タンパク質結晶構造解析入門—ブラックボックスの中身— (2) 位相決定法の裏側をのぞいてみよう

位相決定法の原理

大阪大学蛋白質研究所 中川敦史

Atsushi NAKAGAWA: Fundamentals of Phase Determination for Macromolecular Crystallography

Phase determination is one of the most difficult step for crystal structure determination. Many techniques, such as isomorphous replacement, multiple- or single-wavelength anomalous diffraction, molecular replacement, direct method, and their combinations, are often used for structure determination of biological macromolecules. Basic theories of phase determination of biological macromolecular crystallography are presented for students who want to learn a "Black Box" of phasing.

1. 位相問題を理解しよう

1.1 フーリエ変換を理解しよう

X線結晶構造解析の裏側を理解するためには、フーリエ変換を理解する必要があります。本入門講座の第1回でもフーリエ変換について説明されていましたが、もう一度おさらいをしてみましょう。

フーリエ変換の基礎は、どのような関数もすべて三角関数の組み合わせで表すことができるというところにあります。

周期が1である一次元のフーリエ級数は、次のように表されます。

$$F(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos(2\pi nx) + b_n \sin(2\pi nx)) \quad (1)$$

一方、

$$\exp(i\theta) = \cos\theta + i\sin\theta$$

$$\exp(-i\theta) = \cos\theta - i\sin\theta \quad (2)$$

であるので、

$$\cos\theta = \{\exp(i\theta) + \exp(-i\theta)\} / 2$$

$$\sin\theta = \{\exp(i\theta) - \exp(-i\theta)\} / (2i) \quad (2')$$

となり、この関係から式(1)をより簡単な式で表すことができます(式(3))。

$$F(x) = \sum_{n=-\infty}^{\infty} A_n \cdot \exp(2\pi i n x) \quad (3)$$

図1 ある一次元の関数のフーリエ変換。(Example of 1-dimensional Fourier transformation.) (a) オリジナルの関数。(b) (c) ~ (e) の関数を足し合わせたもの。おおよそ元の関数が再現できている。オリジナルの関数が完全に再現できないのは、十分な数の関数が足し合わされていないため。(f) フーリエ級数の周波数成分の大きさ。n = 2, 3, 5 の成分の寄与が大きいことがわかる。

フーリエ変換を視覚的に理解するために、図1に1つの例を示します。図1aは一見複雑な関数のように見えますが、この関数は、図1c～eの3つの簡単な三角関数の和で近似することができます(図1b)。図1aの関数を厳密に再現するためには、式(1)で明らかなように無限個の波を足し合わせる必要があります。しかし、現実には無限の数の波を足し合わせることは不可能であり、ある程度の周波数の波の足し合わせである関数を近似することになります。有限個の波の重ね合わせで求められた関数と元の関数のずれを、「フーリエ級数の打ち切り誤差」と呼びます。結晶構造解析においても、無限個の構造因子を測定することはできないので、完全に結晶内の電子密度を復元することはできません。

ところでステレオなどで、音に合わせてバーが上下するような表示を見たことがある人も多いと思います(図2)。これは、入力音の中にどのような周波数の成分がどれくらい含まれているかを視覚的に表していますが、このような表示は入力音をフーリエ変換して、それぞれの周波数成分の大きさを表したものです。

図2 音のスペクトルの表示。(Example of spectrum of sound.) 左右の音に対して、どのような音の成分がどれくらい含まれているかをリアルタイムで視覚的に表示している。

図3 結晶の構造因子は、結晶内の各原子からの散乱に相当するベクトルの和で求められる。(The structure factor F is the sum of the scattering by the separate atoms in the unit cell.)

