令和２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 MD による細胞内分子動態の解明」

成果報告書

令和 3 年 5 月 28 日

国立研究開発法人　理化学研究所

杉田　有治

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 MD 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。細胞内で働くタンパク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（タンパク質）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タンパク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、タンパク質間に働く弱い相互作用がタンパク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1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（μs: 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、μm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（名古屋大学）らが開発しているSPICAなどの生体高分子の粗視化分子モデルは、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２年度（報告年度）の実施内容

２－１．当該年度（令和２年度）の事業実施計画

（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）

MDソフトウェアGENESISを「富岳」に向けて最適化・並列化のチューニングを行い、高速な全原子MDの実現に向け研究を推進する。また、統計的に正しい長時間刻みに基づくMDを安定に動作させるための検証を行う。さらに、GENESISに移植された粗視化分子モデルCafeMolを用いた計算の最適化・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（名大・篠田）

定量的粗視化モデルSPICA力場の分子ライブラリの拡充を行い、より広い生体分子を取り扱えるようにする。また、SPICA-FFに基づく粗視化MD計算を「富岳」を用いて行うために、GENESISに移植し、高速かつ高並列計算の実現に向けて研究を推進する。

（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）

MDシミュレーションの結果をマルコフ状態モデルとして解析し、一分子計測実験で得られた時系列データを参照することにより、異なる分子構造の存在確率と遷移確率を最適化する機械学習を行う。この機械学習（隠れマルコフモデル）に必要なツール群を整備するとともに、応用範囲を拡大するために必要な改良を行う。

② 粗視化モデル CafeMol のパラメタ最適化（京大・高田）

粗視化分子モデルCafeMolは、タンパク質・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。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タンパク質・タンパク質相互作用の粗視化パラメタの最適化を行い、より精度の高いMDシミュレーションの実現に向けて研究を推進する。

（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、この環境におけるタンパク質の構造・ダイナミクス・機能の関係を調べる。特に、タンパク質と基質の相互作用がターゲット以外のタンパク質が存在す

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、タンパク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3次元構造モデルを構築し、CafeMolを用いた粗視化MD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等の実験情報を用いて粗い3次元構造モデルを構築する。さらに、粗視化分子モデルCafeMolを用いて詳細な動態シミュレーションを実施し、転写活性におけるタンパク質とDNAの相互作用および立体構造変化に関する知見を得る。

③ 核内タンパク質・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、原子レベルでのクロマチン動的構造を構築する。電子顕微鏡像をベースにヌクレオソーム多量体の原子モデルを構築するとともに、クロマチン結合因子の結合した複合体構造をモデリングし、MDシミュレーションを実施することで、これらの分子の結合安定性や動態を調べる。

④ ウイルス（名大・篠田）

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。本年度は未だ不明なカプシド外部のエンベロープ(ウイルス由来のタンパク質と脂質からなる二重層膜)の構造モデリングを行い、さらに分子シミュレーションにより、その安定性やダイナミクスを調べる。

⑤ 多剤排出トランスポータの MD シミュレーション（東大・篠田）

大腸菌の生体膜は様々な脂質分子が含まれる混合膜である。本年度はこの大腸菌モデル膜の作成とシミュレーションを実行し、生体モデル膜との違いを解析する。また、その大腸菌モデル膜に多剤排出トランスポータAcrAB-TolC複合体を埋め込んだ700万原子系の分子モデルを構築する。AcrAB-TolC複合体の構造は2017年にcryo-EMによって得られた構造であり、世界で初めてリアルな大腸菌内のモデルのシミュレーションを実行する。

（４）プロジェクトの総合的推進

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プロジェクトミーティングを適宜開催する。GENESISを「富岳」にチューニングし、大規模な全原子・粗視化MDシミュレーションで活用する。プロジェクト内外の研究者も交えたワークショップを適宜開催し、成果の公開と研究交流を行う。得られた成果については、学会発表、論文発表などによって積極的に公表すると共に、開発した手法についてはフリーソフトウェアとしてホームページ上で公開する。また、他のプロジェクトとの連携などにより効率的・効果的に研究を推進する。

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）

我々は、理研を中心に開発しているMDソフトウェアGENESISを「富岳」に向けて最適化・並列化のチューニングを行い、高速な全原子MDシミュレーションの実現に向けた研究を推進した。GENESISは「富岳」開発プロジェクトにおけるターゲットアプリケーションの一つであり、「Flagship 2020 Project」においてコデザインがなされた。本プロジェクトはその成果を受けて、新しい非共有結合相互作用アルゴリズムなど「富岳」への最適化・並列化のチューニングがなされたカーネルを含むコードをフルパッケージ（GENESIS 2.0beta）としてまとめて、以下のGithubからリリースした。

https://github.com/genesis-release-r-ccs/genesis-2.0

ここで2.0betaという名前でリリースしたのはいくつかの理由がある。リリース時点にはまだ「富岳」の共用利用が開始されておらず、試験利用であったため、アプリケーションに影響するシステムの変更があり得ると考えたこと、さらに従来のバージョン（GENESIS 1.X）に導入されている機能の全てがこのバージョンでサポートされているわけではないことである。GENESIS 2.0betaでは、「富岳」用に高速化されたSPDYN（空間分割コード）のみが利用できる。実際、多くの「富岳」ユーザーがすでにこのプログラムを利用している。GENESIS 1.XにはSPDYN以外に粗視化MDやQM/MM計算が可能なATDYNと様々な解析プログラム群が含まれている。GENESIS 1.XのSPDYNでは自由エネルギー摂動計算（FEP: Free-energy perturbation）のモジュールが含まれているが、GENESIS 2.0betaのSPDYNではFEPはまだサポートできていない。理研を中心としたGENESIS開発者は、非共有結合相互作用や逆空間相互作用など律速となる主要なエネルギー計算の高度化と並列化を進めるとともに、自由エネルギー計算・粗視化MD・QM/MM計算などを実現する高機能化を、並行して進めている。全ての開発の同期をとることは事実上不可能であり、この数年間はGENESIS 1.XとGENESIS 2.0betaの両方についての開発・利用が必要である。この複雑さを解消するためにはGENESIS 1.Xに含まれる開発をある時点で、全て2.0betaに移行する必要がある。

GENESIS 2.0betaの高速化は様々なアルゴリズムの改良の積み重ねとして実現した。図1-1-1に示すようにMDシミュレーションの計算時間の主要な部分を含む非共有結合相互作用を「富岳」に最適化することは最も重要なポイントであった。そのため、我々はこの部分のプログラムを書き直し、繰り返し長を伸ばすことで、「富岳」のCPUを有効に活用できるようになった。

さらに、MDシミュレーションで用いるニュートンの運動方程式を逐次的に積分する際に必要な時間刻み（Δt）を伸ばしても正確な軌跡を得るアルゴリズム開発にも成功した。ここで我々は従来のMDプログラムでは系の運動エネルギーを記述する方法が二つ（Full-time step kinetic energyとHalf-time step kinetic energy）存在していたことに着目し、これらの理論式に立ち戻って考察した。図1-1-2に示すように、いずれの式もΔtのオーダーで正確な温度を定義しているが、この低い次数が問題となり、大きな時間刻みを用いた場合に運動方程式を正確に記述できなかった。そこで我々は、Full-time step kinetic energyとHalf-time kinetic energyに係数をかけて足し合わせる事で、Δtの3乗のオーダーまで正確な温度を計算する運動エネルギー（Optimized kinetic energy）を定義できることを発見した。実際、このようにして得られた温度や圧力を用いて運動方程式を解く際には従来のΔt（= 2〜2.5fs）の倍くらいまで（5.0fs）伸ばすことができる。この解法はすでにGENESIS 2.0betaに導入されており、適切なパラメタを選択することで利用可能である。

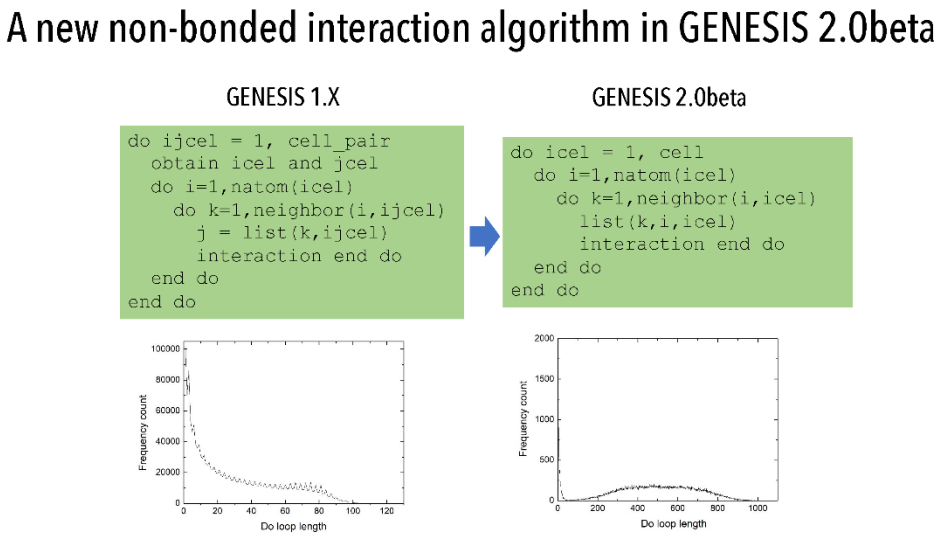


図1-1-1: GENESISにおける非共有結合相互作用計算のアルゴリズム。
従来の計算法（1.X）と新しく「富岳」に最適化された計算法（2.0beta）のアルゴリズムとループ長の比較。Apoa1（約10万原子系）を用いて比較した。

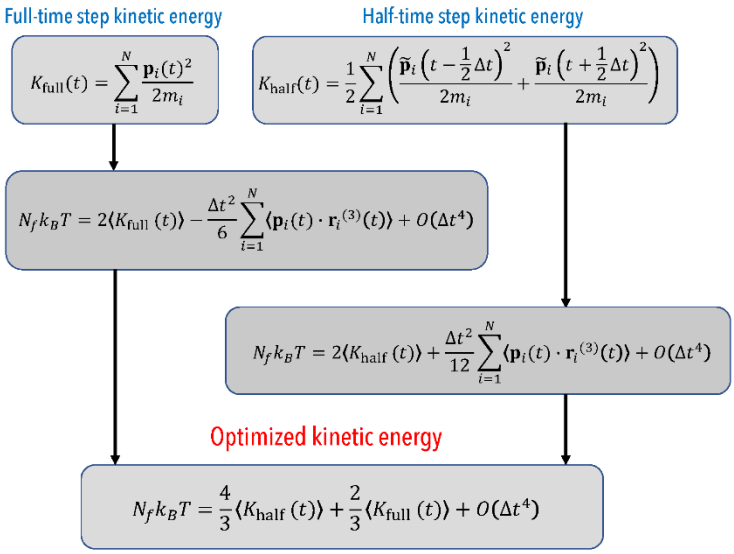


図1-1-2: MDシミュレーションにおける運動エネルギーの定義。
従来の定義（Full-time step kinetic energyとHalf-time step kinetic energy）と、これらを用いて最適化されたOptimized kinetic energyの関係。

「富岳」の試験利用が令和2年度から可能になったため、我々はGENESIS2.0betaを用いて「富岳」上でベンチマークテストを行った。ここで「富岳」の多数の演算ノードを有効に使うために、約1億原子（101.05M）、8億原子（808.43M）、10億原子（1.1511B）を含む非常に大きな生体分子システムについて計算した。この生体分子システムは我々が「京」を用いて行ったバクテリア細胞の分子混雑系と基本

的には同じものであるが、新たにモデリングをやり直し、従来より大きな系（「京」で実施した最大のMDは1億原子系）のシミュレーションを実現した。「富岳」の全系を用いたMDシミュレーションを有効に行うためには、さらに大きな生体分子システムをターゲットにすることが望ましいが、10億を超える要素を含む配列を扱うためにはプログラム全体の大きな変更が必要となるため、約10億原子を最大のターゲットシステムとした。GENESIS 2.0betaのベンチマークの結果、約1億原子系について「京」と比べて「富岳」を利用することで約1/8のノード数でも4倍以上の性能向上を示していることがわかった（図1-1-3）。また、10億原子系についても「富岳」では11.9ns/dayの計算効率を示しており、非常に大きな生体分子システムのMDシミュレーションを行う際にGENESISは世界最高水準の演算性能を示していることがわかった。

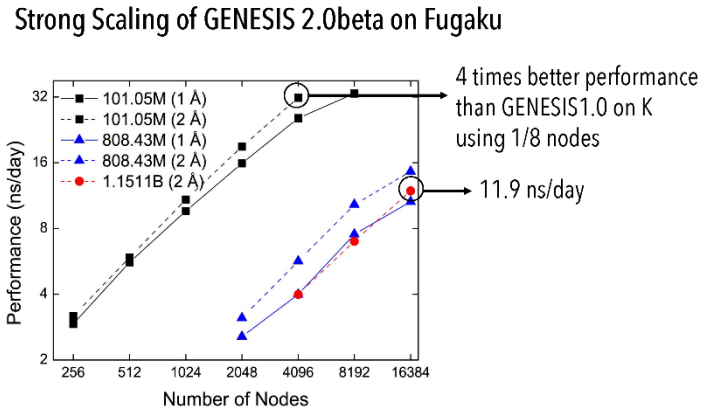
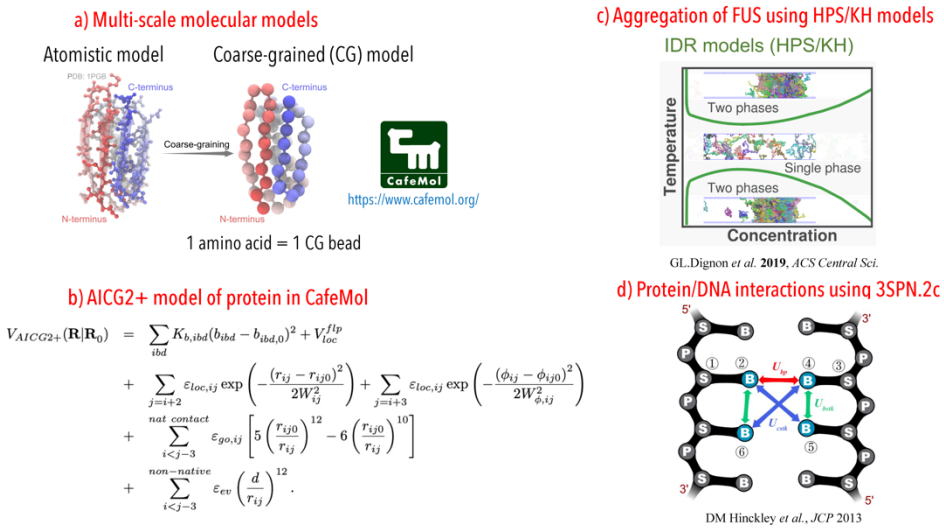


図1-1-3: GENESIS2.0betaを用いて実施した「富岳」上での巨大生体分子系のMDシミュレーションのベンチマーク結果。粒子数が増えた場合も効率的な計算ができていることを示す。

全原子MDシミュレーションに加えて、我々は粗視化MDについての高速化・並列化等の開発も行った。このセクションでは、高田彰二（京都大学）らが長年開発してきたCafeMolをGENESISに導入するために行った共同研究についてまとめる。高田らが開発してきた粗視化分子モデルCafeMolでは、タンパク質について各アミノ酸を1つの粗視化された粒子として表現している。他の生体高分子、例えば、核酸や脂質分子についても大体同じような割合で粒子数の削減ができるように粗視化粒子が定義されている（図1-1-4）。粒子を粗視化することによって失われた情報を補うために、CafeMolのタンパク質モデル（AICG2+）においてはタンパク質立体構造から得られる情報を含むエネルギー関数(Structural based energy function)が用いられている。このエネルギー関数は、郷モデル（あるいはFunnel theoryなど）と呼ばれる統計力学的考察にその理論的基盤を持つ。図1-1-4に示すように結合長、結合角、二面角などの内部自由度を記述するエネルギー項に加えて、Native Contactあるいは Non-Native Contactと呼ばれる非共有結合相互作用を含む。Native Contactはタンパク質の天然構造で見られるアミノ酸間の接触ペアについてカウントし、Non-Native Contactはそれ以外のアミノ酸間に働く排除体積項である。この近似の妥当性は過去20年以上も多くの理論生物物理学者によって様々な検討がなされ、多くの生体分子系において有効であることが示されている。特に、高田（京都大学）らはこの方法を用いて、タンパク質折れ畳み、タンパク質の立体構造変化、分子モーターなどの分子シミュレーションにおいて多く

の成果をあげてきた。彼らは、このモデルをタンパク質・核酸複合体あるいはタンパク質・核酸の相互作用の解析に適用するために3SPN.2cと呼ばれる核酸の粗視化分子モデルと併用し、ヌクレオソーム・クロマチン、さらに転写因子とDNAの相互作用などシミュレーションを行ってきた。

CafeMolは高田（京大）らが開発してきた粗視化分子モデルの総称であるとともに、そのモデルを用いてMDシミュレーションを行うプログラムの名前でもある。我々は粗視化分子モデルとしてのCafeMolをGENESISで利用可能にするために、高田チームと密接な議論を重ね、さらに京大から理研に異動したC.Tan博士が中心となって、GENESIS/ATDYNにこのモデルを導入した。最初に問題となったのはCafeMolに含まれるエネルギー関数やパラメタをどのようにしてGENESISの計算に取り込むかという手続きであった。粗視化分子モデルであっても非常に多くのパラメタを含むため、それを正確に簡単に入力できる仕組みを新たに作ることは容易ではない。試行錯誤の結果、GROMACSと同じ入力ファイル形式を用いることで、CafeMolに含まれる全てのエネルギー項を記述できることがわかり、その方式で統一した。



$$
\begin{aligned}
V_{AICG2+}(\mathbf{R}|\mathbf{R}_0) = & \sum_{ibd} K_{b,ibd}(b_{ibd} - b_{ibd,0})^2 + V_{loc}^{flp} \\
& + \sum_{j=i+2} \varepsilon_{loc,ij} \exp\left(-\frac{(r_{ij} - r_{ij0})^2}{2W_{ij}^2}\right) + \sum_{j=i+3} \varepsilon_{loc,ij} \exp\left(-\frac{(\phi_{ij} - \phi_{ij0})^2}{2W_{\phi,ij}^2}\right) \\
& + \sum_{i<j-3}^{nat\ contact} \varepsilon_{go,ij} \left[5\left(\frac{r_{ij0}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{ij0}}{r_{ij}}\right)^{10}\right] \\
& + \sum_{i<j-3}^{non-native} \varepsilon_{ev} \left(\frac{d}{r_{ij}}\right)^{12}.
\end{aligned}
$$

図1-1-4：粗視化分子モデルの説明。a)　全原子モデルと粗視化モデルでの粒子の取り扱いの違い。b）CafeMolのタンパク質モデルAICG2+で用いられているエネルギー関数。c）天然変性タンパク質を扱う粗視化分子モデルHPS/KH。d）核酸に適用される粗視化分子モデル3SPN.2c。

我々はさらに、天然変性タンパク質が多く含まれる細胞内液滴（膜のないオルガネラ、液液相分離）のモデリングに使われている分子間相互作用（HPS/KHモデル）を、GENESIS/ATDYNで利用可能とした（図1-1-4c））。このように、GENESIS/ATDYNを用いることで、タンパク質単体のみならず、タンパク質・核酸、天然変性タンパク質とその相互作用まで様々な生体高分子を含む系の粗視化MDシミュレーションが可能である。しかし、GENESIS/ATDYNでは空間分割を行なっていないため、並列計算を行う場合も各ノードで全ての粒子の座標等を保持しなければならない。そのため、必要とするメモリが非常に多いこと、並列化率に限界があることなど制限もあった。これらの問題はプログラムコードを改良することで、図1-1-5に示すようにある程度解決し、Intel製計算機などメモリを比較的多く積んでいる計算機（64GB以上）では、100万粒子以上を含む大きな系の粗視化MDシミュレーションが可能となった。また、その並列化効率においてもCafeMolを上回るパフォーマンスを示している。一方、「富岳」は各ノードあたりのメ

モリが48GBという制限があり、GENESIS/ATDYNでは十分に大きな生体分子系の粗視化MDシミュレーションを実行できない。そこで、空間分割を行うことで粗視化MDシミュレーションに必要なメモリ量を減少させ、並列化効率を引き上げることのできる新たなMDコードを開発した。このコード（GENESIS/CGDYN）では全原子MDを行うSPDYNとは異なる並列化アルゴリズムが導入されており、将来的に様々な粗視化分子モデルを用いたMDシミュレーションをサポートしたい。CGDYNの基盤はすでに構築でき、良好な並列化率を示しており、令和3年度には巨大な生体分子系の粗視化MDシミュレーションを「富岳」上で実行できる。

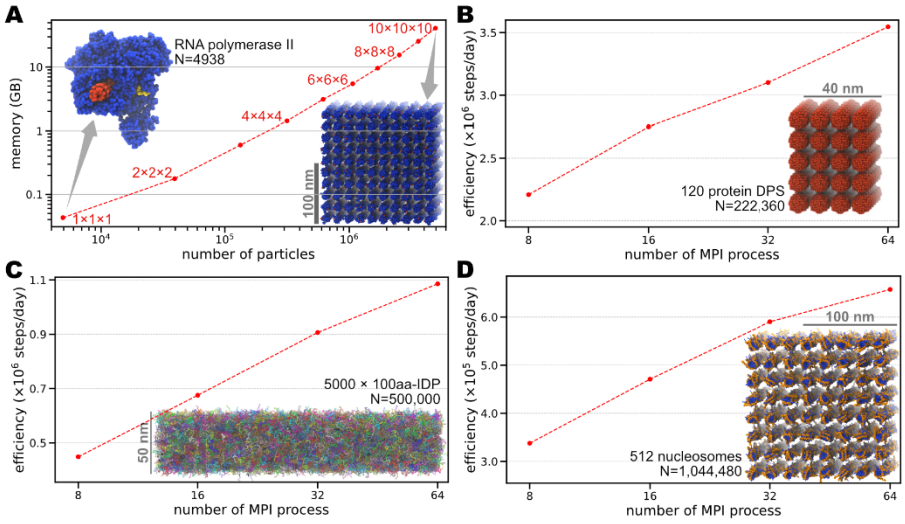


図1-1-5: GENESIS/ATDYNに導入されたAICG2+を用いたベンチマークテストの結果。
A. RNA polymerase （4,938 CG粒子）の数を増やすことによってATDYNで扱うことのできる最大の粒子数の検討。B. 222,360 CG粒子、C. 500,000 CG粒子、D. 1,044,480 CG粒子を含む生体分子系での並列化効率の検討。

