

課題名(タイトル):

実験と協調的な生体高分子の分子動力学シミュレーション

利用者氏名:

森 貴治(1)、○松岳 大輔(1)、新津 藍(1)、Marta Kulik(1)、Hisham Dokainish(1)、George Pantelopulos(1)、杉田 有治(1)

理研における所属研究室名:

(1) 開拓研究本部 杉田理論分子科学研究室

本課題では、実験とシミュレーションの両者の欠点を補い、長所を生かしながら、より信頼度の高い立体構造情報、およびダイナミクスに関する情報を得ることを目的としてシミュレーションを行った。課題 1-3, 6 は結晶を使った構造解析が困難な膜貫通ペプチドに対する分子動力学(MD)シミュレーション、課題 6, 7 は環境により構造を変えることが示唆されているタンパク質に対する enhanced sampling 法を使用したシミュレーション、課題 4, 5 は低温下顕微鏡像構造解析 (cryo-EM) に対する新規の立体構造フィッティング手法の開発についての研究である。

本課題では、各メンバーがそれぞれのテーマに沿って、HOKUSAI を使って研究を進めているので、メンバーが得た結果をそれぞれ記載する。

課題 1. FGFR3 のレプリカ交換法による膜貫通ヘリックスの構造サンプリング (担当: 松岳、杉田)

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

繊維芽細胞増殖因子受容体 FGFR3 は、細胞表面に存在するシグナル伝達蛋白質 Receptor Tyrosine Kinase (RTK) の 1 種であり、細胞増殖や分化に重要な役割を果たす。他の RTK と同様、細胞外ドメインにリガンドが結合するとタンパク質全体の構造変化が生じ、細胞内キナーゼドメインの二量体構造が変化することによって、細胞内のシグナル伝達を引き起こす。細胞外ドメイン、キナーゼドメインの構造は結晶構造解析によってその構造がそれぞれ解かれているが、膜貫通(TM)ドメインについては、結晶構造解析は困難であり、MDシミュレーションによる構造推定が有用な手段の 1 つである。

TM ドメインは、シグナルを細胞外から細胞質側のキ

ナーゼドメインに伝える役割を果たすが、TM ドメインが、どのように活性化に関与しているのかはなお分かっていない。そのため、野生型(WT)の構造のみならず、活性状態時の TM 領域の二量体構造も解く必要がある。FGFR3 TM ドメインに存在する病原性変異体 G380R は、リガンド非結合時においても活性を示すことが知られており、TM ドメインの活性状態の二量体構造を調べるためのモデルとして用いられている。京都薬科大学の佐藤教授らによって行われた脂質二重膜中での固体 NMR 測定によって、WT と G380R 変異体で、TM ドメインの二量体構造が異なることが示唆する結果が得られたが、その 3 次元構造については今なお不明である。

そこで 2016 年度から、レプリカ交換法の一つである "generalized Replica Exchange with Solute Tempering" (gREST) 法 [M. Kamiya & Y. Sugita (2018) *J. Chem. Phys.*, **149**, 072304] を用いて TM 領域の二量体構造に対するサンプリングを行っている (関連課題: G17038)。昨年度の報告で、gRESTシミュレーションを通じて、WT と G380R 変異体で異なる TM 領域の二量体構造を取得できたことを報告した。しかし、そのデータをさらに解析すると、脂質二重膜中の二量体構造が不安定であり、正しく二量体構造をサンプリングできていない可能性が考えられた。同様な議論が、2018 年に R. Best らによっても報告されており、その論文では、脂質-蛋白質間相互作用を従来のものよりも弱くすることが必要であると論じている。そのため、本年度の課題では、脂質-蛋白質間相互作用を弱くした力場パラメータを使って、サンプリングをし直すことになった。

なお本課題は、理研・独創的研究課題“脂質の統合的理解”の 1 サブプロジェクト『脂質-膜タンパク質相互作用』に関係する課題である。

2. 具体的な利用内容、計算方法

本研究では、FGFR3 の膜貫通一膜近傍(TM-JM)領域からなる膜貫通ペプチド二量体を作成し、それを DMPC/DMPG 混合膜からなる脂質二重膜に埋め込んだ系を用意してシミュレーションを行った。力場は CHARMM36 を用いたが、R. Best らの提案に沿って、脂質-蛋白質間相互作用は従来のものを 0.9 倍したパラメータを使用した。

膜中での二量体構造を効率よくサンプリングするために、ペプチド分子および脂質分子を“溶質”として扱った gREST シミュレーションを行った。“溶質”の温度は、310 K - 350 K の範囲で変化させた。WT および G380R 変異体それぞれに対して 300 ns のシミュレーションを行った。上記のシミュレーションおよびその解析は、本研究室で開発しているプログラム“GENESIS”を用いた。

3. 結果

脂質-蛋白質間相互作用を従来のまま使用した場合、gREST シミュレーションで、時間がたつほど、二量体が壊れる傾向があったが、当該相互作用を補正した場合には、300 ns のシミュレーション中、二量体が壊れることはなかった。

この gREST シミュレーションによって、実験結果と一致する結果を得ることに成功した。すなわち、

- 1) G380R 変異体のヘリックス慣性主軸の膜方面に対する角度が、WT のものと比べて小さい。これは TM ヘリックスの傾き角を推定できる FTIR による結果と一致するものである。
- 2) WT では L385 側鎖が、二量体界面にある確率が高いが、たいして、G380R 変異体では、L385 側鎖が二量体界面にある確率は小さい。これは、二量体界面にある残基を推定できる固体 NMR による結果と一致するものである。

gREST シミュレーションで得られた二量体構造の特徴は WT と G380R 変異体で大きく異なっていた。WT では、TM ヘリックスの N 末端側がやや開いている構造が主に得られたが、G380R 変異体では、TM ヘリックスの N 末端側でコンタクトしている傾向があった (図 1)。また、この構造の違いは、他の RTK でも提案され

