

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～ のまとめ

パネルディスカッションにご参加頂いた、パネリストの皆様、また会場やオンラインから議論に参加して下さったり、聞いてくださった皆様、ありがとうございました。

当日のパネルディスカッションでは、司会の力不足や時間の制約などから、ある程度のコンセンサスは得られたものの、議論のまとめは不十分となってしまいました。そこで、パネリストの皆様が無理をお願いして、それぞれの皆様の考える「精密化完了の目安」「他人の論文を review する際に気をつけていること」などを、書いて頂きました。

先生方のお名前の 50 音順にコメントを掲載させて頂きました。また、5 人の先生方のあとに、私自身の考える精密化についてのコメントを掲載しました。併せてご覧ください。このコメントの公開をもちまして、パネルディスカッションのまとめとさせていただきます。

ありがとうございました。

第 4 回タンパク質結晶構造解析中級者向け講習会@東京大学山上会館(2019 年 10 月 31 日)

パネルディスカッション司会: 京都大学 大学院理学研究科 藤橋雅宏

学習院大学 理学部 中村顕さん

精密化は、マニュアルでのモデル修正とソフトウェアでの計算の繰り返しかと思いますが、タンパク質分子やリガンド分子を有意な電子密度に過不足なく置いたら、Ramachandran plot を見ながら修正し、Coot の Validation の項目を上から順に実行して結果に問題がないことを確認しています。結局のところ、「過不足なく」とか「問題がない」という定義自体が曖昧で、判断が人それぞれで異なっているでしょうし、データの分解能次第でその閾値が変わるでしょうから、それが『精密化の完了点』を一義に決めることができない理由なのでしょう。例えば、Ramachandran plot の outlier は周りの環境から説明できれば、そのままにしています。私は比較的低分解能のデータを扱うことが多いため、強めの Restraint で精密化していますが、結果の解釈は甘めかもしれません（これも曖昧な表現ですが）。水分子を置いた後も何サイクルか精密化を行っていますが、そのくらいの段階では R/R_{free} はほぼ横ばいになっているのが普通です。最後の方では、モデル修正してもソフトウェアの計算で元に戻されて、ということを繰り返すのもよくあることです。何とか落としどころをみつけています。PDB の Validation Server の結果も併せて評価し、必要なモデル修正を行った上で、電子密度とにらめっこしてそれ以上の手を加えられないのであれば、そこで精密化の区切りをしています。基本的なスタンスは、講習会でも話題に上った、公開しても恥ずかしくない精密化を目指すということだと思います。もちろ

ん、改善するかどうかは別として、回折強度データの処理を見直したり、ループや末端の電子密度が弱いところのモデル修正を続けたり、細かいところを見ていけばキリがないのも当然です。そういう意味では、精密化に終わりはありません。

東京大学 定量生命科学研究所 野澤佳世さん

- ・結晶データセットの統計値で、高角の $I/\sigma(I)$ (≥ 1.5) や $CC_{1/2}$ (≥ 40) の値を見て、最高分解能が正しく見積もられているか確認する。
- ・ R_{work} と R_{free} の差が 5% くらいの範囲であることを目安に、モデルが Over fitting されていないか確認する。
- ・大まかに分解能が 3.5 Å 以下なら R_{free} が 30% 以下くらいになると見積もって構造を見る。(PDB の Advanced Search の Refinement Details Report から歴代の結晶構造の分解能と R 、 B -factor を見ることが可能。)
- ・Ramachandran plot statistics の outlier がないことを確認する。
- ・Molprobity clash score は 10 以下が理想だが、超分子複合体や分解能が低い構造では値が大きくなってしまっても仕方がない気がしている。
- ・すでに PDB に登録されている構造で基質との結合が疑わしいものは、PDB-REDO (<https://pdb-redo.eu/>) で計算された結果と比較して、電子密度とモデルの整合性を見る。

慶應大学 薬学部・大学院薬学研究科 原田彩佳さん

<精密化の際に気をつけていること>

回折強度データで頻繁に得られるであろう分解能 3Å 前後の R 因子に関しては、 $R_{free} = 25\%$ 、 $R_{work} = 20\%$ 前後を目標に精密化していきます。

