

## 第4回中級者講習会

パネルディスカッション ～精密化はいつ「完了した」といえるのか～

2019/11/25 藤橋雅宏

講習会の実施に先立ち、「構造精密化が完了したとみなすために必要な項目」と「構造精密化の際に便利なソフト」の、アンケートを行いました。ご協力ありがとうございました。アンケートに記載していただいたコメントをそのまま掲載します。パネルディスカッション当日には、類似のコメントをひとまとめにした summary を皆様にお示ししました。この summary については、発表スライドも upload いたしましたので、そちらをご覧ください。

下表で、

黒文字は、回答者が構造精密化が完了したとみなすために必要な項目としてご回答頂いた内容です。  
青文字は、構造精密化を行う際に、この目的にはこのソフトが便利と教えて頂いた内容です。

同一回答者からの回答は、一枠の中に記述しています。

グレーのマーカーは、アンケート回答者が疑問に感じていることを、強調したものです。

なお、回答フォームは *italic* などを受け付けない仕様になっていたため、本来 *italic* 等で記述すべき事柄も *plane style* で入力して頂いています。このまとめでは、皆様に頂いた回答をできるだけそのまま掲載するため、形式の補正はしていません。ご承知おきください。

### 精密化に関するアンケートにいただいたコメント

- ・ポリペプチドを N 末から C 末まで一度はちゃんと見て電子密度に合わせて直し（特に主鎖）、molprobity の His/Asn/Gln flip を参考にして直し、coot の validation のいくつか決めている項目で問題点を消し、difference density peak をチェックして blob (solvent molecule/metal) や拾い忘れの水をチェックし、変な水（水素結合距離、ピーク高さなど）を全部取り払って、自動精密化サイクルで R/Rfree が落ち着いたら、終わりにしています。
- ・本当は phenixの方がよさそうなのだけれどつい使い慣れた Refmac を使ってしまう。
- ・ R 値がどちらも 30%以下
- ・  $\Delta R$  が 5%以下
- ・ ラマチャンドラプロットと rotamer outlier に (map が poor な領域などは除いて) 問題がない
- ・ PDB Validation Service を使ったときに忠告等が出ない
- ・ 有意な電子密度に過不足なくモデルを置くことができた
- ・ Ramachandran Plot などで問題がない（説明ができる）
- ・ Coot の Validation の項目を上から順に実行して結果に問題がない
- ・ PDB の Validation Service を使用して確認"