ており、この TM ヘリックスの二量体構造の違いが、FGFR3 の活性化に重要である可能性がある。

4. まとめ

膜貫通ペプチド二量体のサンプリングのために、脂質とペプチドを“溶質”として扱った gREST シミュレーションを行い、実験結果と一致する結果を得ることに成功した。また脂質-蛋白質間相互作用の補正が gREST シミュレーションにおいて必要であることを確認した。現在、この結果は論文としてまとめているところである。

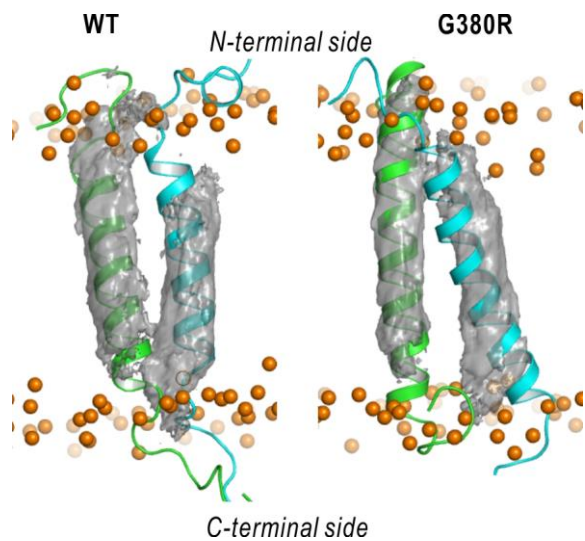


図 1. gREST シミュレーションによってサンプリングされた構造のうち代表的なもの (リボンモデル) とシミュレーションから計算された主鎖原子の分布密度 (灰色の表面モデル)。

課題 2. EGFR TM-JM 領域のシミュレーション (担当: 松岳)

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

上皮細胞増殖因子受容体(EGFR)も RTK の 1 種である。TM-JM 領域は、他のタンパク質と直接相互作用することのない領域ではあるが、その領域は、細胞外ドメインに与えられたシグナルを細胞質側へ伝達する役割を有しており、EGFR の機能に不可欠な領域である。

EGFR TM-JM 領域の脂質二重膜中の二量体構造は、溶液 NMR によって解かれている (PDB ID: 2M20)。その構造では JM 領域は α ヘリックス構造を示し、そのへ

リックスは反平行な二量体を形成する。その NMR 構造は活性状態での構造を表すと考えられている。

JM 領域に存在するスレオニン残基 T654 がリン酸化された場合、EGFR の受容体活性が低下することが知られている。そこで、EGFR JM 領域の構造についての知見を得るために、理化学研究所・佐甲細胞情報研究室において、EGFR TM-JM 領域からなるペプチドの C 末端部位に蛍光分子である cy3, cy5 を付加し、それを脂質二重膜に埋めて FRET 効率を測定する実験が行われた (図 2)。その実験では、FRET 効率分布が、T654 側鎖がリン酸化されている時(pT654)とリン酸化されていない時(T654)で大きく異なっていた。また、膜組成によっても FRET 効率分布が異なっていた。したがって、膜組成や T654 のリン酸化状態によって、EGFR TM-JM 領域の 2 量体構造が変化することが考えられるが、FRET 測定実験では、蛍光分子間の距離情報しか得ることができず、その立体構造情報については何もわからない。そこで FRET 測定で使われた系を MD シミュレーションで再現し、膜組成やリン酸化状態でどのような違いが生じるかを調べることにした。

なお本課題は、理研・独創的研究課題“脂質の統合的理解”の 1 サブプロジェクト『脂質—膜タンパク質相互作用』に関連する課題の 1 つとして取り組んでいるものである。

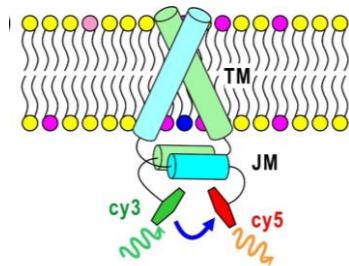


Fig 2. EGFR TM-JM 領域に対する FRET 実験の概念図。 円筒モデルは α ヘリックスを表す。JM 領域のヘリックスが二量体を形成しているときは、蛍光分子 (cy3/cy5) 同士が近くに位置していると考えられるので、得られる FRET 効率は高いと考えられる。逆に蛍光分子が離れている時、得られる FRET 効率分布は低い。

2. 具体的な利用内容、計算方法

EGFR TM-JM 領域の二量体構造が、膜組成や T654 残基のリン酸化状態から、どのような影響を与えられ

るかを議論するため、下記の 5 の系に対して、それぞれシミュレーションを行った：

Sim1: T654; PC:PS:PIP2 = 10:0:0

Sim2: T654; PC:PS:PIP2 = 10:3:0

Sim3: T654; PC:PS:PIP2 = 10:0:0.8

Sim4: pT654; PC:PS:PIP2 = 10:0:0

Sim5: pT654; PC:PS:PIP2 = 10:3:0

EGFR TM-JM の構造は、DMPC からなるバイセルを利用して解かれているので、同様な組成を持つ Sim1 が構造比較の基準となる。それぞれの系に対して、1 μ s の計算をおこなった。また、脂質の初期分布を変えてそれぞれ 3 回ずつ計算を行った。そのため、全部で 15 μ s もの計算が必要になり、また同時並行的に計算を行うことが必要であったので、HOKUSAI だけでなく「京」や研究室計算機クラスターを併用して、計算を行った。

