

令和 2 年 9 月 12 日

海外特別研究員最終報告書

独立行政法人日本学術振興会 理事長 殿

採用年度 2018

受付番号 201860016

氏名 中根 崇智

(氏名は必ず自署すること)

海外特別研究員としての派遣期間を終了しましたので、下記のとおり報告いたします。

なお、下記及び別紙記載の内容については相違ありません。

記

1. 用務地（派遣先国名）用務地：ケンブリッジ（国名：英国）
2. 研究課題名（和文）
構造変化を伴う生体高分子の電子顕微鏡による立体構造解析法の開発
3. 派遣期間：2018 年 9 月 12 日 ~ 2020 年 9 月 11 日
4. 受入機関名及び部局名
MRC Laboratory of Molecular Biology（MRC 分子生物学研究所）

も可)

(研究・調査実施状況及びその成果の発表・関係学会への参加状況等)

(注)「6. 研究発表」以降については様式 10-別紙 1~4 に記入の上、併せて提出すること。

海外特別研究員としての活動期間が始まる前から、筆者は RELION における MultiBody refinement 法の開発に従事していた。本法は、構造変化がある標的をいくつかの剛体の和として近似し、各剛体の方位を粒子ごとに決定するものである。特別研究員としての活動では、本法をいろいろなデータセットに適用し、その限界と改善点を探るところから始めた。その結果、剛体への分割が成立しないような連続的な変形を示す蛋白質があること、各剛体が小さい(およそ 200 kDa 未満)場合にシグナル雑音比が悪すぎて方位を安定的に決定できないことが明らかとなった。

連続的変形に向けた試みと問題点

三次元主成分分析(3D-PCA)は、連続的な構造変化がなす多様体を低次元の反応座標系に埋め込む手段として以前より報告されている方法だが、高分解能データセットへの適用例がほとんどないため、自分でも実装して検討した。具体的には、粒子の部分集合をブートストラップ法で多数抽出し、それぞれで三次元再構成を行ったあと、双対空間で固有値分解を施すことにより、メモリ使用量と計算量を抑えつつ固有マップを計算した。

この方法の弱点は、構造変化に伴うマップ中の各ボクセル値の変化を線形変化で表現することである。例えば、ドメインがあるボクセルを横切って動く状況を考えてみる。最初、ボクセルとドメインは重なっていないので、値はゼロである。ドメインがボクセルの上に重なるにつれてボクセルの値は徐々に増大し、ドメインが通りすぎると再びゼロに戻る。つまり、ボクセル強度は上がって下がるという山形の変化を呈し、明らかに線形変化ではない。注目する領域の分解能が高くなればなるほど、また、動きの範囲が分解能に比べて大きくなればなるほど、変化の非線形性は顕著になる。二次構造レベルの分解能であってすら非線形性は著しく、無理に線形表現して外挿すると構造が破綻したり、非物理的な artifact を生じることが分かった。これは本手法の原理的な限界であり、高分解能に到達するような改善は難しいと判断した。線形な主成分分析の代わりにニューラルネットワーク(変分自己符号化器)を用いるアプローチ(Ellen D. Zhong らの cryoDRGN など)も最近登場したが、現状の成果はやはり低分解能に留まっている。

つまるところ、粒子 1 つ 1 つに含まれる情報が限定的であることが問題の根源である。1 つ目の問題は、投影像は投影方向(厚み)の情報を完全に失っているということ。もう 1 つは、電子線損傷のため 1 つ 1 つの粒子に照射できる電子数が限られており、著しいショットノイズを含むということである。これらは原理的な限界であって改善することは不可能だが、事前知識という形で情報を補える可能性がある。一方、電子顕微鏡の光学系の不完全性(収差)や検出器の性能については、ハードウェアや処理ソフトウェアの面から改善の余地がある。この 2 つの点についてアプローチした。

事前情報のニューラルネットワークによる表現

RELION では三次元再構成の際に、観察されたデータ(= 粒子像) の尤度 likelihood と、生体高分子として妥当な構造かどうかという事前分布 prior distribution のバランスをとる、最大事後確率推定 Maximum A Posteriori (MAP) estimation を行っている。現在の実装では事前情報として、高周波数成分を指数関数的に減衰させるような Wilson 分布を用いている。Wilson 分布はランダムな原子配置から数学的に導出可能であり、結晶学でもよく用いられる分布である。