<ul style="list-style-type: none"> <li>• R-factor が 20%以下になる。</li> <li>• Ramachandran plot から Disallowed がなくなる。disallowed でも目で見て電子密度にはまっていれば許容。</li> <li>• Coot の各種チェックをパスする。パスしなくても目で見て電子密度にはまっていれば許容。</li> <li>• 埋められる電子密度は全て埋める。 "</li> </ul>
<ul style="list-style-type: none"> <li>• R, Rfree の値、 Ramachandran plot</li> </ul>
<ul style="list-style-type: none"> <li>• 分解能に見合った R 値</li> <li>• PDB validation でそこそこの値</li> <li>• crash になるべく少ない</li> </ul>
<ul style="list-style-type: none"> <li>• 2Fo-Fc マップを見て水分子、alternate conformer、末端残基がこれ以上置けない</li> <li>• molprobity を使って全体に構造が正しいかチェックする</li> </ul>
<p>(1) 電子密度マップの解釈 (溶媒分子含む)</p> <p>(2) Rwork/Rfree</p> <p>(3) タンパク質、核酸、リガンド分子についてジオメトリー(結合角、結合距離、ラマチャンドランプロット、側鎖のロータマー、クラッシュスコアなど) が問題ないか。</p> <p>(4) 複数の結晶データを持っている場合には、各データ間で結果の解釈に違いがないことを確認</p> <p>基本的には PDB のバリデーションレポートで明らかな異常値がなくなるまで行っていますが、かなりの労力と時間を費やしてしまいます。どこまで気にしたらよいのかよくわかりません。</p>
<ul style="list-style-type: none"> <li>• Phenix で大きい赤や緑のメッシュが出なくなった時。</li> </ul>
<ul style="list-style-type: none"> <li>• ラマチャンドランプロットの disallowed region を 0 にする。</li> <li>• 水分子と思われる電子密度を全てアサインする。</li> <li>• 分解能にしたがって目標の R-factor を決め、それをクリアする。</li> <li>• R-work と R-free の差がおよそ 5%程度になるようにする</li> <li>• 自動モデリングは Buccaneer が高速で良いモデルを組んでくれるので好きです。</li> <li>• Phenix の Morph model(時間がかかる)や Refmac の jelly body refinement(比較的すぐ終わる)は MR の後にやると大幅に R が下がる場合が(たまに)あるのでとりあえず試します。</li> <li>• ラマチャンドランプロットの改善に取り組んでいて行き詰まった時に coot の Extention, All Molecule, Refine/improve Ramachandran plot... を試すこともあります。coot では Ramachandran restraints にチェックを入れたりしますが(効果があるかはよくわかりません)。</li> <li>• モデリングで Ramachandran が許容範囲になっても、精密化を走らせた後には元に戻ってしまうような場合に、有用な方法があるか知りたいです。</li> </ul>
<ul style="list-style-type: none"> <li>• Coot の「Unmodelled blobs」でヒットしなくなった</li> <li>• Ramachandran plot で極端にハズレたところなくなった</li> <li>• wwPDB Validation Service で構造の誤りを指摘されなくなった</li> <li>• R-work/R-free の値を下げたいときは Refmaac を使用し、Ramachandran plot や Geometry Analysis の結果を改善したときには Phenix.refine を使用しています。</li> </ul>

<ul style="list-style-type: none"> <li>• Rfactor, Rfree, Rmerge, Completeness</li> <li>• 数値に関しては、構造解析の経験が長い先生と相談しながら、その先生の経験で決めています。</li> </ul>
<ol style="list-style-type: none"> <li>1) ラマチャンドラプロットに問題がないこと、</li> <li>2) PDB のバリデーションサーバーでの結果に問題がないこと、</li> <li>3) どの余っている電子密度に原子を置いても Rfree が改善しないこと、</li> </ol> <p>の3つを満たすときに終了としている。</p>
<ul style="list-style-type: none"> <li>• Validation の結果を見て決める。</li> </ul>
<ul style="list-style-type: none"> <li>• Fo-Fc で、ぱっと見でいかにも説明がつくピークが無くなった。</li> <li>• FreeR が下がらなくなった。</li> <li>• Phenix.refine の MolProbity の各タブで指摘されるものをあらかじめ解決した（すべては無理）。</li> <li>• PDB の validation サーバーに投げてみて、「恥ずかしくない」程度になっていれば終了する。</li> <li>• 分子置換直後から水を拾う前ぐらい、または粗く精密化してリガンドが入っているかどうかを判断するだけの段階では Refmac を使っています。計算が速いからです。</li> <li>• 水を拾って、しっかり精密化を終わらせる段階では、Phenix.refine を使っています。計算に時間はかかりますが、水を拾う精度が高いこと、validation との連携が便利なところが利点と思っています。あとは、精密化最終段階にリストレインをいろいろ細かく指定することができるので便利です（Refmac でのやり方を知らないだけ?）。polder map もいいかも。</li> </ul>
<ul style="list-style-type: none"> <li>• Rfree が分解能の 1/10 ぐらいになるまで。</li> </ul>
<ul style="list-style-type: none"> <li>• Ramachandran plot の outlier を可能な限り 0 に</li> <li>• coot の varidation の, geometry と rotamer におかしな箇所がないように</li> <li>• Rfree が 20% 台前半以下になるように（分解能にもよる）</li> </ul>
<ul style="list-style-type: none"> <li>• 最初は Free R 値が 30%以下になるのを目指して、その後は Fo-Fc マップの正と負の部分をしてだけ無くすように頑張るだけ頑張る感じで精密化を進めています。</li> <li>• 最終的に構造バリデーションで指摘される、主鎖の二面角の異常や原子同士のクラッシュしているところを修正して完了させています。</li> </ul>
<ul style="list-style-type: none"> <li>• R, Free R が~27 くらい</li> <li>• ラマチャンドラプロットで disallowed が無い</li> </ul>
<ol style="list-style-type: none"> <li>1. 解釈可能な電子密度にモデルがすべて構築されていること。</li> <li>2. R/Rfree 値が分解能と対応して妥当な値以下まで下がっていること。</li> <li>3. molprobity のモデル評価で outlier がないこと。ただし dihedral angle restraint, rotamer については outlier を許容している。"</li> </ol>
<ul style="list-style-type: none"> <li>• 大体の positive peak が見えなくなる。</li> <li>• PDB に登録されている構造の中で分解能に近いものの R-free の中央値を調べ、それより下回ったら終了とする。</li> </ul>
<ul style="list-style-type: none"> <li>• 分解能に対して、R/Rfree がどの程度になったか。</li> <li>• 質問の答えになっていないのですが、以下知れたらいいなを記入しました。自分が refmac を使っていて R が改善されなかったものが、他の研究室で Phenix.refine を使っている人では落ちたと聞きました。どんな違いがあったのか講習を通して分かったら良いなと思いました。</li> </ul>