我々は、河野（量研機構）チームと共同で、彼らがこれまでに「京」を用いたヌクレオソームの計算に用いてきたABMD（Adaptive Biased Molecular Dynamics）法をGENESISに移植した。通常のMDでは系が自由エネルギー面上の局所点に拘束されてしまうことがあるが、ABMD法では反応座標（Collective Variable: CV）に依存したバイアスポテンシャルを追加することで自由エネルギー面を均一にし、系が局所点に留まることなく幅広いサンプリングが可能となる。ABMD法はメタダイナミクス（Metadynamics: MetaD）法と似ているが、MetaD法ではバイアスポテンシャルをガウス関数型ポテンシャルの足し算で表現し、ある緩和時間ごとに新たにガウス関数を追加していく。そのため計算時間に比例して足し合わせるガウス関数の個数が増えていき、計算負荷も増大するという問題がある。一方、ABMD法ではバイアスポテンシャルをB-spline関数の線形結合で表現し、緩和時間ごとに結合係数のみを更新するため、MetaDと比較し計算時間にポテンシャル計算の負荷が依存しないことが大きな利点である。また、複数のレプリカでABMD法を行うことでバイアスポテンシャルを並列に更新する“Multiple Walker ABMD”やABMD法と温度レプリカ交換法を組み合わせた“Replica-Exchange ABMD”などを用いることで、さらにサンプリング効率を向上できる。現在GENESIS 2.0betaにABMD、“Multiple Walker ABMD”、“Replica-Exchange

ABMD”の導入が済んでおり、Alanine-dipeptideとChignolinを用いたテスト計算を行なっている。Alanine-dipeptideにおいては図1-1-6に示す通り、AMBER16を用いて行われたABMD計算と非常によい一致を示した。また、Chignolinにおけるテスト計算においては、Gaussian Accelerated Replica-Exchange Umbrella Sampling (GaREUS) と計算結果[H.Oshima et al.,*J.Chem.Theory Comput.*15, 5199-5208 (2019)]の比較を行う予定である。

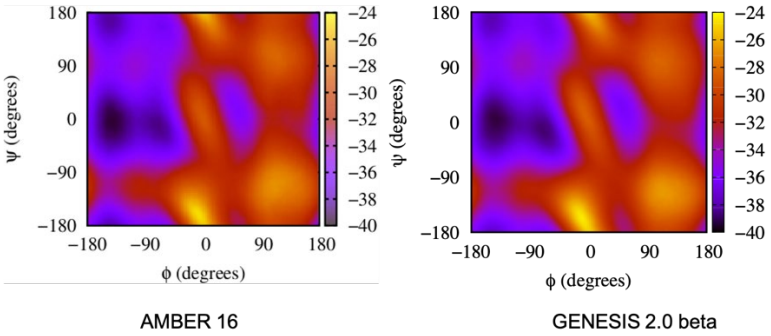


図 1-1-6: GENESIS2.0beta および AMBER16 を用いて Alanine-dipeptide に対して実施した ABMD 計算の結果。ϕ、ψの 2 次元自由エネルギー面がよく一致していることを示す。

② 粗視化モデル SPICA-FF の開発（名大・篠田）

SPICA力場は基本的に重原子3つを一つの粗視化粒子と見なす程度の解像度を持つ粗視化力場であり(図1-2-1)、そのパラメータ決定は、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ータをコンビネーションルールによって求めることをしないため、パラメータ構築に時間がかかる。そのため、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ことが大きなメリットである。今年度は、イオン性のヘッドグループ(PS:phosphatidylserine, PI:phosphatidylinositol, PG:phosphatidylglycerol)を持つ脂質のモデリング(図1-2-1)を行い、分子ライブラリに追加した。ここでは、PSをヘッドグループに持つ、POPS (palmitoyl oleoyl PS)の計算結果を一部報告する。

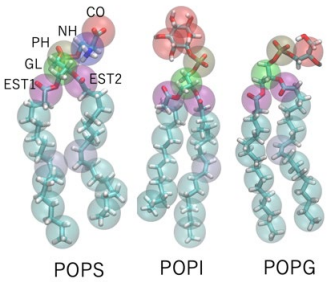


図1-2-1: POPS, POPI, POPGの粗視化モデル(球)。対応する全原子モデルはstickで表示している。各粒子の定義はCO:$-CO_2^-$, NH:$-CH(NH_3^+)CH_2-$, PH:$-PO_4^--$, GL:$-CH_2CH-CH_2-$, EST1/EST2: $-CCO_2-$。

参照データはCHARMM36力場による全原子(AA) MDシミュレーション（323K, 300ns）を行って、分布関数を得た。まず、図1-2-2に粗視化(CG)とAAMDで得られた膜面積データを示す。CGMDの平均値は60.8Å^{2}でありAAMDの結果（60.6 Å^2）をよく再現しているため、正しい界面張力の記述がSPICAでも実現していることがわかる。精度の高い実験値が利用できる場合は、さらにそれを利用して、CGパラメータを作成することもできる。膜面積の一致は、膜中での脂質疎水鎖のパッキングがAAMDとCGMDで類似していることを意味しており、疎水鎖の秩序度を表すオーダーパラメータも、AAMDとCGMDで非常によい一致を示した(図1-2-3)。POPSにはオレオイル基にcis-の二重結合があり、その位置でオーダーが下がることが知られているが、その振る舞いもSPICAで良く再現できている。

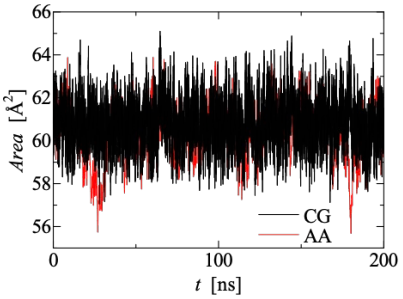


図1-2-2: POPS膜の膜面積変化。
全原子（AA）と粗視化（CG）モデルを用いて行ったMDシミュレーションの結果を比較した。

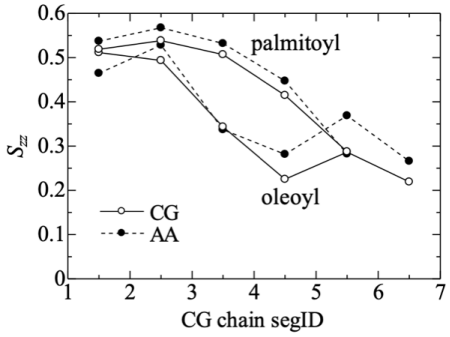


図1-2-3: 疎水鎖のオーダーパラメータ。
全原子（AA）と粗視化（CG）モデルを用いて行ったMDシミュレーションの結果を比較した。

膜垂直軸に沿っての各粗視化セグメントの配置分布についてもCGMDとAAMDの結果はよく一致していた(図1-2-4)。脂質膜の厚みはリン酸基間の距離で測定されるため、膜の厚みも非常に良く再現されている。SPICAでは水3分子を一つの粒子と見立てた粗視化粒子として扱っているため、水のサイズ効果がその分布に多少影響を与えている。また、POPSのカウンターイオンとして導入したNa^+(SOD)の分布も、AAMDをほぼ再現する分布をCGMDで得ており、電気二重層の構造も非常に類似している。SPICAでは静電相互作用を通常のクーロン力の形でカットオフ無しに計算しているため、このようなイオン分布も忠実に再現できる。さらに、膜の弾性係数のような物性も一致することを確認できた。これにより、大規模膜系

に対してCGモデルを用いて計算するのに十分な信頼性があるという確認ができた。

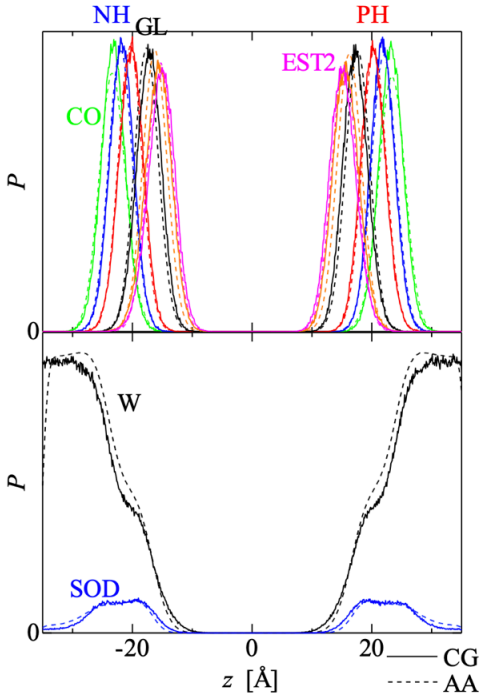


図1-2-4：膜垂直軸に沿っての各粗視化セグメントの分布。脂質のセグメント名は図1-2-1に対応している。SODは水和ナトリウムイオン($Na^+(H_2O)_3$)、Wは水($(H_2O)_3$)である。
全原子（AA）と粗視化（CG）モデルを用いて行ったMDシミュレーションの結果を比較した。

先に述べたとおり、異種粒子間の非共有結合相互作用(LJ関数で記述される)パラメータは、SPICAにおいて、クロス項それぞれについて決める必要があり非常に手間がかかる。現在、ニューラルネットワークを用いた、クロス項の予測ツール(MLベース)を開発中であり、すでに一部の相互作用パラメータはニューラルネットワークで決定した。例えば、タンパク質とコレステロール間相互作用は、この方法で決めたパラメータを用いており、現在までに大きな問題は検出されていない。新しく作成した脂質間相互作用についてもこの手法を適用している段階であり、今後、AAモデルとの比較により、その精度を評価する。

脂質分子間相互作用を記述するパラメータ最適化は、これまで特にコレステロール(Chol)やスフィンゴミエリン(SM)脂質を含む膜系について行ってきた。これらの脂質分子の膜面内相分離挙動について、実験結果を非常に良く再現する結果を得ている。特に、令和2年度は、SMとCholが誘起する相分離構造の脂質二重層間のカップリングに関する大規模MDシミュレーションを行った。SMとChol以外の脂質分子を含む三元系において、SMとCholは多くの場合、よりオーダーの高いLo(liquid order)相を作り、Ld(liquid disorder)相と相分離挙動を示す。このような挙動は実験結果と良く一致するが、脂質二重層膜の片側リーフレットのLo相が他のリーフレットの構造や脂質分布にどのような影響を与えるのかは、いまだに良く理解されていない。SPICA力場を用いたCGMDシミュレーションから、一方のリーフレットで生じたLo相は他方リーフレットの脂質分布や脂質オーダーに影響を与えることが見いだされた。

図1-2-5には、DOPC(dioleoyl phosphatidylcholine)とPSM (palmitoyl SM)、Cholの三元系の計算結果を示す。PSMは上層のみに導入されて4μsのMDシミュレーションを行った結果である。DOPC/Cholだけでは相分離は起こさない。すなわち、上層にPSMが導入されたことにより、上層で相分離が起こるが、その相分離によって生じたLo相が下層に影響を及ぼし、Cholの分布を増やしていることがわかる。その結果、オーダーパラメータも当該領域で高くなっていることが図1-2-5からわかり、脂質リーフレット間の相関が明確に確認されている。このような相分離を観察するためにはCGMDでも数μsが必要であり、実際の物理時間ではこの観測に数十μs程度必要であると見積もられている。上記の現象は、一方の層の堅いLo相が他方の層にLo相を誘起する例であるが、この挙動は長鎖SM (C24)では大きく変わり、長鎖SMによるLo相は他方の層にLd相を誘起することが見いだされている。図1-2-6に非対称膜において、SMの疎水鎖の長さが反対側の脂質リーフレットのドメイン構造に与える影響を示す。LSM系では上層がLo相ドメインに対して下層がLd相に近い、anti-registeredな相分離を起こすことがCGMD計算から予見されている。疎水鎖の長さだけで大きく層間相互作用が変わることを示唆しており、長鎖SMが細胞において膜間の情報伝達を担う可能性が議論されている。SPICA力場は、タンパク質や核酸に対しても拡張中であり、phosphatidylcholine (PC)脂質との相互作用については、ほぼ完成している。

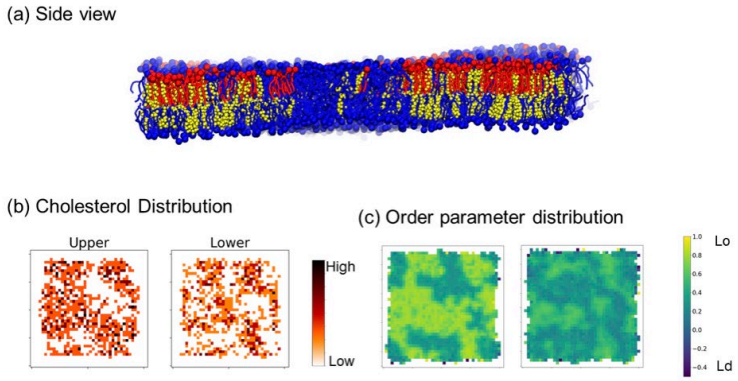


図1-2-5: PSM(赤)/DOPC(青)/Chol(黄)のMDからの(a)スナップショット、 (b)上下脂質層における Cholの分布、(c) 上下層における脂質のオーダーパラメータ。

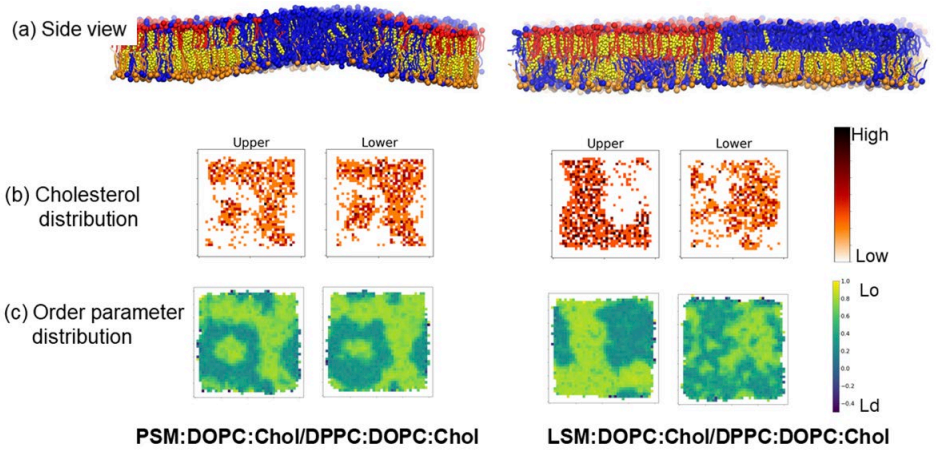


図1-2-6: 上下層とも相分離する膜の層間相互作用。PSM系とLSM系の比較。
(a)側方スナップショット、(b) 上下層のChol分布、 (c) オーダーパラメータ。

また、SPICA 力場のMDソフトウェアGENESISへの組み込みと、「富岳」での最適化・並列化を杉田（理研）チームの協力の下で遂行中である。今年度は、SPICA力場をGENESISへと導入し、Intel CPUの計算機と「富岳」を用いて、LAMMPSを用いた計算結果を再現することを確認した。SPICA力場と全原子CHARMM力場のエネルギー関数における大きな違いは以下の2点である。(1) CHARMM力場では、LJ関数として通常の12-6型が用いられているが、SPICAでは分子種に応じて12-4 、9-6 型の２種類が使われている。

$$U_{LJ12-4}\left(r_{ij}\right) = \frac{3\sqrt{3}}{2}\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{4}\right] \quad \text{water} - \text{water/ water} - \text{other}$$

$$U_{LJ9-6}\left(r_{ij}\right) = \frac{27}{4}\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{9} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad \text{others}$$

(2) SPICAでは、分子内相互作用のangle項の補正として、LJ9-6関数のエネルギー項が計算される。

$$U_{1-3}^{correct} = \sum_{angle} U_{LJ}\left(r_{ij}\right) - U_{LJ}\left(r_{ij} = r_{min}\right) \quad \text{for } r_{ij} < r_{min}$$

すでに述べたようにSPICAでは、長距離クーロン力の計算はカットオフ無しの手法を使用するためParticle Mesh Ewald法が利用できるGENESIS/SPDYNに組み込んだ。

SPICA力場が正しくGENESIS/SPDYNに移植できたかを判定するため、表1-2-1に示すような計算条件で、DMPC脂質二重膜系(図1-2-7)のMDシミュレーションにより膜物性の評価を行った。すでに実績のあるLAMMPSでのMDシミュレーション結果を参照とした。まず、各スナップショットのポテンシャルエネルギー値をGENESISとLAMMPSで比較すると数値精度内で完全に一致することを確認した。さらにそれぞれのプログラムで、MDシミュレーションを実行することにより、物性量は統計誤差内で一致した。例えば、膜面積の値は65.7±1.2Å^2 (LAMMPS)、64.9±0.4Å^2 (GENESIS)とよく一致しており、荷電粒子を含むヘッドグループの角度分布など、静電相互作用の見積もりに最も敏感に変わり得る構造に関しても非常によい再現性が得られた。

Item	Condition
Total Number of atoms	83174
Total number of Water molecules	64741
Total number of POPC molecules	1152
Box size [Å]	193 193 194
Cutoff [Å]	15
Integrator	velocity Verlet
Times step size	10 fs
Temperature Control	Langevin
Temperature	303 K
Pressure isotropy	Semi-iso
Pressure	1 atm

表1-2-1: DMPC膜のMDシミュレーションの条件。

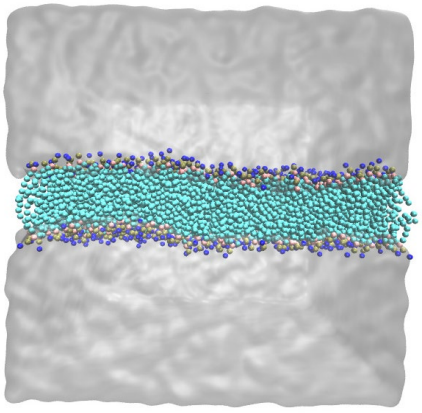


図1-2-7: DMPC脂質膜系のスナップショット。灰色は水分子領域、脂質分子は球で表示。

SPICA力場を用いる場合、タンパク質の二次構造をElastic network model(ENM)を用いて固定する必要があるため、GENESISでENMモデルを取り扱えるように拡張を行った。ENMモデルでは、構造を保つために参照構造の中で、あるカットオフ距離内にある原子間の距離をバネポテンシャルにより拘束する。小さなタンパク質を埋め込んだ脂質膜系をベンチマーク系として選び、GENESISとLAMMPS を用いて得られたスナップショットのエネルギー値を比較した。タンパク質の分子間相互作用とVDW相互作用は、単精度の範囲である7桁程度まで完全に一致した。

「富岳」への並列計算と最適化の前に、まずIntel系のクラスター計算機(九州大学の計算機センター)での計算速度を計測した。ここで示す結果は、「富岳」でのチューニングを行う前のベンチマーク計算結果である。まずMPI の並列性能を確認した。GENESISにおいては、使用できるMPIのプロセス数は、系のサイズに比例しており、特に系を空間分割した時のドメインの個数が最大のプロセス数となる。ドメインの最小サイズは、カットオフ距離によって決まる。よって小さな系では使用できるMPIの個数が大規模系よりも小さくなることを注記する。その上で、タンパク質と脂質二重膜からなる27万原子系で、タイムステップを10fsとしてカットオフ距離を1.5nmとしたときに、NPTアンサンブルの条件下で計算した時の計算速度は、144MPIで27ns/day, 288MPIで50ns/day, 576MPIで 84ns/day、1152MPIで 104ns/dayとなった。この系では1152MPI がおおよそ最大の空間分割と等しくなったため、これ以上MPI並列数を増やすことで高速化を行うことは困難である。計測した数値から、この小規模系では弱い線形の並列性能のスケーリングが最大コア数まで確認できている。

今回計測した場合の最大サイズの系である640万粒子系(図1-2-8) について、5832MPIと6 thread(OpenMP)でのハイブリッド並列(34,992コア)した場合に「富岳」上で計測した演算性能は、101.8 ns/dayであった。このことから、本プロジェクトで標的としているウイルスエンベロープを包含した系に関しても、「富岳」上で10日間計算を実行すると、1μsのMDシミュレーションが実行できることがわかつた。

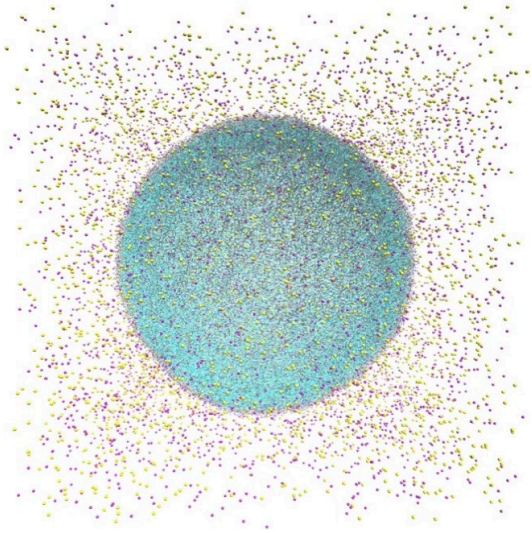


図1-2-8: 640万粒子のベシクル系。
水は割愛しており、中央にベシクル、周囲の球はNa^+とCl^-を表示している。

（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が、用いる力場パラメー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。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、小タンパク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タンパク質へ応用可能にすることが求められている。

今年度は、MD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、より大きなサイズのタンパク質へのデータ同化手法応用の準備として、Protein Gタンパク質のフォールディングを解析するための追加構造サンプリングを主に「富岳」を用いて行った。