我々は「蛋白質らしい構造」「核酸らしい構造」について、より多くの知識を持っている。例えば、蛋白質中の原子はランダムに並んでいるのではない。原子間の結合距離や結合角には許された範囲があり、それらがアミノ酸を構成し、アミノ酸が縮合してペプチド鎖となり、 α -ヘリックスや β -シートといった二次構造を形成し、さらにドメインに折り畳み……といった階層構造を持っている。Wilson 分布は、このような化学的な知識を無視している。構造変化を伴う小さいドメインの方位を決めたり、マップを改善するには、少しでも多くの事前情報を利用することが望ましい。

原子モデルの精密化においては、これらの情報を chemical restraints として利用するのが当たり前になっているが、原子モデルを作る前のマップの段階で、これらの情報をどう表現し、利用したらよいかは難しい問題である。筆者は同グループ所属の Dari Kimanius やスウェーデン KTH の Ozan Öktem らと共同して、ニューラルネットワークを用いて事前情報を表現する方法を検討した。具体的には、ノイズや間違いを含む低分解能マップを、より生体高分子らしい、やや高分解能のマップに修正するようなニューラルネットワークを denoiser として訓練する。その微分を正則化項として RELION の三次元再構築に取り入れた。その結果、人工的に作った(正解の分かっている)データについては、従来法よりも正確な方位を回復することができ、再構成の分解能を向上できた。この結果は IUCrJ において査読中であり、preprint とソースコードを公開済み(doi: 10.1101/2020.03.25.007914)である。現実のデータセットへの適用に向けて作業中である。

局所分解能が著しく異なる標的の安定的精密化

構造変化を伴う蛋白質や界面活性剤ミセルなどに覆われた膜蛋白質では、部位によって局所分解能に大きなばらつきがある。従来の精密化は実空間では分解能が均一であるという前提に基づいているため、このような系では局所分解能が低い部分においてノイズに対する過剰適合が発生することがあった。Imperial College London の Christopher H.S.Aylett らとの共同研究により、実空間において座標依存的なフィルタを施すことでこの問題を回避するプログラム SIDESPLITTER を開発した。この成果は Journal of Structural Biology にて発表した。

高次収差

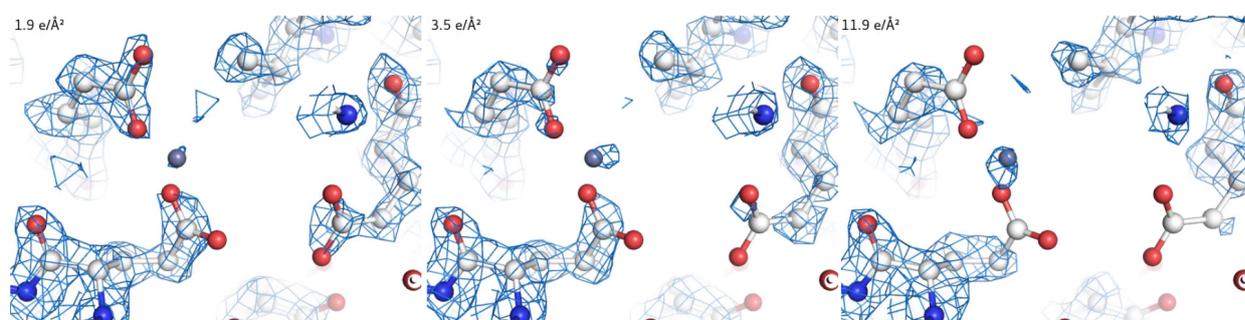
所属グループの Jasenko Zivanov らと共に、電子顕微鏡で生じる高次収差の影響を計算によって補正する方法を開発した。従来の CTF モデルは、強度変調/位相シフト・焦点外し・球面収差のみを扱っていたが、波面収差関数を任意の次数の Zernike 多項式で表現し、データセットごとに最適化できるようにした。

EMPIAR データバンクに登録されている生データを再処理することにより、一部の構造の分解能を大幅に改善することができた。例えば、EMPIAR-10185 の 20S proteasome は、当初 3.1 Å で報告されていたが、2.26 Å まで改善した。この成果は IUCrJ にて発表した。