<ul style="list-style-type: none"> <li>• R 値が、最大分解能の値に近く、かつ R 値と Rfree 値のズレが 5%以内であること。後は、電子密度とモデルの兼ね合い。(ただし、いまだに構造精密化を完了したことがありません。)</li> </ul>
<ul style="list-style-type: none"> <li>• R 因子が分解能に対して妥当な値で下げ止まる。</li> <li>• Ramachandran プロットで Outlier が 0.1%以下。</li> <li>• 温度因子が周りと比べて高すぎる場所がない。</li> <li>• ざっと見て水がすべて拾えている。"</li> <li>• 特殊なリガンドは JLigand で PDB と cif を作ります。</li> <li>• それ以外は殆ど Coot&amp;Refmac5 で出来ます。</li> </ul>
<ul style="list-style-type: none"> <li>• 下記青文字のプロセスにてまずは、Rfree が分解能 2.0Å以下の場合は 0.20 以下を目指す。2.0-2.4Å の場合は 0.24 以下。2.4-3.0 Å 以下の場合は 0.26 程度(ここはかなり曖昧)。</li> <li>• 上記のプロセスを実施した後、タンパク質部分においては ramachandran plot を確認する。低分子化合物やペプチドなどのリガンドの場合は特に重点的に納得がいくようにモデルの置き方を検証する。</li> <li>• リガンド周辺だけでなくモデル全体の DFo - mFc map を描画し、特に不可解な電子密度図がないかを検証する。</li> <li>• 最後は、PDB deposit に投げて、文句を言われなくなるようになるまで修正する。"</li> <li>• 分子置換後の最初の精密化は、早く MR の解が正しいか否かを知りたいので、REFMAC5 を用いる。</li> <li>• その後、解が正しければ続いて phenix.refine (simulated annealing = True、ordered_solvent = False) で構造精密化を行う。この simulated annealing を使えるのが phenix.refine しかない(?)ため、phenix.refine をここでは用いており、以降も default の parameter を用いた場合に phenix.refine の方が良い構造を出力するため、phenix.refine を用いている。</li> <li>• その後、一度、coot にて model 全体を修正した後、phenix.refine (simulated annealing = False、ordered_solvent = True)で構造精密化を行う。複数の chain がある場合は、chain 毎に coot -&gt; phenix.refine を行う。</li> <li>• 低分子リガンドを入れる場合はここで phenix.ligand_fit を用いるのが有効である。</li> <li>• 再び model 全体と水分子を確認した後、phenix.refine (simulated annealing = False、ordered_solvent = False)で構造精密化を行う。</li> <li>• phenix.ligand_fit のための ligand pdb ファイルと cif ファイルの準備には marvin sketch と jligand の組み合わせが便利である。</li> <li>• 初めて homolog 構造から分子置換法によって位相を計算し、構造精密化を行う場合は、buccaneer, arpwrap, phenix.autobuild によるモデル構築 (もしくはその組み合わせ) を用いることで構造精密化の労力を大幅に削減することが可能である。</li> </ul>
<ul style="list-style-type: none"> <li>• Coot の validate、MolProbity 等によって構築した構造の妥当性を評価した結果、それ以上修正できそうな箇所がない場合"</li> </ul>

