

単結晶X線構造解析 基礎講座

第14回 モデルの構築と精密化，結果の評価

松本 崇*

1. はじめに

タンパク質の単結晶X線構造解析において、タンパク質の発現、結晶化、位相決定、分子モデルの構築・修正・精密化等のステップがある。結晶化や位相決定という構造解析の中で高いハードルを越えることができる、モデルの構築および精密化のステップに進むことができる。

位相決定直後に得られる電子密度図は通常ノイズが多い。そこでモデルの構築の前に、“溶媒領域の平滑化”や“ヒストグラムマッチング”等の電子密度修正法によりノイズを減少させ、モデルの構築が行いやすい電子密度図へ改良する。この改良された電子密度図を使用して、モデルの構築のステップを実行することとなる。

モデルの構築はかつてシリコングラフィックス社等の高価なグラフィックワークステーションを駆使して行い、精密化の計算も一晩を超えるものが大半であった。しかし、現在ではコンピュータの高速化により、Windowsマシンや、Macなどのパーソナルコンピュータでも構造解析やモデルの構築が可能となり、精密化の計算時間も早ければ数分で終了するようになってきている。また、ソフトウェアの発達により、分解能の高い質の良いデータを取得することができれば、分子の大半のモデルの構築が自動的に行われるようになってきている。しかしながら、自動構築に必要な十分な分解能が得られない場合には、現在でもアミノ酸残基を手動で一つずつ追加して分子を構築していく必要がある。

今回はX線構造解析におけるモデルの構築と精密化について解説する。

2. 電子密度修正法

以前は、ようやく位相が決定したものの初期位相の精度が悪いために、得られた電子密度図にノイズが多

く、モデルの構築を行うための解釈ができないということが多く存在した。この場合、初期位相の精度を上げるため、別の重原子同型置換結晶を作成しデータ収集を行う等、新たに実験を追加する必要があった。しかしながら現在ではソフトウェアの発達により、“溶媒領域の平滑化”や“ヒストグラムマッチング”等の電子密度修正法を行うことで、ノイズを減少させ解釈可能な電子密度図へ改良することが可能となっている(図1)。このことにより追加実験が減り、構造解析の迅速化につながっている。電子密度修正法に主に使用されている、“溶媒領域の平滑化”や“ヒストグラムマッチング”について記載する。

タンパク質結晶は、非常に多くの溶媒を含んでいる。この溶媒の多くは結晶中에서도動的な性質を保っているため、溶媒領域の電子密度は低く基本的に平坦である。一方タンパク質領域は原子が局在化されているため、相対的に高い電子密度を持つため平坦ではない。しかしデータの質が悪い場合には、電子密度図の質も低下し、本来平坦であるはずの溶媒領域に様々な大きさの密度が含まれ、結果としてピークが生じてしまう。溶媒領域の平滑化では、溶媒領域の電子密度を強制的に平坦にしピークの除去を行う。結果位相が改良されて溶媒領域とタンパク質領域の境界が明確になり、モデルの構築が可能な電子密度図へと改良することができる。

一方これまで構造解析をされたタンパク質の電子密度分布のヒストグラムを見てみると、ほぼ等しくなっている。ノイズを多く含む電子密度分布のヒストグラムはガウシアン(正規分布)となっているのに対し、良質な位相から計算された電子密度分布のヒストグラムは、シャープなピークを持ちかつ左右非対称な分布となっている。ヒストグラムマッチングは、現在得られている電子密度分布のヒストグラムを理想的なヒストグラムに強制的に修正することで、電子密度分布の改善を行う手法である。

