 最終座標 (3 μs 後)におけるカプシド含有エンベロップ系におけるカプシド蛋白質とエンベロップの S 蛋白質の接触箇所のみを示したカプシド全体像と (b)接触部位の、カプシドタンパクのスパイク部分と S 蛋白質二量体の複合構造の代表構造の拡大図を示している。図 3-4-4(a)の全体構造の接触セグメントから、カプシドタンパクと S 蛋白質の接触部分はカプシド先端で起こり、分布は一様ではなく、接触セグメントの多い部位と少ない部位がある。すなわち、実験データから推察されたようなカプシド蛋白質との相互作用により、すべての S 蛋白質が結晶のような秩序構造を持っている訳ではなく、揺らぎの中でも維持はされるが、より緩やかなコンタクトであることが示唆される。図 3-4-4(b)のような

部分構造を解析することによって、より統計的にカプシドとエンベロープ膜蛋白質の複合体構造を分子レベルで明らかにすることができる。図 3-4-4(b)の複合体構造からは、1つのカプシドスパイクは2つのS蛋白質二量体と接触できることがわかった。すなわち、1対1の結合で結晶状の構造化は必要ではない。しかし、蛋白質同士の結合は比較的強く、結合後 μs の間に解離することは見られなかった。一方で、カプシド分子と脂質分子の接触部位の数は、カプシド分子とエンベロープの蛋白質接触部位の数よりも圧倒的に少なかった。これらの事実からわかるのはエンベロープ膜のサイズと形状を規定できる複合体構造は、ウイルス由来の膜蛋白質とカプシド蛋白質の相互作用によりほぼ説明されることが示唆された。脂質とカプシドの直接相互作用は弱く、脂質分子は膜構造の維持に必要で膜蛋白質間を埋める役割が主であると考えられる。

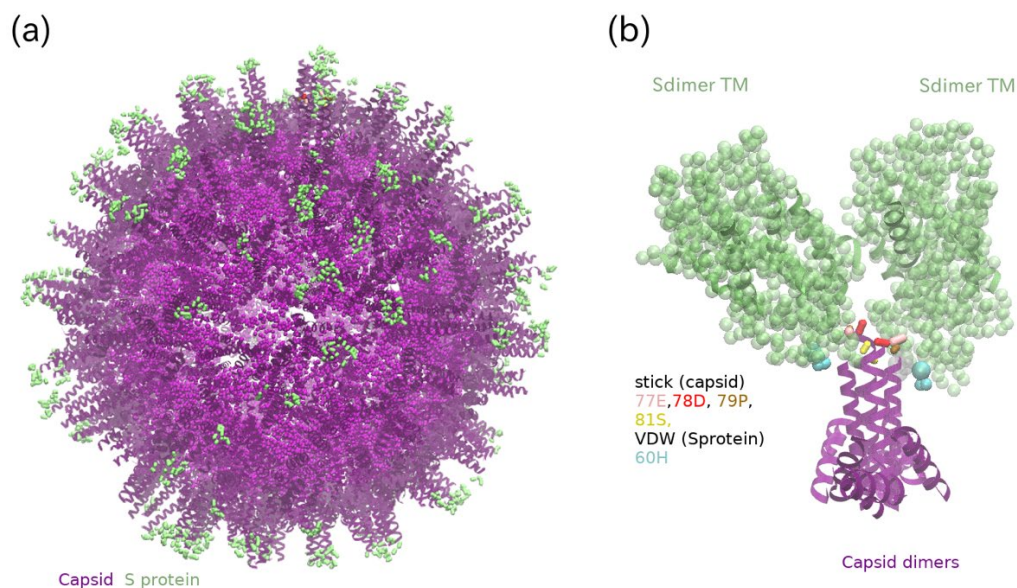


図 3-4-4: 3 μs の時点におけるカプシド含有のエンベロープ系におけるカプシド分子と S 蛋白質の接触構造の (a)全体像と (b)接触構造の中での代表構造。(a)では S 蛋白質のカプシド接触セグメントのみを Licorice(緑)で表示している。紫色のヘリックスはカプシド蛋白質。(b)では、カプシドタンパクのダイマー(紫)と、それに接触するエンベロープ中の S 蛋白質(緑)を表示し、さらにカプシド及び S タンパクの接触残基を Licorice で表している。残基毎に色分けで表示した。

次に、蛋白質の複合体構造における重要なアミノ酸残基を特定するために、計算の最後の 1 μs における S 蛋白質とカプシド蛋白質の間のコンタクトペアの平均個数が多い順に 10 ペアを抽出したものを表 3-4-2 に示す。カプシド蛋白質の残基, 77E, 78D, 79P, 81S はエンベロープ構造の形成や機能に必須部位の残基として実験的に確認されており、これらの残基がエンベロープ膜蛋白質と複合体構造を構成していることは合理的である。一方で、それらのカプシド残基と相互作用する S 蛋白質の残基はほとんどが 60H となっており、この残基が S 蛋白質の重要な部位となることが初めて明らかとなった。以上のように HBV エンベロープ膜において、その構造を形成し、安定化する上での重要な相互作用・因子が分子レベルで CGMD を通じて明らかになった。HBV エンベロープのモデリングと分子シミュレーションの結果より、ウイルス由来の蛋白質が、ウイルス粒子の形成過程において必要な役割を果たすだけでなく、粒子の構造そのもの