動画処理

Thermo Fisher Scientific 社の新しい直接電子検出器である Falcon4 は、Electron Event Representation (EER) という新しい形式で動画を出力できる。従来の検出器は出力データ量を抑えるため、内部フレームを数十枚ごとに加算してから一コマとして出力していたが、EER 形式ではそのようなまとめをすることなく、検出器本来の時間分解能を保ったまますべての電子イベントを記録することができる。

RELION をこの形式に対応させたうえ、フレームごとに再構成するプログラム `relion_movie_reconstruct` を実装した。その結果、以下に示すように、電子線損傷に伴って側鎖のカルボン酸が脱炭酸する様子などを動画として観察できた。



高速・自動処理

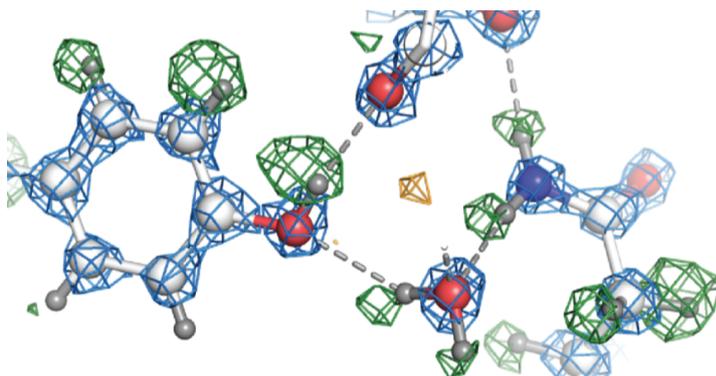
X 線結晶学では処理の自動化が進んでおり、標準的なデータセットなら、生データ(回折像)から構造まで、ほとんど人間の介入なく数十分程度で処理することが可能である。電子顕微鏡画像処理の現状はこれに遠く及ばないが、撮影速度の高速化に伴って、処理の高速化・自動化が要求されている。現在、所属グループの大学院生 Liyi Dong と共に、深層学習を用いて二次元分類後のクラス平均の中から良いクラスを自動で選択するプログラムを開発している。また、前プロジェクトで従事した大規模 XFEL データ処理の経験を生かし、自動処理パイプラインの構築や高速化を目指している。

超高分解能構造解析

これらの工夫を総合して、2019 年にマウス由来アポフェリチン重鎖を 1.54 Å で構造決定することに成功した(EMDB-9865)。低温電界放出電子銃・オメガ型エネルギーフィルタ・K2 検出器を搭載した日本電子の CRYO ARM300 を用いたもので、大阪大学加藤・理研米倉らとの共同研究である。これは、2019 年 9 月初旬時点で EMDB に登録されている単粒子解析によるマップの中で、世界最高分解能であった。

さらに 2020 年春、低温電界放出電子銃・エネルギーフィルタ・Falcon 4 検出器を搭載した Thermo Fisher Scientific 社の新型電子顕微鏡を利用することにより、同蛋白質の分解能を 1.22 Å まで改善し、クライオ電顕単粒子解析法の世界最高分解能記録を再度更新した。

1.22 Å ではほとんどの原子が独立した球として観察でき、**真の原子分解能**とすることができ(下図)。また、水素を含まない原子モデルから計算したマップとの差マップ(図、緑色のメッシュ)を計算することで、明瞭に水素原子を観察することもできた。場所によっては、水分子の水素原子さえ見えていた。



この顕微鏡を利用して、ヒトの塩化物イオンチャネルである GABAA 受容体の構造を、膜蛋白質のクライオ電顕単粒子解析法としては最高の 1.70 Å (局所的には 1.60 Å)で決定することにも成功した。このような分解能では、側鎖やリガンドの向きを確定し、水和状態や水素結合ネットワークの詳細を観察することが可能であり、酵素の反応機構や蛋白質とリガンドの相互作用などを化学のレベルで議論できる。応用面でも、構造に基づいた創薬 (SBDD) に大きく貢献する。この成果は Nature に受理された(本報告執筆時点で *in press*)ほか、Nature News でも取り上げられ(doi: 10.1038/d41586-020-01658-1)、大きな反響を呼んだ。

このデータを元に、電荷の解析など X 線に対する電子線の利点を活かすような解析ができないか、同研究所の Garib Murshudov らと検討している。

他の構造解析

このほか、condensin の構造解析や SARS-CoV-2 スパイク蛋白質の構造解析にデータ処理の面で協力し、それぞれ Nature Structural & Molecular Biology および Nature に発表した。