FRET 実験との比較を行うため、各ペプチドの C 末端に蛍光分子 cy3/cy5 を付加したものを作成した。計算には GENESIS を用いた。力場には CHARMM36 を使用した。

3. 結果

非リン酸化時(T654)とリン酸化時(pT654)では、JM 領域の構造に違いがあった。すなわち、pT654 では、リン酸基と、近傍に位置するもう一方の分子の R662 と塩橋を作り、より二量体構造が安定していた。また、膜が POPC のみで構成されている場合、二量体の揺らぎは小さかったが、酸性リン脂質(POPS, PIP2)が膜に含まれている場合、二量体構造の揺らぎは大きくなった。

蛍光分子間の距離を比較した場合、いずれの系でも高 FRET 効率を与える距離分布であったが、酸性リン脂質が含まれる系では、より短い距離にピークが現れた (図 3)。

4. まとめ

膜組成や JM 領域にある残基のリン酸化状態に応じて、二量体構造の揺らぎの程度が変化することが分かった。

5. 今後の計画・展望

現在、なぜ酸性リン脂質が膜に含まれるときに、構造の揺らぎが大きくなるかを調べているところである。

シミュレーションで得た知見を適宜実験結果と比較し、この結果を論文としてまとめ、提出する予定である。

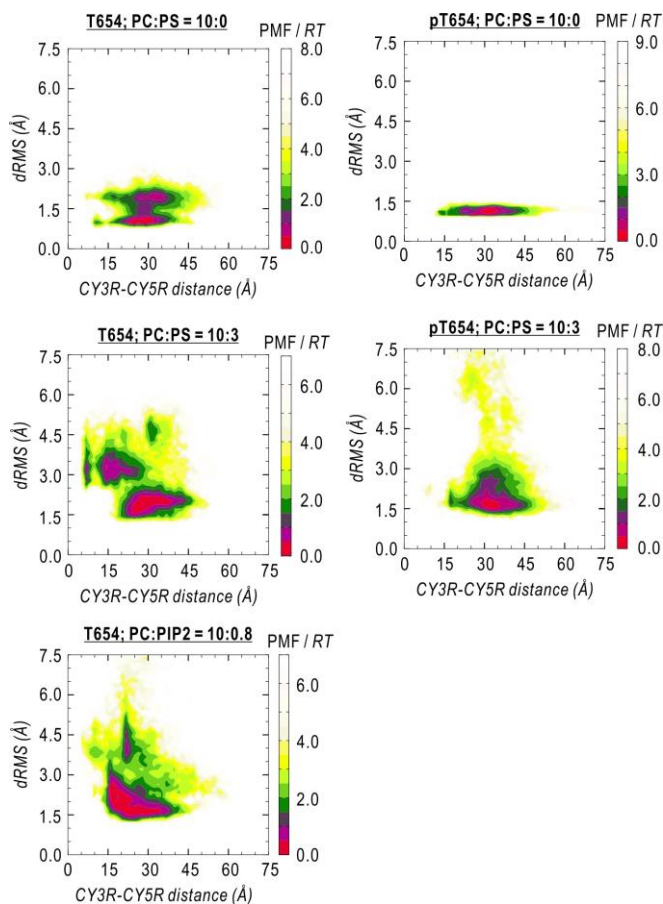


Fig 3. 各系に対するペプチド二量体の構造揺らぎ (dRMS) および蛍光分子間距離 (CY3R-CY5R distance)。dRMS は、各系に対する平均構造からのずれの大きさを表す。

課題 3. cWza ペプチドポアの会合過程の解明 (担当: 新津)

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

チャネルタンパク質は細胞膜内外のイオンや生物活性分子を輸送・伝達する重要な役割を持つ。タンパク質の機能はその立体構造によって決定されるので、チャネルタンパク質の折り畳み機構を理解することは、当該タンパク質の機能発現メカニズムの理解に欠かせない。しかしながら疎水性膜タンパク質であるチャネルタンパク質の折り畳みに関する研究は、水溶性タンパク質に比べ、高解像度の構造情報が限られているために遅れている。そこで我々は単純化されたタンパク質であるペプチド会合体を設計・合成し、その会合機

構を調べることでこの問題に取り組むこととした。設計で標的としたのは膜タンパク質構造において最も基礎的な膜貫通 α -ヘリックス- α -ヘリックス相互作用である。これに関して、これまでに本課題担当者は脂質二重膜中で α -ヘリックス 8 量体を形成する cWza ペプチド (Mahendran *et al.*, *Nat. Chem.*, 2017)、および新規 (*de novo*) 設計による膜ペプチド 5-7 量体を作製した。しかしながらこれらペプチド会合体の形成機構については電気生理実験をもとにしたモデルの提示に留まっており、詳細な機構はわかっていない。そこで本課題では実験では観測できない動的なペプチドの構造変化、会合体形成を、分子動力学計算により解析し、会合機構を明らかにすることを目的とする。

2. 具体的な利用内容、計算方法

cWza ペプチドについて、本年度は水中 \Rightarrow 膜結合に関する準備段階の計算を実施した。伸長ペプチドを水中に配置して脂質二重膜存在下での全原子 MD 計算を行って得た構造を初期構造とした、レプリカ交換 solute-tempering (REST) 法を用いた計算を行った。*de novo* 設計ペプチドについては、設計モデルを脂質二重膜中に配置したものを初期構造として全原子 MD 計算を行い、モデル構造の安定性とダイナミクスを評価した。すべての計算は GENESIS を用い、力場は CHARMM36、水モデルは TIP3P を用いた。