や安定性の付与もこの膜蛋白質により担われていることが初めて明らかとなった。

表 3-4-2: S 蛋白質とカプシド蛋白質の接触している残基ペアについて、1 マイクロ秒で平均された 1 構造あたりのコンタクトペアの個数。コンタクトは粒子間距離が 5Å 以内にあるとき、1 カウントとして定義している。残基ペアは左側が S 蛋白質の残基番号と残基名、右側がカプシド蛋白質の残基番号と残基名である。特に、S 蛋白質の 60H 残基がカプシド蛋白質の様々な残基とコンタクトペアを形成していることがわかる。

ランク	Residue pair [S - Capsid]	ペアの平均個数
1	60H-77E	47
2	60H-78D	47
3	60H-79P	38
4	60H-81S	31
5	35W-78D	30
6	72Y-81S	29
7	60H-75N	26
8	72Y-78D	25
9	72Y-79P	24
10	72Y-77E	23

⑤ 多剤排出トランスポータの分子動力学シミュレーション (東大・篠田)

(1) 大腸菌モデル膜 (LM301 株膜; カルジオリピン無し) に埋め込まれた多剤排出トランスポータ AcrABZ-TolC 複合体の MD シミュレーションの実施と相互作用解析。

グラム陰性菌による薬剤耐性化は細胞膜に存在する多剤排出トランスポータが薬を細胞外へ排出することが主な原因と考えられている。大腸菌多剤排出トランスポータ AcrA-AcrB-AcrZ-TolC (AcrABZ-TolC) は、ポンプ機能を担う AcrB、外部チャネルの TolC、AcrB と TolC をつなぐアダプタータンパクの AcrA、そして基質選択性に関与し排出を調整すると考えられている AcrZ のサブユニットから構成される約 800kDa の巨大な膜蛋白質複合体である (図 3-5-1)。各サブユニットはホモ多量体で、TolC、AcrA、AcrB、AcrZ はそれぞれ、三量体、六量体、三量体、三量体である。この複雑で巨大な膜蛋白質は、膜内でその機能を発揮するため、蛋白質-蛋白質間相互作用のみならず、膜との相互作用が基質排出メカニズムを解析する上で非常に重要である。大腸菌の細胞膜は様々な脂質分子が含まれる混合膜である。今年度は前年度構築した、実在データに基づく大腸菌モデル膜 (LM301 株膜; カルジオリピン無し) に埋め込まれた多剤排出トランスポータ AcrABZ-TolC 複合体に対して、1 μ s の分子動力学 (Molecular Dynamics; MD) シミュレーションを合計 3 本実施し、膜とタンパク質の相互作用を解析した。当初、同じ初期構造を用いて初期運動量を変えることにより初期条件の異なるシミュレーションを進めていた

が、3本目のシミュレーションにおいて膜とタンパクの相互作用に膜の初期構造依存性が見られたため途中で中止し、新たに同じ脂質構成で初期座標が異なる2つ目の膜にトランスポータを埋め込んだ系を構築して3本目として分子動力学シミュレーションを追加実施した。