<ul style="list-style-type: none"> <li>• 全て resolution にもよるが、例えば 2.5Å 程度であれば、 R/Rfree の差 - 目標 3%以内、5%以下 R は低ければ低い方がいいが、より geometry 優先 Ramachandran は Outlier は 0.3%以下。 水を置く場合はこれ以上 occupancy1 の水を置けなくなったら。"</li> </ul>
<ul style="list-style-type: none"> <li>• R 値が十分に下がっている</li> <li>• 解釈可能な差フーリエが残っていない</li> <li>• steric clash を可能な限り除いている</li> <li>• Ramachandran plot, rotamer, peptide omega, bond length/angle 等の outlier になっている箇所を見て、修正可能かどうか確認している（ゼロにすることではない）</li> <li>• PDB の validation を確認している</li> <li>• 論文の表・図をすべて作り終えている</li> <li>• rigid body は phenix.refine が収束半径が大きくて便利</li> <li>• refmac5 の jelly body は分子置換直後に有用</li> <li>• 異常散乱体が含まれているときは refmac5 の SAD function も有用</li> </ul>
<ul style="list-style-type: none"> <li>• 構造精密化が完了したとみなす項目はモデル構築をし直しても R-free が下がらなくなった時だと思っています。どこまで下がるかは分解能やデータの質にもよりますが、3 Å レベルであれば 30%、2.5 Å 以下は 25%、2 Å 以下は 20%を目安にしています。</li> <li>• 15 年前までは CNS/XPLOR や REFMAC を併用して精密化を行っていましたが、PHENIX が出来からはこれ一本で済ませています。時々オーバーフィッティングが気になりますが、COOT や PYMOL との連動も非常によく使いやすいです。</li> </ul>
<ul style="list-style-type: none"> <li>• outlier を 0 にする</li> <li>• R factor が 20%以下、R free が 25%以下になる</li> </ul>
<ul style="list-style-type: none"> <li>• R-work、R-free（下げ止まりまで）</li> </ul>
<ul style="list-style-type: none"> <li>• PDB ファイル登録時の要求が全て満たされた時。</li> </ul>
<p>私が個人的に指標としているのは以下の項目です。</p> <ol style="list-style-type: none"> <li>1. 電子密度に対して適切に蛋白質、リガンド、水分子を配置できているか。</li> <li>2. R/Rfree が分解能に対して十分に低下しているか。</li> <li>3. Molprobity を使って二面角や側鎖の Rotamer が許される範囲に入っているか。</li> </ol> <p>最近 Molprobity のスコアを論文に記載することも多いので、3 を最も大事にしています。 大まかな精密化には Refmac5、仕上げの精密化には Phenix.refine を使用しております。</p>
<p>以下の項目を全て満たすことを基準としています。</p> <ul style="list-style-type: none"> <li>• R work と R free の差が 5%以内</li> <li>• Bond lengths が 0.015 Å 以下</li> <li>• Bond angles が 1.5 °以下</li> <li>• 基質複合体などの場合は、基質等の mFo-DFc SA-omit map が contour level 3 sigma で明確に観察でき、基質とアミノ酸残基、水分子の温度因子に大きな差がないこと"</li> </ul>
<ul style="list-style-type: none"> <li>• すべての電子密度図が解釈できたとき</li> </ul>