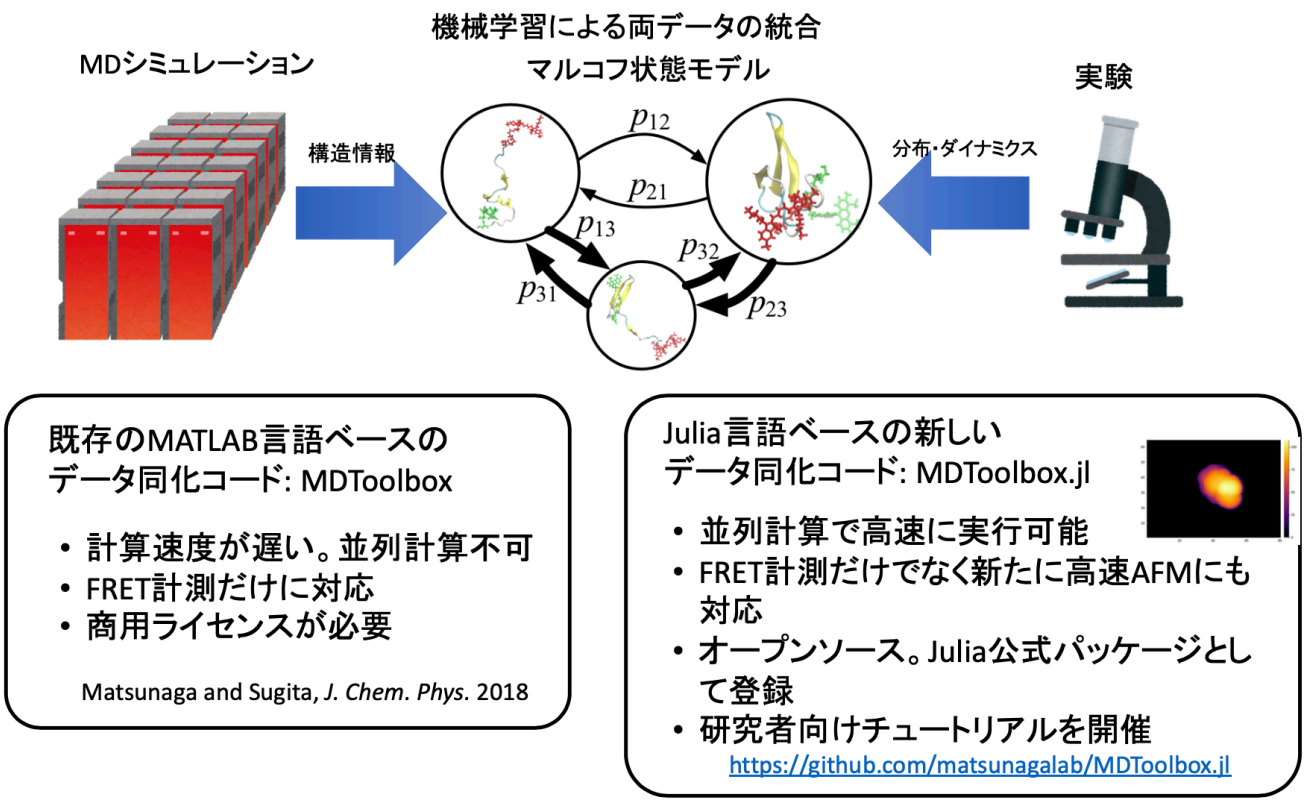


図2-1-1: 生体分子データ同化パッケージの開発。

これまでに我々が開発してきたデータ同化のためのコードはライセンスを必要とするMATLAB言語で書かれており、一般の研究者だけでなく本プログラムに参加しているMD計算を行う研究者も容易に使える環境ではなかった。更に、データ同化のベースとしているマルコフ状態モデルの構築に必要な計算が網羅されておらず、一部の計算を他のPythonパッケージで行ってマルコフ状態モデルを構築し、そのデータを変換してMALTABコードへ読み込ませるなど行う必要があった。そこで、本プログラムへ参加しているMD研究者を含め広くデータ同化を応用できるようにMDデータの読み込みからマルコフ状態の構築、その後の実験計測データと組み合わせたデータ同化までを一つのパッケージ内でシームレスに実行できる環境として、Julia言語をベースとした新たなパッケージ MDToolbox.jl を開発して公開した https://github.com/matsunagalab/MDToolbox.jl 。

一般に、MDデータからマルコフ状態モデルを構築し、データ同化計算を行うには以下の機能が必要となる。

1. MD データの IO: PDB、PSF、DCD、NetCDF といったデータフォーマットを読み込む
2. 特徴量の抽出: 距離マップやコンタクトマップ、さらにその後の次元縮約(主成分解析や、緩和モード/独立成分解析)による特徴量抽出
3. マルコフ状態定義: k-mean や k-centers クラスタリングによる特徴量のクラスタリング
4. 遷移確率の推定: カウントマトリックスの計算と、エルゴード性のチェック。カウントマトリックスからの最尤法による遷移確率推定
5. マルコフ性の検証: 様々な lag-time を用いて推定した遷移確率行列からの implied-timescale の計算

6. 実験計測データを用いた隠れマルコフモデリング: Viterbi アルゴリズム、Baum-Welch アルゴリズム

以上の機能を全て新たなパッケージMDToolbox.jlへ実装し、パッケージ内でMDデータの読み込みからデータ同化までを実行できるようにした。データ同化のための英語ドキュメントとサンプルファイルを整備して、MDを行う研究者が実行できるチュートリアルを作成し、本プログラムの研究者向けにハンズオンチュートリアルを開催した。また、開発したパッケージ自体の英語Referenceマニュアルを整備し、Juliaの公式パッケージとして申請し、承認された。

次に実験計測データとして、FRET計測以外の計測データへ広く対応させるために、開発したパッケージを用いて高速原子間力顕微鏡(High-Speed Atomic Force Microscopy; HS-AFM)データへのデータ同化手法の研究開発を行った。AFMデータは、分子表面の凸凹を捉えた2次元表面形状データである。データ同化を行って精度の良いマルコフ状態モデルを構築するためには、2次元表面形状と、MDシミュレーションで探索された3次元構造との類似度を測る指標が必要である。そこでクライオ電子顕微鏡の画像解析(Cossio and Hummer, *J. Struct. Biol.* 2013)で提案された類似度をAFM向けに修正してMDToolbox.jlへ導入した。類似度を使って元の3次元構造を推定するには、未知パラメータとして、構造の種類、構造の並進、向き(3次元回転)、短針の半径、高さのスケーリングパラメータ、ステージ底面の高さ(オフセット)、ノイズの分散、などについて積分しなければならない。こちらもクライオ電子顕微鏡の画像解析の先行研究の知見を元に、一部のパラメータについては解析的に積分してしまって、残りの構造の並進や向きなどについて数値的に積分するプログラムを開発した。しかし、数値積分するパラメータ数を減らしても、それでも次元が大きく、通常の計算では100フレームの解析におよそ2日かかってしまったため、数値積分を並列化して実行できるように開発した。その結果、2日かかっていた計算がおよそ2時間足らずで終了することができた。

それぞれのAFM画像フレームからの3次元構造推定には、類似度が最大となる構造・並進・向きを選ぶ以下の最尤法の式をベースに推定を行った。

$$\arg \max_{m_t, \phi_t, dx_t, dy_t} = L(I_t \mid m_t, \phi_t, dx_t, dy_t)$$

ここで、 L はAFM画像と3次元構造との類似度、I_t はtフレーム目のAFM像、m_t, ϕ_t, dx_t, dy_t はそれぞれ構造、向き、x 座標と y 座標の並進である。テストデータを用いてこの推定方法の精度を評価した。テストデータを作成するために、味覚レセプターの粗視化モデルを使って50個の構造のうちからランダムに選び、それを更にランダムに並進・回転させた上で、新稲ら(Niina *et al.*, *JCTC* 2019)によって開発された擬似AFM画像生成手法で構造からAFM画像を生成し、最後にガウスノイズを加えるということを繰り返し行い、合計1000フレームから成る擬似AFMデータを作成した。テストデータの各フレームのAFM画像がどの3次元構造から生成されたかを上記の手法を用いて推定した。その結果、1000フレームのうち980フレームで正しい構造・並進・向きを推定できることがわかった。

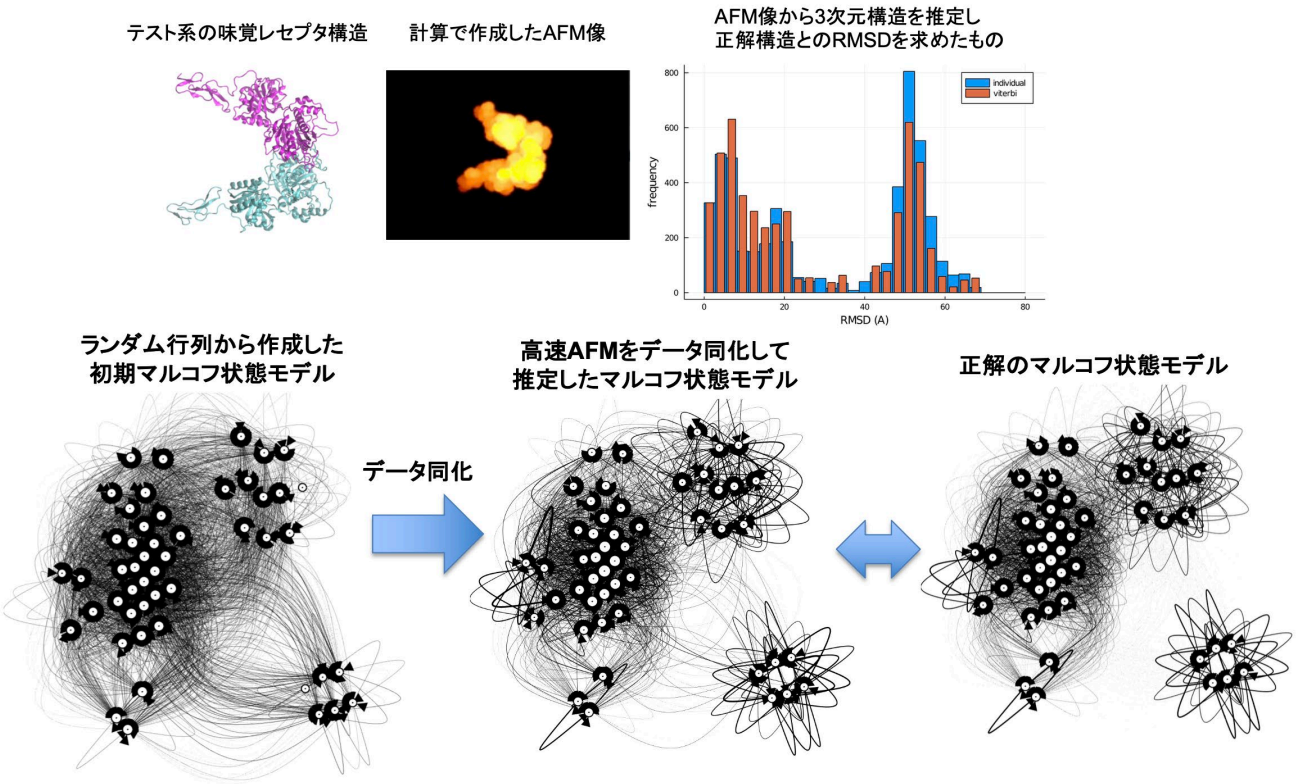


図2-1-2: テストデータを用いた高速AFMデータ同化の検証結果。

AFM画像は短針の半径に強く依存する一方で、計測中に短針の形状を知ることは難しい。そこで我々は構造推定において短針形状の見積もり誤差が構造推定に対してどの程度影響を与えるのか調べるために、テストデータを生成するときの短針半径と、類似度を計算するときの短針半径を意図的に違うものにして、構造推定への影響を調べた。その結果、テストデータ生成と類似度計算の際の半径が一致する場合には980/1000で高精度に正しい構造・並進・向きを推定できていたが、両者に5Åの齟齬がある場合には、特に向きの推定において大きな誤差が生じることがわかった。図2-1-2の右上の青のヒストグラムは、5Åの齟齬がある場合の推定された構造・並進・向きの正解からのRMSDの頻度を示している。テスト系に用いた味覚レセプターはヘテロダイマーであるが、両モノマーの構造は似ており、AFM画像のみからモノマーを区別することが難しい。それでも短針半径が一致する場合は高精度でモノマーを区別できていたが、5Åの齟齬がある場合には両モノマーを区別できずに、RMSDにおいて2つのピークをもつ分布が出た。RMSDが大きい方のピークは向きの推定が180度間違っているために生じる結果である。5Åの齟齬がある場合には、およそ半分のフレームにおいてモノマーが区別できていないことがわかる。

実際のAFMデータへ応用する際には、短針形状の見積もり誤差はどうしても排除することができない。したがって短針形状に誤差があってもロバストに構造・並進・向きを推定する手法が求められる。そこで我々は、1つのフレームだけではなく複数のフレームを用いて構造・並進・向きとの類似度を評価して推定する手法を提案した。複数フレーム(ステップ1～T)を用いた場合の類似度は以下のように定式化される。

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

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。遷移確率は一般には事前に知ることは不可能であるが、ここでは事前に知っているものとして遷移確率を使って構造・向きのデータを作成して、そこからAFM画像の時系列として1000フレームから成るテストデータを新たに作成した。また類似度を評価するときにもテストデータを作成した時と同じ遷移確率を利用した。上記の複数フレームを用いた類似度の最大化にはViterbiアルゴリズムを用いた。そのようにして、従来の1フレームだけを用いて構造・並進・向きを推定した場合と、複数フレームを用いて構造・並進・向きを推定した場合を比べた結果が、図2-1-2の右上の青とオレンジである。複数フレームを用いた場合、有意にRMSDが大きい側のピークが下がっており、これは味覚レセプターのモノマーを区別する精度が向上したことを意味する。これは1フレーム目だけで推定すると推定値の分散が大きい一方で、複数フレームを用いて推定すると分散を下げることができるアンサンブル学習と似た効果が起こっていると予想される。

次に、遷移確率がわかっていないものとして、遷移確率も含めてデータから推定して、1フレームを使った推定に比べて構造・並進・向きの推定が向上するかどうかを調べた。遷移確率の推定にはBaum-Welchアルゴリズムを用いた。その結果、遷移確率が既知とした場合と同様にして、有意にRMSDが大きい側のピークの減少が観測された。ただし、遷移確率が既知の場合と比べてはピークの下り幅は減少した。これは遷移確率の推定に誤差があるためであると思われる。このようにして、データのみから1フレーム毎の推定よりもロバストな構造・並進・向きの推定方法を開発し評価した。

最後に、遷移確率の推定精度を評価するために、短針形状に齟齬がある場合とない場合の推定を行なった。その結果、短針形状が一致している場合には正解の遷移確率(マルコフ状態モデル)に非常に近い結果を得ることができた(図2-1-2下側)。一方で、短針形状に齟齬がある場合には遷移確率の誤差も大きく、短針形状の誤差の影響が確認された。今後は短針形状の推定の精度向上も合わせて手法を改善していくことが求められる。

応用研究として、次年度のデータ同化応用の準備として、Protein Gタンパク質の不安定領域における構造サンプリングを主に「富岳」を用いて行った。対象とするデータは、Chungら(Chung *et al.*, *Science* 2012)に計測されたProtein Gのフォールディングダイナミクスの1分子FRET計測データである。これまでに、FRET計測のプローブ(ドナー、アクセプター蛍光色素分子、図2-1-3左)がラベルされたProtein Gをモデリングし、主に研究室内のPCクラスタを用いて構造サンプリングを行い、そのデータを用いてマルコフ状態モデルを構築しデータ同化まで行ってきたが、自由エネルギーのバリア付近(遷移状態)はサンプルが少なく、統計が不十分であった。そこで、GENESISを用いてProtein Gの通常のMDとgeneralized REST(gREST)計算を追加で行なった。得られたデータを既存のデータに加えて、contact mapによる特徴量抽出、独立成分解析による次元縮約、k-meansクラスタリングを行った後で、マルコフ状態モデリングを構築した。その結果、遷移状態付近のクラスタを複数捉えることができた(図2-1-3右)。

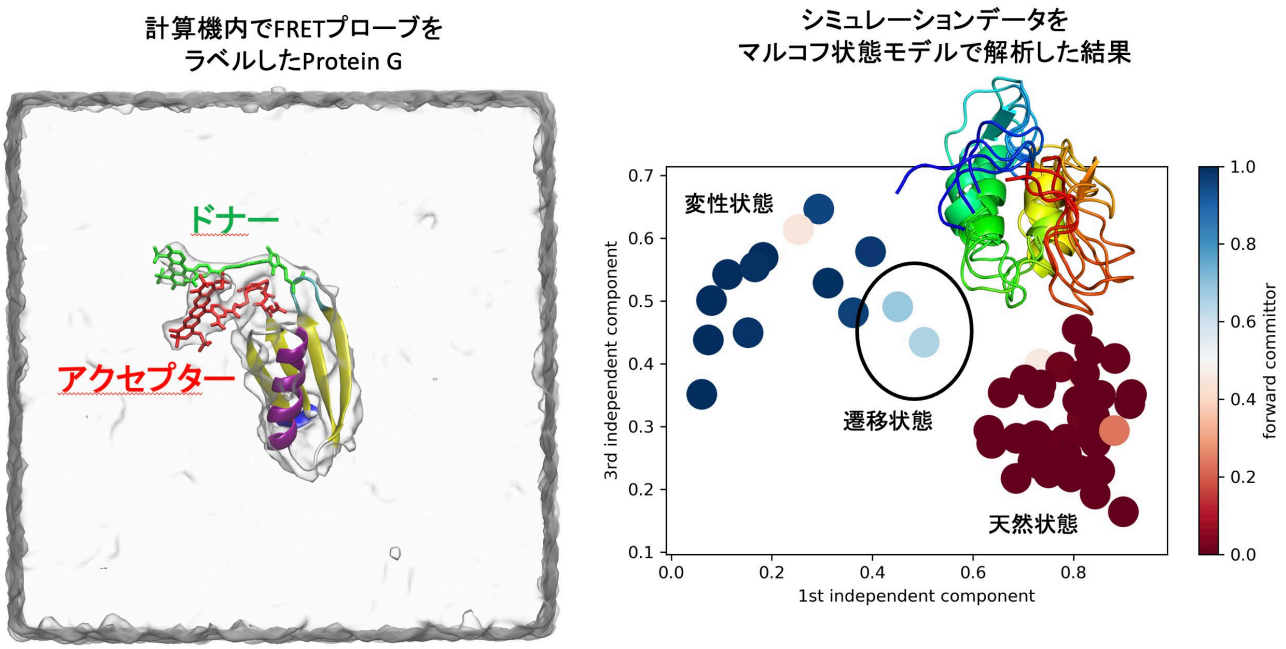


図2-1-3: Protein Gの構造サンプリング（MDシミュレーション）によって得られたマルコフ状態モデリング。

② 粗視化モデル CafeMol のパラメタ最適化（京大・高田）

粗視化分子モデルCafeMolは、タンパク質・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。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タンパク質・タンパク質相互作用の粗視化パラメタの最適化を行い、より精度の高いMDシミュレーションの実現に向けて研究を推進した。

まず、生体膜を構成するリン脂質の高精度な粗視化分子モデルiSoLF、リン脂質とタンパク質との相互作用モデルを開発し、そのパラメタを注意深くチューニングした。我々は、これまでの研究により、タンパク質の粗視化モデルと適合する解像度を持つリン脂質の粗視化モデルiSoLFを構築した(Ugarte La Torre & Takada *J. Chem. Phys.* 2020）。この粗視化モデルは、水分子を陽に取り扱わない陰溶媒モデルであり、リン脂質を5粒子（頭部に2粒子H1, H2、尾部に3粒子T1, T2, T3）で表現する（図2-2-1左上）。一方、タンパク質は各アミノ酸を1粒子で表現する。リン脂質の粗視化粒子とアミノ酸粒子の相互作用には、Kim-Hummerが提案した関数形を採用した（図2-2-1中央上）。このモデルは、排除体積を表す近距離の斥力と、それより長い距離の引力又は斥力を表現することができる。具体的には、引力の場合には、アミノ酸iとリン脂質粒子jの間に標準的なLJ相互作用

$$V_{LPI} = \sum_{i<j}^{n_{int-pairs}} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

を用いる（図2-2-1上中の青線）。ここで、 ε_{ij} は、アミノ酸種とリン脂質の粒子種に依存するエネルギーパラメタである。サイズパラメータσ_{ij} は、アミノ酸およびリン脂質の従来モデルから導かれる。

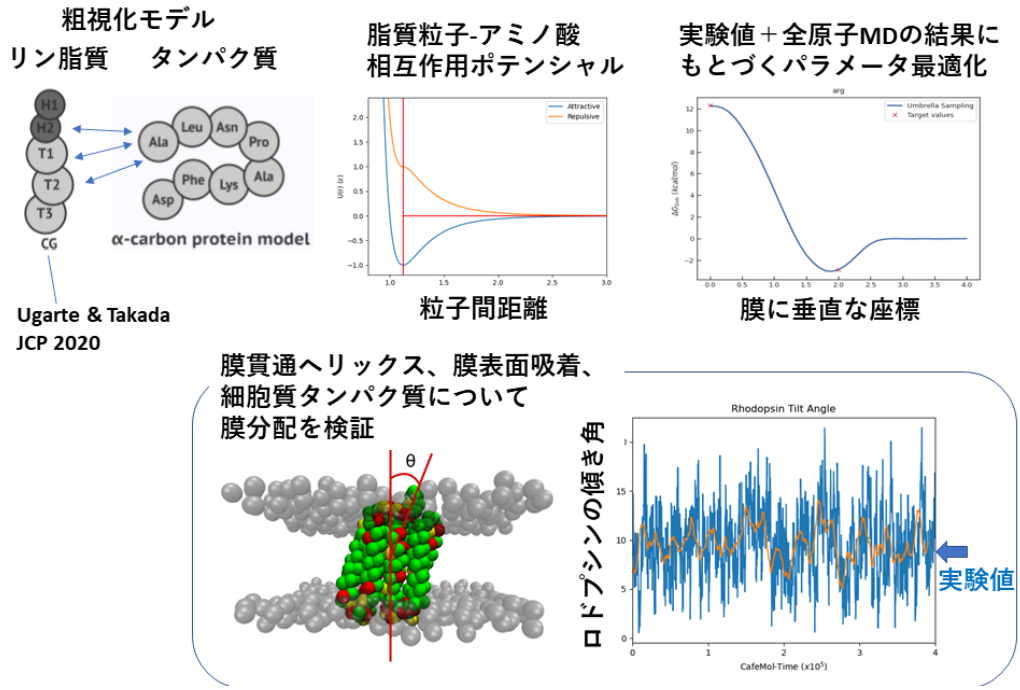


図2-2-1: 脂質分子の新しい粗視化分子モデルの開発。

長い距離でも斥力であるペアについては、Kim-Hummerに従って、

$$V_{LPI} = \sum_{i<j}^{n_{int-pairs}} \begin{cases} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} + \frac{1}{2}\right], & r_{ij} \le \sqrt[6]{2}\sigma_{ij} \\ -4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], & r_{ij} > \sqrt[6]{2}\sigma_{ij} \end{cases}$$