3. 結果

cWza ペプチドの計算では、実験条件に則って脂質 DPhPC(1,2-diphytanoyl-sn-glycero-3-phosphocholine) 160 分子からなる二重膜、ペプチド鎖 2 本、1 M 濃度の KCl (計 74,000 原子) の gREST 法 (6 レプリカ、300-500 K) を用いた計算を実行した。100 ns 後のスナップショットを図 4 に示す。ペプチド鎖同士の相互作用、ペプチド-脂質間の相互作用が見られるが、二次構造は形成されておらず、初期構造の検討および計算時間の延長が必要と考える。

de novo 設計ペプチドでは、ペプチド 5 本、7 本、12 本からなる会合体 3 種について DPhPC 二重膜、1 M KCl 存在下 (計約 140,000 原子) での全原子分子動力学計算を実行した。それぞれ 200 ns の計算を行った。5 量体の構造スナップショットを図 5 に示す。

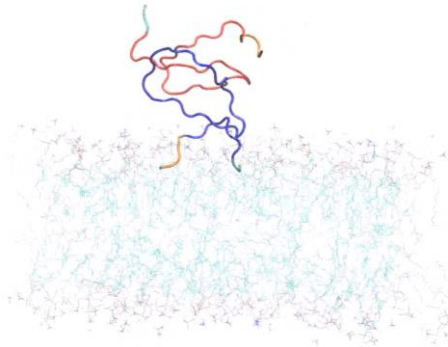


図 4. 膜表面に存在する cWza ペプチドの会合過程のシミュレーションスナップショット構造。細線は脂質分子、リボンモデルはタンパク質構造を示す。

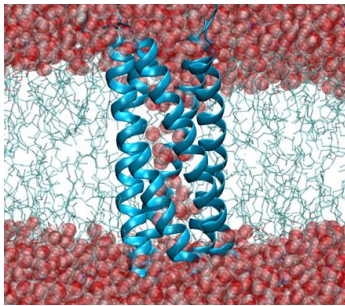


図 5. 脂質二重膜中の *de novo* 設計ペプチド 5 量体シミュレーションスナップ構造。赤球は水分子酸素原子を示す。

このペプチド 5 量体は電気生理実験から、平面 DPhPC 二重膜中に自発的にイオン透過性を持つチャンネルを形成することがわかっている。本計算ではペプチドチャンネル内に水分子が入りこみ、かつ安定に会合体を形成しており、電気生理実験を支持する結果が得られた。

4. まとめ

本課題では、脂質二重膜中でチャンネルを形成するペプチド 2 種について膜近傍での挙動を全原子分子動力学計算を用いて解析した。cWza ペプチドには準備段階の計算が完了した。*de novo* 設計ペプチドでは電気生理実験と矛盾しない計算結果を得た。

5. 今後の計画・展望

cWza ペプチドについては、引き続きレプリカ交換法を用いて膜表面のペプチドの 2 次構造形成を解析するとともに、アンブレラサンプリングを用いてペプチドを膜に挿入した時の自由エネルギー計算を行う予定である。これらの結果を合わせて cWza ペプチドチャネ

ルの会合機構を明らかにして行く。*de novo* 設計ペプチドについては、本課題で得られた結果を電気生理実験の結果と合わせて論文にまとめ、投稿準備中である。

課題 4. クライオ電顕フィッティング計算における全原子モデルと粗視化モデルの比較 (担当: 森)

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

近年、低温下電子顕微鏡像解析(クライオ電顕と略す)を用いた単粒子解析によるタンパク質の立体構造解析法が盛んに行われている。この手法では、精製したタンパク質を含む溶液を極低温にまで冷却し、電子線を照射することで、タンパク質の溶液中での 2 次元画像を観察する。単粒子解析では、これらの 2 次元画像を集め、それらから 3 次元情報を再構成することにより、3 次元密度マップを得る。3 次元密度マップからタンパク質の立体構造をモデリングする手法として、フレキシブル・フィッティングが広く用いられている。これは X 線結晶構造解析法や NMR 分光法などで得られた高分解能のタンパク質の構造を、MD 計算を行いながら 3 次元密度マップに当てはめる方法である。MD 計算を用いて密度マップを再現するように各原子に力を与えるため、対象とするタンパク質の別の状態の構造が予め X 線結晶構造解析などで分かっているならば、機能的構造変化を模倣した自然な構造モデリングが可能である。フレキシブル・フィッティングにおいて、リボソームのような巨大な生体分子を扱う場合、通常の MD 計算同様、膨大な計算コストが必要となる。そのため、溶媒を露に取り入れれない連続体近似モデルや粗視化モデルの利用が有効になるが、このようなモデルを用いた場合、フィッティング後の構造がどれくらい妥当かどうかは自明ではない。

なお本研究は、理化学研究所 Pioneering Project “Dynamic structural biology”の一環として行われた。

2. 具体的な利用内容、計算方法

本研究では、Glucose/Galactose 結合蛋白質やリボソームなどに対して全原子モデルと粗視化モデルを用いてフレキシブル・フィッティングを行い、得られた構造の比較を行った。全原子モデルには CHARMM C36 力場を用いて溶媒分子は露に取り扱い、粗視化モデルには全原子型郷モデルを用いた。これは全原子解像度を持つ粗視化モデルであり、重原子のみを扱い、通常の郷モデル同様、ポテ

ンシヤルエネルギーには native contact と non-native contact energy 項を含めている。本研究では、X線結晶構造から生成した人工的な電顕密度マップを用いることで正解構造を予測出来るか検討した。ソフトウェアには我々の研究グループで開発している GENESIS を用いた。

