

令和 4 年 6 月 14 日現在

機関番号：82401

研究種目：基盤研究(C) (一般)

研究期間：2019～2021

課題番号：19K06532

研究課題名(和文) 連成解析によるクライオ電顕構造モデリング法の開発

研究課題名(英文) Cryo-EM structure modeling based on the multiple score analysis

研究代表者

森 貴治 (Mori, Takaharu)

国立研究開発法人理化学研究所・開拓研究本部・専任研究員

研究者番号：90402445

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：本研究では、クライオ電子顕微鏡を用いた単粒子解析により得られた実験データから、タンパク質の立体構造を精密にモデリングするための方法として、SAUA-FFR法を開発した。これにより従来の方法と比べて、タンパク質の二次構造をより正確に再現でき、また、タンパク質らしい構造を候補構造モデルの中から選択することが可能になった。本手法は分子動力学計算ソフトウェアGENESISに実装され、無償で利用できる。また、本手法に関するわかりやすいチュートリアルについても GENESIS のウェブサイト内で公開した。

研究成果の学術的意義や社会的意義

蛋白質科学を医学あるいは薬学に応用するためには、タンパク質の構造と機能を正しく理解する必要がある。そのためには、タンパク質の立体構造を実験データから精密にモデリングする必要がある。現在の構造生物学において、単粒子解析を用いたタンパク質の立体構造モデリングは世界的に広く行われているが、得られた構造の妥当性の検証法、あるいは妥当な構造を選ぶ方法論の開発は未だ不十分である。本研究で開発したプロトコルは、合理的に妥当な構造モデルを選ぶ方法でもあり、本研究は蛋白質科学の発展に大きく貢献したと考えている。

研究成果の概要(英文)：In this study, we developed the SAUA-FFR method for efficient modeling of protein 3D structures from experimental data obtained by single-particle analysis using cryo-EM. Compared to conventional methods, this method more accurately reproduces the secondary structure of a protein and enables the selection of a protein-like structure from among candidate structural models. The method is implemented in the molecular dynamics software GENESIS and is available free of charge. An easy-to-understand tutorial on this method is also available on the GENESIS website.

研究分野：計算生物物理学

キーワード：クライオ電顕 タンパク質 分子動力学計算 立体構造予測

1. 研究開始当初の背景

近年、タンパク質の立体構造解析において、クライオ電子顕微鏡を用いた単粒子解析が盛んに行われている。単粒子解析とは、タンパク質を含む溶液を極低温にまで急速凍結したあと、透過型電子顕微鏡によって撮影されたタンパク質の多数の2次元像を再構成し、3次元密度マップを得る方法である。単粒子解析により、ウイルスやリボソームなどの巨大タンパク質複合体だけでなく、膜タンパク質や核酸複合体などの比較的小さな系に対しても近原子解像度で密度マップが得られるようになり、いくつかのケースでは原子解像度が達成されている。一方、タンパク質の構造は溶液中ではダイナミックに揺らいでいるため、大規模構造変化を起こすような柔らかいタンパク質をクライオ電顕の対象とした場合、密度マップ中で柔らかいドメインにおいて局所的に解像度が低下するなどの問題が起こる。従って、側鎖がはっきりと見えないような低解像度のマップを扱う場合、タンパク質の立体構造を慎重にモデリングする必要がある。

2. 研究の目的

密度マップから分子構造をモデリングする手法として、剛体ドッキング、フレキシブル・フィッティング、*De novo* モデリングがある。剛体ドッキングでは、タンパク質を構成するドメインを電顕密度マップに当てはめ、構造がマップに一致するように各ドメインの最適な位置や配向を網羅的に探索する。フレキシブル・フィッティングでは、タンパク質分子が電顕マップに一致するように分子動力学計算に基づいて原子にバイアスを加えながら立体構造を最適化する。*De novo* モデリングでは、アミノ酸配列に基づき電顕マップ中の密度をトレースしながら立体構造を構築する。実験マップから構造を精密にモデリングする場合、これらの方法を組み合わせた方法が用いられ、特にフレキシブル・フィッティングは最終構造のリファインメントにも重要である。しかしながら、局所低解像度領域やノイズが多く含まれる部分に対するモデリングは一般的に困難であり、効率の良い立体構造探索だけでなく、*overfitting* をなるべく抑えること、言い換えれば、フレキシブル・フィッティングのバイアスの力の定数 k をどうやって決めるか、また、得られた構造をどのように評価して最も確からしい構造を選ぶかが重要になってくる。そこで本研究ではこのような問題を解決するために、新たな分子構造モデリングプロトコルの開発を目指した。