とする（図 2-2-1 上中の赤線）。

リン脂質とアミノ酸の相互作用で鍵となるのは、エネルギーパラメタε_{ij}である。パラメタチューニングを行う際に、実験的に求められている水・有機溶媒間の分配係数と、全原子MD計算から得られている各アミノ酸の膜垂直方向に関する自由エネルギー曲線を参照することとした。前者を、アミノ酸が膜中央に位置する時の自由エネルギー値と一致させること、後者の膜中央から20Å離れた層（膜内の親水層）の自由エネルギー値を粗視化モデルの対応する値と一致させることを目指してパラメタチューニングを行った。前者が主にアミノ酸とリン脂質尾部の相互作用、後者がアミノ酸とリン脂質頭部の相互作用を規定する。得られたパラメタによるアミノ酸の膜垂直方向の自由エネルギー曲線は図2-2-1右上のようになった。参照した実験値および全原子MDからの値が、対応する膜垂直方向の位置でよく一致していること、および全体として滑らかな関数となっており、狙い通りのモデルが出来たことが分かる。

次に、得られた相互作用モデルを検証するために、リン脂質膜への膜タンパク質等の挿入シミュレーションを実施した。ロドプシンなど幾つかの膜貫通ヘリックスについて実験と誤差範囲で整合性のある、深度、傾き角を得ることを確認した（図2-2-1下）。次に、ミオグロビンなどいくつかの水溶性タンパク質として知られるタンパク質について類似のMDシミュレーションを行った。水溶性タンパク質は膜に侵入せず、また膜表面にも吸着しないことを確かめた。さらに、膜表面に吸着することが知られているタンパク質についてもMDシミュレーションを実施し、確かに膜表面に吸着することを確認した。これによ

り、膜を構成する脂質分子と、タンパク質アミノ酸分子の基本的な相互作用を精度よく再現する粗視化モデルを構築できたと考える。

次に、非特異的タンパク質・タンパク質相互作用に関して、機械学習を用いた粗視化パラメタの最適化を実施した。例えば、アミノ酸配列に沿った局所相互作用のひとつ結合角ポテンシャル（図2-2-2左上）として、従来の粗視化モデルは中央アミノ酸の種類にのみ依存するデータに基づくポテンシャルになっている。すなわち20種類のパラメタをもつ。しかし、結合角は、その前後のアミノ酸種にも強く依存することが知られている。そこで、新しい結合角ポテンシャルとして、配列にそった前後のアミノ酸を含む3アミノ酸種に依存するものに更新した。これにより、全部で20^3=8000種類のパラメータセットを持つことになり従来よりも高精度な局所ポテンシャルとなることが期待される（図2-2-2右上）。しかし、パラメタ最適化に利用できるデータ数を、タンパク質立体構造データベース（αヘリックスやβストランドを形成する部分を除いたループ領域に限定）で調べたところ、3アミノ酸種の並びによってはその立体構造データ数が不足し満足なポテンシャルを構築することが出来ないことが判明した。

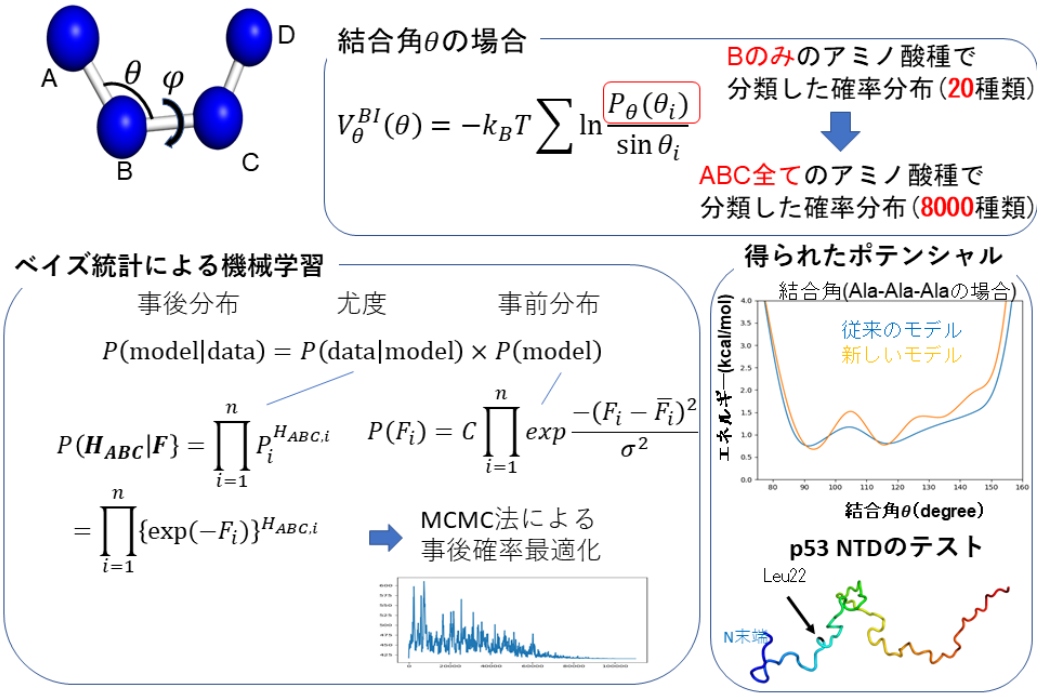


図2-2-2：機械学習による非特異的タンパク質相互作用の粗視化モデルの開発。

そこで本研究開発では、ベイズ統計学に基づく機械学習によって、データ不足を解消することとした（図2-2-2左下）。

$$P(\text{model}|\text{data}) = P(\text{data}|\text{model}) \times P(\text{model})$$

具体的には、従来の20種類に分類したポテンシャルを平均とするポテンシャルの分布を事前分布$P(\text{model})$とする。この事前分布に対して、8000種類に分類したデータを尤度$P(\text{data}|\text{model})$として利用する。上式により得られる事後分布$P(\text{model}|\text{data})$は、3アミノ酸種の並びで規定される8000種類のポテンシャルについて、データ数が少ない場合には事前分布が支配的になる一方、データ数が多い場合には尤度によって決定されることになる。具体的には、事後分布$P(\text{model}|\text{data})$は、マルコフ連鎖モンテカル

ロ法によってサンプリングすることができ（図2-2-2左下）、それによって新しい粗視化モデルを構築することができた。

同様の手法を、2面角ポテンシャル等にも適用した。2面角ポテンシャルは、直接かかわる4アミノ酸並びのうち中央の2アミノ酸種に依存するものと定義した（従来法は22種類に分類していた）。また、従来、フーリエ級数で表現されていたが、より詳細な特徴を捉えるために今回新たにスプライン補間を採用した。結合角の場合と同様、マルコフ連鎖モンテカルロ法によってポテンシャルを求めることができた。さらに、自由度間の相関を考慮に入れるために、繰り返し逆ボルツマン法によって、結合角ポテンシャルおよび2面角ポテンシャルを精密化した。

以上により、アミノ酸配列に依存した弱い非特異的な相互作用モデルを構築することができた。得られた新しい非特異的相互作用モデルを検証するために、p53のN末端の天然変性領域などでテストシミュレーションを行い、従来モデルよりも高い配列依存性を示している様子を確認した。一方、非局所相互作用については、さらなる最適化が必要となる。

（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）

細胞質内部は多くの生体分子で混み合った環境であり、この環境におけるタンパク質の構造・ダイナミクス・機能の関係を調べることは生物学における重要な研究課題である。我々はこれまで、「京」を用いたバクテリア細胞質の全原子MDシミュレーションや、タンパク質混雑系における水和の影響、タンパク質構造安定性、回転・並進拡散の影響、生体膜との相互作用などについてMDシミュレーションを用いて調べてきた。バクテリア細胞質のMDシミュレーションで得られた知見は複数あるが、そのうち興味深いのは、タンパク質と基質や代謝物との相互作用である。細胞内分子混雑環境はタンパク質などの生体高分子だけで構成されているわけではなく、ATPなどの様々な代謝物やイオンがその溶液中に大量に溶け込んでいる。我々は、細胞質中に含まれるタンパク質の一つであるACKEに注目し、その基質ATPのタンパク質周囲の分布を調べた。その結果、ACKEのタンパク質周囲の分布密度は希薄溶液中の方が遥かに高く、混雑環境にあるACKEではタンパク質表面の大部分が他のタンパク質と接触しかかっているため基質ATPが占める濃度はあまり高くないことがわかった。にもかかわらず、基質結合部位はくぼみ（ポケット）を作っており他のタンパク質が接触できない領域であるため、希薄溶液中とタンパク質分子混雑環境のいずれにおいても高い密度でATPが存在していた。

細胞内分子混雑環境におけるタンパク質の機能を考察するために、この計算で明らかになったタンパク質と基質の相互作用・分布密度に関する情報は本質的であると思われた。ではこの現象はどのように一般化できるだろうか？そのような問題意識を我々が抱いている中で、2017年に興味深い実験の論文がScienceに発表されていることを見つけた。ATPは細胞内のエネルギー通貨と呼ばれるように、モータータンパク質などが駆動する化学エネルギーの供給源（通貨）であると考えられている。すなわち、モータータンパク質はATP加水分解で生じる化学エネルギーの利得を使ってその分子構造を大きく変化させることで機能を発現している。このよく知られたATPの機能だけを考えるならば、細胞内で測定されているATPの濃度は高すぎる。Patelらはそのような問題意識をもち、ATPの隠れた役割として細胞内分子混雑環境において多数のタンパク質が凝集しないためにATPが「可溶剤」として働いているのではない

かという興味深い提案をしている。この実験から導かれた提案は我々にとって大変魅力的なものであった。なぜなら、バクテリア細胞質のMDシミュレーションのトラジェクトリの詳細を解析すると、ATPが多く存在するのは、体積の約60-70%を占める水中でいずれかのタンパク質に近い部分であったからである。Patelらの実験と我々の計算結果をまとめると、ATPが細胞内濃度で存在する溶液中では分子混雑環境でもタンパク質の凝集は起きない。一方でATPが存在しない分子混雑環境では凝集は促進されるという仮説が導かれる。この仮説を検証するために我々は約100万原子を含む分子混雑系をモデリングし、1μsの全原子MDシミュレーションを異なる条件で実施した。それらは、ATPを細胞内環境と同じ濃度で含む系とATPを含まずに、その代わりにイオンで系全体を中和した系である。もし、ATPが凝集を防ぐ効果が静電相互作用だけであれば、同じ電荷量のイオンを含む溶液ではやはり凝集が阻害されているはずである。現在、その解析を継続しているところであるが、統計的に有意な差としてATPが含まれている溶液中で凝集の効果が阻害されていた。その原因を現在調査中である。

細胞内分子混雑環境におけるタンパク質と基質の相互作用に関わるもう一つの課題は、ターゲットタンパク質と基質の強い結合がどのように生じているかというものである。この問いは、特に計算創薬において今後重要になるかもしれない。なぜなら、現在行われているIn-silico創薬の多くは希薄溶液中でタンパク質と基質候補化合物の結合定数を予測することに費やされているが、実際の我々の体（すなわち細胞内環境）では多くのオフターゲットタンパク質が存在しているからである。このような問題意識の中で、我々はc-Src Kinaseとその阻害剤PP1の結合・解離のMDシミュレーションを希薄溶液中とタンパク質混雑物であるBSAを含む系で実施した。特に希薄溶液中のMDシミュレーションでは、その結合と解離のイヴェントを多数観察するために、効率の良い構造探索手法の一つである2次元レプリカ交換法（gREST/REUS）を用いて精度の高い自由エネルギー解析を実施した。その際、タンパク質c-Src Kinaseの構造に位置拘束条件をつけて立体構造があまり揺らがないようにした条件と拘束条件をつけない計算の2種類を実施した。その結果、拘束条件の有無で、すなわち、タンパク質の構造揺らぎを許すかどうかで基質の結合経路は大きく異なることが明らかになった（図3-1-1(b)）。この結果自身は2019年に報告済みであるが、我々はこの自由エネルギー解析の結果に対して、今回長時間のMDシミュレーションで得られた複数の基質結合を含む軌跡を射影してみた（図3-1-1(c),(d)）。その結果、希薄溶液中のPP1の結合軌跡はタンパク質構造が揺らいでいる条件での自由エネルギー計算と、混雑環境の軌跡はタンパク質構造を拘束した自由エネルギー計算とよく一致していることが明らかになった。この結果を素直に解釈すると、希薄溶液中と混雑環境を比較するとその排除体積の違いから、タンパク質の揺らぎが異なる（希薄溶液中の方が大きい）。そのため、タンパク質構造に拘束条件をかして揺らぎを抑えた計算条件の方が細胞内環境を模倣しているという考え方もありえる。しかし、MD計算のトラジェクトリを詳細に解析すると、分子混雑中と希薄溶液中でc-Src Kinaseの構造揺らぎはそこまで大きな差はなかった。この問題にはっきりとした答えを見つけることはこの計算結果だけでは難しいが、鍵となるのはタンパク質周囲の基質分子の分布の違いではないだろうか。バクテリア細胞質の計算やATP有無の計算で明らかになってきたように、分子混雑環境と希薄溶液中ではタンパク質周囲の基質濃度が結合ポケット（くぼみ）を除いて大きく異なる。すなわち、真の結合ポケットに至る経路の出発点の確率がそもそも異なっているだろう。実際、我々が行ったc-Src KinaseとPP1の自由エネルギー解析においてもタンパク質表面に到達して結合ポケットに至る経路の出発点（過渡的複合体：Encounter Complex）の位置は大体いつも同じである。その出発点からスタートしたとしても、分子の配向の違いによって、真の結合状態

に至る場合と至らない場合が生じるのである。この考察から、c-Src Kinaseと PP1の過渡的複合体が希薄溶液中と分子混雑環境で異なっていることがこの問題と大きく関わっていると考えられる。さらに、この仮説を詳しく理解するためにはより多くの系について、タンパク質と基質を変えた複数の計算結果とそれに関係する実験が必要であると思われる。この計算は「富岳」とHPCIリソース、さらに米国の共同研究者が申請したAnton2のリソースを用いて行われた。

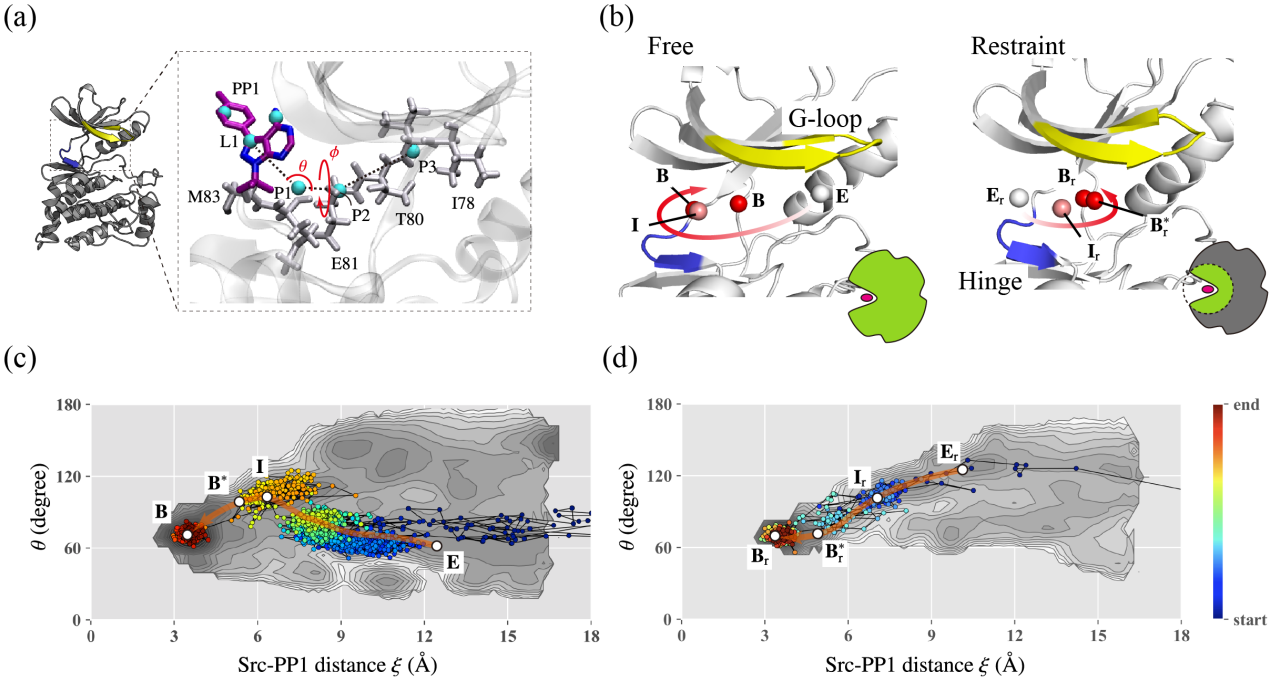


図 3-1-1: c-Src Kinase と阻害剤 PP1 の結合・解離シミュレーション。(a) タンパク質 c-Src Kinase と PP1 結合部位のクローズアップ。(b)希薄溶液中およびタンパク質に拘束条件をつけないレプリカ交換法で観測された結合経路と、混雑溶液中およびタンパク質に拘束条件をつけたレプリカ交換法で予測した結合経路の違い。(c) 希薄溶液中（およびタンパク質に拘束条件をつけないレプリカ交換法）での自由エネルギー面と MD 計算での軌跡。(d) 混雑環境中（およびタンパク質に拘束条件をつけたレプリカ交換法）で得られた自由エネルギー面と MD の軌跡。

連携研究者である東工大・北尾は、細菌べん毛のモデリングを行った。細菌べん毛は多くの病原性細菌、例えば大腸菌やサルモネラ菌などが移動するために用いる長さ10μmに及ぶ巨大な分子機械であり、30種類以上のタンパク質が3万分子程度集合することで構築される。細菌べん毛は、輸送装置を用いて巨大な分子構造体を構築し、イオン勾配エネルギーを回転力へと変換し、自在継手として働くフックを介して回転力を伝達し、スクリューであるべん毛繊維を回転して細胞運動を行う。更に、走化性シグナルによって回転を制御し、べん毛繊維のらせん性を変化させることで細胞運動を制御している。世界の乳幼児死亡原因の約半分はウイルスや細菌が引き起こす感染症によるものであり、特に発展途上国では未だに大きな問題である。細菌の運動性にかかわる細菌べん毛の機能が明らかになれば、将来的に細菌の増殖を抑え病原性に対処する科学技術の発展が期待され、また感染症へのより効果的対策によって人類の健康に寄与する可能性がある。

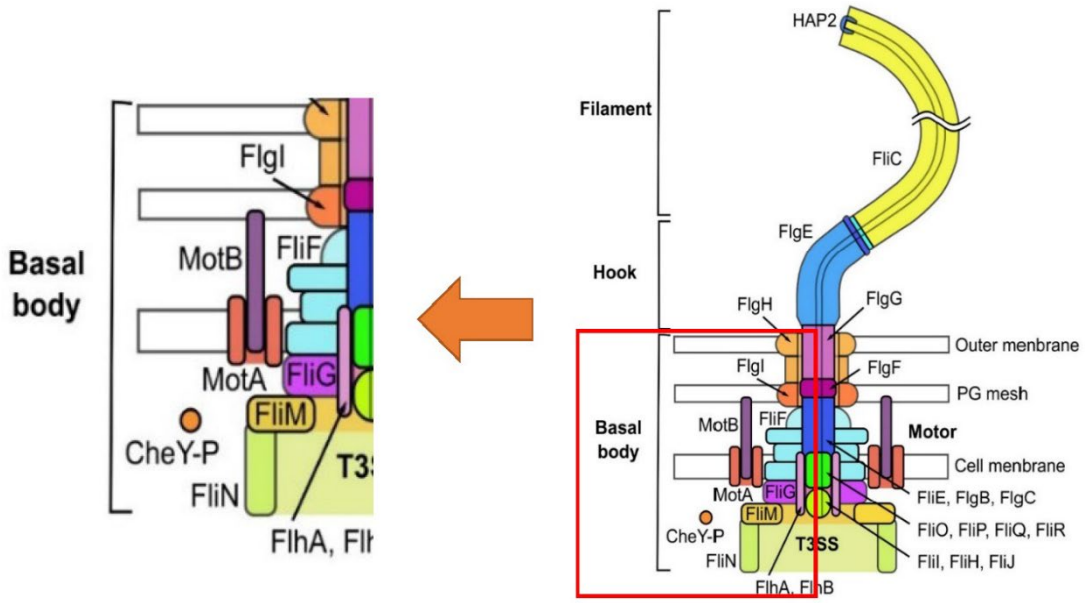


図3-1-2：細菌べん毛の全体構造の模式図（右）と固定子付近の拡大図（左）。

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。その回転はシグナルタンパク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。モーターの部分構造は徐々に決定され始めているが、イオン透過の心臓部である固定子の膜貫通部位の立体構造はごく最近まで未決定であった。しかし、2020年になって固定子MotABやPomAB膜貫通部位の立体構造がいくつか実験的に決定されるに至った。これによりMotAまたはPomAが4分子と、MotBまたはPomBが2分子によって固定子の1ユニットを構成するという従来の予想とは異なり、MotA（PomA）の5分子がMotB（PomB）2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た（図3-1-2）。また構造的な特徴から、イオン透過によってMotB（PomB）の周りをMotA（PomA）が回転しているのではないかというモデルが提唱されるに至った。