ここで式(3)を、三次元に拡張し、パラメータを結晶構造解析で使われるものに置き換えると式(4)、式(4')のようになります。

$$\rho(\mathbf{r}) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} (F(hkl) \cdot \exp\{-2\pi i(hx + ky + lz)\}) \quad (4)$$

$$= \frac{1}{V} \sum_{\mathbf{k}} F(\mathbf{k}) \cdot \exp(-2\pi i \mathbf{k} \cdot \mathbf{r}) \quad (4')$$

ここで、 $\mathbf{k} = (h \ k \ l)$ 、 $\mathbf{r} = (x \ y \ z)$ とする

また、式(4')からフーリエ変換の性質を用いて、

$$F(\mathbf{k}) = \int_0^1 \int_0^1 \int_0^1 \rho(x, y, z) \cdot \exp\{2\pi i(hx + ky + lz)\} dx dy dz \quad (5)$$

$$= \int_{\mathbf{r}} \rho(\mathbf{r}) \cdot \exp(2\pi i \mathbf{k} \cdot \mathbf{r}) d\mathbf{r} \quad (5')$$

という関係式を得ることができます。結晶構造解析では、式(5)はフーリエ変換、式(4)は逆フーリエ変換と呼ばれています。これら2つの式はX線結晶構造解析で最も重要な式なので、必ず覚えておいてください。

式(5)は、結晶内の電子密度を積分する代わりに、結晶の単位格子中の位置 $\mathbf{r}_n$ に原子散乱因子 f_n をもつ仮想的な原子 n が存在すると考えて

$$F(\mathbf{k}) = \sum_{n=1}^N f_n \cdot \exp(-B_n \sin^2 \theta / \lambda^2) \cdot \exp(2\pi i \mathbf{k} \cdot \mathbf{r}_n) \quad (6)$$

(B_n は各原子 n の温度因子)

と表すこともできます。式(6)は、構造因子がそれぞれの原子からの寄与に相当するベクトルの和で求められることを意味しています(図3)。

1.2 位相問題とは？

式(4')から、構造因子 $F(\mathbf{k})$ がわかれば結晶内の電子密度分布 $\rho(\mathbf{r})$ を簡単に求めることができます。

一方、回折実験で得られるX線の強度 $I(\mathbf{k})$ と構造因子の大きさ $|F(\mathbf{k})|$ は式(7)で示されるような関係にあるので、観測値である回折強度 $I(\mathbf{k})$ にLPAの補正(L：ローレンツ因子、P：偏光因子、A：吸収因子)を適用することにより、構造因子 $|F(\mathbf{k})|$ の大きさを求めることができます。

$$I_{\text{obs}}(\mathbf{k}) \propto \text{LPA}\{F(\mathbf{k}) \cdot F^*(\mathbf{k})\} = \text{LPA}|F(\mathbf{k})|^2 \quad (7)$$

しかし、構造因子 $F(\mathbf{k})$ は、式(8)に示すように、大きさ $|F(\mathbf{k})|$ と位相 $\alpha(\mathbf{k})$ をもった量(複素量)であるのに対して、回折強度の測定で得られるのはその大きさ $|F(\mathbf{k})|$ だけであり、位相 $\alpha(\mathbf{k})$ は失われることを意味しています。

$$F(\mathbf{k}) = |F(\mathbf{k})| \cdot \exp\{i\alpha(\mathbf{k})\} \quad (8)$$

X線結晶構造解析というのは、この失われた位相を求めることにほかなりません。位相を求めることは「位相問題を解く」と呼ばれることもあります。

ところで、実験では位相 $\alpha(\mathbf{k})$ は失われていますが、構造因子の大きさ $|F(\mathbf{k})|$ は測定で得られています。そこで、構造因子に適当な位相角を与えた場合、どうなるで

しょうか. 簡単のために, 二次元のフーリエ変換を考えてみましょう. 図4aに示したある球状タンパク質の図をフーリエ変換すると図4bのような像が得られます. 図をフーリエ変換するというは少しわかりにくいかもしれませんが, 図4a上のある位置 (x, y) における色の濃さを $\rho(x, y)$ とすると, そのフーリエ変換は, 式(5')と同様,

$$F(h, k) = \sum_x \sum_y (\rho(x, y) \exp[2\pi i(hx + ky)]) \quad (9)$$

となります. この $|F(h, k)|$ を二次元で表したものが図4bとなります(このとき, $|F(h, k)|^2$ を表示すると, 1分子からの回折像と同じものになります). 位置 (h, k) における値 $F(h, k)$ は複素数ですが, 図では位相 $\alpha(h, k)$ を表現できないので, その大きさ $|F(h, k)|$ のみを表示しています. 計算機上ではそれぞれの位置に対して大きさと位相をもった量として扱っていることに注意してください. 図4b