3. 研究の方法

本研究では、*De novo* 構造モデリングにより得られた $C\alpha$ モデルのデコイを効率よく最適化するプロトコルとして「SAUA-FFR 法」を開発した (図 1A)。これはまず、 $C\alpha$ モデルからアミノ酸の側鎖を生やした後、融合原子モデル (United-atom model; UA) を用いてシミュレーテッド・アニーリング (SA) を複数回繰り返しながら構造をフレキシブル・フィッティングで最適化する方法 (Flexible fitting refinement; FFR) である。融合原子モデルとは炭化水素部分 (CH_n) の水素を炭素に取り込むことで水素を省略する、全原子解像度を持った粗視化モデルである。PULCHRA などのソフトウェアを用いて $C\alpha$ モデルから全原子モデルを生成すると、キラリティーエラーや *cis* ペプチド、芳香環貫通がランダムな場所で多数発生してしまうため、これらを効率的に除去する必要がある。我々はこのような問題を解決するために新しいアルゴリズムを開発した。キラリティーエラーを検知すると、該当箇所水素を反転させ、また、芳香環貫通を検知した場合、貫通を取り外すようにエネルギー極小化計算を行う。*Cis* ペプチドについては該当箇所の二面角を MD 計算でバイアスをかけながら回転させる。これらをアミノ酸や核酸に対して自動的に実行する機能を分子動力学計算プログラム GENESIS に実装した。SAUA-FFR 法では、最終的に融合原子モデルを全原子モデルに変換して構造最適化を行う。

本研究では、得られた多数のデコイの中から最も確からしい構造モデルを選ぶ方法として、*c.c.*、RWplus、MolProbity スコアを組み合わせることでデコイをフィルターリングするプロトコルについても提案した (図 1B)。*c.c.* は実験マップと理論マップのグローバルな一致度、RWplus は側鎖の配向に関する統計的なスコア、MolProbity スコアは原子間衝突や主鎖、側鎖の分布に関するスコアであり、これらを順に解析することで、多数のデコイの中から、マップに一致し、かつ、タンパク質らしい構造を選ぶことができる。まず、強弱の異なる様々なバイアスの力の定数 k を用いたフレキシブル・フィッティングを複数実行し、得られたデコイの中から *c.c.* が高いものを選ぶ。次に RWplus スコアに基づいてタンパク質らしい構造を選び、最後に、MolProbity スコアに基づいて歪みが少ないモデルを選択する。従って、得られた構造は、 k への依存度が抑制され、密度マップに一致し、かつ *overfitting* が少ないものとなる。本研究では、解像度が 3\AA 程度の 8 つのシステムを対象として、融合原子モデル (SAUA-FFR 法) と全原子モデル (SAAA-FFR 法) を用いて最適化の比較を行った。