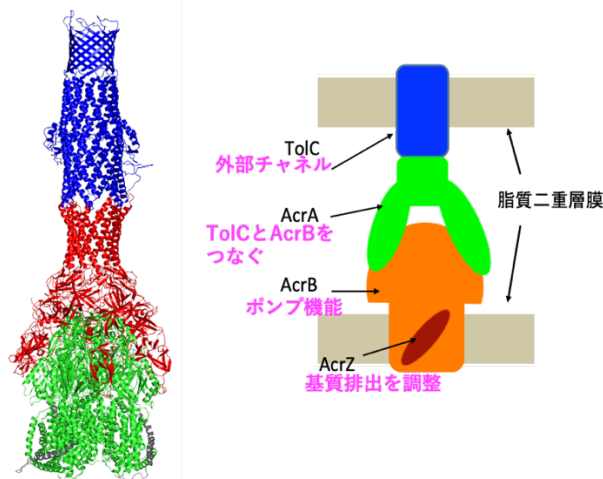


図 3-5-1: 細胞膜と外膜を持つ大腸菌膜に存在する多剤排出トランスポータ AcrA-AcrB-AcrZ-TolC (AcrABZ-TolC) 複合体。

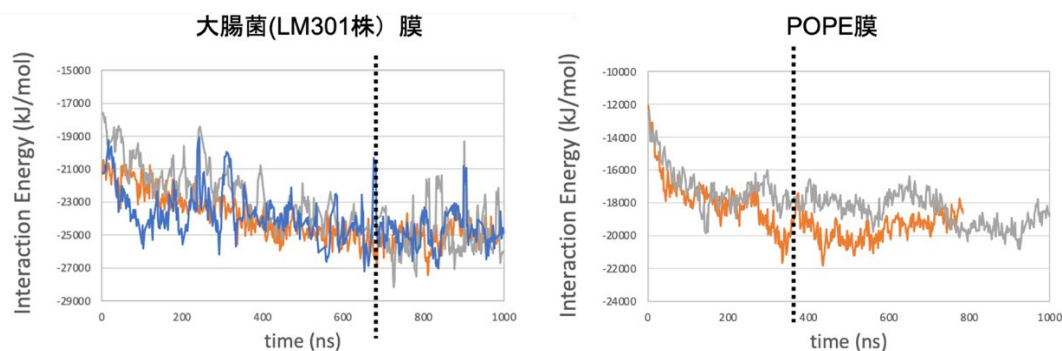


図 3-5-2: 多剤排出トランスポータ AcrABZ-TolC 複合体と膜との相互作用エネルギーの時間変化。左図は大腸菌 (LM301 株膜) 膜、右図は POPE 膜である。グラフの色の違いは、トラジェクトリの違いを示す。黒の点線は相互作用が安定化するおおよその時間の位置を示している。

図 3-5-2 に多剤排出トランスポータ AcrABZ-TolC 複合体と膜との相互作用エネルギーの時間変化を示す。比較のために、以前行った生体モデル膜 (POPE 膜) の結果も合わせて示した。多成分脂質膜である大腸菌膜は、単成分脂質膜である POPE 膜に比べ、膜とトランスポータ複合体との相互作用が安定化するまでに倍近く時間を要することがわかった。このことより、この規模の系において解析まで含めた全体のシミュレーション時間には $1 \mu\text{s}$ は必要であることを見積もることができた。

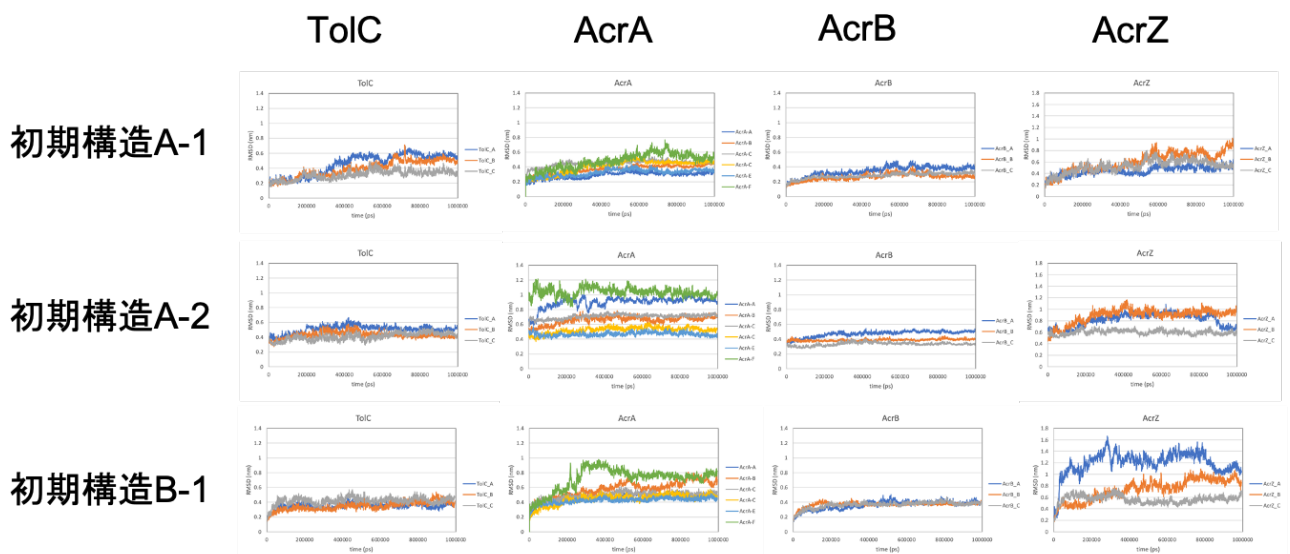


図3-5-3: 多剤排出トランスポータAcrABZ-TolC複合体の各サブユニットのRMSD。初期構造はAとBは、膜の初期座標の違いを示し、A-1とA-2は、初期運動量の違いを示している。グラフの色の違いは、サブユニット内のモノマーの違いを示す。(TolC: 三量体、AcrA: 六量体、AcrB: 三量体、AcrZ: 三量体)

図3-5-3は多剤排出トランスポータAcrABZ-TolC複合体サブユニットの各モノマーにおけるRoot Mean Square Deviation (RMSD)を示す。これらを見ると、AcrAとAcrZの各モノマーのRMSDのばらつきが大きいことがわかる。AcrAについては、値の大きいモノマーと小さいモノマーに分かれることが見て取れる。それらのモノマーを調べると、値の小さなモノマーは、N末端領域が膜に刺さり、揺らぎを抑えていることがわかった(図3-5-4)。また、AcrZの各アミノ酸と膜との相互作用を調べたところ(図3-5-5 (a))、以前の我々の研究で調べたAcrZの各アミノ酸とAcrBとの相互作用(図3-5-5 (b))と同様に、AcrZの大部分のアミノ酸は、膜との相互作用が小さいことがわかった。AcrZの中間部はほとんどnon-polarなアミノ酸で構成されているが、これらのアミノ酸が、AcrB、膜どちらともほとんど相互作用しておらず、一部のpolarなアミノ酸とのみ相互作用していることがわかった。