を位相も考慮して逆フーリエ変換すると図4cが得られます. このように, ある図をフーリエ変換し, さらにそれを逆フーリエ変換することによって元の図を再現することができます. 同様に図4dに示したDNAの絵をフーリエ変換すると図4fのような像が得られ, その逆フーリエ変換により元の図を再現することができます. ここで, 図4bの大きさと図4eの位相を使ってフーリエ変換するとどうなるでしょう? 答えは図4gのようになります. ノイズの多い図ではありますが, 見た目は図4dのようなDNA様の構造が現れています. 逆に図4eの大きさと図4bの位相を使ってフーリエ変換すると, 図4aのような球状タンパク質の構造が表れてきます(図4h). つまり, フーリエ変換において得られる図は, 振幅よりも位相に大きく引きずられることを意味します. このことは, 大きさの違いより, 位相が違うほうが, 複素平面上で大きく違った場所を

図4 二次元のフーリエ変換において位相を入れ替えたときのシミュレーション. (Simulation of 2-dimensional Fourier.) (a) 球状タンパク質の図. (b) 図4aのフーリエ変換. 図4aをフーリエ変換すると各点 (h, k) に対して大きさ $(|F(h, k)|)$ と位相 $(\alpha(h, k))$ をもった複素数 $(F(h, k))$ が得られるが, その大きさ $(|F(h, k)|)$ のみを表している. (c) 図4bの逆フーリエ変換. 図4aを再現することができる. (d) DNAの図. (e) 図4dのフーリエ変換と(f) その逆フーリエ変換. (g) 図4bの大きさの情報と図4eの位相の情報を用いて逆フーリエ変換で得られた図. (h) 図4eの大きさの情報と図4bの位相の情報を用いて逆フーリエ変換で得られた図. 逆フーリエ変換において, 得られる結果は位相情報に大きく引きずられる.

示すことから明らかです。

回折実験で失われた位相を直接回復することは不可能なので、構造解析のためには何らかの方法で位相問題を解決する必要があります。低分子結晶の場合には、直接法が利用されることが多いのですが、タンパク質結晶の場合には原子数が多くまた分解能も十分に高くない場合がほとんどなので、直接法を適用することができません。そこで、さまざまな位相決定法が開発されました。

2. タンパク質結晶構造解析における位相決定法

2.1 分子置換法

分子置換法は、ネイティブ結晶の構造因子のみで構造決定できるため、対象とするタンパク質の構造に類似の構造が知られている場合に非常に有効な方法です。結晶系の違う結晶が得られた場合や、基質複合体や変異体の構造解析の多くの場合に利用されますが、新規のタンパク質であっても一次構造の似たタンパク質の構造が知られている場合には、まず試してみるのがよいでしょう。分子置換法の詳細に関しては、本入門講座の「(2.2) 分子置換法の原理と実際」を参考にしてください。

2.2 重原子同形置換法 (MIR 法, SIRAS 法)

2.2.1 多重原子同形置換法

重原子同形置換法は、1950 年代、Perutz らにより初めてのタンパク質の構造解析に利用されて以来、現在に至るまで広く利用されている位相決定法です。この方法は、ネイティブ結晶以外に、結晶中のタンパク質の特定の部位に重原子を結合させた結晶を作製し、それらの回折強度データの強度差から位相を決定していく方法です。

