

大規模計算を用いた X 線構造解析

—マキシマムエントロピー法によるタンパク質の精密 X 線構造解析—

中 村 高 洋 青 柳 忍
西 堀 英 治 坂 田 誠

I. はじめに

X 線構造解析は、実験・解析のプロセスのいたるところで計算機の利用が欠かすことのできない研究分野である。例えば、一般的な 2 次元検出器を用いた単結晶構造解析では、結晶の向き、検出器の方向・位置が計算機で制御され、1 枚数十 MB のデータが何十枚も計算機による自動測定で収集される。このデータ量だけでも多い場合には数 GB に達する。この数 GB のデータから単位格子の決定、Bragg 反射強度の抽出、構造の決定、構造の精密化で構造解析は完了するが、そのあらゆるプロセスで計算機が利用される。

計算機性能の向上は、X 線構造解析に多大な進歩をもたらし続けている。その顕著な例として、タンパク質の立体構造決定が挙げられる。国際的なタンパク質の立体構造のデータベース (PDB) に登録された立体構造の数は、2007 年 10 月の時点で 46000 に達しており、1996 年の 5000 から約 10 年で 10 倍に増加している。この増加には、SPring-8 等の第三世代放射光 X 線の登場などの実験技術の進歩も寄与しているが、計算機性能の向上が大きな役割を果たしていることはいうまでもない。10 年前には専用のワークステーションを用いなければ行えなかったタンパク質 X 線構造解析が、現在では個人の所有する PC で十分に行えるようになってきた。このような状況の基で、大規模計算機を利用して、これまでは行うことができなかった構造研究も行われるようになってきた。このような研究の例としては、複数のタンパク質が結合した超分子複合体の構造解析⁽¹⁾ やタンパク質の原子間の結合形態を X 線回折から明らかにする研究⁽²⁾ などがある。

我々のグループでは、X 線構造解析にマキシマムエントロピー法 (MEM)⁽³⁾ や遺伝的アルゴリズム (GA)⁽⁴⁾ を適用し、従来の構造解析法では解明できない構造を明らかにする研究を進めてきた。MEM による研究では、回折データから原子配列だけでなく原子間の結合電子や電荷移動などを含めた詳細な電子密度を求めてきた。また、GA では既存の直接法などの構造決定法では構造決定が困難な物質の原子配列を求めてきた。それぞれの研究で対象としたのは、単位格子中に数～50 原子、反射本数が～1000 本程度の無機物や低分子化合物であった。

最近になって、巨大分子であるタンパク質の MEM、GA による X 線構造解析を開始した。タンパク質の X 線回折データを MEM で解析することにより機能と密接に関連した水素結合の様子や機能部の Disorder 構造の解明が期待できるためである。また、GA は、これまで明らかにできなかったタンパク質の構造を解明できる可能性がある。この研究を進める上でボトルネック

となったのは、解析における計算量であった。タンパク質は、これまでの MEM や GA の解析対象と比較して原子数が～100 倍以上、反射本数が～100 倍以上となるためである。幸いなことに、名古屋大学の超強力 X 線実験室に設置された SX-6i や理化学研究所 SX-7 などのベクトル計算機を利用することによりタンパク質の MEM 解析の研究が行えるようになった。また、GA についても理化学研究所の専用計算機 MDM や MD-GRAPE3 を利用することで実用的な時間で計算ができるようになってきている。本稿ではこうした我々の研究の中からタンパク質の MEM 解析について述べる。

II. MEM 電子密度分布解析

X 線回折実験で測定される回折強度 I は構造因子 F の二乗に比例する。

$$I \propto |F|^2 \quad (1)$$

構造因子 F は物質中の電子密度のフーリエ変換として次式で表される。

$$F(\mathbf{k}) = \int \rho(\mathbf{r}) \exp(-2\pi i \mathbf{k} \cdot \mathbf{r}) d\mathbf{r} \quad (2)$$

ここで \mathbf{k} は散乱ベクトル、 \mathbf{r} は位置ベクトル、 $\rho(\mathbf{r})$ は \mathbf{r} における電子密度である。構造因子 F を逆フーリエ変換すれば、電子密度 $\rho(\mathbf{r})$ が求められる。

$$\rho(\mathbf{r}) = \int F(\mathbf{k}) \exp(2\pi i \mathbf{k} \cdot \mathbf{r}) d\mathbf{k} \quad (3)$$