リガンドなどが結合している場合は、ある程度精密化が進んだ時点で simulated annealing を用いて omit map を作成します。 $F_o - F_c$ map で 2.5~3 σ 程度でも電子密度が見えているかを確認しています。あまり見えていない場合は他のデータで再度確認するか、またはデータ取り直し。

Coot での修正における電子密度図は分解能に関係なく $2F_o - F_c$ map: 1.2 σ , $F_o - F_c$ map: 3.5 σ に設定。基本的には「見えない部分には残基を入れない」という基準で修正を行なっています。構造の妥当性は Coot の Validate の中にある項目を上から順にチェックしています。上から順にチェックしていけば、パッと見わからないような修正すべき箇所がわかるのでいつも使っています。分解能 3Å 程度であるとあまり水は拾えないことが多いと思いますが、変な形の電子密度や小さすぎる電子密度に無理に水は置かないようにしています。さらに、Ramachandran plot では outlier に入らないように修正していきますが、周辺残基を含めていくら修正しても戻ってしまう場合もあるので、それはそのまま登録してしまいます。ただ、どう修正しているか、どこを修正したか、などの履歴については word で記録を残しています。(なので Coot を開いている時はいつも word も開いて修正箇所を記載していきます。)

<他人の論文を review する際に気をつけていること>

他人が自分や自分のラボでの基準で精密化をしたり、モデルを組んでいるとは限らないので、必ず電子密度図も見るようにしています。リガンド複合体などの場合は特に注意しています。

東京大学 大学院薬学系研究科 平野良憲さん

精密化はいつ『完了した』のかは難しい問題だと思います。私の場合は「解釈可能な電子密度にモデルがすべて構築されている」時点でそろそろ精密化を終了しようかと判断します。ただし構造データを PDB へ登録するには一定の基準を満たす必要があります、また学術論文へ投稿する際に $R_{\text{work}}/R_{\text{free}}$ 値、RMSD bonds、angle や ADP (B -factor) を統計値として載せるので、これらが分解能なりに適切かチェックします。分解能によって具体的な値は変わりますが、PDB 登録データを見ると 3.8-4.0 Å 分解能でも R_{free} は 30%以下に収束しているものが半数程度で、35%を超えるものは全体の 10%未満です。また、RMSD bonds、angles はそれぞれ 0.02 Å、2.0°を超えることはまれで、分解能が高くなれば値はもっと小さくなります。この最低限度の基準を満たさないと査読でもデータの質やモデルに問題があると判断されます。またご存知の通り PDB の validation report は論文投稿の際に必須になりつつあるので、こちらの統計値も重要です。ジオメトリーは微調整して outlier が全くない状態が理想的ですが、実際にはなかなかそうもいきません。outlier は個別に再確認して電子密度とモデルに自信があれば too-close contact や rotamer はそれほど気にしませんし、Ramachandran についても水素結合形成によって outlier になることもあります。逆に電子密度が不明瞭ならモデルを修正するか削除します。RSRZ について、以前に論文投稿した際に reviewer から outlier が多いと指摘されたことがあります。outlier はループや末端の ADP の高い領域にあるが、 $2|F_o|-|F_c|$ マップで十分トレース可能でモデルもそこそこフィットしていること等を示して納得してもらえました。この辺は個々の事案によって対応が変わると思います。PDB で公開されると誰でもデータを見ることができるので、恥ずかしくない状態まで精密化したいです。

東京大学 大学院理学系研究科 山下恵太郎さん

精密化では結局のところ電子密度（妥当な方法で計算されたという前提で）と合っているかどうかが一番重要なので、なるべく全ての残基を丁寧に見ること。また Coot の Validation 機能を活用して差フーリエピークをチェックし、steric clash や各種 outlier (Ramachandran plot, rotamer, peptide omega, bond length/angle 等)の項目をマップと見比べて修正可能か検討することである。精密化計算を何度か繰り返して、修正可能な箇所も無く、 R 値も十分に低い状態なら、論文の図表等を作り終えて（作成中に何らかミスに気づくことも多いので）PDB validation report もチェックし終えて問題なければ、精密化は一旦一区切りついたと考える。