3. 結果

Glucose/Galactose 結合蛋白質について、Closed 状態の X 線結晶構造から始めて Open 状態の電顕像にフィッティングしたところ、全原子モデルと粗視化モデルともにターゲットとする電顕像にフィットした構造が得られ、正解構造との RMSD は 0.5 Å 以下であった。しかしながら、MolProbity score と呼ばれる立体構造の指標に基づいてタンパク質構造らしさを確認したところ、全原子モデルの構造はスコアが非常に良く、原子同士の衝突がほとんどないのに対して、粗視化モデルは衝突が多く、主鎖や側鎖の二面角の分布も理想から外れていることが分かった。

次にリボソームに対して Classical E/E state (図 6 左上)から Hybrid P/E state (図 6 右上)へのフィッティングを試みたところ、全原子モデルと粗視化モデルともにフィットした構造が得られた。しかしながら、途中構造を見てみると、全原子モデルではフィッティング初期の段階に Stalk ドメインの head 領域が大きく歪んでしまい、5 回行った計算のうち1つは正解構造に辿り着けなかった。一方、粗視化モデルを用いた場合では歪まずにフィットし、多くが正解構造に近づくことが分かった (図 4)。これは郷ポテンシャルの native contact energy 項によってドメインが固くなったことに由来していると考えられ、フィッティングの際にドメイン内部の構造が変わらない場合には、郷モデルが有用であることが分かった。

4. まとめ

全原子モデルと粗視化モデルを用い、フレキシブル・フィッティングにおける構造モデリングの予測精度を比較した。全原子モデルは得られる構造の質は高いが、タンパク質が大規模に構造変化をする場合、構造が崩れてしまうケースもあるため注意が必要である。粗視化モデルでは予測構造の解像度は低くなってしまいが、ドメインを固めてフィッティングさせることなどにも有効である。

5. 今後の計画・展望

本研究により、フレキシブル・フィッティングにおいて、全原

子モデルと粗視化モデルともに一長一短があることが分かった。今後、二つを組み合わせたマルチスケールなアプローチにより、さらに実験データの解析に応用する予定である。

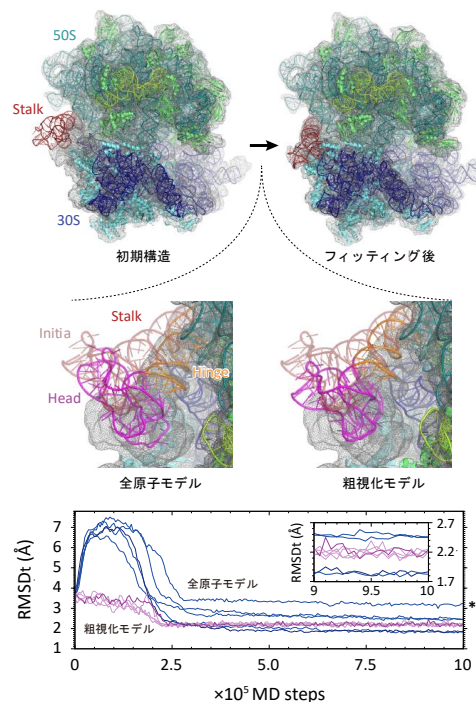


図 6. リボソームのフレキシブル・フィッティング。グラフの縦軸は正解構造からの root-mean squared difference (RMSD)であり、*で示すフィッティング計算は途中で構造が引っかかり、正解に辿り着けなかったものを示している。

課題 5. Investigation of Conformational change in Ca²⁺ ATPase. (担当 : M. Kulik)

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

This research project was focused on employing the flexible fitting algorithm implemented in GENESIS to investigate a membrane protein system of calcium ATPase of sarcoplasmic reticulum (筋小胞体). The purpose of the flexible-fitting MD simulation is to obtain a final model perfectly fitted to a density map, obtained from cryogenic electron microscopy (Cryo-EM) experiments. The Ca²⁺ ATPase is an especially interesting system for this kind of

simulation, since it undergoes a series of non-trivial large conformational transitions during its physiological cycle. Additionally, a range of accurate experimental structures from X-ray crystallography is available for this system in protein data-bank (PDB).

2. Specific usage status of the system and calculation method

For this project, circa 3,700,000 CPU hours were used, mainly on BW-MPC (the estimated number for this project in application was 3,532,000 CPU hours on BW-MPC). However, significant changes of the research plan were made in the course of calculations, since the flexible fitting simulations using the all-atom model with a system containing around 300000 atoms were proven to be very slow and the fitting results were not satisfactory. Thus, we focused on coarse-grained simulations, using Go-like $C\alpha$ atom model. With a kind of Replica-Exchange method in order to automatically adjust the force constant which was driving the fitted structure towards the map. Next, we joined the coarse-grained level with all-atom level in implicit solvent, by performing a targeted MD simulation towards the $C\alpha$ positions of the best model from coarse-grained simulations. Subsequently, another flexible fitting simulation was done with an all-atom model obtained from the targeted MD simulation in order to refine the model. We name this protocol “Multi-scale protocol” (Fig. 7).

3. Result

In the course of simulations, we have developed a new protocol for flexible fitting simulations. The scheme of this protocol is shown in the Fig. 7.

