タンパク質結晶構造解析入門—ブラックボックスの中身— (3) モデリングの裏側をのぞいてみよう

分子モデルの構築

北海道大学大学院先端生命科学研究院 姚 閔

Min YAO: Model Building

Protein model building is a process in which the electron density map is interpreted and protein model is constructed. Resolution limit of protein crystal, inherent thermal vibration of atoms, and experimental errors in data collection and in phasing, make interpreting electron density more difficult. Nevertheless, it is possible to build protein model with the help of sequence information and the stereochemical knowledge of amino acids and peptide bond. Recently, automatic methods for model building have been developed based on the pattern recognition techniques. In this manuscript, we learn how the electron density map is interpreted and how the building process is automated.

1. はじめに

タンパク質の構造解析にはいくつかの記念日があります。1つ目は、「初めて結晶ができた日」、2つ目は、「初めて反射が撮影できた日」、そして3つ目は、「初めて電子密度図が計算できた日」です。3つめの記念日は、「初めて構造が見えた日」ということもできるでしょう。ですから3つの記念日の中でも最も重要な記念日です。近い将来自動プログラムが発展して結晶にX線をあてたら直接分子モデルが表示されるような日が来るでしょう。そうなったら3つめの記念日は、「分子モデルができ上がった日」として代わられるかもしれません。しかしそうなったとしても、X線結晶構造解析から直接得られるものは、分子モデルではなく電子密度図であることを忘れるべきではありません。分子構造を決めるうえで、X線回折法が分光学的な方法と決定的に異なるのは、電子密度(つまり分子構造)が直接見えるところなのですから。

電子密度図は構造を最も忠実に表しています。金属などの重原子の電子密度は高く、水素原子の電子密度は低く表現されます。そればかりでなく、タンパク質鎖の末端や分子表面に突き出ているところのように、ゆらぎの大きな場所では、それを反映して電子密度は低くなります。しかし電子密度図がいったん分子モデル変換にされてしまったら、そのような情報は失われてしまわれがちです。それにもかかわらず電子密度図を基に分子モデルを組み立てるのは、もちろん、分子モデルを組み立てて分子の働きを議論するためです。例えば酵素反応を考える場合には、

一般には、分子モデルのほうが電子密度図よりもずっとわかりやすいでしょう。モデルを組み立てることのもう1つの意義は、構造解析を進展させることにあります。X線結晶構造解析では位相が測定できないために、実験値である回折強度に位相をつけてフーリエ変換をして構造を導き出します。したがって構造は位相の質によって大きく左右されてしまいます。電子密度図を解釈して分子モデルを組み立てることができれば、それを基にさらに位相を改良してよりよい電子密度図を得ることができます。この稿ではタンパク質結晶の電子密度図から分子モデルを組み立てるための方法に焦点をあてて説明します。

2. 電子密度図 ($\rho(r)$) の計算

電子密度図は構造因子の逆フーリエ変換として次の式で計算されます。

$$\rho(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{k}} m(\mathbf{k}) |F_{\text{obs}}(\mathbf{k})| \exp\{i\alpha_{\text{best}}(\mathbf{k})\} \exp(-2\pi i \mathbf{k} \cdot \mathbf{r})$$

ここで、 $\mathbf{k} = (h, k, l)$ は反射の指数であり、 $\mathbf{r} = (x, y, z)$ は結晶周期をユニットとする実空間の座標です。 $|F_{\text{obs}}|$ は構造因子の振幅で測定値です。 α_{best} はこの講座の前の回で紹介されているように、MIR/MAD/SADなどの方法で求められた最良位相角です。 m は Figure of Merit と呼ばれる指標で、計算された位相の確からしさを表します。MIR/MAD/SAD法で得られた位相を用いて計算した電子密度図は実験マップと呼ばれます。一方、MR法の場合では、位相はモデルから計算されたもの(α_{cal})を使い、電子密度図は次の式で求められます。(実際の計算では、使った

図1 単位格子内のグリッド点。(The grid points in a unit cell.)