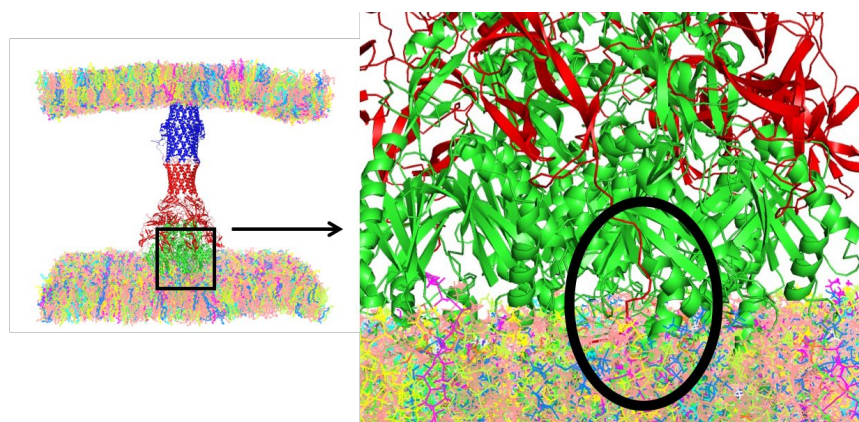


図3-5-4: (左) 1 μ s後のAcrABZ-TolC複合体。青色はTolC、赤色はAcrA、緑色はAcrBを示す。AcrZは膜内にあるためこの図では見えない。膜を構成する8種類の脂質分子は種類によって色分けされている。(右)

左図の拡大図。AcrAのN末端領域が膜に刺さっている様子を示す。

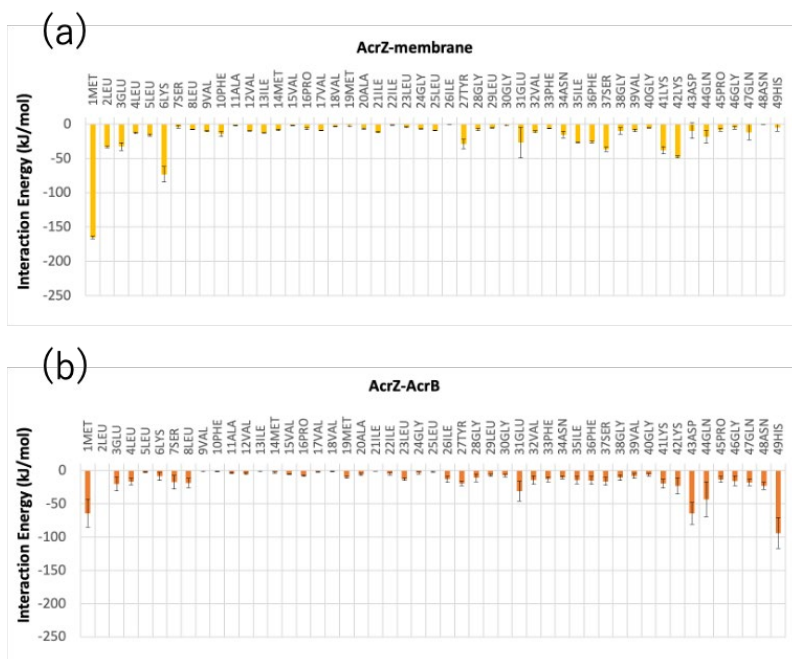


図3-5-5: AcrZの各アミノ酸と膜全体 (a) およびAcrB全体(b)との相互作用エネルギー。横軸はAcrZの各アミノ酸の番号と名前、縦軸は相互作用エネルギーである。

次に、多剤排出トランスポータ複合体の基質を取り込み排出するポンプの役割をするAcrBに対して、膜がどのような相互作用をしているのかを調べるために、POPE膜、PMPE膜（大腸菌の主要脂質）、大腸菌LM301株モデル膜に多剤排出トランスポータ複合体を埋め込んだ3つの系に対して、それぞれAcrBの各アミノ酸と膜全体との相互作用エネルギーを計算した。図3-5-6は大腸菌LM301株モデル膜の3つのトラジェクトリのうちの1つに対して計算した例を示す。3つの系で比較したところ、相互作用の様子が異なるアミノ酸が見出された。具体的には、Asp711、Glu839、Glu842、C末端領域で、図3-5-6の赤で示した領域にある。これらの残基は全て膜と接している。