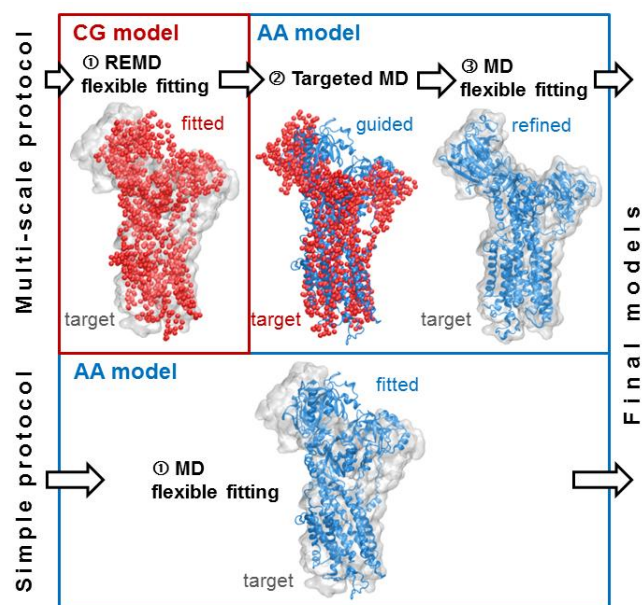


Fig. 7. Flexible-fitting protocols. Red spheres show $C\alpha$ atoms and blue-ribbon models show protein backbone in AA models. White surface in this figure shows a density map surface computed from a protein 3D structure deposited in PDB for the test simulation.

In Fig. 7, a fitting of Ca^{2+} ATPase is shown at two levels: coarse-grained CG level (presented as red dots at $C\alpha$ positions) and all-atom AA level (shown in blue in cartoon style). For comparison, a Simple protocol was used, which is a longer fitting simulation using only the all-atom MD simulation. This kind of protocol was tested for many other protein systems of various size, also including systems with experimental density maps. The quality of the final models obtained using the Multi-scale protocol and the Simple protocol was compared in terms of the Root Mean Square Deviation towards the target structure, correlation coefficients towards the density maps, the MolProbity score, CaBLAM outliers, Ramachandran outliers, secondary structure similarity and many others. Overall, using the Multi-scale protocol gives much better fitted structures than using the Simple protocol for almost all tested systems, especially in the cases involving challenging domain rotation.

4. Conclusion

Performed simulations allowed us to propose a new protocol for flexible fitting simulations, called the Multi-scale protocol. This protocol was tested for many systems and was shown to perform significantly better than a single long flexible fitting simulation.

5. Schedule and prospect for the future

The results of the work were presented as a poster at one international conference and are currently prepared for publication in a journal.

課題 6. Characterization of monomer structures of full-length C99 peptide (担当 : Pantelopulos)

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

Recently, many efforts to accurately determine and understand the conformational ensembles of intrinsically disordered proteins (IDPs) were devoted. IDPs often take on particular folds and functions based on molecular interactions with proteins and other biomolecules, which can endow them with various emergent functionalities or which can cause them to malfunction. In Alzheimer's disease (AD), the aggregation of specific lengths of Amyloid Beta ($A\beta$), an IDP, causes for the formation of Amyloid plaques conventionally implicated in AD. Potentially toxic or non-toxic $A\beta$ are produced based on the cleavage of Amyloid Precursor Protein's 99-residue C-terminus (C99), by γ -secretase in the membrane. Membrane properties and localization of γ -secretase and C99 select whether toxic or non-toxic lengths of $A\beta$ are produced from this cleavage. Thus, as C99 is an IDP, it is critical to understand its molecular ensemble and changes in response to its

environment in order to understand the genesis of AD.

2. Specific usage status of the system and calculation method

Sampling of the full C99 sequence, which contains intrinsically disordered regions in its N- and C-terminus likely critical to determining its interactions with γ -secretase, $A\beta$, and other C99 proteins, requires the careful application of enhanced sampling techniques in MD simulation. Recently, Sugita group developed gREST method, in which parts of the molecular interactions in a system are selectively attenuated (by an effective "temperature") in several parallel simulations to allow for rapid exploration of conformational space. These parallel simulations exchange simulation conditions periodically using a Metropolis-Hastings type exchange conditions which allows us to sample the conformational space of systems at "real", equilibrium conditions despite employing these attenuated molecular interactions.

3. Result

Simulations of full-length C99 in a simple, POPC-only model of a cell membrane were initiated from initial conformations of C99 identified as representatives of populous conformational states in our previous implicit membrane simulations. [Pantelopulos, G.A., J.E. Straub, D. Thirumalai, and Y. Sugita. 2018. *Biochim. Biophys. Acta - Biomembr.* **1860**: 1698–1708.] During FY 2018, we have parameterized a suitable set of effective "temperatures" for gREST simulation. In the simulations, to use a flat-well pair distance restraints was necessary to reduce computational cost and to avoid periodic boundary condition (PBC) artefacts.

We have found that attenuating only the Lennard-Jones, dihedral, and CMAP (coupled

protein dihedral angle) potential terms of C99 residues outside the membrane (residues 1-28 and 53-99) in addition to lipid head groups (up to the ester group) in gREST will allow for rapid sampling and nearly-ideal behavior in the exchange of replicas over time, in which each trajectory should equally sample all "temperatures". We have found that employing 16 linearly-spaced effective temperatures from 310 K ("real" conditions) to 340 K allows for C99 to sample many secondary and tertiary structure transitions, but that these can cause for C99 to self-interact over the PBC at temperatures over 334 K. To address this, we have applied flat-well pair distance restraints from C99 residues 1, 87, and 99 to residue 44 (from the termini to the center of the membrane) (Fig. 8). This restrains the ensemble to within a "dry" or "molten" globule-like state, which is expanded at higher temperatures to permit secondary and tertiary structure transitions while avoiding PBC artefacts. The use of these restraints allows us to simulate the C99 in POPC system as 120,000 atoms rather than 270,000 atoms, which would otherwise be required to prevent PBC artefacts.