図2 直線で描いた近似的な円。(Approximate circle drawn by straight lines.) 左：8グリッド；右：16グリッド。

図3 異なるグリッド間隔で計算した2.0 Å分解能での電子密度図。(Electron density maps calculated by different grid intervals in 2.0 Å resolution.) 上：グリッド点の間隔は1.0 Å, 下：グリッド点の間隔は0.6 Å。

モデルの寄与を考慮するための重みが $|F_{\text{cal}}|$ に掛かります.)

$$\rho(r) = \frac{1}{V} \sum_k \{2|F_{\text{obs}}(k)| - |F_{\text{cal}}(k)|\} \exp\{i\alpha_{\text{cal}}(k)\} \exp(-2\pi i k r)$$

電子密度は、通常、単位格子を整数等分したグリッド点 (x, y, z) で計算されます(図1)。「点を直線でつないで円を描くときに、点の数が多いほど円の形により近い」ように(図2)、電子密度図を滑らかに表現するためには計算用のグリッド点の間隔は十分小さくする必要があります。しかしグリッド点の間隔を必要以上に小さくすると、計算時間が長くなり電子密度図の容量も大きくなってしまいます。一般に、グリッド点の間隔は、実際の回折データの分解能の3分の1程度が適当とされています。図3に示すのは、分解能2.0 Åの電子密度図をグリッド点の間隔を1.0 Åと0.6 Åにして描いた電子密度図です。

3. 電子密度図の等高線表示

電子密度の高さは、一般には、等高線で表されます。等高線と言えば地図上の等高線を思い浮かべることと思います。地形の高低を表すための等高線は地表面、すなわち二次元の座標 (x, y) に対して地形の高さを表すものです

図4 二次元の断面で表示した電子密度図(ミニマップ)。(Electron density map displayed as two-dimensional sections (mini map).) 左：一断面の等高線表示。右：断面を重ね合わせて下からライトを照らすと分子構造が浮かび上がる。

が、電子密度は三次元の座標 (x, y, z) に対して計算されるものです。ですから地図の場合と比べると1つだけ次元が高くなります。この場合、同じ高さの点を結んでできるのは等高線ではなく等高面です。

三次元の座標に計算された電子密度を等高線として表

すための1つの方法は、 (x, y) (y, z) (z, x) の3つの断面のそれぞれについて通常の二次元の等高線を引く方法です。図4には、そのように表示した断面の1つと、それを積み重ねて表示したマップを示しました。一般にはミニマップと呼ばれています。最近ではミニマップがつくられることはほとんどなくなりましたが、三次元コンピュータグラフィックスが登場する以前には、電子密度図が計算されると必ずミニマップがつくられました。現在でも、電子密度の広い範囲を一度に表示するには最適なマップです。しかし、この表示法は、見る方向が限られてしまうこと、またどうしても高さ方向の精度が曖昧になるという欠点があります。

三次元コンピュータグラフィックスを使えば、3つの断面の等高線を同時に表示できます(図5)。この場合、等高線は電子密度を計算したグリッド点を通るように(すなわち内そうをしないように)して引かれます。グリッド間隔を十分に小さくすることで内そうをしない等高線でも不自然を感じることなく見ることができます(図3)。このように表示した等高線をコンピュータグラフィックスに表示して回転させながら見れば、等高線は空間で閉じた

面のように見えます(すなわち等高面です)。もちろん等高線といっても三次元表示する場合には複数のレベルを同時に表示したら複雑過ぎるので、通常は1レベルだけ表示し、複数のレベルを同時に表示する必要がある場合には異なる色を使います。

雑誌の紙面に電子密度図を描く場合には、3つの断面で計算した等高線を二次元に投影して表示することが多いようです。この場合にも、一般には1レベルだけを表示します。また、上下を適当な厚さで切り取ることによって電子密度の重なりを避けてより見やすくする工夫をします。このような電子密度図は、約6～10度回転させた2つの図を横に並べることでステレオ図となります(図6)。

4. 電子密度図と分解能

もしも、まったくエラーのない十分に分解能の高い電子密度図が得られたならば、その図を基に分子モデルを組み立てるのは簡単です。図7に示すように等高線の中心に原子をおくことで分子モデルを作製できるでしょう。しかし、原子の熱振動などによってタンパク質結晶の回折データの分解能は制限されています。さらに、構造因子の振幅

図5 コンピュータの利用。(Using computer to display electron density map.)