ダイヤモンドや水晶、塩化ナトリウムなどのような分子性・イオン性結晶とは違って、タンパク質の結晶では、分子間に非常に大きな「すき間」ができています。通常そのすき間は水などの溶媒分子で占められていて、一般的な水溶性タンパク質結晶の場合その体積の 27 ~ 65 % が溶媒領域であると言われています。²⁾ 言い換えると、固体であるタンパク質結晶も、実は非常に濃い溶液の特殊な状態であるとも言えます。この溶媒領域を通して、結晶の外から結晶内にいろいろなイオンや低分子化合物を導入することが可能となります。タンパク質結晶を水銀や白金などの重原子化合物の入った溶液 (結晶が壊れないような組成・条件が必要ですが) に数秒から数日程度浸けておくと、重原子化合物はこの「すき間」から結晶全体に浸透していきます。このとき、重原子化合物がタンパク質に特異的に結合すれば、ネイティブ結晶とほとんど変わらず、限られた数の重原子が特異的な位置を占める結晶を作ることが可能になります。

低分子化合物で利用される重原子法は、結晶内の分子全体の限られた数の重原子を初期構造として近似し、そこから逐次的に全体へと拡張していく方法ですが、タンパク質

結晶の場合には、重原子以外の軽原子の数が膨大であるためにこの重原子法は適用できません。そこで、重原子同形置換法では、ネイティブ結晶と重原子結晶の強度差を利用して位相決定を行っています。

重原子同形法では「同形」という言葉どおり、重原子のあるなし以外は、結晶内でのタンパク質の位置・配向および構造がまったく同じであることが前提となっています。実際には両者は多少なりとも異なっていますが、あまり大きく異なる (同形性が崩れる) と重原子同形置換法を適用することはできません。同形性が保たれているとき、ネイティブ結晶の構造因子を $F_P(\mathbf{k})$ 、重原子誘導体結晶の構造因子を $F_{PH}(\mathbf{k})$ 、重原子からの寄与を $F_H(\mathbf{k})$ とすると、これらは、式 (10) で関係付けられます。

$$F_{PH}(\mathbf{k}) = F_P(\mathbf{k}) + F_H(\mathbf{k}) \quad (10)$$

この関係式は、複素平面上で図 5 のように表すことができます。

実験的には、ネイティブ結晶、重原子誘導体結晶それぞれの構造因子の大きさ $|F_P(\mathbf{k})|$ 、 $|F_{PH}(\mathbf{k})|$ を得ることができます。これらの値から複素平面上での作図によって指数 $\mathbf{k}$ の構造因子に対する位相を求める方法を考えてみましょう。まず、複素平面上の原点 O を中心に半径 $|F_P(\mathbf{k})|$ の円を描きます。もし、重原子のみからの寄与 $F_H(\mathbf{k})$ がわかっていれば、終点を原点 O とするベクトル $F_H(\mathbf{k})$ を描くことができます。このベクトル $F_H(\mathbf{k})$ の始点を中心とする半径 $|F_{PH}(\mathbf{k})|$ の円を描くと、2つの円の交点 A に向かうベクトル $\overrightarrow{OA}$ が、ネイティブの構造因子 $F_P(\mathbf{k})$ となります。

図 5 重原子同形置換法におけるネイティブ (F_P)、重原子 (F_H)、重原子誘導体 (F_{PH}) の構造因子の関係を示す Argand ダイアグラム。(Argand diagram showing the native protein (F_P) and heavy atom (F_H) contributions to the structure factor for the heavy atom derivative (F_{PH})).

図6 重原子同形置換法の原理.(Harker diagram to illustrate phase determination by single isomorphous replacement (SIR).) 終点を原点 O とするベクトル F_H を描き、その始点を中心とする半径 $|F_{PH}|$ の円を描く。この円と原点 O を中心とする半径 $|F_P|$ の大きさとの円の交点 A に向かうベクトル $\vec{OA}$ が、ネイティブの構造因子 F_P となる。しかし、重原子同形置換体が1つしかない場合には A' でも2つの円が交わり、別の解 ($\vec{OA'}$) が得られてしまう。これは物理的には意味のない偽の解である。