<ul style="list-style-type: none"> <li>• まず Coot でこれ以上モデルが改善できないところまで手直しする。</li> <li>• Phenix.refine の MolProbity で以下の項目をできるだけ改善する ; Basic geometry (outlier 0), Ramachandran (outlier 0), C-beta (outlier 0), clash score (分解能にもよるが 2-3 以下)。</li> <li>• PDB の validation server でチェック ; major issues を指摘されれば修正する。</li> <li>• 一般的なソフトのみを使用していますが参考までに。主に分子置換法で位相決定した構造の精密化を行なっています。位相決定後の初期段階の精密化では Coot と Refmac の組み合わせで大体の手直しを行ない、水分子を入れるところから最終段階まで Phenix.refine を用いています。使い分けの理由は以下の通りです。Refmac (処理速度が速い) Phenix.refine (Refmac に比べて時間がかかるが、geometry/統計値がより改善する。MolProbity の結果を参照できる)</li> </ul>
<ul style="list-style-type: none"> <li>• PDB に登録されている同等分解能のデータから、Rwork の偏差値を算出し、40 以上であること。</li> </ul> <p>Rfree、Rfree-Rwork も同様に評価。</p> <p>RMS bond angle が 1.7 台に乗っていること。</p> <p>Ramachandran outliers &lt; 0.2%</p> <p>Ramachandran favored &gt; 98%</p> <p>Rotamer outliers &lt; 0.3%</p> <p>Rotamer favored &gt; 95%</p> <p>Clashscore &lt; 4</p> <ul style="list-style-type: none"> <li>• 上記の評価項目をポイント化し、独自の関数を用いて、タンパクモデルの妥当性を点数化している。解析のスループット上、全てを満たすことは困難であるため、一定の点数以上あればよしとしている。</li> <li>• リガンドモデルについては、omit map とモデルの rsc 妥当性評価に用いている。"</li> <li>• 同形性のある結晶を大量に解く場合、dimple が良い。早い、かつジオメトリーもボロボロにしない程度で結果を返してくれる。</li> </ul>
<ul style="list-style-type: none"> <li>• 精密化の統計値。</li> <li>• 電子密度とそれに対するモデルの妥当性。</li> <li>• 特にこだわりはありません。</li> </ul>
<ul style="list-style-type: none"> <li>• phenix.refine に入っている MolProbity を指標に使っています。そこで ramachandran plot が 0.2-0.3%以下になっていることが前提で、その後 PDB の validation server を使って最終的な修正を行っています。</li> <li>• 基本的に全ての精密化を phenix.refine を GUI を使って行っています。</li> </ul>
<ul style="list-style-type: none"> <li>• Rfree/Rwork がこれ以上頑張っても下がらない</li> </ul>
<ul style="list-style-type: none"> <li>• Molprobity を参考にしています。Ramachandran plot は outlier が無いようにしています。Rotamer outliers は 1%です(できれば 0 にする)。R factor は、下がりきるまで refine しています。</li> </ul>

- 精密化の終盤は、R, Rfree と molprobity の評価を記録しています。
- 気になるところ(validation の値が悪いとか、FoFc map にピークが見えるなど)については、それぞれ可能な改善を行ってから **refmac** で **refinement** を行い、Rfree と validation が良くなったらその改善を採用、悪くなったら不採用としてすすめ、修正可能なところがなくなったら完了と見なしています。

• 占有率の低い分子を **omit map** で可視化するために、**phenix** の **polder** が便利

• 分解能にもよりますが **refinement** 後の R-fac が 20% くらい、free R が 26% くらい、Ramachandran Outlier が 0.5% 以下くらい。Ligand、水を置いて適宜修正して完了とします。

• まだ経験が浅いので便利なソフトがあれば教えていただきたいです。回折データ→**XDS**→**AIMLESS**→**MOLREP**→**REFMAC** 適宜 **COOT** の流れしかほとんど経験がありません。

• R 25% 以下, Rfree 20% 以下

• Ramachandran outliers がない

• Rotamer outliers がない

• 余っている電子密度マップがない (COOT の「Unmodeled blobs」基準)

• 異常に高い温度因子や、マップが **poor** な部分が削除されている