4. Conclusion

To characterize the conformational ensemble of the membrane IDP C99, we have developed a simulation scheme that will allow us to accomplish the otherwise impossible task of describing its conformational ensemble. This scheme for enhancing secondary and tertiary conformational states of C99 using gREST can serve as a basis for future efforts in sampling proteins and other polymers that require this type of expansion-collapse transition to quickly sample states.

5. Schedule and prospect for the future

Simulation of C99 in a POPC membrane using gREST will resume using the simulation scheme we

have developed. We plan to perform these C99 in POPC membrane calculations until achieving 500 ns of sampling. In the future, following convergence of this simulation, we will use representative conformational states from its ensemble to study the effect of changing membrane environments on C99 conformation and functionality.

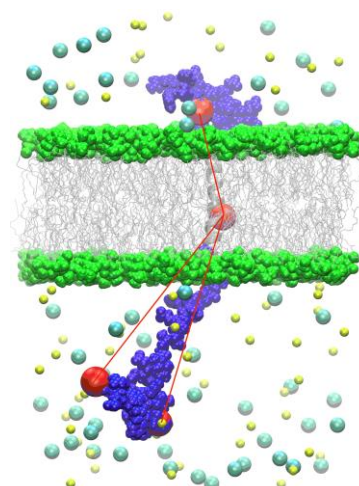


Fig. 8. C99 in POPC system. Flat-well pair distance restraints of 6.5 nm in length drawn in red from residues 1, 87, and 99 to residue 44 of C99. Space filling green (POPC) and blue (C99) represent atoms selected for molecular interaction attenuation by effective temperature are displayed. Gray atoms of C99 and POPC and unperturbed during gREST. Na⁺ and Cl⁻ ions are displayed in cyan and yellow to show the borders of the PBC. Water molecules are not displayed.

課題 7. Computational investigation of Order to Disorder Transition in Drk N-SH3 Domains (担当 : H. Dokainish)

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

Drk (*Drosophila adapter protein*) is a small protein that mediate signal transduction from RTK to RAS G-protein. It contains a Src homology 2 (SH2) domain and two SH3 domains. In which, the isolated

N-terminal SH3 (N-SH3) domain has been experimentally shown to exist in 1:1 equilibrium of folded and unfolded states. On contrary, the C-terminal SH3 (C-SH3) domain is believed to form stable folded structure. Notably, N-SH3 domain folded/unfolded equilibrium was found to shift toward unfolding in the presence of the other two domain (SH2 & C-SH3) in Drk.

In this project, we aimed to elucidate the origin of such instability in isolated N-SH3 domain in comparison to stably folded C-SH3 domain using enhanced sampling techniques. In addition, since the structure of Drk was experimentally solved, the interaction between the three domains was elucidated via classical MD simulations. Finally, the unfolding pathways were investigated at higher temperatures.

2. Specific usage status of the system and calculation method

For this project, Drk and isolated N/C-SH3 domains, 5,794,500 CPU hours were used, mainly on GW-MPC. In which, after modeling Drk system with homology modelling, a total of 2.5 μ s classical MD were performed in order to investigate the interdomains interactions. Furthermore, the isolated N/C-SH3 domains were simulated using gREST for a total of 13 μ s. gREST simulations were performed using multiple replica in combination with multiple temperature ranges to simulate unfolding process and elucidate unfolding pathway. In addition, preliminary mutations were also tested using the same approach.

3. Result

Classical MD results shows different interdomain interactions in comparison to Drk mammalian homologue Grb2 crystal structure. Wherein, C-SH3 and SH2 domain were found to stably bind throughout the simulation. On the contrary, the

unstable N-SH3 domain were found to exist in three states in which, it interacts with either of the other two domains or exist free in the solution, see Fig. 9. Multiple important regions for interactions were elucidated and would be tested in the future upon mutation.

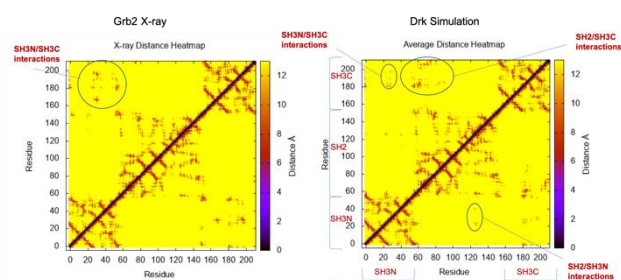


Fig 9. Heatmap of interdomain interactions in Drk simulation vs Grb2 crystal structure.

gREST results of the isolated domains shows unfolding of N-SH3 domain, at higher temperature, due to destabilization via intradomain interactions. Two negatively charged regions in N-SH3 domain (hot spot) were identified to play an important role in the domain instability (Fig. 10). However, unfolding has not been observed at 300 K and would require longer simulation in the future.

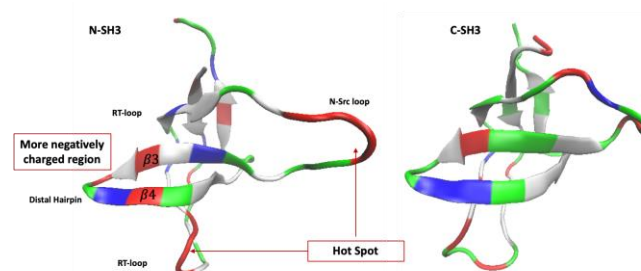


Fig. 10. Distribution of charged residues in N-SH3 domain (left) and C-SH3 domain (right).