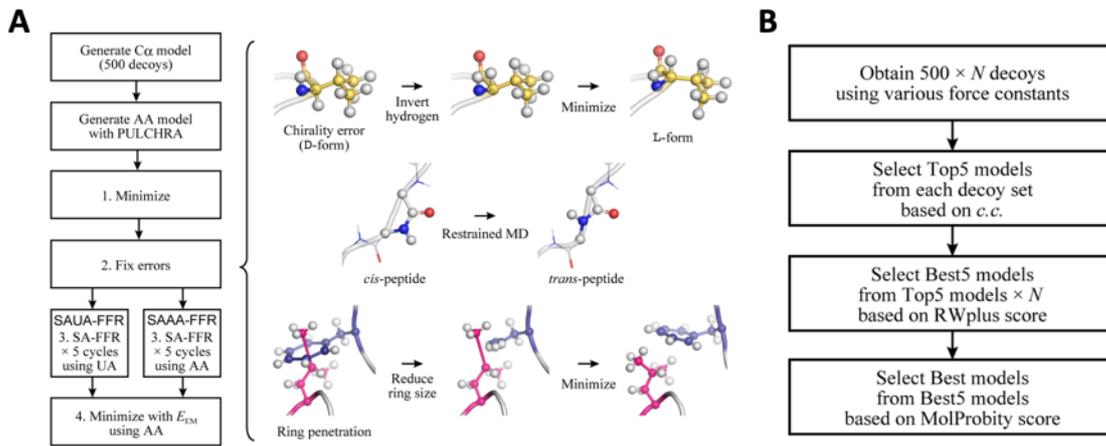


図1 本研究で開発した手法の概要 (A) SAUA-FFR 法、(B) 最適な構造デコイを選ぶ方法

4. 研究成果

De novo モデリング法の1つである MAINMAST (Terashi and Kihara, *Nat. Commun.*, 2018) で得られた 500 個のデコイに対して、SAUA-FFR 法 および SAAA-FFR 法を用いた構造最適化を行った。融合原子モデルについては CHARMM C19 力場と EEF1 モデル、全原子モデルについては CHARMM C36m 力場と GB/SA モデルを用いた。その結果、SAUA-FFR 法は SAAA-FFR 法よりも二次構造の形成が促進され、正解構造により近づくことが分かった (図2)。最も顕著な結果として、テスト系の1つである *Bombyx mori cypovirus 1* (BmCPV-1) に対する SAUA-FFR 法では、 α ヘリックスを 93.2%、 β ストランドを 64.9% の割合で正しく再現していた。また、SAUA-FFR 法では、MDFF や Phenix, SAAA-FFR 法に比べて、ほとんどのテスト系において、RWplus スコアと MolProbity スコアの両方が改善されていた。この理由として、融合原子モデルでは水素が省略されているため、全原子モデルに比べてタンパク質内部の構造に隙間があり、最適化の際に原子が動きやすいことが考えられる。また、Phenix や MDFF などの他のソフトウェアについては、デフォルトのオプションではキラリティーエラーや *cis* ペプチドが十分に除去できないことが分かった。全原子モデルを用いた通常の構造最適化法では、正解構造から大きくずれていると安定な構造の予測が難しいが、融合原子モデルを用いることで効率の良い構造探索ができることが示唆された。

本研究では、すべてのテスト系に対してバイアスの力の定数を $k = 2000, 4000, 6000, 8000, 10000$ kcal/mol に固定して計算を行った。力の定数は *overfitting* を極力抑える最適な値を選ぶことが好ましいが、これはシステムサイズやマップの解像度にも依存することが我々の最新の研究から分かってきた。今後は、そのような最適な力の定数を選ぶ経験則と SAUA-FFR 法を組み合わせることで信頼性の高い構造最適化に有効であると考えている。

本研究成果は、Takaharu Mori, Genki Terashi, Daisuke Matsuoka, Daisuke Kihara, and Yuji Sugita, “Efficient flexible fitting refinement with automatic error fixing for de novo structure modeling from cryo-EM density maps”, *J. Chem. Inf. Model.*, 61, 3516–3528 (2021). (オープンアクセス) にて発表した。本研究で開発した手法は、分子動力学計算ソフトウェア GENESIS に実装されており、フリーソフトウェアとして利用できる (<https://www.r-ccs.riken.jp/labs/cbrt/>)。また、SAUA-FFR 法に関するチュートリアルについても当ウェブサイト内で公開した。