(3) 式の積分範囲は実質的に $-\infty$ から $+\infty$ となるため、フーリエ変換により物質全体の電子密度を連続量として求めるには無限個の F が必要となる。しかし実験で測定されるデータの数には有限である。また、実験的に測定された F は誤差を含む。これらの影響によりフーリエ変換で得た電子密度には実際には存在しないゴーストピークやマイナスの電子密度が現れるため、水素結合を検出するような精密な構造解析に利用するのは困難である。

実験データから精度の高い電子密度を計算する方法として、我々はマキシマムエントロピー法 (MEM) を用いた方法を開発した⁽³⁾。MEM は情報理論から発展した推定法であり、限られた実験データから、データの誤差も考慮して精密な電子密度を求めることができる。以下に MEM による電子密度計算について説明する。

単位格子を 3 次元にピクセル分割する。 h 番目のピクセルの電子密度を ρ_h 、及び事前電子密度を τ_h とする。 ρ_h, τ_h を確率として扱うために、単位格子内の総電子数 Q で規格化する。

$$\rho'_h = \rho_h / Q, \quad \tau'_h = \tau_h / Q \quad (4)$$

この ρ'_h, τ'_h を用いて、事前事象がある場合の情報エントロピー S を次式で定義する。

$$S = - \sum_h \rho'_h \log \frac{\rho'_h}{\tau'_h} \quad (5)$$

実験で得た j 番目の反射の構造因子 $F_{\text{obs}}(\mathbf{k}_j)$ と ρ_h から計算された構造因子 $F_{\text{MEM}}(\mathbf{k}_j)$ が誤差 σ_j の範囲内で一致するという束縛条件を付加し、 S が最大となる ρ'_h を計算する。ここで $F_{\text{MEM}}(\mathbf{k}_j)$ は (2) 式に基づき次式で表される。

$$F_{\text{MEM}}(\mathbf{k}_j) = V \sum_h \rho_h \exp(-2\pi i \mathbf{k}_j \cdot \mathbf{r}_h) \quad (6)$$

V は単位格子の体積である。この $F_{\text{MEM}}(\mathbf{k}_j)$ を用いると、束縛条件は次式で表される。

$$C = \frac{1}{N} \sum_j \frac{1}{\sigma_j^2} |F_{\text{obs}}(\mathbf{k}_j) - F_{\text{MEM}}(\mathbf{k}_j)|^2 < 1 \quad (7)$$

N は構造因子の数である。ラグランジュの未定定数法により S が最大となる ρ'_h を計算する。 $C = 1$ の条件の下で S が最大となるのは、 ρ'_h が次の式を満たすときである。

$$\frac{\partial}{\partial \rho'_h} \left[S - \frac{\lambda}{2} (C - 1) \right] = 0 \quad (8)$$

λ はラグランジュの未定定数である。これを ρ_h について解くことにより次式が得られる。

$$\rho_h = \tau_h \exp \left[\frac{\lambda Q}{N} \sum_j \frac{\{F_{\text{obs}}(\mathbf{k}_j) - F_{\text{MEM}}(\mathbf{k}_j)\}}{\sigma_j^2} \exp(-2\pi i \mathbf{k}_j \cdot \mathbf{r}_h) \right] \quad (9)$$

この式は右辺にも $F_{\text{MEM}}(\mathbf{k}_j)$ の中に ρ_h を含む超越方程式である。そこで以下の近似を用いて、逐次計算により ρ_h を求める。

$$F_{\text{MEM}}(\mathbf{k}_j) = V \sum_h \tau_h \exp(-2\pi i \mathbf{k}_j \cdot \mathbf{r}_h) \quad (10)$$

MEM では、事前電子密度 τ_h を与え、(10) 式により求めた $F_{\text{MEM}}(\mathbf{k}_j)$ を用いて (9) 式で電子密度 ρ_h を計算する。ここで束縛条件を満たさなければ、 ρ_h を τ_h に置き換えて再び (9) (10) 式を計算する。この逐次計算を $C < 1$ となるまで繰り返す、電子密度を得る。

(2) 式より構造因子 F は振幅と位相からなる複素数である。しかし X 線回折実験では振幅 $|F|$ しか得られない。MEM では構造モデルから計算された位相を振幅に与えて電子密度を計算する。MEM の計算の中で、位相は改良される。改良された位相に基づき、構造モデルを改良し、MEM と位相改良のサイクルを繰り返す。本研究では、以上のアルゴリズムに基づきタンパク質構造解析用に開発された MEM 解析プログラム ENIGMA⁽⁵⁾ を用いた。