C末端領域は内膜の細胞質側の界面に位置しておりフレキシブルで特定の構造を持たないため、脂質と多様な相互作用をされると考えられる。Glu839とGlu842は、膜の種類のみならず、シミュレーション・ランによっても相互作用が異なっていた。例えば図3-5-7のように、Glu842は、POPE膜ではどのモノマーも負の相互作用エネルギーだが、PMPE膜では1つのモノマー、大腸菌モデル膜では2つのモノマーが膜と正の相互作用をしていることがわかった。また、Asp711では、大腸菌モデル膜でのみ正の相互作用が見られた。3つの系の違いを調べると、それぞれのアミノ酸と相互作用する脂質分子の種類や、同じ種類の脂質分子でも相互作用する部位が異なることがわかった。図3-5-8にPOPE膜系と大腸菌モデル膜系におけるGlu842と脂質分子の様子を示した。これを見るとPOPE膜系では、POPE分子のヘッドグループの正電荷を有するエタノールアミンがGlu842のカルボキシル基の近くに位置して負の相互作用をしているのがわかる。一方、大腸菌モデル膜系では、Glu842のカルボキシル基の近くに負の電荷を持つPMPGのヘッドグループが位置しているだけでなく、一部のPMPE分子がエタノールアミンではなく負の電荷を持つリン酸基をGlu842の方へ向けている。これらは共に正の相互作用となる。PMPE膜系では、一部のPMPE分子のリン酸基が、大腸菌モデル膜系と同様にGlu842と正の相互作用をしていた。Asp711、Glu839、

Glu842は、この多剤排出トランスポータの基質取り込み経路の一つの入り口に位置していたことから、膜との相互作用が基質取り込み制御に影響を及ぼすことが示唆された。



図3-5-6: 大腸菌LM301株モデル膜におけるAcrBの各アミノ酸（1モノマー1049残基）と膜全体との相互作用エネルギー。横軸はAcrBの各アミノ酸の番号と名前、縦軸は相互作用エネルギーである。3つのモノマーに対して色を変えて示している。初期構造B-1のトラジェクトリによる結果である。赤で囲まれたアミノ酸（Asp711、Glu839、Glu842、C末端領域）における相互作用が、膜の種類やシミュレーションのランによって大きく異なっていた。

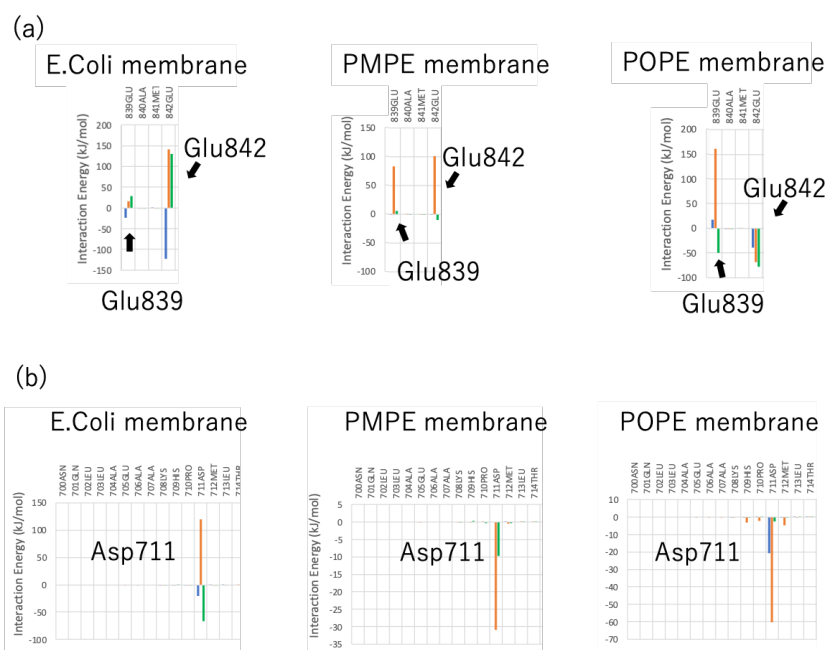


図3-5-7: (a) AcrBのGlu839とGlu842と膜との相互作用エネルギー。(b) AcrBのAsp711と膜との相互作用エネルギー。3つのモノマーについての値を色分けして示している。