4. Conclusion

- Drk simulation suggest different interdomain interactions from GBR2 X-ray structure, where C-SH3 and SH2 binds.
- SH3N domain simulation suggest unfolding due to presence of charged regions at distal hairpin/ β 4, N-src loop and RT-loop.

5. Schedule and prospect for the future

The obtained results will be further investigated using multiple independent MD of Drk and simulating unfolding of isolated domains at 300 K as

well as mutations will be tested theoretically and experimentally. We plan to represent the current results within this year in international conference. Finally, after finalizing all simulations, the results will be submitted for publication.

平成 30 年度 利用研究成果リスト

【雑誌に受理された論文】

- Marta Kulik, Takaharu Mori, Yuji Sugita and Joanna Trylska “Molecular mechanisms for dynamic regulation of N1 riboswitch by aminoglycosides.” *Nucleic Acids Research*, **46** (19), 9960-9970 (2018) (Published: 20 Sep. 2018).
- Takaharu Mori, Marta Kulik, Osamu Miyashita, Jaewoon Jung, Florence Tama and Yuji Sugita “Acceleration of cryo-EM Flexible Fitting for Large Biomolecular Systems by Efficient Space Partitioning.” *Structure*, **27** (1), 161-174 (2019) (Published: 18 Oct., 2018).

【口頭発表】

- 森貴治、杉田有治 “クライオ電顕フィッティング計算における全原子モデルと粗視化モデルの比較”, , 第 32 回分子シミュレーション討論会 (つくば) 2018/11/30
- George Pantelopulos “Understanding an Intrinsically Disordered Membrane Protein in Alzheimer's Disease via Molecular Simulation”, 14th Rencontres du Vietnam “Computational Biophysics at the Molecular and Mesoscales”, Quy Nhon, Vietnam, 2018/10/29
- 森貴治 “クライオ電顕フィッティングによる大規模生体分子系の構造精密化”, 蛋白研セミナー (大阪) 2018/9/26
- 森貴治、江原晴彦、関根俊一、杉田有治 “クライオ電顕フィッティングによる RNA ポリメラーゼ-DNA 複合体の構造精密化”, , 第 56 回日本生物物理学会 (岡山) 2018/9/16
- 松岳大輔、松永康佑、杉田有治
“MD シミュレーションによって調べた脂質二重膜の組成が EGFR 膜近傍領域の二量体構造に与える影響”
日本物理学会 2019 年秋季大会、同志社大学京田辺キャンパス、2018 年 9 月 9 日
- 森貴治、江原晴彦、関根俊一、杉田有治 “クライオ電顕と分子動力学計算による RNA ポリメラーゼ-DNA 複合体の構造最適化”, , 日本物理学会 2018 年秋季大会 (京都) 2018/9/9
- 森貴治 “Acceleration of cryo-EM flexible fitting for large biomolecular systems by high-performance molecular dynamics simulations”, , 第 18 回日本蛋白質科学会年会 若手奨励賞シンポジウム (新潟) 2018/6/27
- 松岳大輔、神谷基司、杉田有治
“REST シミュレーションを用いた FGFR3 野生型と G380R 変異体の膜貫通ドメインの構造研究”
日本物理学会 2018 年年次大会、東京理科大学野田キャンパス、2018 年 3 月 23 日

【ポスター発表】

- 森貴治 “クライオ電顕フィッティング計算における全原子モデルと粗視化モデルの比較”, , 第 32 回分子シミュレーション討論会 (つくば) 2018/11/28

平成 30 年度 利用報告書

- 森貴治、杉田有治 “Cryo-EM flexible fitting for large bimolecular systems”, Takaharu Mori, 4th Rencontres du Vietnam “Computational Biophysics at the Molecular and Mesoscales”, Quy Nhon (Vietnam), 2018/10/28
- 松岳大輔、松永康佑、杉田有治
“Investigation of the correlation between lipid composition and the structure of EGFR JM region.”
第 56 回日本生物物理学会年会、岡山、2018 年 9 月 15 日 (土)
- George A. Pantelopulos, Yuji Sugita, John E. Straub "Understanding Intrinsically Disordered Membrane Proteins in Alzheimer's Disease via Molecular Simulation", RIKEN Summer School, Tsukuba, 2018/9/27-28.
- Takaharu Mori, Marta Kulik, Osamu Miyashita, Jaewoon Jung, Florence Tama, and Yuji Sugita
“Acceleration of cryo-EM flexible fitting for large biomolecular systems by high-performance molecular dynamics simulations”, , 第 18 回日本蛋白質科学会年会 (新潟) 2018/6/27
- 松岳大輔、松永 康佑、杉田有治
“T654 残基のリン酸化が EGFR 膜近傍領域の二量体構造に与える影響”
第 18 回日本蛋白質科学会年会、新潟、2018 年 6 月 27 日 (水)
- 森貴治 “クライオ電顕のフィッティング高速化”、理研シンポジウム「細胞システムの動態と論理 X」(埼玉) 2018/4/13
- Marta Kulik, Takaharu Mori, Yuji Sugita, “Flexible fitting simulations of P-type ATPases to low, medium and high resolution density maps”, 256th ACS National Meeting, August 19-23, 2018, in Boston, MA, USA

【その他(著書、プレスリリースなど)】

- クライオ電顕像からのタンパク質構造モデリングを高速化ー巨大生体分子複合体の構造解析を加速ー 理化学研究所プレスリリース (2018/10/19)
- “研究室メンバーで一丸となってソフトウェアを開発するには?”, 森貴治, 蛋白質科学会アーカイブ エッセイ #07.
- “タンパク質構造のモデリング クライオ電顕像を用い高速化”, 薬事日報 8 面 (2018/10/26)