(図6)。

では、重原子の構造因子 $F_H(\mathbf{k})$ はどうすればわかるのでしょうか。最も簡単な方法としては、 $\Delta F_{iso}^2 = (|F_{PH}(\mathbf{k}) - F_P(\mathbf{k})|^2)$ を係数とするパターンソン関数法がよく使われます。パターンソン関数は構造因子の二乗を係数とするフーリエ変換で、原子間ベクトルを直接得ることができます(式(12), 式(13))。 ΔF_{iso}^2 を係数とした場合には一般的には重原子間ベクトルを得ることができます。

$$P(\mathbf{u}) = \frac{1}{V} \sum_{\mathbf{k}} |F(\mathbf{k})|^2 \cdot \exp(-2\pi i \mathbf{k} \cdot \mathbf{u}) \quad (11)$$

$$= \frac{1}{V} \int_V \rho(\mathbf{r}) \rho(\mathbf{r} + \mathbf{u}) d\mathbf{r} \quad (12)$$

パターンソン関数から、重原子位置を知る方法はいくつか知られていますが、最も簡単な方法は、ハーカ面上のピークを利用して求める方法でしょう。例えば、空間群の対称性により b 軸上に2回軸がある場合を考えます。この場合、 (x, y, z) に原子があれば、必ず $(-x, y, -z)$ にも原子が存在します。パターンソン空間上では原子間のベクトルに相当する位置に、それぞれの原子位置での電子密度の積に相当する高さのピークが得られるので、上の例では $(0, 0, 0)$, $(2x, 0, 2z)$, $(-2x, 0, -2z)$ の位置にピークが得られます。すなわちこの場合、 $v=0$ の面に必ずピークが現れることになります。このときのピークの座標を $(u, 0, w)$ とすると、重原子の座標の (x, z) は、 $x = u/2$,

$z = w/2$ と求めることができます(この例では、 y の値を求めることはできません)。このようにパターンソン関数で自己ベクトルが必ず現れる面はハーカ面と呼ばれ、重原子位置を得る最初の手がかりとなります。ハーカ面の存在しない空間群もありますが、多くの空間群の場合において、特に重原子の数があまり多くない場合には、重原子位置を簡単に求めることができる有効な方法です。

$|F_{PH}| - |F_P|$ から重原子位置を求めるために、このほかに、shelxs³⁾ や solve⁴⁾ で使われているベクトルの重ね合わせによる最小値関数法、あるいは、低分子の構造解析で広く使われている直接法 (shelxs,⁵⁾ rantan⁶⁾ など) やその発展形である shake-and-bake 法 (SnB,⁷⁾ shlexd⁸⁾ など) がよく使われています*1。非対称単位中の重原子の数が多くなってくるとその解釈が指数関数的に難しくなってきますが、2.3 で述べるセレノメチオニンを利用した多波長異常分散法の場合には、100 個以上の重原子位置を決めることも報告されています。

重原子位置が求められたら、次に重原子パラメータ (位置、占有率*2、温度因子) やその他のパラメータ (相対的なスケール因子、相対温度因子など) を精密化します。古い教科書を見ると、

$$|F_{PH}(\text{calc})| = |F_P(\text{calc}) + F_H(\text{calc})| \quad (13)$$

として、

$$\epsilon^2 = \sum w (|F_{PH}(\text{obs})| - |F_{PH}(\text{calc})|)^2 \quad (14)$$

を最小とするように重原子パラメータを精密化すると書いてあります*3。しかし、より厳密には、最小自乗法による精密化では、データのもっている情報をすべて引き出すことはできないため、最近のプログラムでは最尤法を適用するようになってきました(付録参照)。Bricogne らによるプログラム SHARP¹⁰⁾ は、尤度 (likelihood) の取り扱いをより厳密に行い、位相角とネイティブの構造因子の大きさもパラメータとして最尤度を求めるため、従来のプログラムに比べて非常に多くの計算量を必要としますが、アルゴリズムの改良と計算機の性能向上により、十分に実用的な計算時間で結果を得ることができるようになっています。

先ほど、重原子の構造因子がわかれば、ネイティブの位相がわかると書きました。しかし図6からわかるように、ネイティブと1つの重原子誘導体の構造因子からだけでは、別の解も同じ確率で得られてしまうため、一義的に