一方で、構造決定された固定子複合体はそのメカニズムがよく研究されていないことから、既に回転速度やアミノ酸変異の影響などの様々な知見が蓄積されているビブリオ菌とアクウィフェクス菌のホモロジーモデリングを行い、システムを平衡化したのちにイオン透過とMotAやPomAの回転が相関しているかを大規模MDシミュレーションで調べることを計画した。令和2年度はまずアクウィフェクス菌のMotABとビブリオ菌のPomABのモデリングとその後の平衡化を行った。具体的には、まず脂質分子POPEとPOPGが3対1で構成される脂質二重膜を生成し、それにホモロジーモデリングで生成したMotABまたはPomABの1ユニットを埋め込んだ（図3-1-3）。MotABは約22万原子、PomABは約21万原子のシステムになった。MotABとPomABのMDシミュレーションをそれぞれ10回行った。MotABについてはMDシミュレーションが1.5μs程度まで終了し、PomABのMDシミュレーションについても1.2～1.5μs程度まで進んでいる。現在、立体構造の収束状況を調べている。

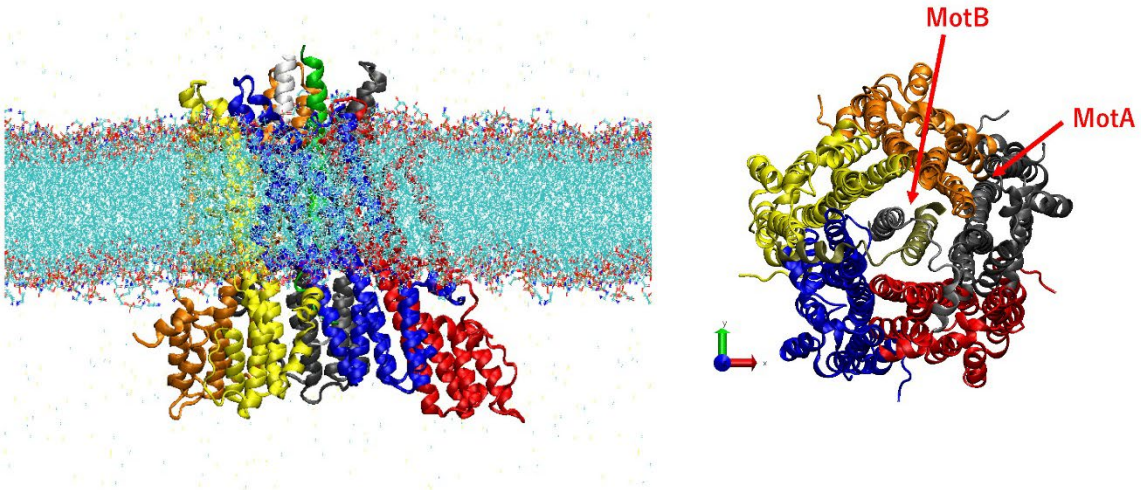


図3-1-3: モデリングしたMotABを含むシステムの全体像（左）とペリプラズム側から見たMotABの立体構造（右）。

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等の実験情報を用いて粗い3次元構造モデルを構築した。さらに、粗視化分子モデルCafeMolを用いて詳細な動態シミュレーションを実施し、転写活性におけるタンパク質とDNAの相互作用および立体構造変化に関する知見を得た。

ヒトを含む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染色体（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。本来と異なるクロマチン3次元構造への折りたたみは、その部分にコードされた遺伝子の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（図3-2-1左）。*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*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ーと、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ー(SE)が存在する。プロモーターおよび3つのSEには、Oct4, Sox2, およびNanogタンパク質の結合モチーフが複数存在し、これらタンパク質の結合が*Nanog*遺伝子の転写を活性化する。プロモーターと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（図3-2-1左下）。また、近距離になった状態で、コアクティベーターと呼ばれるMediatorやBrd4などが、プロモーターと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、プロモーター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、コアクティベーター、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。

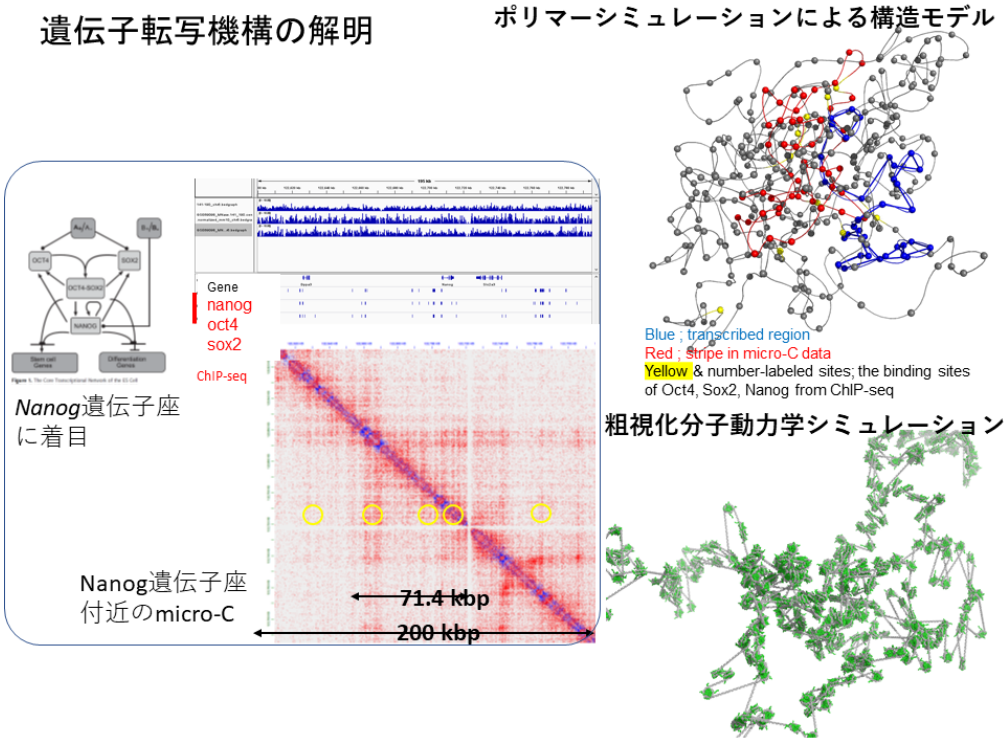


図3-2-1: 遺伝子転写機構の解明にむけた生体分子系モデリングとシミュレーション。

最初に、マウスES細胞についての最新のmicro-C法のデータから、*Nanog*遺伝子座の71.4kbの領域の染色体折りたたみ構造をモデリングした。micro-C法は200bpの分解能で、ゲノムセグメント間の接触情報を与える。これをもとに1ビーズが200bpに相当するポリマーのメソスコピックバネビーズモデルを用いて、分子シミュレーションによって、その3次元構造アンサンブルを得た（図3-2-1右上）。興味深いことに得られた構造アンサンブルでは、*Nanog*遺伝子本体・プロモーターに対して、3つのSEが接触する確率が高かった（図中の赤い粒子、黄色のラベル）。一方で、ゲノム構造は細胞毎に異なり多様性をもつことが分かっているがmicro-C法はその平均量を与えるため、micro-Cデータの取り込み方、アンサンブルの構造多様性の取り扱いには解決すべき課題があることも見えてきた。

次に、*Nanog*遺伝子座のDNA上のヌクレオソーム配置を決定した。ゲノム上のヌクレオソーム配置は、実験的にはMNase-seqあるいはより精度の高い化学マッピングなどによって分析される。大域的にはこれを利用できるが局所的な分解能などに課題がある。相補的に、ヌクレオソーム配置は局所的なDNA塩基配列にも依存する。この配列依存性は、塩基配列の関数としての有効エネルギー関数として表現できる。そこで、大域的には実験データに基づきつつ、局所的には塩基配列による特性を反映させるスコア関数をつくり、メトロポリスモンテカルロ・アニーリング法によって、スコアを最適にするヌクレオソーム配置を求めた。これにより、実験データと塩基配列情報をバランスよく考慮しつつ、ヌクレオソー

ム配置を決定することができた。

以上2つの準備の後、200bp分解能のメソスコピックモデルの3次元構造上で、決定されたゲノム配列位置にヌクレオソームを配置して、CafeMol分解能のモデルをつくる、いわゆるバックマッピング法を開発した。まず、単一ヌクレオソーム構造のCafeMol分解能粗視化モデルとして異なる程度のDNAアンラッピングを伴う多数の構造サンプルを、また隣接するヌクレオソームをつなぐ間のDNA2重らせん（リンカーDNA）にも多様な構造サンプルを、それぞれ用意した。ポリマーメソスコピック構造をもとに、一方の端から順次その上に、粗視化モデルのヌクレオソームサンプルとリンカーDNAサンプルを交互に繋いでクロマチン構造を構築していく。その際、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。こうして得られた粗視化モデル3次元構造は、メソスコピックモデルの構造に概ね沿った構造モデルとなった。

最後に、このようにして得られた*Nanog*遺伝子座の71.4kbのクロマチン構造を初期構造として、「富岳」上で、GENESISを用いて粗視化モデルによるMDシミュレーションのテストを実施した（杉田チームとの共同研究開発）。最初、原子同士の重なりやその他いくつかの技術的な問題が発生したが、これらを順次、解決し、最終的に安定したMDシミュレーションを実現した。これは、約10原子を1粒子として扱う粗視化モデル（CafeMolモデルを含む）のMDシミュレーションとしては、世界的に前例のない大規模シミュレーションとなった。図3-2-1右下に示すようにテストシミュレーションから、クロマチン構造は一様ではなく、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スタを形成する傾向が見て取れた。これは、近年、実験などから提唱されているヌクレオソームクラッチと呼ばれる構造に対応しているものと考えられる。このような折りたたみ構造とヌクレオソーム配置から決まる相互作用様式が転写活性制御の基礎を作っていると考えられる。

③ 核内タンパク質・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、原子レベルでのクロマチンの動的構造を明らかにすることが重要である。真核生物のDNAは、ヌクレオソームという構造体の形で細胞の核内に収納されている。ヌクレオソームはヒストンタンパク質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何千何万と存在する。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状態では大部分のDNAには分子がアクセスすることができない。本研究では、ヌクレオソーム多量体構造を安定化するHP1タンパク質とヌクレオソームに結合して転写を活性化するGATAタンパク質について調べる。今年度は、それぞれの原子モデルの構築を行った。

ヌクレオソーム多量体

ヌクレオソームどうしがコンパクトな構造をとる時は、タンパク質HP1が結合している。2018年に、2量体のヌクレオソームに2量体（ダイマー）のHP1が結合した分子の電子顕微鏡による単粒子構造が報告されたが、HP1に該当する電子密度が悪く、HP1の分子構造は依然としてわからない（図3-3-1）。そこで、分子モデリングとHP1のクロスリンキング実験により、その構造体の原子モデルを作ることに取り組んでいる。HP1は、クロモドメイン(CD)とクロモシャドードメイン(CSD) の2つの安定なドメインを天然変性領域であるリンカー（HR）で連結されたトポロジーを持ち、これが CD もしくは CSD を介して2

量体を形成し、ヌクレオソームに結合する（図3-3-2）。HP1ダイマーの構造を解明するため、実験研究者と共同してクロスリンキング実験（HP1内のリジン残基間を低分子で化学的に結合させ、その後タンパク質をフラグメント化し、分解されたフラグメントの質量分析から化学結合したリジンを同定することにより、リジン残基間の空間的な距離情報を得る）により取得した距離情報を拘束条件に取り込み、拡張サンプリング法などを用いたMD計算を行い、HP1ダイマーの安定な構造を求めた。結果、図3-3-3に示すように、HP1の分子配向まで絞り込むことができた。

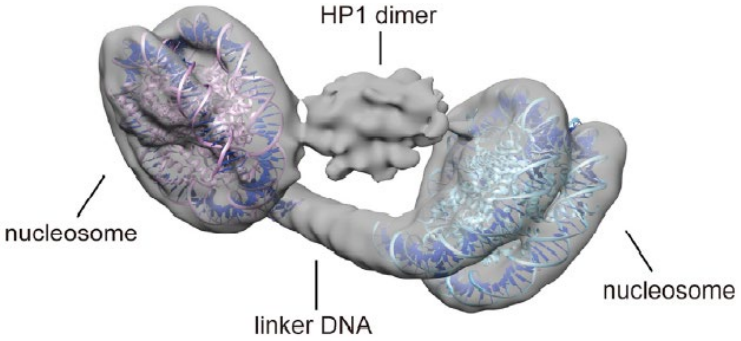


図3-3-1：ヌクレオソームとHP1の電子顕微鏡像。

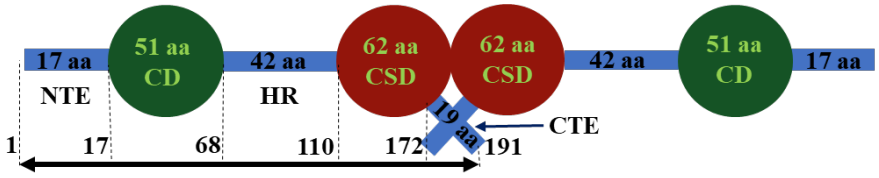


図3-3-2：HP1のトポロジー。

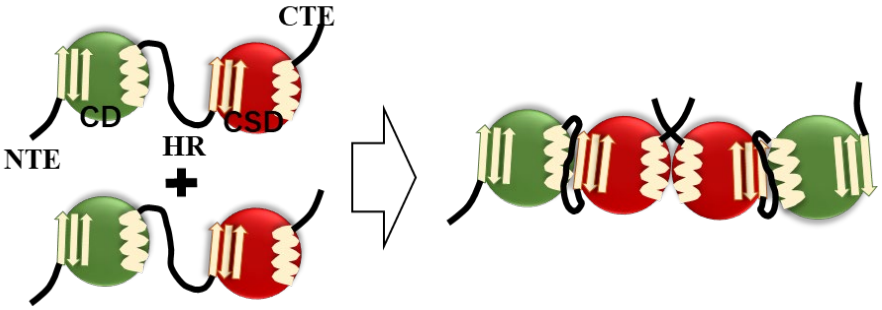


図3-3-3：推定されたHP1ダイマー構造。

また、HP1が結合していない状態でのヌクレオソームダイマーについて、ヌクレオソームのヒストンテールがダイマー構造に与える影響について調べた。ヒストンテールは高頻度で翻訳後修飾を受け、その修飾とクロマチンの構造状態や遺伝子発現状態に高い相関があることが知られている。しかし、ヒストンテールは、特定の安定な構造をとらない天然変性領域であるため、その構造的な解析や修飾の影響解析はほとんど進んでいない。本研究では、まずテールの有無によってダイマーの構造やダイナミクスが

どう変化するかを調べた（図3-3-4）。種々のイオン濃度やテールの有無で、それぞれ500nsのシミュレーション計算を行った結果、テールが無いと2つのヌクレオソームはコンパクトな構造をとることができないことが分かった。一方、テールが有るとそれを介在してヌクレオソームが近接した構造をとることができることが分かった。また、塩濃度を変化させてシミュレーションを行った結果、低濃度は近接構造が見られず高濃度で近接構造が観察されることから、ヒストンテールの静電的な相互作用により近接構造が安定化することが示された（図3-3-5）。

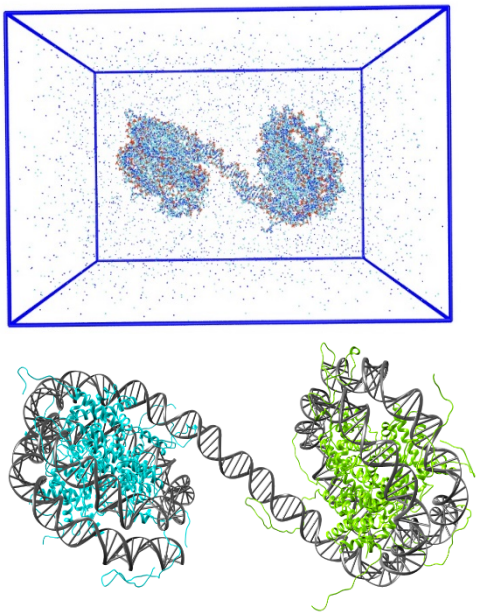


図3-3-4: ヌクレオソームダイマーの系（約100万原子）。全体の系（上図）と
ヌクレオソームダイマーのリボンモデル表示（下図）。

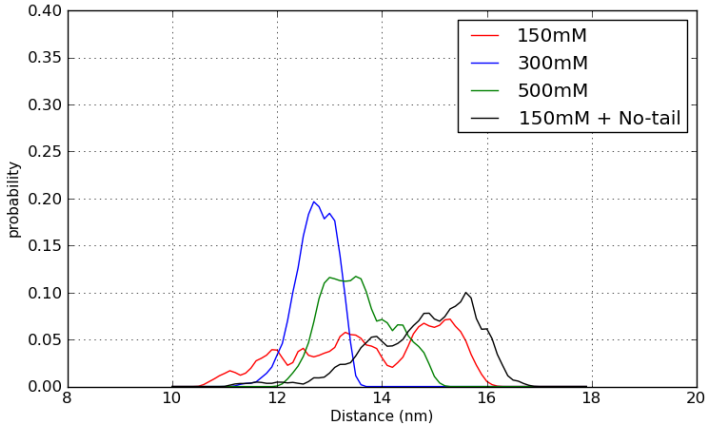
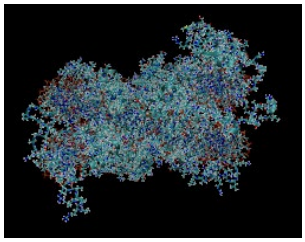


図3-3-5: 2つのヌクレオソームの重心間距離分布。
テール有の系では、塩濃度 150, 300, 500 mM で計算を実施。
テール無の系では、塩濃度 150mM で実施。

GATA3とヌクレオソーム複合体

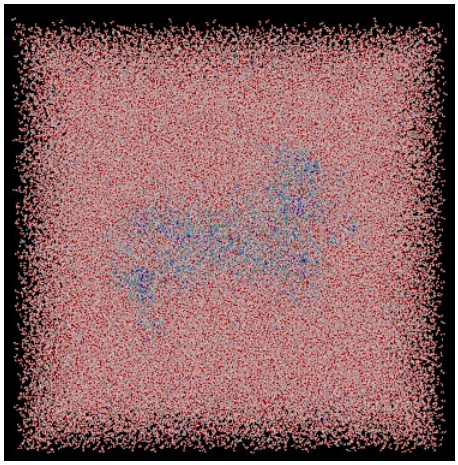
GATA3は遺伝子の転写を活性化するクロマチン結合因子であり、GATA3の標的DNA配列がヌクレオソームによって巻き付いている場合でもDNAと結合することが知られているが、その構造やダイナミクスの詳細についてはよく分かっていない。

そこで、GATA3とオリゴDNA複合体、ヌクレオソームそれぞれのＸ線結晶構造を、クロマチン結合因子GATA3とヌクレオソーム複合体の単粒子電子顕微鏡解析によって得られた電子密度に適合することで、原子モデルを構築した。まずGATA3、ヌクレオソームをそれぞれ剛体として扱い、オリゴDNAとヌクレオソームDNAの構造アライメントにより、大まかな複合体構造を作成した。次に、電子顕微鏡像解析による単粒子構造データに合致するように、MDシミュレーションを用いて構造変形させ、複合体の原子モデルを構築した。生成した原子モデル構造を評価し、構造的な見地からおかしなひずみのない複合体について全原子MDシミュレーションを実施した。そして、GATA3とヒストンテールの相互作用やGATA3とヒストンDNAに対する結合強度の違いを調べた。以下、その詳細である。



初期モデル

溶媒和



計算に用いた系

図3-3-6: フィッティングに用いられた水分子を含んだクロマチン結合因子GATA3とヌクレオソームが結合した複合体構造モデル。系の原子数は、複合体が約3万原子、水分子が約12万分子で、合計約39万原子である。系のサイズは164Å× 164Å× 164Å。塩濃度は150mMとした。

実験研究者から提供された3次元電子顕微鏡像マップ（現在は、EMデータバンク（Electron Microscopy Data BankにEMD-0783として登録。解像度3.15 Å、Voxelサイズ：1.05×1.05×1.05 Å、グリッドポイント数：180×180×180）を用いた。複合体の電子顕微鏡像に、X線結晶構造解析等で決定された各構成分子単体の構造をフィッティングする方法には、さまざまな方法が開発されている。その多くは、複合体を構成する複数の単体分子をそれぞれ剛体として並進回転することで電子顕微鏡像にフィッティングするため、正しい分子の配向であっても分子どうしが衝突するという問題が頻繁におこる。それを解決するには、複合体の界面で起こりうる構造変形を考慮したフィッティング手法が必要であり、MD計算を用いる手法ではその実現が可能である。ただし、MD計算では、原子間力計算と電子密度に対するフィッティングのためのフィッティング力計算に多大な計算コストがかかる。そこで、「富岳」に対してチューニングされたMDシミュレーションプログラムGENESISを用いて、フィッティングを行った。できるだけ自然な形で構造変形をさせて複合体の構造を得るため、水分子も含めた系（図3-3-6）で実施した。電

顕に対するフィッティング力は、安定なドメイン構造をとる領域に対してのみかけ、電顕像の電子密度が小さいところや単体の構造解析において原子座標がない部分に対してはかけなかった。

最初に、フィッティング力の最適値を探した。他のフィッティング方法と同様に、電顕に対するフィッティング力の大きさは調整パラメータであり、あらかじめ最適な値がわからない。また、電子顕微鏡像の解像度にも依存する。強くし過ぎると、分子構造が局所的にゆがんでしまう。そこで、フィッティング力の強さ(5k, 10k, 15k, 20k, 50k, 100k 任意の単位）を６パターン用意し、フィッティング計算を実施した。

図3-3-7にフィッティングの良し悪しを定量的に表す相関係数（CC）を示す。CCはフィッティングシミュレーションを1ns（1,000 ps）実行するとほぼ一定値に収束したと判断し、計算を打ち切った。1ns後の６つの立体構造モデルを、タンパク質や核酸の立体構造の妥当性（2次構造やペプチドの2面角、Cα原子の空間配置等）を評価するメタサーバMolProbityを用いて調べた。その結果、電子顕微鏡像に最も強くフィッティングしたモデル1と2（CCが0.80以上）では、GATA3の2つの安定ドメインの一つであるC末側のジンクフィンガードメインのαヘリックス領域がゆがんで不自然な構造になることがわかった。また、弱いフィッティング力で構築したモデル5と6（CCが0.78以下）では、GATA3の一部が電子顕微鏡像にフィットしておらず、フィッティングが不完全であった。最終的にGATA3がほぼ電子顕微鏡像に当てはまり、自然な構造を保持したモデル3と4（CCが0.79程度）の構造を、最適なフィッティング立体構造とした（図3-3-8）。