III. MEM による putative acylphosphatase の構造解析

タンパク質は生体を構成する主要な成分の一つであり、生体内で様々な機能を担う。例えば赤血球中に存在するヘモグロビンは酸素を全身に運搬し、唾液に含まれるアミラーゼはデンプンを分解する機能を持つ。タンパク質は 20 種類のアミノ酸がペプチド結合で数十～数千個繋がった物質であり、それが水素結合などの相互作用により折り畳まれて立体構造をとり機能を発現する。

立体構造と機能の関係を明らかにするために、タンパク質の構造解析が広く行われている。

本研究では MEM 電子密度を用いて立体構造の詳細な決定を行った。タンパク質の詳細な電子密度を求めるには単位格子を細かく分割しなければならない。また構造モデルの改良をするたびに電子密度計算が必要となるため、一つの物質の解析で MEM 計算を何度も繰り返して行う必要がある。ここでは MEM を用いたタンパク質構造解析の例として putative acylphosphatase の構造解析を報告する。

putative acylphosphatase は高度好熱菌 HB8 由来のタンパク質であり、単位格子体積は約 68000 \AA^3 、非水素原子の数は 677 個である。putative acylphosphatase の構造はデータとモデルのずれを示す指標 $R=17.85\%$ のモデルが PDB に登録されている (PDB ID: 1ulr, Fig.1)。この 1ulr の回折データを、MEM により解析した。MEM 計算に用いた構造因子 $F_{\text{obs}}(\mathbf{k}_j)$ の数は 15801、ラグランジュの未定定数 $\lambda = 0.50$ 、ピクセル分割数は $128 \times 256 \times 256$ である。II の (10) 式で示したように、MEM 計算の信頼度因子 R は、束縛条件 C からデータの誤差 σ_i によって決定される。MEM 計算の信頼度因子は、すべての解析の場合で 7% 程度であった。

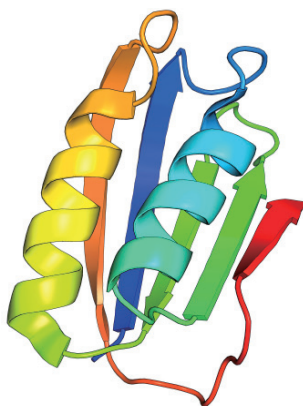


Fig.1. 1ulr の全体構造

1. 水の構造の精密化

結晶中のタンパク質分子間には隙間が存在する。この領域には水分子などの溶媒分子が存在しており、溶媒領域と呼ばれる。溶媒領域では溶媒分子が動き回るため、電子密度は弱く広がったものとなる。従来解析ではこのような電子密度の確認は困難であるため、溶媒領域に一定の電子密度を仮定し補正する。しかし実際には溶媒領域の電子密度は一定ではなく、溶媒分子の拡散経路などを示す濃淡がある。

MEM 電子密度の溶媒領域に弱い電子密度の濃淡が確認された。この様子を Fig.2 に示す。濃淡に合わせて水分子を配置し、溶媒分子の構造モデルを構築した。MEM 電子密度を用いることにより、溶媒領域の構造を精密化できることがわかった。

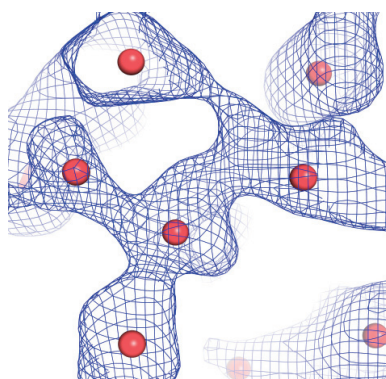


Fig.2. 溶媒領域の電子密度

2. タンパク質の構造の精密化

タンパク質の構造も MEM を用いて修正した。タンパク質分子の一部が複数の異なった立体構造をとる場合がある。このような構造は Disorder もしくは multi-conformer と呼ばれ、タンパク質の機能と密接に関連している。Disorder 構造を発見するために、MEM を用いた omit map による解析を行った。omit map とは構造の一部を削除した位相を与えて求めた電子密度である。MEM 計算は構造モデルから計算された位相を用いるため、間違った構造を仮定してもそれを支持する電子密度が得られる場合がある。omit map はこの影響を抑えることができる。