どちらの手法も簡単な手順で実行でき、大きな効果

*株式会社リガク X線機器事業部 応用技術センター

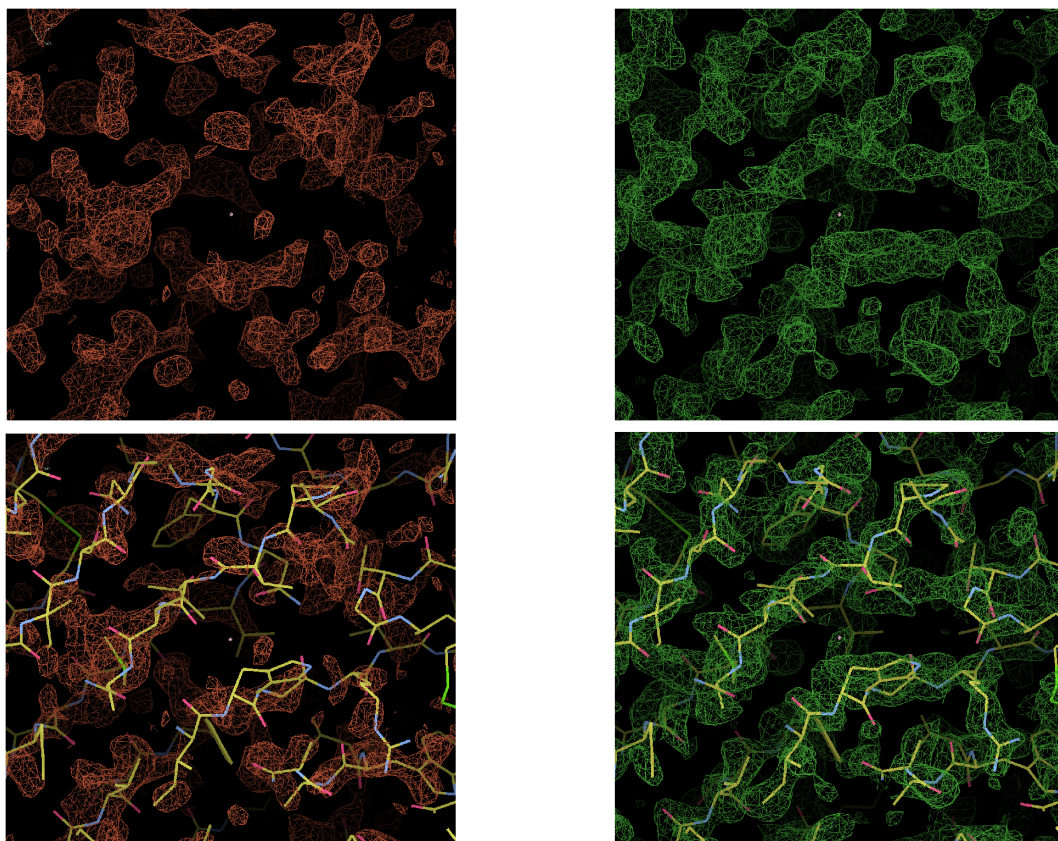


図1. 電子密度修正法の実行前（上段左）と実行後（上段右）の電子密度図の比較および各電子密度図へモデル構造の重ね合わせ（下段左：実行前，下段右：実行後）。

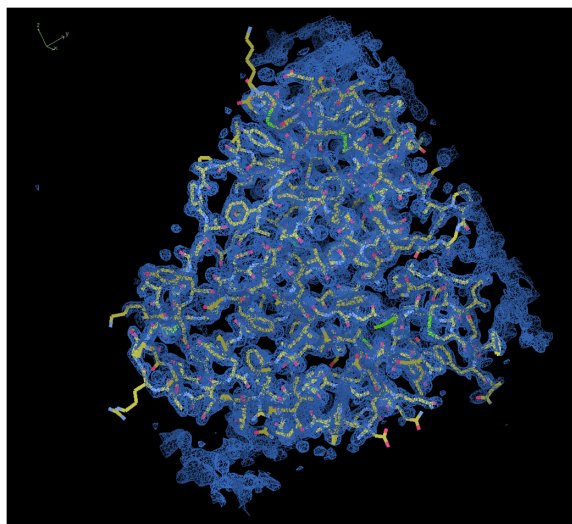


図2. Buccaneerにて自動構築後のサウマチン。80%以上が自動構築されている。

が得られることが多いため、実行することを勧める。

3. モデルの構築

位相が決定され、電子密度修正法を実行すると、改良された電子密度図を得ることができる。得られた電子密度図に対して、目的タンパク質のアミノ酸配列に従い分子を構築していく。ここで、高い分解能を持つ