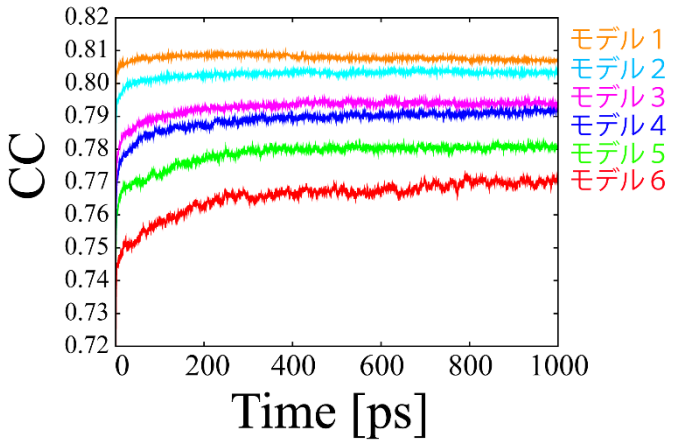


図3-3-7: MDシミュレーションによるフィッティング。横軸はシミュレーション時間、縦軸は電子密度に対する相関係数（CC）。CC＝1.0のとき、複合体の立体構造から計算される理論的電子顕微鏡像と完全に一致する。MDシミュレーションはCCがほぼ一定値となる1 nsまで実施。モデル1から6は、フィッティング力 100k, 50k, 20k, 15k, 10k, 5kに対応。

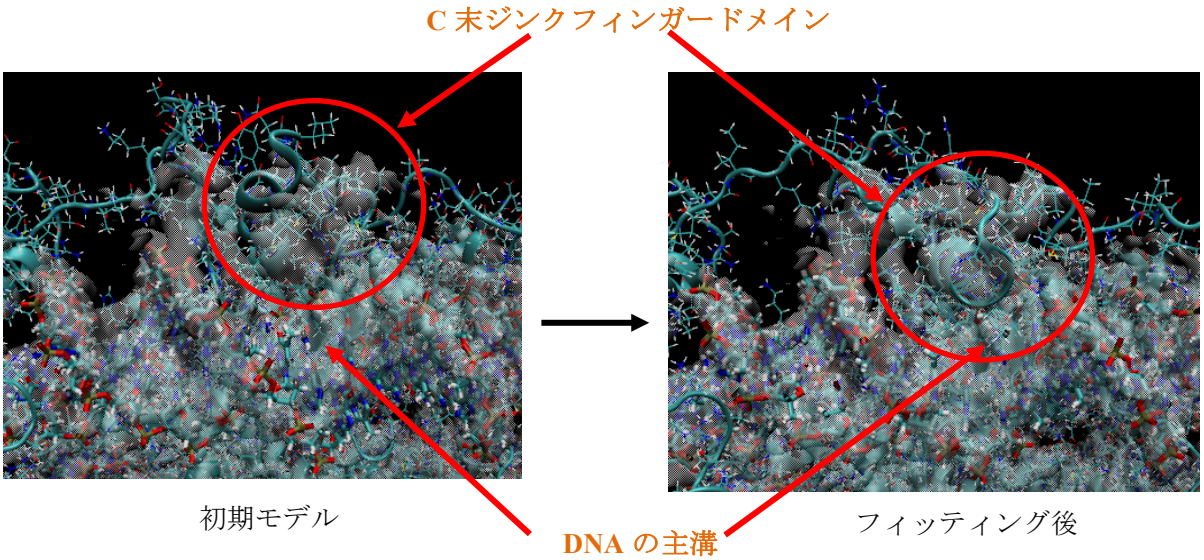


図3-3-8: MDシミュレーションにより構築された複合体モデル。初期モデルでは、GATA3のC末ジンクフィンガードメインが電子顕微鏡像から大きく外れている。フィッティング後には、C末ジンクフィンガードメインが電子顕微鏡像にフィットし、DNAの主溝と相互作用するようになった。

この構造をもとに、GATA3とヌクレオソームの複合体（モデル3と4）と、GATA3がない単体のヌクレオソームの3つの系について、GENESISを用いた全原子MDシミュレーションを「富岳」にてそれぞれ1μs以上の計算を行った。

この長時間のシミュレーション計算の結果、クロマチン結合因子とヒストンテールが動的に相互作用する様子を観測することができた（図3-3-9）。H3のN末端のヒストンテールがGATA3のN末ジンクフィンガードメインと強く相互作用することがわかった。また、H3のN末端のヒストンテールはC末ジンクフィンガードメイン近傍まで伸び、C末ジンクフィンガードメインが結合するDNAの主溝の大きさを変化させることがわかった。さらに、H2BのN末端のヒストンテールがGATA3のC末ジンクフィンガードメインと相互作用することがわかった。これまで、ヌクレオソームとその結合因子との相互作用はほとんど報告されておらず、本研究で初めて明らかに示されたことである。またGATA3結合因子とヒストンのDNAに対する結合については、予備的な解析ではあるが、N末ジンクフィンガードメインはDNAの主溝と強く相互作用して非常に安定していたが、C末ジンクフィンガードメインはDNAの主溝に対して揺らぎが大きかった。このことはN末ジンクフィンガードメインの方が、C末ジンクフィンガードメインよりヌクレオソームDNAに強く相互作用していることを示唆している。

以上、クロマチン結合因子GATA3とヌクレオソームの複合体の複合体モデルを構築し、MDシミュレーションを用いて構造変形させることで、電子顕微鏡像による単粒子構造データに合致する原子モデルを構築する計画を達成することができた。そして、全原子MDシミュレーションを行い、電子顕微鏡像では見ることのできなかったGATA3とヒストンテールの相互作用を観測することができた。GATA3の原子揺らぎを解析することで、GATA3とヒストンDNAの結合強度を説明できる可能性がシミュレーション結果から示唆された。

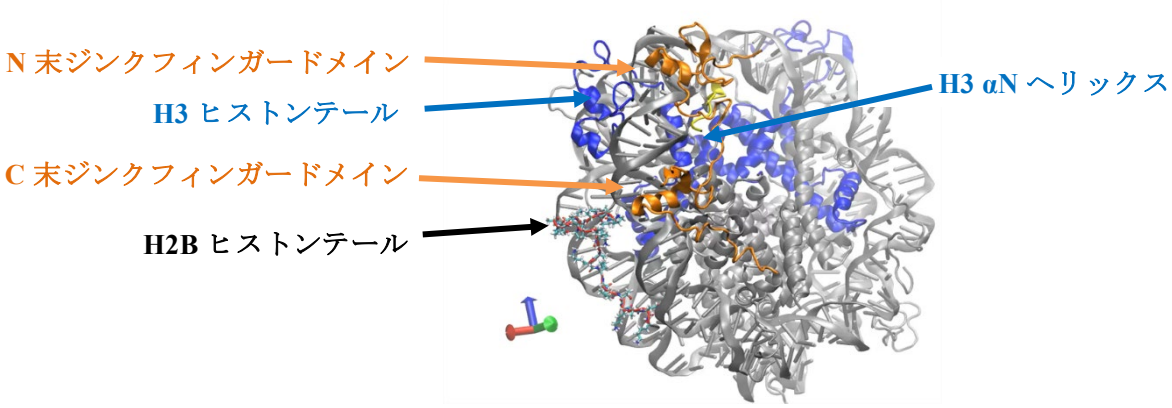


図3-3-9: GATA3（リボンモデル、橙色）のN末ジンクフィンガードメインがH3ヒストンテールと、C末ジンクフィンガードメインがH2Bヒストンテールと相互作用する様子。H3ヒストンテールはリボンモデル、青色で表示、H2Bヒストンテールは原子モデル、主鎖を赤色で表示。

④ ウイルス（名大・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、B型肝炎ウイルスの細胞侵入の膜融合時におこる受容体タンパク質との相互作用の解明を目指している。令和2年度は、エンベロープ単体の構造モデリングを進めた。構造モデリングのプロトコルは未確立であり、HBVエンベロープの分子モデリングは前例がなく、初めての試みとなる。

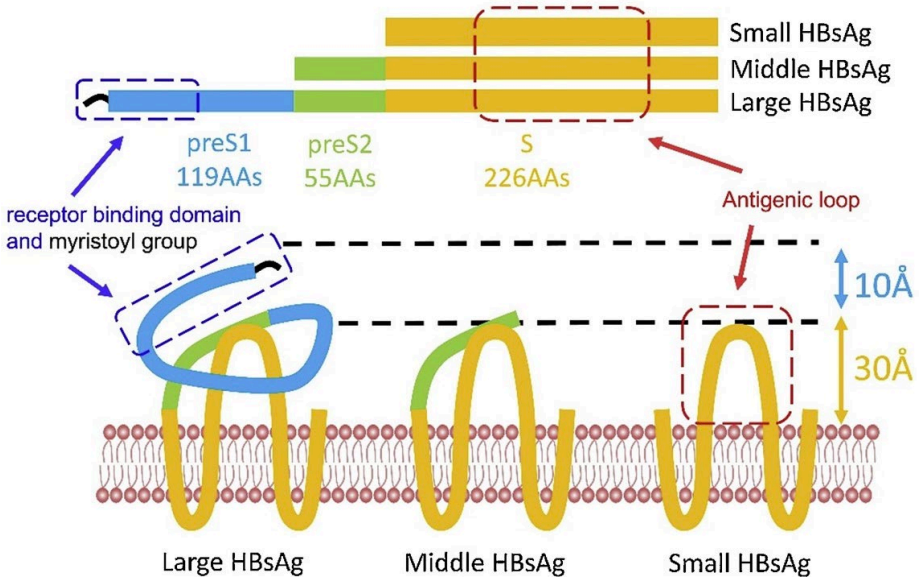


図3-4-1: B型肝炎ウイルスのエンベロープタンパク質3種類 Small (S), Middle (M), Large (L)の比較。上はアミノ酸配列の比較。下は膜貫通領域が共通で、細胞外領域に違いがあることを示している。Cao *et al.*, *Virus Research,* 2019 より改変。

HBVエンベロープ中には図3-4-1にあるように、大きさの異なる3種類のウイルス由来の膜タンパク質

(S, M, L)と脂質分子が含まれている。この3種類の膜タンパク質は共通の膜貫通領域を持っている。これらの膜タンパク質の中で最も小さいのがSタンパク質であり、MはpreS2と呼ばれる領域を追加で持ち、L は preS2に加えてpreS1と呼ばれる領域がさらに含まれたアミノ酸配列を持っている。エンベロープのモデリングを行うためには、これらの膜タンパク質単体の構造が必要であるが、現在までにこれらの実験構造は報告されてない。そこで、まず、膜貫通領域の構造モデリングを実行した。Sタンパク質単体の膜貫通領域の膜中で安定な構造を求めた。

最初に、2次構造予測を行うために、図3-4-2のようにバイオインフォマティクスサーバーConstrained Consensus Topology(CCTOP)を用いて、アミノ酸配列をもとに膜貫通領域の予測を行った。CCTOPでは複数の予測サーバーの結果を同時に用いて最も妥当な膜貫通領域を予測する。図3-4-2からわかるように、4本の膜貫通領域ヘリックスがほとんどの予測法により予測されており、その領域も10-30, 80-100, 170-190, 200-220 の4つの領域で重なっている。異なる予測法により共通の結果が得られたことから、これらの結果は信頼性が高いと期待できる。さらに、4つの膜貫通ヘリックスが存在することは実験でも報告されているため、4本の結果を採用することにした。さらに細胞外領域の2次構造については、水溶タンパク質用の2次構造予測バイオインフォマティクスサーバーI-TISSERを用いて構造予測を行った。以上の結果をもとに、初期構造を作成したものが、図3-4-3(A) である。

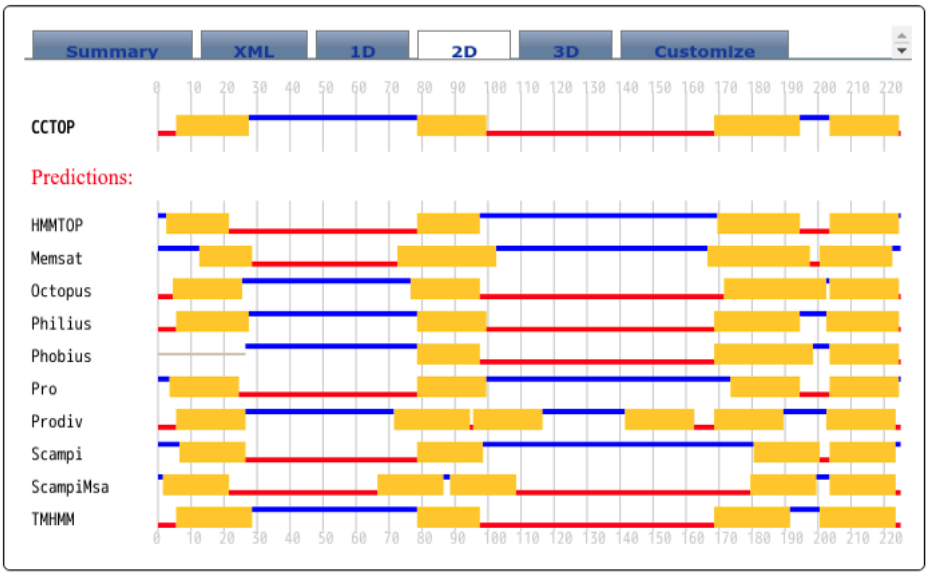


図3-4-2: 膜タンパク質の膜貫通領域予測構造サーバーConstrained Consensus Topology (CCTOP) より、HBV S タンパク質の構造予測の結果。黄色の領域が膜貫通領域を表し、青が膜外内側、赤が膜外外側を示している。一番上のCCTOPが最終的な予測配列である。

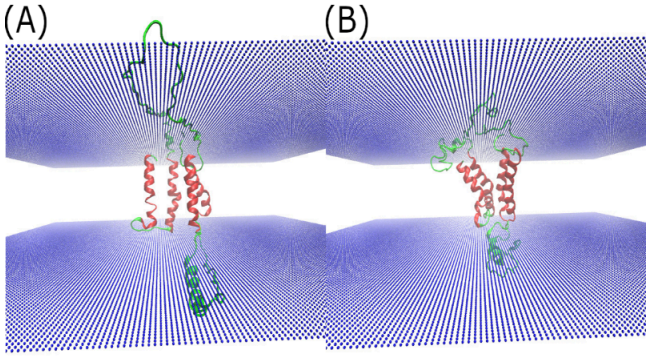


図3-4-3: HBV S単量体のモデリングと膜陰溶媒モデルによるMDで得た安定構造の探索結果。(A) はバイオインフォマティクスサーバの結果より膜内外の2次構造をモデリングした直後のスナップショット。(B) は MD 80 ns実行後の構造。

次に、この初期構造に対して、脂質膜の陰溶媒モデルを用いてMDシミュレーションを行った。これにより膜内領域での膜貫通ヘリックスが凝集し、安定した膜内パッキング構造が得られる。約80nsのMDシミュレーションにより、図3-4-3(B)のような凝集構造を得た。この構造は最初の20ns の間で膜貫通領域のパッキングの緻密化が起こった後、残り約60nsの間、ヘリックス構造に大きな変化がなかったため、少なくとも局所安定な構造を得ることができたと考えられる。また、陽溶媒中でのヘリックスの構造安定性を確認するため、CHARMM-GUIを用いて、POPC二重層膜中に埋め込んだ後、100 ns 以上のMD計算を行った。その結果、モデリングされたSタンパク質は構造変化をせず、安定であることが確認された。このように、S単量体の安定な膜貫通領域構造を得た。

エンベロープ中ではS, M, L の膜タンパク質は、それぞれ別のSタンパク質と結合した2量体構造を持つことが報告されている。よって、2量体構造の予測が必要である。最初に、S2量体構造の予測は真空中のMDシミュレーションにより行った。真空中で行った理由は、2量体会合構造において脂質が間に介在しない限り、そのタンパク質間の相互作用は真空中での相互作用とほぼ等しくなるからである。したがって、チャネル構造のような中空構造をもつ凝集体を作らない場合には、真空中で2量体構造を予測することが計算資源の観点と構造探索の高速さから効率的となる。しかし、完全にフリーな真空中での通常のシミュレーションは非効率的であるため、膜貫通領域の上端、下端の高さをバネで留め、ヘリックス構造に拘束された膜貫通領域間の会合構造を近距離に保つバイアスを付けて探索した。複数のシミュレーションを繰り返し、2量体構造の候補をリストアップした。得られた候補構造のうち、膜貫通領域のヘリックス構造が凝集後に保持され、また、膜貫通領域が膜内に収まり、他の部位が界面から膜外に存在する構造を採用した。以上のように生成されたS2量体構造の1つが図3-4-4である。得られた構造が実際に脂質二重膜中で安定であることを確認するため、CHARMM-GUIを用いて二重膜脂質中への埋め込みを行い、50ns のシミュレーションを行った。図3-4-5に示すように実際の脂質二重膜中でも安定であることがわかった。

図3-4-4: HBS 2量体を真空中で拘束ありの条件下でMDを行い凝集した構造。赤色は膜貫通領域と予測された領域、緑色は膜外領域と予測された領域である。

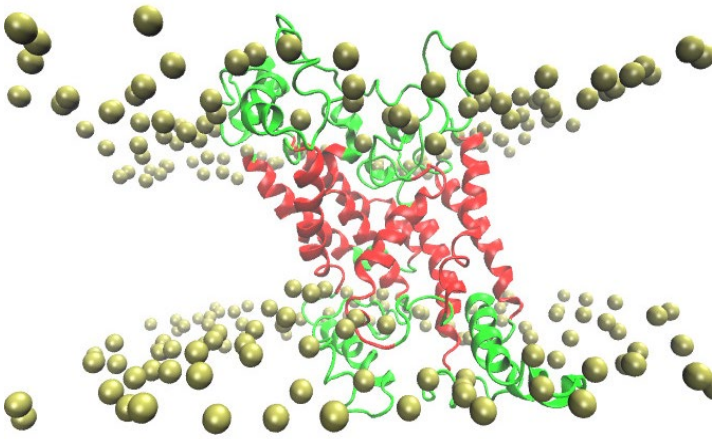


図3-4-5: 真空中で得た2量体構造を脂質二重層膜に入れて30 nsのMDシミュレーションを行った後の構造。黄色の分子は脂質頭部のリン原子である。その他の脂質の原子、水、イオンは省略した。

次に、得られたS2量体構造をベシクル中に埋め込むことでエンベロープ構造の構築を試みた。MやLタンパク質は、Sタンパクと膜貫通領域が共通であるため、Sタンパクの膜への埋め込み後、残基の追加・置換によって導入することができる。エンベロープの概形構造は電子顕微鏡観察より、タンパク2量体間の平均距離や粒子サイズが報告されている。また、エンベロープ中にある膜タンパク質2量体の概数は140-170との報告に従い、POPCベシクルへ140の2量体の埋め込みを行った。2量体の配置後、物理的に接触している脂質分子やイオン、水分子は削除し、電気的中性条件を満たすようにイオンの濃度と個数を調整した。図3-4-6は埋め込んだ直後の構造であり、粒子の重なりなどの構造上の問題は取り除いた。

この構造に対して、エネルギー最小化、温度平衡化、圧力平衡化を含めて30 nsのMD計算を行った後の構造が図3-4-7(左)である。タンパク質の埋め込みの際に、粒子の重なりを防ぐために脂質分子を過剰に削除していることが原因で、同サイズのベシクルを保つのに十分な数の脂質分子が不足して穴が空いていると考えられる。一方で、このような穴の空いた構造が数十ns もの間安定に存在することは、埋め込みを行った膜タンパク質に、初期配置で仮定した曲率を持つベシクルを安定化する効果があることが示唆される。通常、ベシクルの細孔は、高い線張力（～30 pN ）のために、非常に不安定であり、すぐに閉じるか他の高次構造(ディスクなど)へと転移することが予測されるためである。濃厚なSタンパク質の存在により、カプシドを覆うのによいサイズのエンベロープ膜の曲率が選ばれる可能性が示唆されており、興味深い。HBVカプシドサイズを包含するエンベロープの作成には、不足した脂質分子を補う必要があり、細孔の領域に対して、脂質分子を追加で配置した。図3-4-7(中) は実際に青色で表示される脂質分子を細孔の領域に配置した状態の構造である。この後、エネルギー最小化と平衡化MDを30ns行った。得られた構造が図3-4-7(右)である。これにより実際に、細孔のない脂質エンベロープのモデル構造が得られた。図3-4-8は、HBVカプシドをエンベロープ構造の内部に配置することによって作成した、ヌクレオカプシド含有のHBVビリオン構造の分子モデルである。この構造では、カプシドがエンベロープと一部物理的に接触しているため、エネルギー最小化と平衡化計算がさらに必要である。また、膜タンパクとカプシドスパイク領域の相互作用により、膜タンパクが再配置されることも考えられるため、長時間のMD計算が今後必要となる。

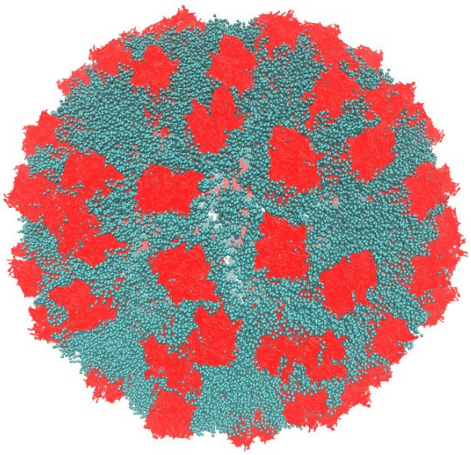


図3-4-6: POPC（空色）ベシクル中にS２量体（赤色）構造を140個埋め込んだ状態の構造。
重なりのある脂質分子を削除したため局所的に脂質分子が疎なところが生じている。