図7 分離原子の電子密度図。(Electron density contour for resolved atoms.)

図6 電子密度とモデルのステレオ図。(Stereo diagram showing electron density map at 2.03 Å resolution superposed with molecular structure.)

図8 電子密度と分解能. (Electron density maps drawn in different resolutions with atomic model.)

$|F_{\text{obs}}|$ を測定するときの実験誤差や位相の計算誤差などがあるため、現実には、図7のような理想的な場合はほとんどありません。

図8にモデルから計算した6.0, 4.0, 3.0, 2.5, 2.0, 1.5 Åの実験誤差のない理想的な電子密度図をモデルに重ね合わせて示しました。6.0 Å分解能の電子密度図では α ヘリックスはただのチューブ状に見えます。4.0 Åの電子密度図では、二次構造の区別ができるでしょう。しかし、側鎖の形はよくわかりません。3.0 Åでは、大きな側鎖(Trp, Tyr, Phe)の確認はできますが、小さな側鎖の確認は困難です。2.5 Å以上の高分解能の電子密度が得られれば、側鎖を含めてモデルの組み立てが可能になります。1.5 Å分解能になると、個々の原子が分離する様子が見えます。もちろん、ここに示した例は、構造因子の振幅と位相に誤差がない理想的な状態であり、それらの誤差を考慮すると、2.5 Å分解能の場合でも、側鎖の区別ができないことがあるし、1.5 Å分解能でも、部分的に見えないこともあります。

5. 電子密度図から分子モデルを組み立てる

個々の原子が分離しているような高分解能の電子密度図から分子モデルを組み立てるのは簡単です。原子が分離していなくても、ある程度高分解能のきれいな電子密度が得られたら、タンパク質分子の鎖を追うのは比較的簡単になるでしょう。しかし、分解能が低くなってくると原子の位置がわからなくなってきます。また、同じ分解能でも位相が不正確であると電子密度図は不鮮明になってきます。そうなってくると分子モデルは簡単に組めなくなってきます。タンパク質の構造解析では一般に最初の分子モデルを組み立てる際にはそのような場合が多いのです。そのような電子密度を使って部分的にでも分子モデルを組むことができるならば、その座標からより正確な位相を求めることができ、より誤差の少ない電子密度を描くことができます。したがって不鮮明な電子密度図から分子モデルを組むのは、解析の成否をも左右する重要な技術です。

図9 2.3 Å 分解能の二次構造と電子密度. (The 2Fo-Fc electron density maps of the secondary structure regions drawn in 2.3 Å resolution.) 左: α ヘリックスの電子密度図, 中: α ヘリックスの電子密度図と構造, 右: β シートの電子密度図と構造.

位相や強度測定の見誤差, さらに分解能の限界によって不鮮明になった電子密度図からでも分子モデルを組み立てることができるのは, 分子構造についてあらかじめ多くの情報が蓄積されているからです. 電子密度図からタンパク質分子構造を構築する際には, タンパク質分子について私たちがもっているすべての情報を使います. 一番重要な情報はもちろん分子構造つまりアミノ酸配列情報です. また, ペプチド結合や特徴的な側鎖の構造についても私たちは十分な情報をもっています. 例えば, ペプチド結合は一般にはリジッドな平面構造と考えて差支えありませんし, Phe, Tyr, Trp は特徴的な芳香環をもっています. さらにペプチド鎖は, α ヘリックスや β シートなど二次構造と呼ばれる規則的な構造をとることが多いことも知られています.

分子モデルを組み立てるには, まず, 電子密度図からタンパク質主鎖の流れを求めます. このためには, 比較的高い等高線レベルで表示した電子密度図を用意し, そこから長く続く領域を探します. 主鎖の中でも α ヘリックス, β シートの領域は主鎖水素結合によって固定されているためにランダムな領域よりも電子密度が強く, また規則的な繰り返し構造であるために比較的容易に構造を決めることができます (図9). 3.0 Å 以下の分解能の電子密度図で, 側鎖の認識ができなくても, α ヘリックスの領域では C_β がN末方向に突き出るので, N末端を下, C末端を上にした図を描くと, ちょうどクリスマスツリーのように見えます (図9). これを利用すれば主鎖の極性 (N末端とC末端の方向) を決めることができます.