*1 直接法に関しては、日本結晶学会誌 Vol.38 No.5 入門講座「X線構造解析を始めよう (6. 直接法で構造を解く)」を参考にしてください。

*2 一般的には重原子は結晶内のすべての単位格子中において同じ位置に存在しているとは限らないので、単位格子中の特定の位置に原子がどれくらいの割合で存在しているかを占有率として取り扱います。

図7 多重原子同形置換法の原理. (Harker diagram to illustrate phase determination by multiple isomorphous replacement (MIR).) 2つの重原子同形置換体のデータを用いて図6と同様の作図を行うと, 2つめの誘導体のデータ(点線で示す)では図5で見られた偽の解($\overrightarrow{OA}$)は条件を満足しないため, 結果的に3つの円が交わるのは一個所のみとなり, 正解が一義的に得られる.

解を求めることができません(単一同形置換法(Single Isomorphous Replacement: SIR法)). そこで, 1つめの重原子誘導体とは別の重原子試薬を用いて2つめの重原子誘導体を作製し, その構造因子の大きさも利用することで, 一義的に位相を決定することができます. これが, 多重同形置換法(Multiple Isomorphous Replacement: MIR法)と呼ばれる位相決定法です(図7).

実際の構造解析では, ネイティブ結晶と重原子置換体は完全に同形であるわけではなく, また重原子のパラメータのあいまいさや観測量の誤差などにより, 式(10)を満足しません(式(15), 図8).

$$\mathbf{F}_{PH}(\mathbf{k}) = \mathbf{F}_P(\mathbf{k}) + \mathbf{F}_H(\mathbf{k}) + \boldsymbol{\varepsilon} \quad (15)$$

この場合は, 図7のように1点で円が交わることはありません. そこで, 実験的には一義的な位相ではなく, 位相の確率分布を得ることになります.

j 番目の重原子誘導体に対する $|F_{PH}(\mathbf{k})|$ の観測値と計算値のずれ, “lack of closure error” は

$$\varepsilon_j = k_j \cdot |F_{PH,j}(\text{obs})| - |F_{PH,j}(\text{calc})| \quad (16)$$

k_j は相対尺度因子

で表すことができます. 一方, 構造因子の大きさの誤差の平均を $\langle \delta \rangle$ とすると, 全体の誤差は

$$\langle E \rangle^2 = \langle \delta \rangle^2 + \langle \varepsilon \rangle^2 \quad (17)$$

図8 同形性が崩れた場合の, ネイティブ($\mathbf{F}_P$)と重原子誘導体($\mathbf{F}_P$)の構造因子の関係. (Argand diagram showing the native ($\mathbf{F}_P$) and heavy atom ($\mathbf{F}_H$) contributions to the structure factor ($\mathbf{F}_{PH}$) for the heavy atom derivative considering errors ($\boldsymbol{\varepsilon}$).) $\boldsymbol{\varepsilon}$ の分だけベクトルが交わらない.

となります.

誤差が正規分布に従うと仮定すると, j 番目の誘導体の位相 α での位相の確率 $P_j(\alpha)$ は,

$$P_j(\alpha) = \exp\{-\varepsilon_j(\alpha)^2 / 2E_j^2\} \quad (18)$$

となり, 複数の同形置換体からの寄与を考えると全体の位相確率は,

$$P(\alpha) = \prod_j P_j(\alpha) = \exp[-\sum_j \{\varepsilon_j(\alpha)^2 / 2E_j^2\}] \quad (19)$$

となります. では, 位相確率分布 $P(\alpha)$ から, 最も良い位相を求めるにはどうすればよいでしょうか. 逆フーリエ変換で得られる電子密度の誤差を最小にするには,

$$\mathbf{F}(\text{best}) = \mathbf{F}_P \int_0^{2\pi} P(\alpha) \cdot \exp(i\alpha) d\alpha / \int_0^{2\pi} P(\alpha) d\alpha \quad (20)$$