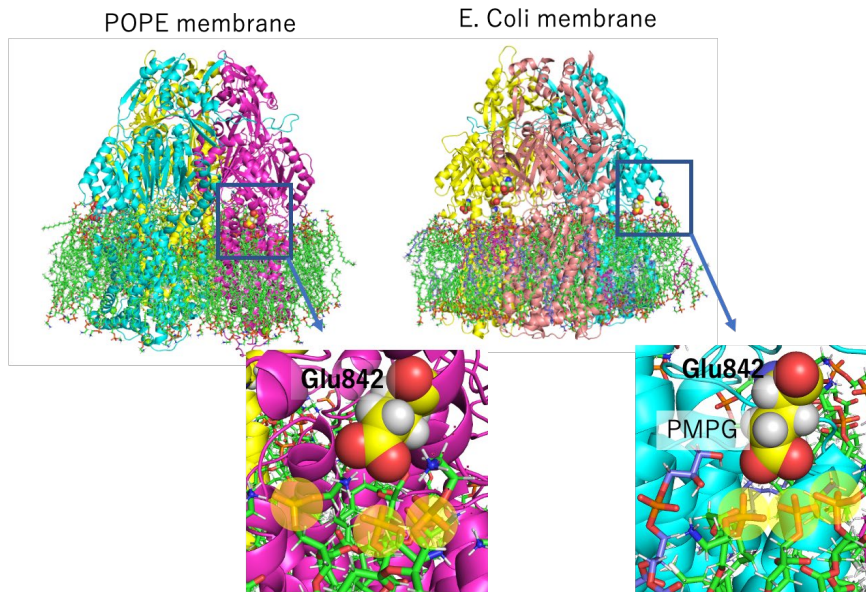


図3-5-8: AcrABZ-TolC複合体のAcrBの部分と、Glu842付近の拡大図。(左)POPE膜 (右) 大腸菌膜。脂質分子はAcrBの周りの分子だけスティック表示で示している。Glu842はsphere表示で、AcrBはモノマー毎に色分けして示している。黄色の半透明の円はリン酸基である。

(2) カルジオリピンの力場作成とカルジオリピンを含む大腸菌膜(W3110S株モデル)の構築及び大腸菌W3110S株モデル膜に埋め込んだAcrABZ-TolC複合体のMDシミュレーションの実施

カルジオリピンは大腸菌の細胞膜に存在する、リン酸基2つに脂肪酸が4本あるリン脂質である (図3-5-9)。カルジオリピンは疎水性かつ2価の負電荷を帯びた特徴的な構造から、大腸菌では総リン脂質の5%程度であるにもかかわらず、膜の湾曲への関与や膜タンパク質の二量体化の制御など、さまざまな細胞内のダイナミクスに関わることが知られている。

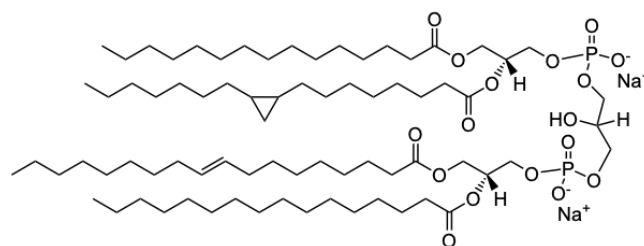


図3-5-9: 大腸菌のカルジオリピンの分子構造。

今年度は、カルジオリピンを含む大腸菌 (W3110S株) 膜を作成するために、まず大腸菌の実際に市販されているカルジオリピン分子の力場を作成し、脂肪酸組成を測定した実際の実験データに基づいた6種類の脂質分子で構成されたW3110S株モデル膜を作成した (図3-5-10)。このモデル膜にAcrABZ-TolCを埋め込んだ系を構築し、エネルギー最小化、水やイオン、膜の構造緩和を実行した後、MDシミュレーションを実施した。図3-5-11には構築した系と、現在までの膜-トランスポーター複合体間の相互作用エネ

