

課題名 (タイトル) :

電子顕微鏡による単粒子解析

利用者氏名 : ○竹本 千重, 横山 武司, 加茂 友美, 赤坂 領吾, 長内 隆

所属 : ライフサイエンス技術基盤研究センター 構造・合成生物学部門 タンパク質機能構造研究チーム

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

生命科学の分野において、生体高分子の複合体の構造情報を得ることは、それらの生理的機能や制御機構を知る上で重要である。電子顕微鏡解析は、NMR 解析や X 線結晶構造解析と並んで、主要な手法のひとつである。生体試料の電子線損傷を軽減するために、液体窒素やヘリウムによる試料の保冷が必須であるが、近年、測定系の大幅な改良と、新たな検出器の装備により、著しく感度と分解能が向上し、クライオ条件下で取得した生体分子複合体の単粒子解析は、原子分解能に迫ろうとしている。また、試料調製からデータ収集において、飛躍的なハイスループット化が可能になり、試料・測定条件さえ整えば、数日で必要なデータが取得できる環境が整いつつある。一方で、解析データの分解能は、扱う画像枚数 (ファイル数) やアライメントの精度に依存するため、迅速にデータ解析を行って、精度の高い構造を得るための計算環境を如何に確保できるかが、重要な課題となってきた。

当グループでは、平成 26 年 3 月に設置されたクライオ電子顕微鏡を使って測定を始めている。得られたデータをローカル PC (8core, 2cpu, 2台) で計算させているが、現時点で必要と想定されるデータの 10 分の 1 のテストスケール計算でさえ 60 時間かかってしまい、計算環境の改善は必須要件となっている。今後、高分解能構造解析が迅速に行える環境を実現するために、RICC の計算ノードに必要なソフトのインストールし、解析手順を含めて検討を行いたい。

2. 具体的な利用内容、計算方法

2-1. 使用したプログラム

・ CTFFIND3

<http://grigoriefflab.janelia.org/ctffind4>

・ RELION

[http://www2.mrc-lmb.cam.ac.uk/relion/index.php/Main\\_Page](http://www2.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page)

2-2. 解析手順 (ステップ)

1) Preprocessing: 取得画像の非点補正等に用いる CTF 値を算出し、フーリエ変換関 (Thon ring) を表示する。これを基に質の悪い画像を除く。

2) Particle Picking: 1) で選択した画像から粒子像 (particle) を選択・座標計算・抽出する。

3) 2D classification: 粒子像を、2D 画像として似ているもの同士をクラスにまとめ、アライメント・アベレージを実行する。明らかに間違った像 (クラス) を除く。

4) 3D classification: 3D projection (投影像) としてアライメントし、異なる構造由来の分子像を仕分ける。

5) 3D reconstruction and refinement: 4) で選択したクラスに属する分子像を使って 3D 構造を構築し、back-projection (逆投影) とより細かな角度でのアライメントを繰り返すことで、構造を精密化する。

\*ファイル数、クラス数、アライメント精度の設定によって、また、ステップによっても、必要となる計算資源が変わる。

2-3. データ

近い将来、測定方法の改良によって必要なデータ量が減る可能性もあるが、現時点での解析手法では、分子像の数に依存して分解能が向上する傾向にある。目標とする分解能を得るために必要と予想されるデータサイズ・量は以下の通りである。

- ・画像 1 枚あたり 100 分子像として、2000-4000 枚 (200-400k particles)
- ・画像サイズは、約 33Mbyte/枚

\*粒子像のファイルサイズは、撮影時の倍率や分子の大きさによって異なる。

3. 結果

- ・ステップ 2 の Particle picking は、画像を見ながら判断する部分があるので、ローカル PC での作業とした。
- ・最初は、ローカル PC に近い環境で使用するために多目的クラスタ(upc)を使用していたが、待ち時間も合わせると、全体の作業が効率化しないというジレンマに直面し、超並列クラスタ(mpc)を使用する環境整備に取り組んだ。
- ・1600 枚程度の異なるデータセットについてステップ 1 を行ったところ、upc では 32 core で 30 時間程度、mpc では 128core で 8 時間だった。
- ・mpc は待ち時間が少なく、実用的なので、コア数を倍にして、加速するかどうか等、テストしてみたい。
- ・ステップ 3 の 2D classification 以降は、使用メモリ量が心配だったので、ローカルでテストしてから移行することにした。
- ・8 月に RELION ver.1.3 がリリースされたので再度インストールを行った。
- ・ステップ 3 の 2D classification は、RAM を thread に対して割り当てる造りになっていて、正しく解釈できていないので、まだ必要容量が解明できていない。

4. まとめ

- ・研究室でのローカルの環境とほぼ同じ感覚で使うことができた。
- ・プログラムが年度途中でバージョンアップしたため、対応に時間がかかってしまい、RICC の運用変更についていけず、新しいプログラムを試すことはできなかった。
- ・計算時間を要するステップのメモリ使用量を正確に把握して、今後の利用に備える必要がある。

5. 今後の計画・展望

- ・新しいシステムに対応できるか不安ではあるが、引き続き使用する事を希望している。
- ・神戸の京を試用した結果では、あまり相性が良くない印象だったが、プログラムもバージョンアップしたので、状況が違うのかもしれない。
- ・12 月に CTFFIND4 がリリースされており、ス

テップ 2 の処理速度が向上しているので、インストールしたい。

- ・解析手法としては、GPU も活用できそうだが、RELION は正規には対応していないので、他のプログラムを試すかどうかなど、検討する予定である。