を用いたフーリエ合成を行えばよいことが知られています. さらに,

$$\mathbf{m} = \mathbf{m} \cdot \exp(i\alpha_{\text{best}}) = \int_0^{2\pi} P(\alpha) \cdot \exp(i\alpha) d\alpha / \int_0^{2\pi} P(\alpha) d\alpha \quad (21)$$

式(21)で定義される $\mathbf{m}$ は位相の確からしさを表す目安となり Figure of merit (FOM) と呼ばれ, $P(\alpha)$ が鋭いピークをもつとき1に近い値をとり, 一方 360° にわたって一様に分布する場合には0に近い値となります. また, 位相角の重心 α_{best} は最良位相角と呼ばれています. SHARP では, ネイティブの構造因子もパラメータとして最適値を求めていますので, $|F_P|$ の代わりに $|F_P(\text{best})|$ を係数とし

た逆フーリエ変換が使われています。

重原子同形置換法の結果の良否を判定するためのいくつかの指標があります。そのうち、重原子パラメータの精密化の一致度を示す指標としては、

$$R_{\text{Kraut}}(\text{iso}) = \frac{\sum \|F_{\text{PH}}| - |F_{\text{PH}} + F_{\text{H}}(\text{calc})\|}{\sum |F_{\text{PH}}|} \quad (22)$$

および

$$R_{\text{Cullis}}(\text{iso}) = \frac{\sum \|F_{\text{PH}} \pm F_{\text{P}}| - |F_{\text{H}}(\text{calc})\|}{\sum |F_{\text{PH}} \pm F_{\text{P}}|} \quad (23)$$

($R_{\text{Cullis}}(\text{iso})$) は、セントリック反射のみを使って計算する。符号は F_{PH} と F_{P} の位相が逆向きのときプラスに、 F_{PH} と F_{P} の位相が同じ向きするときマイナスになる)

が使われ、また、重原子の位相への寄与を示す指標としては、

$$\text{Phasing Power} = \langle |F_{\text{H}}(\text{calc})| / \varepsilon \rangle \quad (24)$$

ε は lack of closure error

がよく用いられます。

重原子パラメータの精密化と平行して、 $|F_{\text{PH}}| - |F_{\text{P}}|$ を係数とし α_{P} を位相とする差フーリエ図や $|F_{\text{PH}}| - |F_{\text{P}} + F_{\text{H}}|$ を係数とし α_{P} を位相とする二重差フーリエ図 (double difference Fourier), あるいは Bijvoet 差 (後述) を係数とし $\alpha_{\text{P}} - \pi/2$ を位相とする異常分散差フーリエ図や Log-likelihood gradient map^{*3}などを手がかりにマイナーサイトを見つけていきます。このようにパラメータの精密化—マイナーサイトの同定のサイクルを繰り返し、 $F_{\text{H}}(\mathbf{k})$ を求めます。

2.2.2 異常分散を考慮した単一同形置換法 (SIRAS 法)

重原子置換体の作製は、1 mg/ml 程度の濃度で白金や水銀化合物を数日ソーキングすることから始める場合が多いですが、結晶が壊れたり、分解能が大きく低下したりする場合には、濃度や時間をいろいろ変えながら最適な条件を見つけていく必要があります。Br⁻イオンやI⁻イオンの場合には、数秒から数分程度の短時間のソーキングでうまくいく場合が多いと言われています。このほか、キセノンを1~2 MP程度の圧力でかけることによって、タンパク質分子中の疎水ポケットにキセノン分子を導入できる方法も提案されています。いずれの方法でも、重原子の選択も含め条件決定に試行錯誤を必要とする場合が多いことが、重原子同形置換法の欠点であると言えます。

2つ以上の重原子誘導体を作製する必要がある多重

*3 SHARPで計算されるマップで、通常の差フーリエ図より重原子の寄与に対して敏感であると言われています。

図9 異常分散項の例。鉄、セレン、水銀の異常分散項を示す。(Wavelength dependence of anomalous scatterings term of iron, selenium, and mercury.) 水銀の3つの吸収端は短波長側からそれぞれL_I, L_{II}, L_{III}吸収端を示す。それ以外はK吸収端。