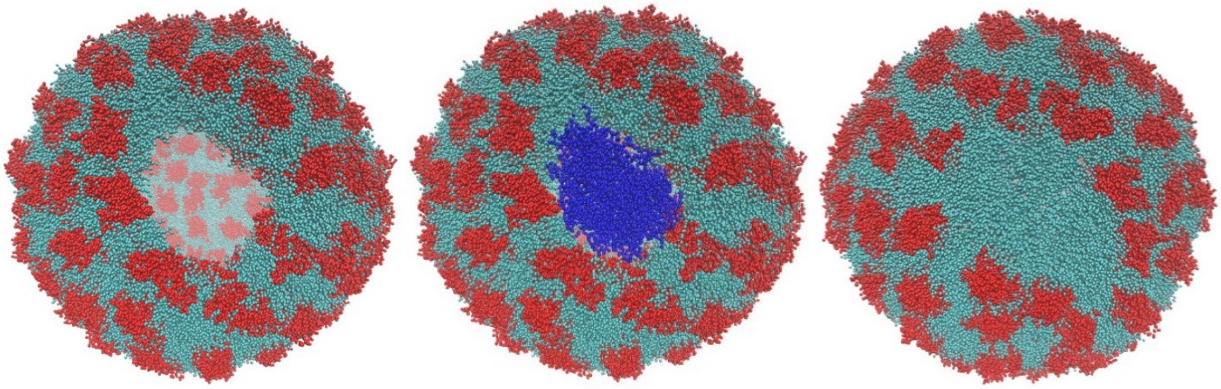


図3-4-7: (左)埋め込み後、30 nsの平衡化MDシミュレーションを行った後の構造。
脂質膜に細孔が形成されている。（中）細孔部分に、脂質パッチ（青色）を追加した状態。（右）脂質パッチの配置後、再度、平衡化MD計算を30 ns行ったあとのスナップショット。細孔のないエンベロープ構造が作成された。

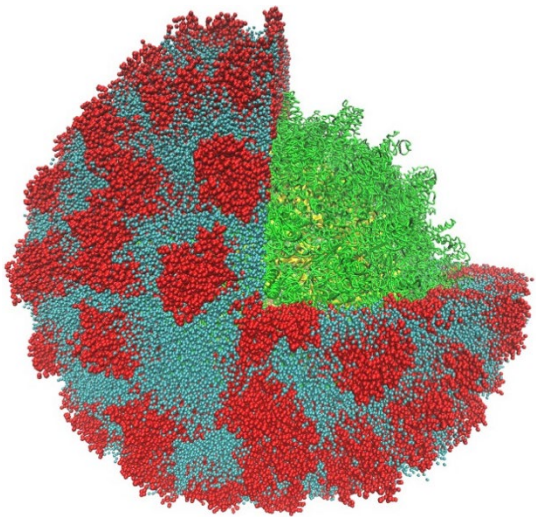


図 3-4-8: S2量体(赤色)を含んだエンベロープ構造に対して、内部にRNA(黄色)含有カプシド(緑色)を配置したスナップショット。内部構造を表示するため、エンベロープの1/4を非表示にした。

⑤ 多剤排出トランスポータの MD シミュレーション（東大・篠田）

大腸菌の生体膜は様々な脂質分子が含まれる混合膜である。本年度は実験の実在データに基づいた新たな大腸菌モデル膜の構築と、そのための新しい脂質分子の力場の開発、そして全原子MDシミュレーションを行い、生体モデル膜との違いを解析した。また、大腸菌モデル膜に多剤排出トランスポーターAcrAB-TolC複合体を埋め込んだ800万原子系の分子モデルを構築し、全原子MDシミュレーションをテストし、安定にシミュレーションが実行できることを確認した。具体的には以下の通りである。

他の生物と同様に大腸菌も環境や細胞周期に応答して細胞膜の脂質組成が変化することが知られているが、その脂質組成の変化がどのように膜タンパク質の機能に影響を及ぼすのかまだ不明な点が多い。今年度、本研究では脂質組成による膜の性質の違いを調べた。実験では、シクロプロパン環をもつ脂質分子の比率で膜タンパク質の輸送能が変化することが近年報告されている。そこで、本研究では、まず大腸菌の野生株（W3110S）の細胞膜について、実験データより得られた脂肪酸組成に基づき、脂質分子の特定とその比率を決定した。大腸菌の主な脂質分子の構造は、極性の高い親水性のヘッドグループと疎水性の2本の脂肪酸から成る。大腸菌の主要なヘッドグループはPEとPGでその比はおよそ8:2である。また脂肪酸は①パルミチン酸(C16:0)、②パルミトレイン酸(C16:1)、③C17　シクロプロパン脂肪酸(C17:0)、④ cis -バクセン酸(C18:1)、⑤C19 シクロプロパン脂肪酸(C19:0)の5種である。これらの脂肪酸の組成比率は、順に47%、5%、38%、2%、8%である。これらのことから表3-5-1のような7種の脂質分子の組成で膜を構築することとした。これらの中でPMHPE　(1-palmitoyl-2-cis-11,12-methylene-hexadecanoic-acid-sn-glycero-3-phosphoethanolamine　)　と　DMHPE　(　1-cis-9,10-methylene-hexadecanoic-acid-2-cis-11,12-methylene-hexadecanoic-acid-sn-glycero-3-phosphoethanolamine) の力場が無いため、まずこれらの分子の力場の作成を行った（図3-5-1）。

野生型（W3110S 株）	
脂質分子	比率（%）
PMPE	52
PMPG	20
PMHPE	12
DMHPE	4
VYPE	2
PVPE	2
PYPE	8

表 3-5-1: 大腸菌モデル膜（W3110S 株）の構成脂質分子とその比率。

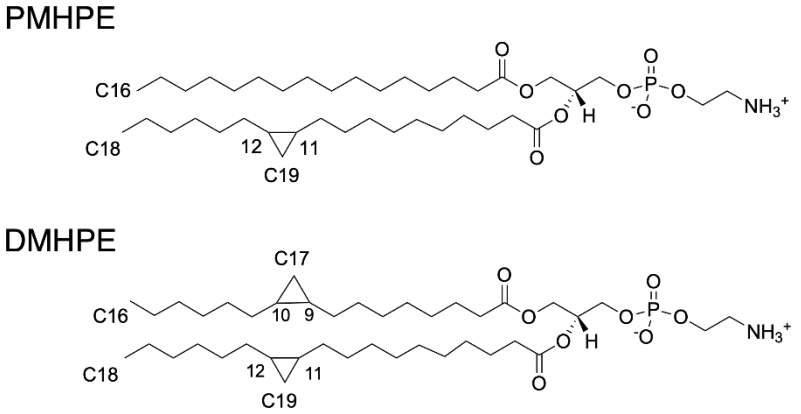


図 3-5-1: 新たに力場開発した脂質分子。

次にCHARMM-GUIプログラムを用いて二重膜の構築を行った。CHARMM-GUIではPMHPEやDMHPEのテンプレートとなる脂質分子の構造が存在しないため、これらの脂質分子のかわりに二本の脂肪酸が炭素数18のDSPEを用いて脂質二重層膜をモデリングし、後でDSPE をPMHPEやDMHPEに置換した。初めに、表3-5-1の比率でトータル100脂質分子を1unitとした膜をモデリングし、それを16倍することで1600分子の膜を構築した。今回、膜の構築にはCHARMM-GUIの他に、MemBuilderIIやMemGenプログラムなども検討した。MemBuilderIIはCHARMM-GUIと同様に、用いることのできる脂質分子が限られているが、その数はCHARMM-GUIと比べてかなり少ないため今回の膜モデルでは使用しなかった。一方、MemGenは新しい脂質分子の構造をpdbなどの構造ファイルから読み込むことができるためモデリング自体は簡便にできることがわかったが、CHARMM-GUIやMemBuilderIIで行われる、初期の構造緩和の工程はない。このことから、今回はCHARMM-GUIに存在しない脂質分子の比率（PMHPE、DMHPEの16%）が小さかったため、CHARMM-GUIを用いたが、比率が大きい場合は、MemGenでモデリングし初期の構造緩和のプロセスから行って膜モデルを構築する方が効率が良いことがわかった。

野生型（LM301 株）	
脂質分子	比率（%）
PMPE	39.4
PMPG	11.1
PVPE	15.4
PYPE	4.0
PYPG	8.9
PVPE	4.0
QMPE	9.0
VYPE	8.2

表3-5-2:　大腸菌モデル膜（LM301株）の構成脂質分子とその比率。

モデリングが完了した大腸菌膜（W3110S株）はenergy minimization後、構造緩和のシミュレーションを経て、1μsの計算を初期運動量を変えて3本行った。

最初に、膜全体の電子密度プロファイル（図3-5-2）を計算し、これまでに行ったMDシミュレーションによって得られたPOPE膜、PMPE膜、大腸菌膜（LM301株）の場合と比較した。POPEは生体膜のモデル脂質分子、PMPEは大腸菌の主要な脂質分子（シクロプロパン環をもつ）、大腸菌膜（LM301株）はシクロプロパン環をもつ脂質分子比率が全体で50%の大腸菌膜（表3-5-2）である。大腸菌膜（W3110S株）のシクロプロパン環をもつ脂質分子比率は88%である。図3-5-2をみると、POPE膜と比べてシクロプロパン環をもつ脂質分子を含む膜は全体的にブロードになっていることがわかる。

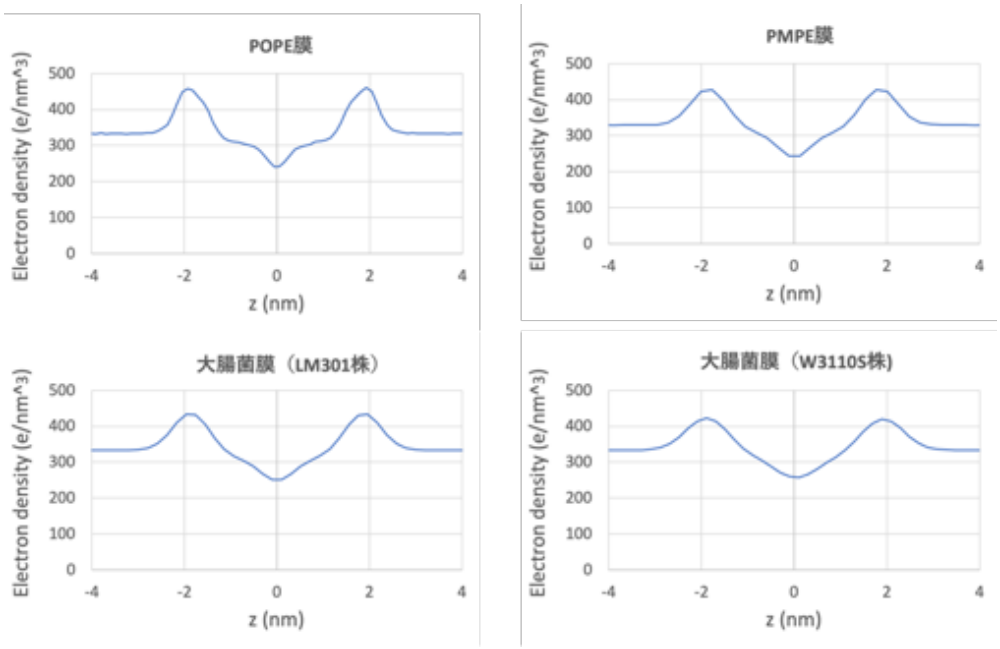


図3-5-2: 膜の電子密度プロファイル。横軸zは膜に垂直な方向を示す。膜の中心をz=0 とする。

膜	シクロプロパン環比率 (%)	膜の厚さ (nm)
POPE 膜	0	3.92±0.01
PMPE 膜	100	3.55±0.01
大腸菌膜 野生型（LM301 株）	50	3.93±0.01
大腸菌膜 野生型（W3110S 株）	88	3.78±0.07

表3-5-3: 膜の厚さとシクロプロパン環比率。

次にこの電子密度プロファイルから膜の厚さ（表3-5-3）を計算した。膜の厚さは電子密度の最大密度間の距離である。表3-5-3を見ると、大腸菌モデル膜ではシクロプロパン環の比率が上がると膜の厚さが小さくなっていることがわかる。

最後に各膜において、PMPE分子の拡散係数を調べた。拡散係数はMean square displacement（MSD）の傾きからアインシュタインの関係式により計算することができる。各膜に対してPMPE分子のMSDの時間変化を図3-5-3に示す。参考のためPOPE膜についても示した。どの膜においてもMSDがなめらかに得られおり、このグラフから傾きを求めることができる。傾きは、最初の速い拡散を除くため30～50nmの区間から計算した。計算した拡散係数は、POPE膜（POPE分子）は2.34±0.06 ×10^{-8} m^2/s、PMPE膜では2.28±0.06 ×10^{-8} m^2/s、LM301膜では1.76±0.02 ×10^{-8} m^2/s、W3110S膜では1.92±0.02 ×10^{-8} m^2/sとなった。PMPE分子を含む膜で比較すると、わずかではあるが有意差が見られ、シクロプロパン環の比率が上がるにしたがって拡散係数が大きくなっていること、つまりPMPE分子が動きやすくなっていることがわかった。大腸菌は細胞周期によってシクロプロパン環を含む脂質の比率が上がることが実験から示唆されており、シクロプロパン環を含む脂質の運動性の変化が細胞分裂へ向かう周期の膜の膨張に関係する可能

性が示唆された。

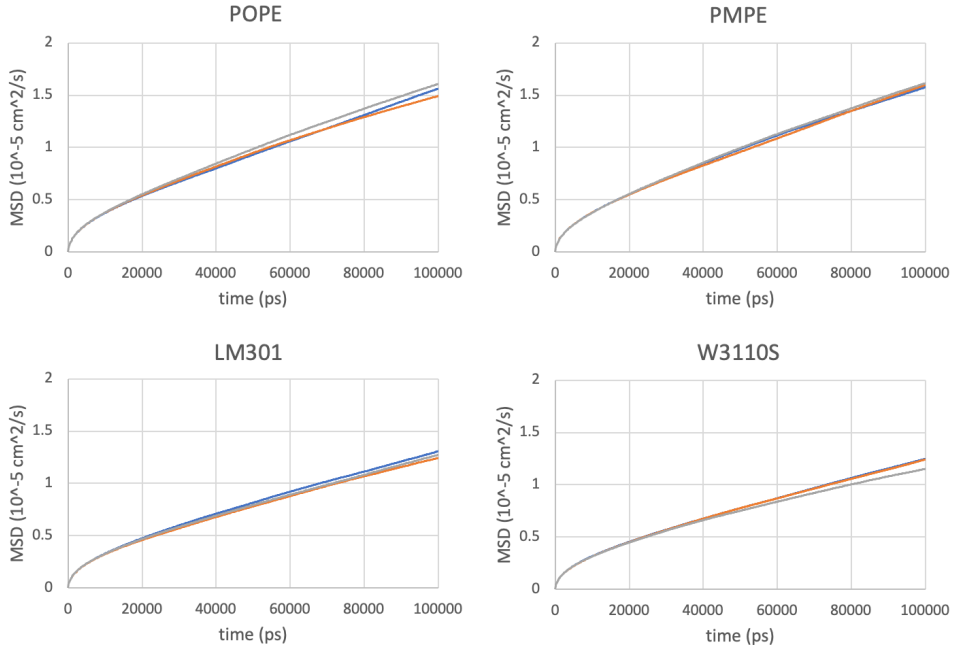


図3-5-3: Mean square displacement (MSD) の時間変化。3本のトラジェクトリによる結果を色分けした。

多剤排出トランスポーター複合体ー大腸菌モデル膜系（800万原子）の構築は以下のように行った。

多剤排出トランスポーター複合体の構造はcryo-EMにより取得された構造（PDBID：5NG5）を元に、AcrBの部分をより分解能の高いX線結晶構造（PDBID：4DX7）に置き換え、missing residueを補って作成した。基質は削除した。大腸菌膜は6000分子からなる二重膜を作成し、トランスポーターを埋め込む前に、平衡化した構造を得るための1μsのMDシミュレーションを行った。平衡化した二重膜を細胞質側の内膜と細胞質外側の外膜とし、トランスポーター複合体は、外膜にTolC+AcrAを、内膜にAcrA-AcrB-AcrZ複合体をそれぞれ別に埋め込んでから、AcrAを重ね合わせるように統合した。Na^+、Cl^-は150 mM、Mg^{2+}は約7 mM加えた（図3-5-4）。

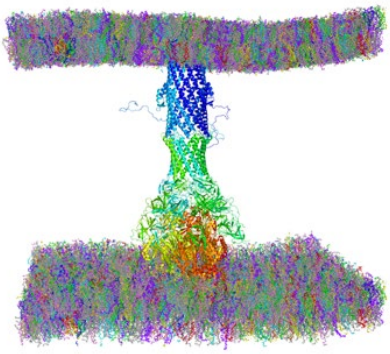


図3-5-4: 大腸菌モデル膜に埋め込んだ多剤排出トランスポーターAcrAB-TolC複合体。大腸菌モデル膜は脂質分子の種類によって色分けして表示している。

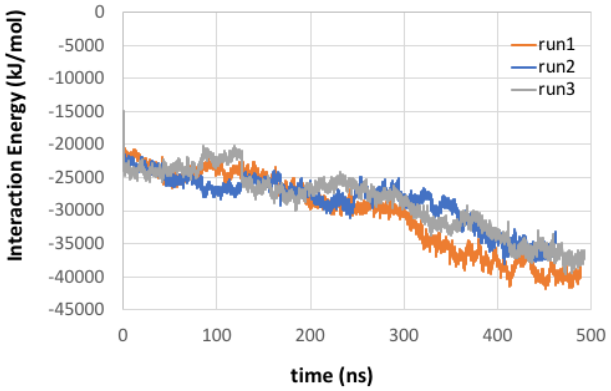


図3-5-5: 多剤排出トランスポーターAcrAB-TolC複合体と大腸菌モデル膜との相互作用。
初期運動量を変えた3つのシミュレーションの結果を表示している。

当初、以前行ったトランスポーター複合体-POPE膜と同様の手順で構造緩和を行なったが、脂質分子とタンパク質との衝突により計算がストップしてしまうことがあるため、タイムステップを4fsから2fsに変更し、更に、構造緩和に向いているBerendsenの方法による圧力制御も2段階行い、タンパク質の動きに制限をかけた脂質分子と水分子の構造緩和100nsの計算の後、全系においてもBerendsen法で100nsのシミュレーションを行った。またその後のParrinello-Rahman法による解析データ取得のためのシミュレーション（プロダクションラン）もタイムステップを2fsに変更した。図3-5-5はプロダクションランにおける、多剤排出トランスポーター複合体と膜全体との相互作用エネルギーの時間変化である。ゆるやかに減少し、まだ収束していないことが見て取れる。

（４）プロジェクトの総合的推進

本年度は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 2 年度は、5 月 26 日（火）、6 月 16 日（火）、7 月 14 日（火）、9 月 8 日（火）、10 月 13 日（火）、11 月 10 日（火）、12 月 8 日（火）、2 月 9 日（火）、3 月 9 日（火）に開催した。1 月は後述する Online Workshop を 6 日と 7 日に開催したため、プロジェクトミーティングは開催しなかった。ここでは、プロジェクト参加者（PI だけでなく参加希望する学生・ポスドクも含めて）を集めて、毎月 1 チームずつが各自の研究（計算手法・モデル・応用計算）について紹介し、その内容について議論した。オープンなセッションの後で、PI だけを集めて短く、事務連絡とプロジェクトの総合的推進に必要な議論を行った。

Slack 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。ミーティングや各自 Scientific Talk のタイトル・要旨などの連絡だけでなく、「富岳」や HPCI 課題での計算リソースの利

用についての深い議論もできた。この報告書も含めて、様々な事務書類の提出・修正等の連絡も主にはSlack を用いて行った。

Slack の利点はテキストベースの議論が速やかにできることで、何らかの相談があった場合にチャネル上で議論を進めて結論を出すことができる。一方で、書類等を整理することには向いていないので、重要な書類等の共有は理研が利用する Box を通じて行った。ここでは、プロジェクト参加者が全員見ることができる権限を与えているので、秘匿すべき情報（例えば研究員の給与など）が含まれる書類は別の形で共有するなど、情報管理についても注意を払った。すでに公開・発表している研究成果等については、プロジェクトの Box フォルダーで共有しているが、進行中の共同研究等については関係者のみが閲覧・記載できる Box フォルダー（共同研究フォルダー）を用意して情報管理を徹底した。

HPCI や「富岳」の計算リソースの利用については複数のチームで共有しているため、管理が難しい。そのため、計算リソースの利用量を毎月確認・報告する約束をして、その情報を 1 枚のシートに集約した。4 月から翌年 3 月までの利用を見据えて複数の計算資源を利用してきたため、最終的には計画した計算リソースのほとんどを利用することができた。令和 3 年度は「富岳」の計算リソースに集約されるため、利用量の管理はかなりやりやすくなると考えられる。

本研究課題では、GENESIS に MD の機能・モデルを集約して「富岳」での大規模計算を行うという計画であるため、理研以外のチームがこれまでに開発してきた計算手法・モデルを GENESIS に導入するという作業が必要になった。主には、

・高田（京都大学）チームが開発してきた粗視化分子モデル CafeMol と粗視化 MD 計算

・篠田（名古屋大学）チームが開発してきた粗視化モデル SPICA と粗視化 MD 計算

・河野（量研機構）チームが開発してきた ABMD（Adaptive Bias Molecular Dynamics）計算

を GENESIS に導入した。それぞれについては、各研究課題の実施内容の中に詳細を記したため繰り返さない。いずれにおいても理研における GENESIS 開発担当者が CafeMol, SPICA, ABMD それぞれの開発者とミーティングや Slack での議論を繰り返して、問題を特定して一つ一つ解決するという地道な作業を行った。令和 2 年度の成果としてこれらの計算手法・モデルがほぼ「富岳」上で利用可能になっている。細かいチューニングやバグの修正は必要であるが令和 3 年度中に速やかに解決して、まずはプロジェクトの研究成果を達成することを目指す。その上で、GENESIS の次期バージョン等に導入してプロジェクト外の研究者・学生・企業研究者等にも広く公開していく。

まとめると、新型コロナウイルス禍ではあるもののオンラインのツールを効率的に利用することで情報交換・情報共有を実現し、プロジェクトの総合的推進につなげることができたと考えられる。令和 3 年度も状況は大きく変化することはないと思われるため、令和 2 年度の経験を生かしてプロジェクトを推進していく。

２－３．活動（研究会の活動等）

令和 2 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サイトでは実施できなかった。そのため、オンラインで以下の 2 つのイヴェントを開催した。

・The 1st Fugaku Bio-supercomputing Workshop on Cellular-scale Molecular Dynamics Simulations (2021 年 1 月 6, 7 日)