ルギーの時間変化を示した。LJ相互作用、クーロン相互作用共に安定状態には至っていないのが見てわかる。現在は平衡化へ向けてMDシミュレーションを進めている。

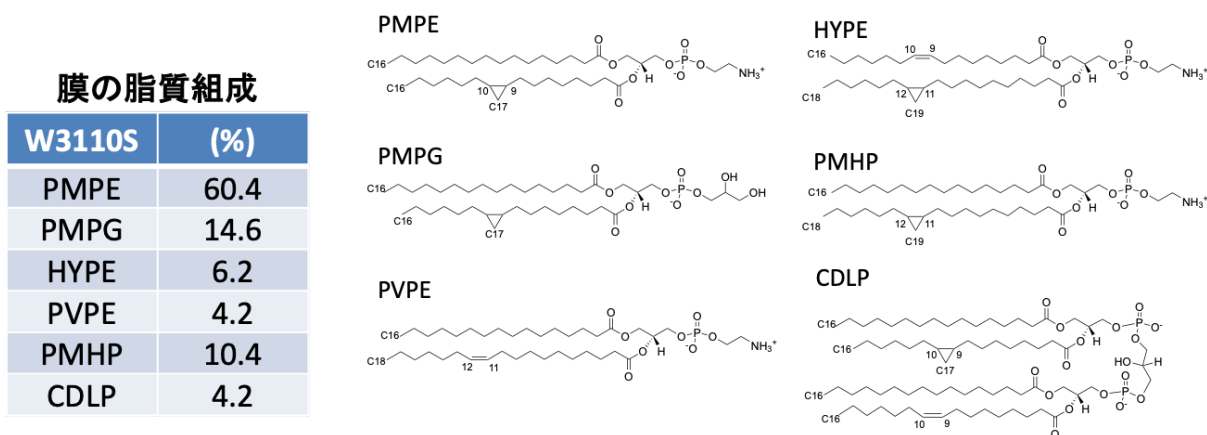


図3-5-10: 膜の脂質組成と、構成脂質分子の分子構造。PMPE [1-palmitoyl-2-cis-9,10-methylene-hexadecanoic-acid-sn-glycero-3-phosphoethanolamine]; PMPG [1-palmitoyl-2-cis-9,10-methylene-hexadecanoic-acid-sn-glycero-3-phosphoglycerol]; PVPE [1-palmitoyl-2-vacenoil-sn-glycero-3-phosphatidylethanolamine]; PMHP [1-palmitoyl-2-cis-11,12-methylene-hexadecanoic-acid-sn-glycero-3-phosphoethanolamine]; HYPE [1-palmitoleoyl-2-cis-11,12-methylene-hexadecanoic-acid-sn-glycero-3-phosphoethanolamine]; CDLP [1-(1-palmitoyl-2-cis-9,10-methylene-hexadecanoic-acid-sn-3-phosphatidyl)-3-(1-palmitoyl-2-oleoyl-sn-3-phosphatidyl)-sn-glycerol]

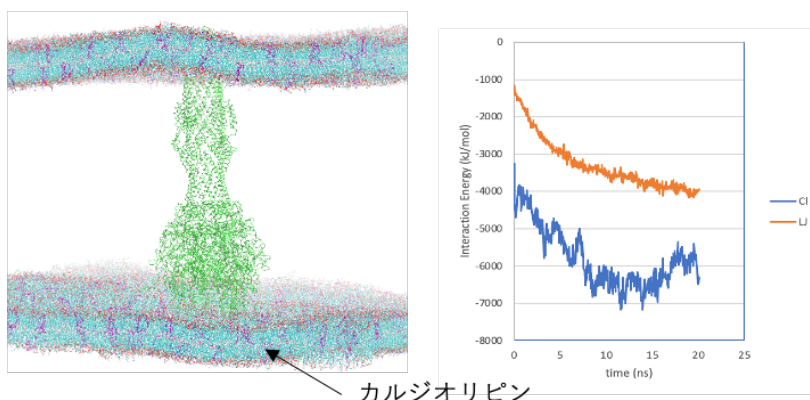


図3-5-11: カルジオリピンを含んだ大腸菌W3110S株膜に埋め込んだ多剤排出トランスポータ複合体 (左) と膜とトランスポータ複合体との相互作用エネルギー (右)。膜内のマゼンダで示した脂質分子は、存在比率が4.2%のカルジオリピンである。相互作用エネルギーはLJ相互作用 (オレンジ) とクーロン相互作用 (青) に分けて表示した。

(4) プロジェクトの総合的推進

本年度も新型コロナウイルスの影響で、オンサイトでのミーティング等を開催することは困難であったため、インターネットのツールを駆使して共同研究、打ち合わせ、議論等を活発に行った。プロジェクトの総合的推進に必要な情報交換は以下のようにして行った。

- ・ Zoom を用いたプロジェクトミーティング (月 1 回)
- ・ Slack でプロジェクト専用のチャンネルを開設し情報交換を行う
- ・ Box を用いたファイル共有

プロジェクトミーティングは、Zoom を用いて毎月 1 度行った。令和 3 年度は、4 月 13 日 (火)、5 月 11 日 (火)、6 月 8 日 (火)、7 月 20 日 (火)、8 月 5 日 (木)、9 月 14 日 (火)、10 月 12 日 (火)、11 月 16 日 (火)、12 月 14 日 (火)、1 月 18 日 (火) に開催した。本年度は、我々のプロジェクトの計算に深く関係する実験家を招き、オンラインでセミナーをしていただき、細胞内環境での生体分子動態に関する深い議論をすることができた。講演していただいた研究者は以下の通りである。

7 月 20 日 新海創也 (理研)

8 月 5 日 柴田幹太 (金沢大)

9 月 14 日 落合博 (広島大)

10 月 12 日 乙須拓洋 (埼玉大)

11 月 16 日 村上聡 (東工大)

12 月 14 日 本間道夫 (埼玉大)

1 月 18 日 前仲勝美 (北大)

また、プロジェクトミーティングの最後に、「富岳」の有効な利用方法についてメンバーで情報公開することで、利用率を上げたり、無駄な計算を行わないようにした。これらもプロジェクトの推進に有効であったと思われる。12 月 28 日には、スペシャルミーティングとして、この 1 年間の進捗について各 PI から研究紹介をしていただいた。前年度からの計画に基づき、「富岳」を利用することで大規模なモデリングとシミュレーションが実現してきたことを実感した。また、Slack と Box を用いた情報交換も新型コロナウイルス禍では大変に有効だったが、前年度の報告書に詳しく記載したためここでは詳細を割愛する。

本プロジェクトでは、理研を中心に開発している分子動力学ソフトウェア GENESIS の開発と公開をおこなってきた。「富岳」に最適化されたカーネルを有する GENESIS 2.0b については細かいバグフィックス等を行い、利用者の便宜をはかった。一方、多くの機能を有する GENESIS 1.X については最新版 GENESIS 1.7 を公開して、京大高田らと開発してきた残基レベルの粗視化モデル(とそのツール GENESIS-cg-tools)が利用できるようになった。今後、この 2 つのバージョンの GENESIS を統合してさらに使いやすいソフトウェアに進化させていく計画である。それに伴い、チュートリアルの変更、講習会も積極的に行っていく。また、本プロジェクトで開発してきた ABMD (量研機構河野グループと連携)、SPICA/pSPICA (岡山大学大篠田グループと連携)についても既に「富岳」で良好なパフォーマンスを示しているため、統合された新しいバージョンの GENESIS に近日中に導入する予定である。