一度に大量の構造を削除すると詳細な omit map が得られないため、1ulr の原子を約 3.4% ずつ削除しながら、タンパク質全体で 29 個の omit map を作成した。この omit map から 9 箇所 で Disorder 構造を発見した。Fig.3 にその一例を示す。Fig.3 (a) の omit map では、Fig.3 (b) のように先端が 2 つに分かれた電子密度が確認された。これに合わせて Disorder の構造モデル

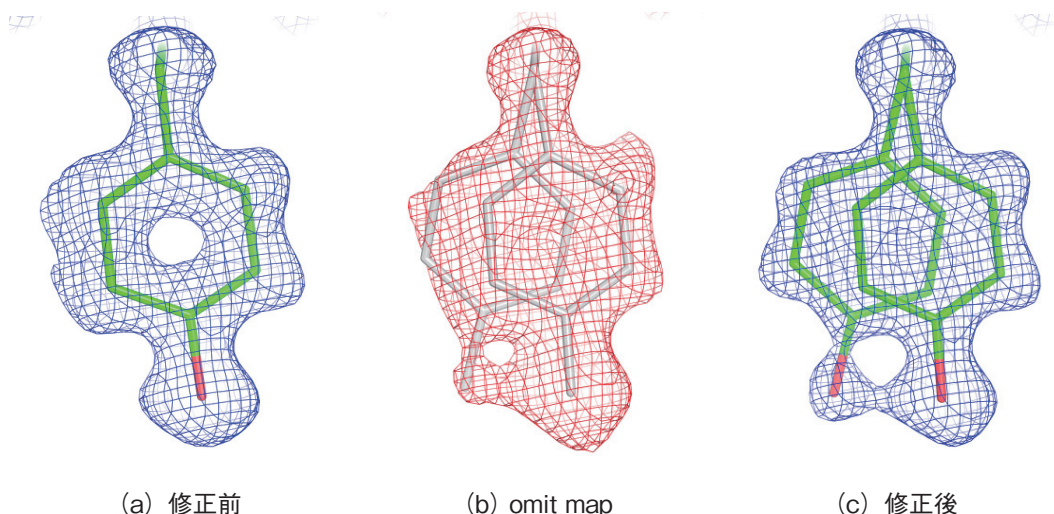


Fig.3. omit map を用いた Disorder モデルの構築

を構築したところ、Fig.3 (c) のように2つの位置で構造に一致する電子密度が得られた。このように omit map を用いることで、最終的に9箇所の Disorder 構造を発見した。

発見した Disorder 構造を Fig.4 に示す。MEM によりタンパク分子の外側に多数の Disorder が存在することが明らかになった。

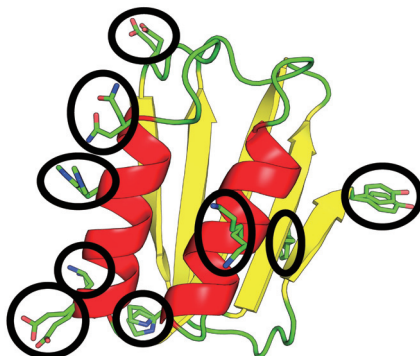


Fig.4. putative acylphosphatase の Disorder 構造

3. 水素原子及び原子の異方性の検討

水素原子は電子を1個しか持たずX線の散乱能が低いため、X線回折データへの寄与が小さい。このため水素原子を確認することは難しく、1ulr の構造には水素原子は加えられていなかった。MEM 電子密度では Fig.5 に示すように、多くの場所で水素原子が確認された。このため構造に水素原子を加えた。

タンパク質の構成原子の多くは異方的に熱振動しており、非等方的に広がった電子密度として観測される。この熱振動もX線回折データへの寄与が小さいため、従来の解析では確認することができなかった。MEM 電子密度では Fig.6 に示すように多くの箇所で非等方的に広がった電子密度を確認することができたため、構造に非等方的な原子モデルを使用した。

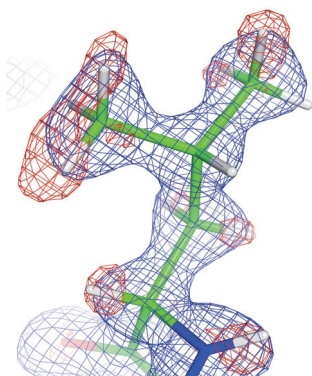


Fig.5. 水素原子を示す電子密度 (赤)