まず、前者については、国内外のアクティブな研究者を集めて細胞スケールのMDシミュレーションの現状と将来を議論した。オンラインで国際ワークショップを開催する場合、やはり気を遣うのは日本と米国、あるいはヨーロッパの大きな時差である。そこで今回はこの問題を回避するために主に日本とアジアの研究者でミーティングを開催することに決定した。講演者は以下の通りである。

Xuhui Huang (Hong Kong University Science and Technology, Hong Kong)
Juyong Lee (Kangwon National University, Korea)
Yuji Sugita (RIKEN)
Yi Qin Gao (Peking University, China)
Jejoong Yoo (Sungkyunkwan University, Korea)
Shoji Takada (Kyoto University)
Peter Bond (A*STAR, Singapore)
Yi Wang (Chinese University of Hong Kong, Hong Kong)
Wataru Shinoda (Nagoya University)
Wenning Wang (Fudan University, China)
Shigeyuki Matsumoto (RIKEN)
Wei Han (Peking University Shenzhen Graduate School, China)

（下線は海外の招待講演者）

最終的に85名が参加登録をして、科学的にハイレベルの講演・議論がなされるとともに大きなトラブルなしに2日間のワークショップを終了することができた。日本以外のアジアの研究者のレベルもどんどん向上していて今後もこのような交流を継続していくことが重要であると痛感した。今回海外から招待した講演者はいずれも米国やヨーロッパの有名研究室でポスドクや大学院学生として所属していた経歴をもっている。従って、彼らの研究室からも今後優秀な若手研究者が育成されていくことは間違いないので人材獲得の意味でもアジアの交流は重要であるし、日本の同世代の研究者・大学院生にも強い刺激を与えたようである。

・GENESIS講習会（筑波大学Cygnusを用いたハンズオン）(2021年1月29日)

GENESIS講習会については、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の開発者であるJ. Jung博士、岩橋(小林)千草博士（ともに理研R-CCS）が参加したため、ここに記載する。ハンズオンのワークショップであるため12名の定員を設けたが、定員数の多数の申し込み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、大学院学生からの申し込みも半数近くあった。GENESISはフリーソフトウェアとして公開しており、誰でも利用することができる。しかし、初めて利用する人、スパコン等に慣れていない人にとっては、我々開発者に見えない障壁があるため、このような機会を通じてそれが解消され、さらなるGENESISの普及につながること期待される。

２－４．実施体制

実施項目	実施場所	担当責任者
（1）全原子・粗視化MDの最適化と並列化	〒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 〒464-8601 愛知県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	杉田有治 篠田渉
（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	〒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 〒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	高田彰二 松永康佑
（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	〒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 〒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 〒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 〒464-8601 愛知県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	杉田有治 高田彰二 河野秀俊 篠田渉

	〒153-8904 東京都目黒区駒場4-6-1 国立大学法人　東京大学	篠田恵子
（４）プロジェクトの総合的推進	〒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

学会等における口頭・ポスター発表（国際会議）

1　Use of enhanced conformational sampling for the analysis of protein-ligand binding processes, Yuji Sugita (RIKEN Cluster for Pioneering Research, RIKEN Center for Computational Science, RIKEN Center for Biosystems Dynamics Research), Virtual Workshop "Biomolecular Interactions in Cellular Environments", in Telluride Science Research Center,2020 年 8 月，国際

2　Integrative modeling of protein dynamics from single-molecule experiments and molecular dynamics simulations, Yasuhiro Matsunaga(Saitama University), Tomonori Ogane(Saitama University), 第 58 回日本生物物理学会年会, 2020 年 9 月，国際

3　Integrative modeling of protein dynamics from experimental data and molecular dynamics simulations, Yasuhiro Matsunaga(Saitama University), 第 43 会日本分子生物学会年会, 2020 年 12 月，国際

4　Multi-scale MD simulations using GENESIS on Fugaku, Yuji Sugita(RIKEN CPR, RIKEN BDR, RIKEN R-CCS), The 1st Fugaku Bio-supercomputing Workshop on Cellular-Scale Molecular Dynamics Simulations, 2021 年 1 月，国際

5　Nucleosome and transcription factor dynamics, Shoji Takada(Kyoto University), The 1-st Fugaku Bio-supercomputing Workshop on Cellular-scale Molecular Dynamics Simulations, 2021 年 1 月，国際

6　Computer simulations of biomacromolecular self-assembly using the SPICA coarse-grained model, Wataru Shinoda(Nagoya University), The 1st Fugaku Bio-supercomputing Workshop on Cellular-scale Molecular Dynamics Simulations, 2021 年 1 月，国際

7　Intrinsic Conformational Flexibility of SARS-CoV-2 Spike Protein Simulated on Fugaku, Yuji Sugita(RIKEN CPR, RIKEN BDR, RIKEN R-CCS), 2021 R-CCS Symposium, 2021 年 2 月，国際

8　Replica-exchange simulations on the conformational dynamics of spike protein on the surface of SARS-CoV-2, Yuji Sugita(RIKEN CPR, RIKEN BDR, RIKEN R-CCS), Molecular Basis of Proteinopathies, 2021 年 2 月，国際

9　Diverse nucleosome dynamics revealed by molecular dynamics simulations, Shoji Takada(Kyoto University), International Webinar Multiscale Simulation & Mathematical Modelling of Complex Biological Systems, 2021 年 3 月，国際

学会等における口頭・ポスター発表（国内学会）

1　Coarse-grained local potential for disordered proteins optimized via Bayesian inference, Azuki Mizutani(Kyoto University), Giovanni Brandani(Kyoto University), Shoji Takada(Kyoto University), 第 20 回日本蛋白質科学学会年会, 2020 年 7 月, 国内

2　Bayesian analysis of high-speed atomic force microscopy images towards integrative modeling of biomolecular dynamics, Tomonori Ogane(Saitama University), Yasuhiro Matsunaga(Saitama University), 第 20 回日本蛋白質科学会年会, 2020 年 7 月, 国内

3　Infinite-swap limit in generalized replica exchange with solute tempering, Hiroaki Nakayama(Saitama University), Yasuhiro Matsunaga(Saitama University), 第 20 回日本蛋白質科学会年会, 2020 年 7 月, 国内

4　Protein-protein complex dissociation simulated by Parallel Cascade Selection Molecular Dynamics, Yoshiki MIyazawa, Dun P. Tran, Kazuhiro Takemura, Akio Kitao(Tokyo Institute of Technology), 第 20 回日本蛋白質科学会年会, 2020 年 7 月, 国内

5　Simulating the association/dissociation process of flexible protein complexes, Tran P. Duy, Akio Kitao(Tokyo Institute of Technology), 第 20 回日本蛋白質科学会年会, 2020 年 7 月, 国内

6　Unguided Binding MD of Protein-Protein Complexes by PPI-ColDock, Kazuhiro Takemura, Akio Kitao(Tokyo Institute of Technology), 第 20 回日本蛋白質科学会年会, 2020 年 7 月, 国内

7　Bayesian inference and Iterative Boltzmann approach to coarse-grained local potential of disordered proteins, Azuki Mizutani(Kyoto University), Shoji Takada(Kyoto University), 第 58 回日本生物物理学会年会, 2020 年 9 月, 国内

8　Molecular dynamics simulation of the diffusion of membrane proteins on vesicles, Diego Ugarte(Kyoto University), Shoji Takada(Kyoto University), 第 58 回日本生物物理学会年会, 2020 年 9 月, 国内

9　計測とシミュレーションを統合したタンパク質ダイナミクスのモデリング, 松永康佑(埼玉大学), 第 93 回日本生化学会大会, 2020 年 9 月, 国内

10　DNA binding mechanisms of the p53 C-terminal domain elucidated by MD simulation, Yuta Taira, Dun Tran Akio Kiao(Tokyo Institute of Technology), 第 58 回日本生物物理学会年会, 2020 年 9 月, 国内

11　Investigating the dissociation process of DBD-p53/DNA complex by PaCS-MD and MSM, Sobeh Mohamed Marzouk, Akio Kitao(Tokyo Institute of Technology), 第 58 回日本生物物理学会年会オンライン学会, 2020 年

9 月, 国内

12　Kinetic rate evaluation for protein complexes by PaCS-MD/MSM, Yoshiki Miyazawa, Duy Phuoc Tran, Kazuhiro Takemura, Akio Kitao(Tokyo Institute of Technology), 第 58 回日本生物物理学会年会, 2020 年 9 月, 国内

13　コンピュータで体をつくる分子をみる, 松永康佑(埼玉大学), 埼玉大学ハイグレード理数高校生育成プログラム(HiGEPS), 2020 年 9 月, 国内

14　タンパク質―リガンド結合の自由エネルギー計算法の開発と応用, 尾嶋拓(理化学研究所生命機能科学研究センター), CBI 学会 2020 年大会, 2020 年 10 月, 国内

15　細胞環境はどのようにタンパク質の構造・ダイナミクス・機能に影響を与えるか？, 杉田有治(理化学研究所開拓研究本部、理化学研究所計算科学研究センター、理化学研究所生命機能科学研究センター), CBI 学会 2020 年大会, 2020 年 10 月, 国内

16　中性子散乱データを活用した超分子のモデリング, 河野秀俊(量子科学技術研究開発機構量子生命科学領域), CBI 学会 2020 大会, 2020 年 10 月, 国内

17　All-atom molecular dynamics simulations of spike protein on the surface of SARS-CoV-2 in solution, Yuji Sugita(RIKEN Cluster for Pioneering Research, RIKEN Center for Computational Science, RIKEN Center for Biosystems Dynamics Research), RIKEN-NRC HPC Workshop, 2020 年 10 月, 国内

18　新型コロナウイルス 表面のタンパク質動的構造予測, 杉田有治(理化学研究所開拓研究本部、理化学研究所計算科学研究センター、理化学研究所生命機能科学研究センター), 第 9 回 JCAHPC セミナー, 2020 年 10 月, 国内

19　相分離への計算科学的アプローチ, 河野秀俊(量子科学技術研究開発機構量子生命科学領域), 第 6 回蛋白質工学研究会ワークショップ, 2020 年 11 月, 国内

20　計算機シミュレーションで細胞の中を観る, 杉田有治(理化学研究所開拓研究本部、理化学研究所計算科学研究センター、理化学研究所生命機能科学研究センター), 京都大学 MACS コロキアム, 2020 年 11 月, 国内

21　New parallel computing algorithm of molecular dynamics for extremely huge scale biological systems, Jaewoon Jung(RIKEN R-CCS, RIKEN CPR), 小林千草, Cheng Tan(RIKEN R-CCS), 笠原健人(RIKEN BDR), Michael Feig(Michigan State University), 杉田有治(RIKEN R-CCS, RIKEN CPR, RIKEN BDR), 第 34 回 分子シミュ

レーション討論会, 2020 年 12 月, 国内

22　全原子分子動力学シミュレーションを用いた、ヌクレオソーム安定性に対するねじれストレス影響の解析, 石田恒、河野秀俊(量子科学技術研究開発機構量子生命科学領域), 量子生命科学会第 2 回大会, 2020 年 12 月, 国内

23　Understanding the Dinucleosome Structural Dynamics at All Atom Resolution, Kumar, Amarjeet(QST), Matsumoto, Atsushi(QST), Kono, Hidetoshi(QST), 量子生命科学会第 2 回大会, 2020 年 12 月, 国内

24　光活性化アデニル酸シクラーゼの量子/古典ハイブリッド分子シミュレーション, 田口　真彦(量子科学技術研究開発機構), 櫻庭　俊(量子科学技術研究開発機構), Chan Wai Soon(量子科学技術研究開発機構), 河野　秀俊(量子科学技術研究開発機構), 量子生命科学会, 2020 年 12 月, 国内

25　定量的粗視化力場 SPICA の DNA モデルへの拡張: 脂質ナノ粒子の構造予測, 田中 裕貴, 宮崎 裕介, Akhil Pratap Singh, 篠田 渉(名大院工), 第 34 回分子シミュレーション討論会, 2020 年 12 月, 国内

26　薬剤分子の B 型肝炎ウイルスへの透過機構の探索, 弦巻 周平, 浦野 諒, 藤本 和士, 篠田 渉(名大院工), 岡崎 進(東大院新領域), 第 34 回分子シミュレーション討論会, 2020 年 12 月, 国内

27　スーパーコンピュータが明らかにする蛋白質の動的構造と機能, 杉田有治(理化学研究所開拓研究本部、理化学研究所計算科学研究センター、理化学研究所生命機能科学研究センター), 理研・東北大学連携シンポジウム「計測科学が拓く生命科学の新展開」, 2020 年 12 月, 国内

28　ヌクレオソームポテンシャルからクロマチンポテンシャル, 河野秀俊(量子科学技術研究開発機構), 金沢大学異分野融合セミナー, 2021 年 2 月, 国内

29　膜蛋白質の分子動力学シミュレーション, 篠田恵子(東京大学生物生産工学研究センター),第 5 回新学術「高速分子動画」オンラインセミナー, 2021 年 2 月, 国内

30　新型コロナウイルス表面のタンパク質動的構造予測, 杉田有治(理化学研究所開拓研究本部、理化学研究所計算科学研究センター、理化学研究所生命機能科学研究センター), HPCI フォーラム, 2021 年 3 月, 国内

31　「富岳」を用いた新型コロナウイルス表面のスパイクタンパク質動的構造予測, 杉田有治(理化学研究所開拓研究本部、理化学研究所生命機能科学研究センター、理化学研究所計算科学研究センター), 近畿化学協会コンピュータ化学部会, 2021 年 3 月, 国内

学会誌・雑誌等における論文掲載

1　Free energy profile of permeation of Entecavir through Hepatitis B virus capsid studied by molecular dynamics calculation, Kazushi Fujimoto(Department of Materials Chemistry, Nagoya University, Nagoya, Japan), Motohiro Fukai(Department of Materials Chemistry, Nagoya University, Nagoya, Japan), Ryo Urano(Department of Materials Chemistry, Nagoya University, Nagoya, Japan), Wataru Shinoda(Department of Materials Chemistry, Nagoya University, Nagoya, Japan), Tetsuya Ishikawa(Department of Radiological and Medical Laboratory Sciences, Nagoya University, Nagoya, Japan), Katsumi Omagari(Department of Virology and Liver Unit, Nagoya City University, Nagoya, Japan), Yasuhito Tanaka(Department of Virology and Liver Unit, Nagoya City University, Nagoya, Japan), Atsushi Nakagawa(Institute for Protein Research, Osaka University, Suita, Japan), Susumu Okazaki(Department of Materials Chemistry, Nagoya University, Nagoya, Japan), Pure and Applied Chemistry, Volume 92, Issue 10, Pages 1585−1594(2020), 2020 年 4 月

2　Molecular dynamics simulation of proton-transfer coupled rotations in ATP synthase F-O motor, Shintaroh Kubo(Kyoto University), Toru Niina(Kyoto University), Shoji Takada(Kyoto University), Scientific Reports, vol.10, 8225, pp.1-6(2020), 2020 年 5 月

3　A new lipid force field (FUJI), Nozomu Kamiya(Fujitsu Limited Bio-IT R&D Office, Research Center for Advanced Science and Technology), Megumi Kayanuma(Research Center for Computational Design of Advanced Functional Materials), Hideaki Fujitani(Laboratory for Systems Biology and Medicine, Research Center for Advanced Science and Technology, The University of Tokyo), Keiko Shinoda(Laboratory for Systems Biology and Medicine, Research Center for Advanced Science and Technology, The University of Tokyo), Journal of Chemical Theory and Computation, Vol 16 (6), pp.3664-3676 (2020), 2020 年 5 月

4　Nucleosomes as allosteric scaffolds for genetic regulation, Shoji Takada(Kyoto University), Giovanni B Brandani(Kyoto University), Cheng Tan(RIKEN), Current Opinion in Structural Biology, Vol.62, pp.93-101(2020), 2020 年 6 月

5　Exact long-range Coulombic energy calculation for net charged systems neutralized by uniformly distributed background charge using fast multipole method and its application to efficient free energy calculation, Ryo Urano(Nagoya University), Wataru Shinoda(Nagoya University), Noriyuki Yoshii(Nagoya University), Susumu Okazaki(Nagoya University), The Journal of Chemical Physics, Vol.152, pp.244115 (2020), 2020 年 6 月

6　Pivotal Role of Interdigitation in Interleaflet Interactions: Implications from Molecular Dynamics Simulations, Sangjae Seo(Nagoya University), Michio Murata(Osaka University), Wataru Shinoda(Nagoya University), J. Phys. Chem. Lett. 11, 5171-5176 (2020), 2020 年 7 月

7　Nucleosome allostery in pioneer transcription factor binding, Cheng Tan(RIKEN), Shoji Takada(Kyoto University), Proceedings of the National Academy of Sciences, Vol.117, pp.20586-20596 (2020), 2020 年 8 月

8　CHARMM-GUI Free Energy Calculator for Absolute and Relative Ligand Solvation and Binding Free Energy Simulations, Seonghoon Kim(Department of Biological Sciences, Chemistry, Bioengineering, and Computer Science and Engineering, Lehigh University), Hiraku Oshima(RIKEN Center for Biosystems Dynamics Research,), Han Zhang(Department of Biological Sciences, Chemistry, Bioengineering, and Computer Science and Engineering, Lehigh University), Nathan R. Kern(Department of Biological Sciences, Chemistry, Bioengineering, and Computer Science and Engineering, Lehigh University), Suyong Re(Laboratory for Biomolecular Function Simulation, RIKEN Center for Biosystems Dynamics Research), Jumin Lee(Department of Biological Sciences, Chemistry, Bioengineering, and Computer Science and Engineering, Lehigh University), Benoit Roux(Department of Biochemistry and Molecular Biology, The University of Chicago), Yuji Sugita(RIKEN Center for Biosystems Dynamics Research, RIKEN Center for Computational Science, RIKEN Cluster for Pioneering Research), Wei Jiang(Leadership Computing Facility, Argonne National Laboratory), Wonpil Im(Department of Biological Sciences, Chemistry, Bioengineering, and Computer Science and Engineering, Lehigh University), J. Chem. Theory Comput. 2020, 16, 7207-7218, 2020 年 10 月

9　Particle Filter Method to Integrate High-Speed Atomic Force Microscopy Measurements with Biomolecular Simulations, Sotaro Fuchigami(Kyoto University), Toru Niina(Kyoto University), Shoji Takada(Kyoto University), Journal of Chemical Theory and Computation, 16, 10, 6609−6619 (2020), 2020 年 10 月

10　Coarse-grained implicit solvent lipid force field with a compatible resolution to the C-alpha protein representation, Diego Ugarte La Torre(Kyoto University), Shoji Takada(Kyoto University), The Journal of Chemical Physics, Vol.153, pp.205101 (2020), 2020 年 11 月

11　The HMGB chromatin protein Nhp6A can bypass obstacles when traveling on DNA, Kiyoto Kamagata(Tohoku University), Kana Ouchi(Tohoku University), Cheng Tan(RIKEN), Eriko Mano(Tohoku University), Sridhar Mandali(University of California), Yining Wu(Tohoku University), Shoji Takada(Kyoto University), Satoshi Takahashi(Tohoku University), Reid C Johnson(University of California), Nucleic Acids Research, Vol.48, Issue 19, 10820−10831 (2020), 2020 年 11 月

12　Position-Dependent Diffusion Constant of Molecules in Heterogeneous Systems as Evaluated by the Local Mean Squared Displacement, Tetsuro Nagai(Department of Advanced Materials Science, Graduate School of Frontier Sciences, The University of Tokyo), Shuhei Tsurumaki(Department of Materials Chemistry, Graduate School of Engineering, Nagoya University), Ryo Urano(Department of Materials Chemistry, Graduate School of Engineering, Nagoya University), Kazushi Fujimoto(Department of Materials Chemistry, Graduate School of Engineering, Nagoya University), Wataru Shinoda(Department of Materials Chemistry, Graduate School of Engineering, Nagoya University), Susumu Okazaki(Department of Advanced Materials Science, Graduate School of Frontier Sciences,

The University of Tokyo), Journal of Chemical Theory and Computation, 16(12), 7239-7254　(2020), 2020 年 11 月

13　Group-based evaluation of temperature and pressure for molecular dynamics simulation with a large time step, Jaewoon Jung(RIKEN R-CCS, RIKEN CPR), Yuji Sugita(RIKEN R-CCS, RIKEN BDR, RIKEN CPR), The Journal of Chemical Physics, Vol. 153, 234115 (2020), 2020 年 12 月

14　New parallel computing algorithm of molecular dynamics for extremely huge scale biological systems, Jaewoon Jung(RIKEN R-CCS, RIKEN CPR), Chigusa Kobayashi, Cheng Tan, Akiyoshi KUroda, Kazuo Minami, Yutaka Ishikawa(RIKEN R-CCS), Kento Kasahara(RIKEN BDR), Yuji Sugita(RIKEN R-CCS, RIKEN BDR, RIKEN CPR), Shigeru Ishiduki, Tatsuo Nishiki, Hikaru Inoue(Fujitsu), Michael Feig(Michigan State University), Journal of Computational Chemistry, Vol. 42, 4, 231-241(2021), 2021 年 2 月

15　Torsional stress can regulate the unwrapping of two outer half superhelical turns of nucleosomal DNA, Hisashi Ishida(Institute for Quantum Life Science, National Institutes for Quantum and Radiological Science and Technology), Hidetoshi Kono(Institute for Quantum Life Science, National Institutes for Quantum and Radiological Science and Technology), Proceedings of the National Academy of Sciences of the United States of America, Vol.118, 7, e2020452118(2021), 2021 年 2 月

16　Linker DNA Length is a Key to Tri-nucleosome Folding, Hiroo Kenzaki(RIKEN), Shoji Takada(Kyoto University), Journal of Molecular Biology Vol.433, Issue 6, 166792 (2021), 2021 年 3 月

17　Case Report: Bayesian Statistical Inference of Experimental Parameters via Biomolecular Simulations: Atomic Force Microscopy, Sotaro Fuchigami(Kyoto University), Toru Niina(Kyoto University), Shoji Takada(Kyoto University), Frontiers in Molecular Biosciences, Vol.8 (2021), 2021 年 3 月

書籍

1　分子シミュレーションと実験データを統合する機械学習手法, 松永康佑(埼玉大学), 羊土社,実験医学増刊,Vol.38,No.20,機械学習を生命科学に使う！, 2020 年 12 月

新聞/TV/WEB 配信/広報誌/雑誌/等のメディア

1　計算による顕微鏡, 松永康佑(埼玉大学), 埼玉新聞 サイ・テクこらむ, 2020 年 9 月