電子密度（2.0 Å 程以上）が得られた場合には、ARP/wARP⁽¹⁾、RESOLVE⁽²⁾ および Buccaneer^{(3),(4)} を使用すると、ソフトウェアが与えられたアミノ酸配列にしたがって電子密度図を自動的に解釈し、分子を構築してくれる。高分解能かつ良質な電子密度の場合には、分子の約80%以上のモデルの構築が自動的に行われる（図2）。特に Buccaneer は、データの質が悪くノイズが多くやや分解能が低い電子密度図（2.5 Å～3.0 Å 程度）についても、自動構築が成功した例が散見される。Buccaneerを用いた自動構築のステップを下記に記載する。

1. C α 位置候補の探索を行う。
2. C α を繋ぎフラグメント鎖候補を構築する。
3. フラグメント鎖候補をマージする。
4. マージの結果得られたフラグメントのN末とC末を関連付けしさらに連続性の高い大きなフラグメントに繋いでいく。
5. アミノ酸鎖を割り当てる。
6. 正しいアミノ酸鎖の挿入および削除を行う。
7. 接触しているアミノ酸鎖を削除する。
8. 再構築を行う。

Buccaneerでは、C α の探索から構築までに多くのステップに分けて実行すること、および既知のタンパク質構造から計算した主鎖の電子密度の特徴を、ター

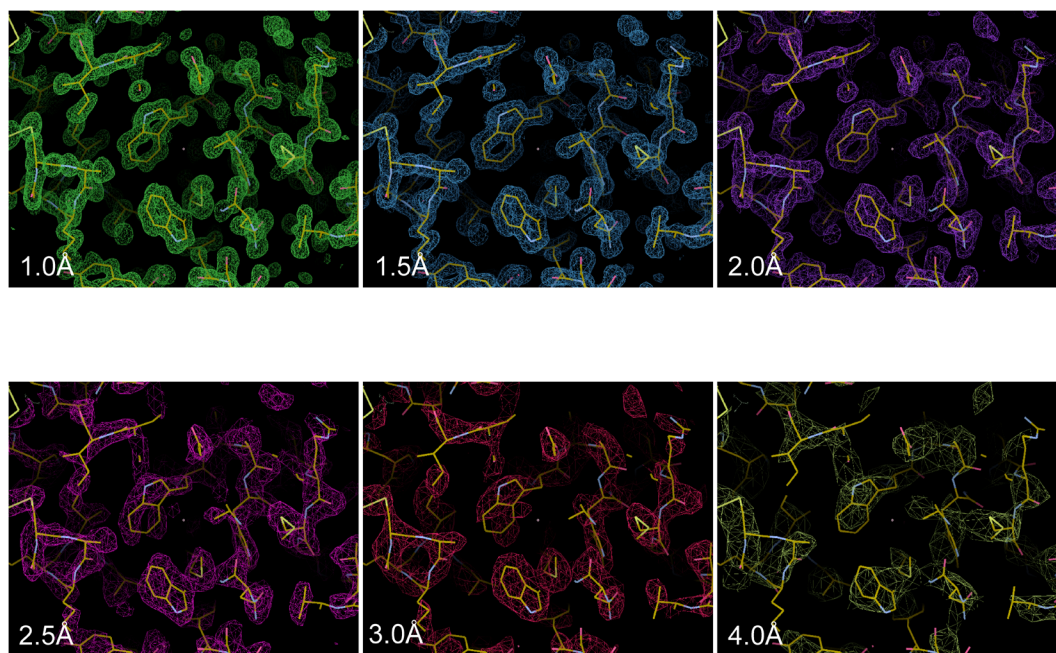


図3. 各分解能でのリゾチームの電子密度.