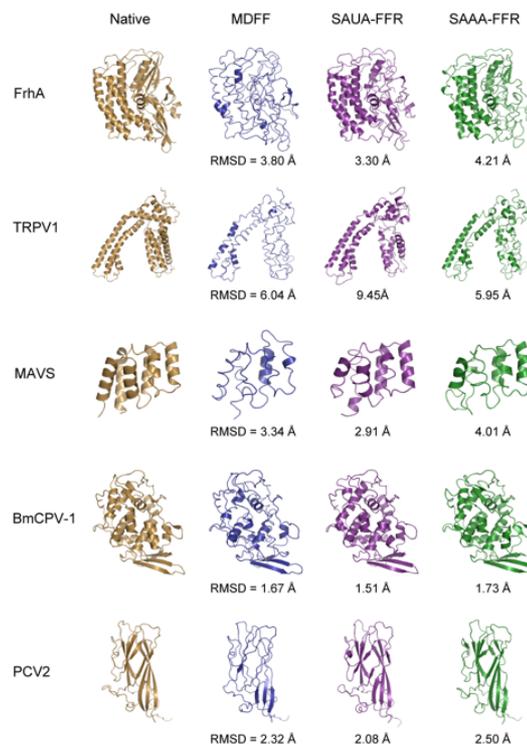


図2 MDFF, SAUA-FFR, SAAA-FFR による最適化構造と正解構造 (天然構造) との比較。8つのテスト系のうち代表的な4つを示す。

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件/うち国際共著 1件/うちオープンアクセス 3件）

1. 著者名 Mori Takaharu, Sekine Shun-ichi	4. 巻 12
2. 論文標題 Overview of the “1SBA: integrative approaches towards understanding of gene expression” session at the 57th BSJ meeting	5. 発行年 2020年
3. 雑誌名 Biophysical Reviews	6. 最初と最後の頁 253 ~ 254
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s12551-020-00644-1	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Kulik Marta, Mori Takaharu, Sugita Yuji	4. 巻 8
2. 論文標題 Multi-Scale Flexible Fitting of Proteins to Cryo-EM Density Maps at Medium Resolution	5. 発行年 2021年
3. 雑誌名 Frontiers in Molecular Biosciences	6. 最初と最後の頁 631854
掲載論文のDOI（デジタルオブジェクト識別子） 10.3389/fmolb.2021.631854	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Mori Takaharu, Terashi Genki, Matsuoka Daisuke, Kihara Daisuke, Sugita Yuji	4. 巻 61
2. 論文標題 Efficient Flexible Fitting Refinement with Automatic Error Fixing for De Novo Structure Modeling from Cryo-EM Density Maps	5. 発行年 2021年
3. 雑誌名 Journal of Chemical Information and Modeling	6. 最初と最後の頁 3516 ~ 3528
掲載論文のDOI（デジタルオブジェクト識別子） 10.1021/acs.jcim.1c00230	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕 計7件（うち招待講演 2件/うち国際学会 1件）

1. 発表者名 森貴治
2. 発表標題 高速クライオ電顕フィッティング法の開発と応用
3. 学会等名 CBI学会2020年大会
4. 発表年 2020年

1. 発表者名 森貴治、寺師玄記、松岳大輔、木原大亮、杉田有治
2. 発表標題 Improved Structure Refinement in MAINMAST for de novo Modeling from Cryo-EM Maps
3. 学会等名 第20回日本蛋白質科学会年会
4. 発表年 2020年

1. 発表者名 Takaharu Mori
2. 発表標題 Modeling complex biological systems
3. 学会等名 Workshop on molecular dynamics based binding free energy calculations, post-processing of trajectories and force field development (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 森貴治、江原晴彦、 関根俊一、杉田有治
2. 発表標題 Dynamic structures of the RNA polymerase II elongation complex by cryo-EM and MD approaches
3. 学会等名 第57回 日本生物物理学会年会 (招待講演)
4. 発表年 2019年

1. 発表者名 森貴治、江原晴彦、 関根俊一、杉田有治
2. 発表標題 Dynamic structures of RNA polymerase II revealed by cryo-EM and MD approaches
3. 学会等名 第19回日本蛋白質科学会年会
4. 発表年 2019年

1. 発表者名 森貴治、八木清、尾嶋拓、Jaewoon Jung、小林千草、松永康佑
2. 発表標題 GENESIS 1.4 の開発と生体分子系への応用
3. 学会等名 スーパーコンピュータワークショップ2019
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関			
米国	パデュー大学			