主鎖のトレースを行いながら徐々に側鎖を加えていきます. 電子密度の特徴からアミノ酸を同定するのは, 分解能が非常に高い例外的な場合を除いて, 一般には困難です. そこで, 側鎖の同定は, まずマーカーとなる残基から行います. 例えばSe-Metを位相決定に利用した場合には, Seの座標を利用してMet残基を特定することができます. 特徴的な芳香環をもつTrp, Tyr, Pheなどの残基や, Leu,

Ileなどの脂肪族側鎖は, 分子内部にあるために電子密度図に鮮明に現れていることが多く, 比較的容易に同定することができます. 親水性の長い側鎖をもつArg, Lys, Glu, Glnは, 分子表面で熱振動していることが多く, 一般には不鮮明です. Gly, Alaなどの小さな残基が多い部分やループ部分の構築では, 残基の挿入あるいは欠失が起こりやすいので注意しましょう. アミノ酸側鎖を同定する際には, 1残基を見て決めるのではなく, 確実に同定できる複数の側鎖を含むフラグメントを, 配列情報と比較しながら, まるごと同定する方法がお勧めです.

6. 自動モデル構築

前項で述べたような方法をプログラミングすることによって自動でモデル構築を行うことができるようになりますが, 一般にパターン認識は広い範囲を一度に見ることのできないデジタルコンピュータの苦手とするところで, す. しかしゲノム解析に続く時代の要請として, 自動モデル構築のためのアルゴリズムが考案され, プログラムが開発されてきました.

6.1 分子の骨格を自動的に決めるための方法¹⁾

自動モデル構築で最初にすべきことは, タンパク質鎖のつながりを見つけることです. そのための方法として, 電子密度図からグリッド点を段階的に減じて, 最終的にタンパク質の鎖 (主鎖および側鎖) をグリッド点のつながり (bones) として表すためのアルゴリズムが考案されています. この方法では, 最初に電子密度図からあるしきい値以下の領域 (溶媒領域と考えられるところ) を除きます. 残されたグリッド点はタンパク質領域であるとして, そこから, 鎖を切断することなく極限まで細くしていきます (図10). そのためには, 残された各グリッド点に隣接する26個のグリッド点を調べ, つながりを保つのに不要と判定されるグリッド点を, 電子密度値の低い方から1つつ除いていきます. 最終的に残されたグリッド点のうち, つながりの長いものは主鎖, そこから横に伸びるものは側

図10 分子の骨格の作製の二次元表示。(Making bones from electron density in a two-dimensional section.) 左から右まで、しきい値を徐々に上げながら、鎖を切断しないように、しきい値以下の領域を除いていきます。

図11 フラグメント構造の拡張⁴⁾ (The structural extension from fragment.) (1) 構築途中のモデル (実線) と電子密度図 (輪郭). (2) 未構築部分の電子密度図の切り出し. (3) 未構築部分 (点線) の構築. a. 両端が既知の場合, それをつなぐようにモデルを構築していく. b. 始点のみが既知の場合, 最初に終点を求める。

鎖とみなすことができます。

6.2 自由原子を設定して分子を構築する方法²⁾

ある程度高い分解能の電子密度図が得られたときに使える方法として、電子密度の高い位置に原子を置くことで分子を構築していく方法が考案されています。この方法では、まず電子密度図をサーチして、そのピーク位置に原子を置きます (結合とは無関係に置かれることから自由原子と呼ばれます)。その自由原子の位置を精密化した後に、自由原子の中から妥当な原子間距離でつなぐことのできる対を選び、さらにはそのような原子対について、ペプチド結合 ($\text{C(=O)-N-C}\alpha$ や $\text{N-C}\alpha\text{-C(=O)}$) を形成し得るかどうかを調べていきます。このときに、リジッドなペプチド面では $\text{C}\alpha$ 間の距離が約 3.8 Å であるというルールを使うことができます。 α ヘリックスと β シートのような二次構造を形成しているものから初めて主鎖を確定し、その後側鎖を付け加えます。