ゲットのタンパク質の主鎖の電子密度から探索し、取りうる可能性のある主鎖の電子密度図を得ることで、分解能が低い電子密度図においても、ターンやループの領域を含むモデルの自動構築の成功率が高いと考えられる。

図3にリゾチームの一部の電子密度を異なる分解能で示す。分解能が低くなるにつれ側鎖の電子密度が不明瞭となり、最終的には主鎖の電子密度も切れ切れとなるなど、電子密度の質が大きく下がっていくことがわかる。低分解能の電子密度の解釈は難しく、モデルの構築が困難であることが実感できるだろう。

参考のため、昔ながらの手動によるモデルの構築の手順を記載する。最初に構築の足掛かりとして、特徴のある配列部分（トリプトファン（Trp）、フェニルアラニン（Phe）、チロシン（Tyr）等の側鎖の大きいアミノ酸が密集している箇所）の電子密度の部分を探す。特徴を持つ電子密度よりどちらの方向がN末であるかを決定し、アミノ酸を配置する。側鎖が明瞭でない領域ではとりあえずアラニン（Ala）を配置してC α のトレースを行い、精密化が進み側鎖が見えてきた段階で順次Alaから本来のアミノ酸へ置換する。手動によるモデルの構築は、このように非常に根気のいるステップとなっている。このため現在ではまず自動構築を実行してフラグメント構造を構築し、その後手動で残りの部分の構築・フラグメントの結合を行っていくことが一般的となっている。

4. 精密化

モデルの構築が終了したら、次は精密化のステップとなる。位相決定に分子置換法を使用したときにはモ

デルの構築は不要であるため、位相決定後すぐに精密化を実行することになる。

分子モデルから構造因子の振幅：Fcを計算することができるため、Fcを実測の構造因子の振幅：Foに一致するように座標、温度因子を最小二乗法等を用いて修正していく。精密化の実行により改良された位相を用いて新たに計算した電子密度図を基に、モデル構造の修正を行う。構造の精密化とモデルの修正を繰り返していくことでより確からしい構造を構築していくことになる。このステップを構造の精密化と呼ぶ。

精密化のプログラムには最小二乗法の他に、分子動力学の手法を応用し、仮想的に高温（2500 K程度）にしたモデルを徐々に室温まで下げることで、局所的最小値（Local minimum）を脱出し、より安定な構造（大域的最低値：Global minimum）に収束させるSimulated annealing（SA）法がある。SA法は最小二乗法に比べて精密化の収束半径が広いいため、初期のモデル構造に多少の間違いを含んでいたとしても、精密化の過程で修正することができるという特徴を持つ。しかし、計算時間は最小二乗法に比べて圧倒的に長くなっている。SA法は、CNS⁽⁵⁾やPhenix⁽⁶⁾で使うことができる。

低分解能データを用いての精密化にもソフトウェアの革新があった。かつては、低分解能（4 Å程度）のデータを用いて精密化をする際に、精密化をするパラメータに対して反射数が少ないため、過剰なモデル構築をしてしまう（オーバーモデリング）ことでR_{work}を下げてしまい、R_{free}との差が大きくなってしまいう問題があった（R_{work}とR_{free}については「5. 結果の評価」参照）。この問題を解消するため、Refmac5⁽⁷⁾