2-3. 活動（研究会の活動等）

令和3年度の研究会の活動等も、やはり新型コロナウイルスの影響が大きくオンラインでは実施できなかった。そのため、オンラインでシンポジウムを開催した。

・Symposium on Computer Simulation and cryo-EM/ET of Complex Biomolecular Systems (2021年11月18, 19日) : https://bfs.riken.jp/misc/symp_complex_biomolecular_systems/

前年度のシンポジウムでは、アジアと日本のシミュレーション研究者に限定して深い議論を行った。本年度は、ヨーロッパ、オセアニア、アジア、日本に拡大して招待講演者を募った。さらに計算だけでなく、実験、特に近年の発展が著しい cryo-Electron Microscopy/Tomography の研究者に講演をしていただき、実験と計算の連携についても深く議論することができた。招待講演者は以下の通りである。

Toby Allen (RMIT University, Australia)

Kei-ichi Okazaki (Institute for Molecular Science, Japan)

Ben Corry (Australian National University, Australia)

Bert de Groot (Max Planck Institute for Biophysical Chemistry, Germany)

Masahide Kikkawa (University of Tokyo, Japan)

Sai Li (Tsinghua University, China)

Joanna Trylska (Warsaw University, Poland)

Atsunori Oshima (Nagoya University, Japan)

Zhenli Xu (Shanghai Jiao Tong University, China)

Syma Khalid (University of Oxford, UK)

Yuko Okamoto (Nagoya University, Japan)

Xin Zhou (University of Chinese Academy of Sciences, China)

Akio Kitao (Tokyo Institute of Technology, Japan)

（下線は海外の招待講演者、イタリックは実験研究者）

最終的に32か国405名が参加登録をするという非常に大規模なオンラインミーティングとなった。これには理由があり、オーガナイザーとして、中国北京大学のSong Chen教授に参加していただき、非常に多くの中国からの参加者がいたためである。その熱意を考えると、日本の若い学生・ポストクの育成が急務であることを痛感した。日本からは理研の杉田と京大の高田がオーガナイザーとしてシンポジウムを主催した。

・GENESIS講習会（筑波大学Cygnusを用いたハンズオン）（2021年9月28日）

令和2年度に引き続き、GENESIS講習会を、筑波大学計算科学研究センター、RISTと連携して実施した。このプロジェクトとして共催等の手続きは踏んでいないが、講師としてGENESISの開発者の1人である岩橋（小林）千草博士（理研R-CCS）が参加したため、ここに記載する。前回より定員を増やし、21名の参加がありチュートリアルは盛況であった。大学・研究所の研究者だけでなく、民間企業の研究者からの参加者も4割近くあった。フリーディスカッションの時間も設けることで、参加者の実際の研究上での利用方法などより深い議論を行う事ができた。この機会によってさらなるGENESISの普及につながる事が期待される。開発者側にとっても、利用者のニーズや難しいと感じる所を知ることができ、今後の研究開発にも反映していく予定である。

2-4. 実施体制

実施項目	実施場所	担当責任者
<p>(1) 全原子・粗視化分子動力学の最適化と並列化</p>	<p>〒650-0047 兵庫県神戸市中央区港島南町 2-2-3 国立研究開発法人 理化学研究所・生命機能科学研究センター</p> <p>〒700-8530 岡山県岡山市北区津島中 1-1-1 国立大学法人岡山大学</p>	<p>杉田有治</p> <p>篠田渉</p>
<p>(2) データ科学的手法による粗視化モデルパラメタの最適化</p>	<p>〒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</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>高田彰二</p> <p>松永康佑</p>
<p>(3) 生体分子系のモデリングと「富岳」を用いたシミュレーション</p>	<p>〒650-0047 兵庫県神戸市中央区港島南町 2-2-3 国立研究開発法人 理化学研究所・生命機能科学研究センター</p> <p>〒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</p> <p>〒263-8555 千葉県千葉市稲毛区穴川四丁目9番1号 国立研究開発法人 量子科学技術研究開発機構</p> <p>〒700-8530 岡山県岡山市北区津島中 1-1-1 国立大学法人岡山大学</p> <p>〒113-8657 東京都文京区弥生1-1-1</p>	<p>杉田有治</p> <p>高田彰二</p> <p>河野秀俊</p> <p>篠田渉</p> <p>篠田恵子</p>

	国立大学法人 東京大学	
(4)プロジェクトの総合推進	〒650-0047 兵庫県神戸市中央区 港島南町 2-2-3 国立研究開発法人 理化学研究所・生命機能科学研究センター	杉田有治

別添 1 学会等発表実績
別紙の通り