6.3 フラグメントライブラリーを利用する方法³⁾

既知タンパク質構造をライブラリー化し、それを利用して分子モデルを組む方法もあります。この方法では、主鎖を構築するためには、二次構造ライブラリー (α ヘリックスと β シート構造) とショートライブラリー (数残基分の長さのランダム構造) という2つのフラグメントライブラリーを準備します。モデル構築時には、最初に二次構造

ライブラリーから電子密度図に合う二次構造フラグメントを選択し、主鎖の α ヘリックスと β シート部分を構築します。次に、ショートライブラリーから残りの電子密度図に合うフラグメントを選択し、先に構築した二次構造部位の両端に延長します。ライブラリーを利用するので分解能が多少低くても構築が可能です。

6.4 フラグメント構造の拡張^{4),5)}

自動モデル構築を行う際に、 α ヘリックス、 β シートのような規則的な構造をもつ部位の構築は比較的容易で、電子密度の質にもよりますが 80 ~ 90 % まで構築できることもあります。一方、ターンやループ、末端部分は多様な構造をとるために、また一般には電子密度も低いために、モデルの自動構築は困難で、30 ~ 50 % しか構築できないことも多いのです。このため、最初に構築された構造フラグメントから、ターンやループ、末端部分の構造を拡張するための方法が別途考案されています。フラグメントを延長するためには、まず電子密度図から既知の部分を取り除き、未構築部分の電子密度図を切り出します。切り出した電子密度の島の中でタンパク質鎖の始点と終点がわかっている場合、その始点から終点に向かって1残基ずつ伸ばしていく方法や (図11)、データベースを利用して最大12 ~ 15残基ぐらいのフラグメントを始点と終点を固定して精密化することで一気に構築する方法があ

ります。分子の末端のようにフラグメントの終点がない場合には、電子密度の島で始点から最も遠い位置を終点とすることができます。未構築で残された部分はどうしても電子密度が低く、主鎖と側鎖の区別が難しい場合も多いので両者を混同しないように注意を払う必要があります。

7. モデル構築と精密化⁶⁾

分子モデルは電子密度図に基づいて構築されるので、構築した分子モデルの完成度は電子密度図の質に依存します。電子密度図は、構造因子の振幅 $|F_{\text{obs}}|$ の測定誤差に加えて、位相の計算誤差を含んでいます。したがって、それ

から構築した構造は一時的なものとみなすべきです。より正しい原子座標を求めることができれば、その座標を使ってより正しい位相を計算することができ、より誤差の少ない電子密度を得ることができます。一般に、最終的なモデルを得るためには、このようにして求めたモデルの座標を精密化プログラムで精密化します。精密化された原子座標を使えば、さらに上質の電子密度図を計算することができます。このように原子座標の精密化とモデルの構築とは互いに深く関係しています。モデル構築を自動化することによって精密化とモデル構築とを一体化したプログラムを開発することができます(図12)。

文 献

- 1) J. Greer: *J. Mol. Biol.* **82**, 279 (1974).
- 2) R. J. Morris, A. Perrakis and V. S. Lamzin: *Acta Cryst.* **D58**, 968 (2002).
- 3) T. C. Terwilliger: *Acta Cryst.* **D58**, 1937 (2002).
- 4) Y. Zhou, M. Yao and I. Tanaka: *J. Appl. Cryst.* **39**, 57 (2006).
- 5) van den H. Bedem, I. Lotan, J.-C. Latombe and A. M. Deacon: *Acta Cryst.* **D61**, 2 (2005).
- 6) M. Yao, Y. Zhou and I. Tanaka: *Acta Cryst.* **D62**, 189 (2006).

プロフィール

姚 閔 Min YAO

北海道大学大学院先端生命科学研究院

Faculty of Advanced Life Science, Hokkaido University

〒060-0810 札幌市北区北10条西8丁目北海道大学理学部5号館313号室

Kita10, Nishi8, Kita-ku, Sapporo, Hokkaido 060-0810, Japan

e-mail: yao@castor.sci.hokudai.ac.jp

最終学歴：北海道大学大学院理学研究科博士課程後期修了

研究分野：構造生物学、タンパク質X線結晶学

現在の研究テーマ：翻訳の制御と分子機構の研究

図12 モデル構築と精密化の一体化。(Model building with refinement.)