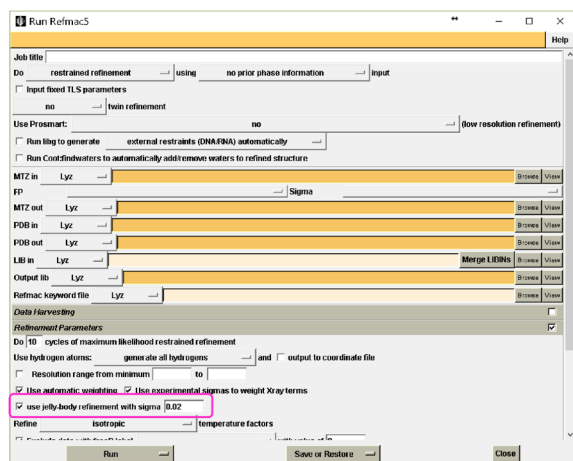


図4. Refmac5の低分解能用精密化用モード“jelly body refinement”。



図5. 半導体検出器 HyPix-6000HE (左) と PILATUS 200K (右)。

では精密化用プログラムに低分解能用のモード，“jelly body refinement”が搭載されている（図4）。Jelly body refinementでは、通常の精密化よりも多くの束縛を使用することで R_{work} と R_{free} の値がかけ離れないように精密化をすることができるようになっている。

5. 結果の評価

精密化の結果の評価には、一般的に R_{work} と R_{free} が使用される。 R_{work} は、モデル構造の妥当性を評価する指標である。 R_{free} は、精密化には使用しないテスト反射（独立な反射のうち5%を使用することが多い。）のみを使用して計算したR値である。 R_{free} は精密化の過程でオーバーフィッティングなどにより分子モデルが偏りを受けていないかを評価する指標となっている。精密化が適切に行われているときには、 R_{work} と R_{free} の両方が減少してくるが、 R_{free} が少し高い値となる。（5%程度 R_{work} より高い値が多い）。ノイズ由来の電子密度に間違えてモデルをアサインしてしまったような場合や、低分解能データでオーバーフィットしてしまった場合には、 R_{free} が下がりずに R_{work} のみが下がり R_{work} と R_{free} の値に大きな剥離（10%以上）が生じる。このような状況が生じたときには、構造全体を精査し、モデルの違いを修正していく必要がある。

6. おわりに

モデルの構築と精密化では、PCの高速化により、普段持ち歩くノートパソコンでも構造解析ができるようになってきた。さらにソフトウェアの発達により、これまではあきらめていたような電子密度図でも電子密度修正法を使用することで、十分に解釈可能な電子密度図を得ることができるようになった。また、モデルの自動構築を有効に利用することで、精密化に費やす時間も大幅に削減することができるようになった。これは、同様なタンパク質を研究対象とする競争相手も同じように構造解析が迅速化していることや、競争相手が増えていくことを同時に意味している。このため、以前と比較してより競争が激化していると感じる。この競争を勝ち抜くには、いかに分解能の高い質の良いデータを収集できるかにかかってくる。そのためには、できるだけ身近なシステムで手軽に測定することで、移動時間や待ち時間、旅費を大幅に節約することができる。高速高感度なHyPix-6000HE（リガク）やPILATUS（DECTRIS）に代表される半導体検出器（図5）を搭載したシステムでは、測定時間や対応できるタンパク質の大きさや結晶の質も大幅に改善されている。

また、質の良いデータの取得には、質の良い結晶を得ることが必須であることから、結晶化に適した純度の高いタンパク質を、結晶化に適した溶液状態で結晶化することが重要となる。結晶化に適したタンパク質の溶液状態の評価には、単結晶X線構造解析 基礎講座の第11回で紹介したBioSAXSが便利である。

参考文献

- (1) G. Langer, S. X. Cohen, V. S. Lamzin and A. Perrakis: *Nature Protocols.*, **3** (2008), 1171–1179.
- (2) T. C. Terwilliger: *Acta Cryst.*, **D59** (2003), 38–44.
- (3) K. Cowtan: *Acta Cryst.*, **D62** (2006), 1002–1011.
- (4) K. Cowtan: *Acta Cryst.*, **D64** (2008), 83–89.
- (5) A. T. Brunger, P. D. Adams, G. M. Clore, P. Gros, R. W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, N. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson and G. L. Warren: *Acta Cryst.*, **D54** (1998), 905–921.
- (6) P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart: *Acta Cryst.*, **D66** (2010), 213–221.
- (7) G. N. Murshudov, A. A. Vagin and E. J. Dodson: *Acta Cryst.*, **D53** (1997), 240–255.