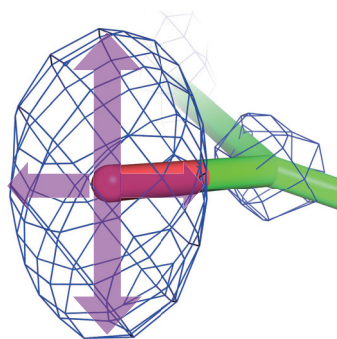


Fig.6. 電子密度の非等方的な広がり

4. MEM により決定された構造

これまでに述べたモデルの修正を繰り返し、修正箇所がなくなったところで最終の MEM 電子密度分布を得た。データとのずれを示す指標 R は、初期構造の 17.85% から 9.56% まで飛躍的に向上した。この指標の値は、PDB に登録された構造の中でもトップ 0.1% に入る値である。最終構造に基づく MEM 電子密度を Fig.7 に示す。原子間の電子密度が繋がっており、水素結合を形成している様子が電子密度で確認できる。このように、MEM を用いた構造解析により putative acylphosphatase の詳細な構造を明らかにすることに成功した。

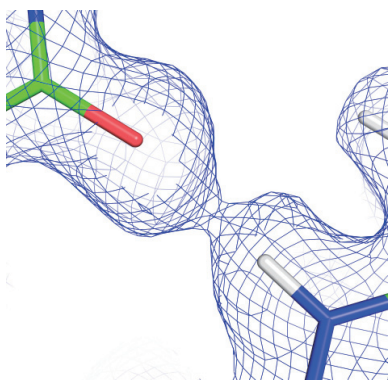


Fig.7. 水素結合の MEM 電子密度

IV. おわりに

MEM を用いたタンパク質構造解析として putative acylphosphatase の研究を報告した。MEM を用いることで、機能と関連の深い、水の構造・水素結合・Disorder 構造を明らかにできることがわかった。今後は、この手法を薬剤開発のターゲットとなるタンパク質などに適用していきたいと考えている。

今回の解析では、最終の解にたどり着くまでに 500 回近い MEM 計算が必要となった。構造解析では、構造モデル（構造因子の位相）を少しずつ修正しながらそのつど電子密度を求めることが要求される。MEM を用いた場合、水素や Disorder 等の微細な構造が明瞭に電子密度で観測できるため、モデル修正が終了するまでに通常の解析よりも多数回の電子密度計算が必要となった。現状では 1 回あたりベクトル計算機で数時間の MEM 解析を数百回繰り返すのは困難であるが、計算機性能の向上によりいずれは 1 台の PC で計算が簡単に行えるようになることを期待したい。

本研究は、理化学研究所播磨研究所宮野雅司主任研究員、吾郷日出夫専任研究員、理化学研究所成崎俊一主任研究員との共同研究である。putative acylphosphatase のデータは SPring-8 にて測定された。MEM 解析プログラム ENIGMA を作成していただいた島根大学田中宏志准教授、理化学研究所播磨研究所高田昌樹主任研究員に感謝いたします。MEM 計算には、理化学研究所

RSCC システムを使用しました。

参考文献

- (1) Klein, D. J., Schmeing, T. M., Moore, P. B. & Steitz, T. A. (2001). *Embo J.* **20**, 4214-422 など
- (2) Afonine, P. V., Lunin, V. Y., Muzet, N. & Urzhumtsev, A. (2004). *Acta Cryst.* **D60**, 260-274
- (3) Sakata, M. & Sato, M. (1990). *Acta Cryst.* **A46**, 263-270
- (4) Kariuki, B. M., Serrano-Gonzalez, H., Johnston, R. L. & Harris, K. D. M. (1997). *Chem. Phys. Lett.* **280**, 189-195
- (5) Tanaka, H., Takata, M., Nishibori, E., Kato, K., Ishii, T. & Sakata, M. (2002). *Acta Cryst.* **A35**, 282-286

(なかむら たかひろ：名古屋大学大学院工学研究科マテリアル理工学専攻博士前期課程)

(あおやぎ しのぶ：名古屋大学大学院工学研究科マテリアル理工学専攻)

(にしほり えいじ：名古屋大学大学院工学研究科マテリアル理工学専攻)

(さかた まこと：名古屋大学大学院工学研究科マテリアル理工学専攻)