当日の議論では、Ramachandran, rotamer 等の outlier は必ずしも間違いを意味するものでなく、マップを見るべきというコンセンサスがあったのは良かったと思う。ただ数値的指標に関しては分解能によって議論を分けるべきだが、そこに踏み込むには時間の制約もあり難しかった。

また、 R 値は非常に重要な指標であるものの、原子モデル以外の要素(つまり bulk solvent や overall scaling)が与える影響も大きい。これは特にプログラム間で扱いが大きく異なっており、例えば Refmac5 で R 値 28%だったものを phenix.refine で評価するだけで 26%になったりする(原子モデルの精密化をしなくても)。これをもって十分に R 値が下がったから良いと思うのは早計であるし、またプログラム自体も変化していく中で、過去の R 値の分布と比べて十分に低いからよしとするのもまた何か違う。極端に大きな R 値から下がらないというのも何らかの異常である(双晶や空間群の間違い、その他データの異常)ことは明らかだが、そのような状況を除けば、やはりマップとモデルが十分に一致しているかどうかは重要である。ただしこの点はなかなか数値的評価では見えてこないことが多い。

論文をレビューするときは数値や図表のみから判断するしか無いので、結局一番重要な点については確認できない。分解能の切り方が適正かどうか、精密化・Validation に使ったプログラム等が明記されているかどうか(特に複数の構造が出てくるとき)、あまりに outlier が多くないか、重要な箇所について電子密度図が示されているか、金属イオンがある場合はその根拠が示されているか...などに注意している。

京都大学 大学院理学研究科 藤橋雅宏

精密化完了の見極めは簡単ではありませんが、結局のところ geometry に異常がないことを前提に、 R_{free} が改善できなくなったところをもって、完了とするのが良いのかなと感じています。

「geometry に異常が無い」ことの見極めは、「molprobity や pdb の validation server で指摘される異常のうち、説明できないものがない」とするのが良いと、いまは感じています。

例えば CNO 原子の clash は、その存在が reasonable であるという理由を探ることが基本的に難しいので、取り除くように精密化をやり直すことが自然だと思います。Ramachandran plot の outlier が存在することは稀だと思いますが、電子密度から明らかにその構造をしていると言えるなら、outlier でもかまわないと思います。個人的にはこれら 2 種の error を取り除くためには、 R_{free} が多少悪くなくても仕方ないと思っています。bond length の outlier についても、原子間距離の正しさを主張できるほどの分解能のデータを扱っているのでなければ、精密化における束縛が甘いことを示唆している気がします。

rotamer の異常や ramachandran plot で allowed となる残基は、少ない方が好ましいと思いますが、0 にするのは難しいと思います。これらの異常は、そもそも数%の残基に存在するとされているので、自分が精密化した構造についてもいくつかは存在するのが自然です。このことから、これらの異常については、一通り改善を試みても改善されなければ、諦めています。

「 R_{free} が改善できなくなる」ということを見極めるためには、記録が重要と考えています。いつも、model のどこを改良したのかを、改良前後の file 名とともに記録したうえで、双方を software で精密化しています。精密化の結果の統計値と、改良した部分が精密化によりどうなったかを記録します。この記録はノートに手で取ると大変なので、コピペをうまく使って word file

として残しています。このようにして精密化を進めると、改善するところが残っているか否かがわかりやすくなります。

終わりの見極めが特に難しいこととして、水分子と **alternate conformation** が挙げられると思います。個人的にはこれらについては、1-3 サイクル程度のチェックを行った段階で切り上げています。 F_o-F_c や mF_o-DF_c map を見ていると、ピークがいつまでも残る傾向にありますが、精密化が進むほど同じ σ で表示した電子密度におけるノイズの割合が大きくなるので、適当に切り上げることにしています。

自分が論文 reviewer になったときの結晶構造の評価ポイントは、議論の中心になっている分子が電子密度に正しく入っているかと、説明の出来ない **geometry** の異常が観察されていないか、の2点です。電子密度は、基本的には $2F_o-F_c / 2mF_o-DF_c$ map ではなく、 F_o-F_c / mF_o-DF_c map を使うべきだと考えています。特に、議論の中心になっている小分子リガンドが、 $2F_o-F_c / 2mF_o-DF_c$ map で示されている場合、 F_o-F_c / mF_o-DF_c map を出すようにコメントをつけるようにしています。