図10 異常分散を考慮したときの構造因子を示す Argand ダイアグラム。(Argand diagram showing structure factor of heavy atom derivative with contribution of anomalous scattering factor.) 異常分散の虚数項(f'')の位相は、 F_{H} に対して必ず $\pi/2$ だけ進むため、異常分散効果が無視できない場合にはFriedel則が成り立たなくなり $|F_{\text{PH}}(+)| \neq |F_{\text{PH}}(-)|$ となる。

原子同形置換法では、重原子置換体の作製のための試行錯誤をより多く行わなければならないため、解析に時間がかかる場合があります。そこで、単一同形置換法に異常分散効果を組み合わせる方法 (Single Isomorphous Replacement with Anomalous Scattering: SIRAS 法) が利

図11 SIRAS法の原理。(Harker diagram to illustrate phase determination by single isomorphous replacement with anomalous scattering (SIRAS).) 単一同形置換法では正しい位相を得ることができないが異常分散効果を考慮することにより、正しい位相を求めることができる(図では簡単のために f'' は無視している)。

用されます。

通常原子散乱因子は波長に依存しない量として取り扱えますが、異常分散効果が無視できないとき、原子散乱因子は波長に依存する2つの補正項(f' , f'')を用いて次のように表すことができます。

$$f = f_0 + f' + if'' \quad (24')$$

異常分散項は、原子種に特有の波長近傍で大きく変化します(図9)。例えば、セレン原子の場合、吸収端の波長は0.9792 Åですが、これは、K殻の電子の共鳴吸収によるものであり、K吸収端と呼ばれています。実際には、どの波長・原子種であっても異常分散効果を有していますが、水素、炭素、窒素、酸素といったタンパク質を構成している軽原子は通常のX線回折実験を行う波長では非常に小さな値であるので、無視することができます。しかし、水銀や白金といった重原子では異常分散項が大きくなり、無視することができなくなります。特に異常分散項の虚数部(f'')の寄与により、通常同じ大きさを示す $|F(\mathbf{k})|$ と $|F(-\mathbf{k})|$ の値が違ってきます(図10)。この強度差(Bijvoet差： $\Delta F^\pm(\mathbf{k}) = |F(\mathbf{k})| - |F(-\mathbf{k})|$)を利用することにより、 $F(\mathbf{k})$ と $F(-\mathbf{k})$ を独立なデータと考えることができ、1つの重原子誘導体からでも一義的に位相を決定することができます(図11)。

いろいろな原子種のいろいろな波長における異常分散項の値は、International Tables for Crystallography Volume Cや佐々木による表¹⁰⁾に載っていますし、あるいは、CCP4¹¹⁾

のcrosssecで計算することもできます。

放射光の波長選択性を利用して回折実験を行う波長を対象とする原子種に最適化することにより、異常分散効果を最大限に利用することが可能となりますが、吸収端近傍では共鳴吸収により異常分散項は大きく変化し、計算値からずれてくるので、SIRAS法やMAD法(2.3)で最適な波長を選択したいときには、結晶自身の吸収係数、あるいは蛍光X線の強度を基に f' , f'' を計算し、最適な波長を求めることが重要です。

異常分散を考慮したときに重原子パラメータの精密化の一致度を示す指標としては、

$$R_{\text{Kraut}}(\text{ano}) = \frac{\sum (||F_{\text{PH}}^+(\text{obs})| - |F_{\text{PH}}^+(\text{calc})|| + ||F_{\text{PH}}^-(\text{obs})| - |F_{\text{PH}}^-(\text{calc})||)}{\sum |F_{\text{PH}}^+(\text{obs})| + |F_{\text{PH}}^-(\text{obs})|} \quad (25)$$

および

$$R_{\text{Cullis}}(\text{ano}) = \frac{\sum ||\Delta F_{\text{PH}}^+(\text{obs})| - |\Delta F_{\text{PH}}^+(\text{calc})||}{\sum |\Delta F_{\text{PH}}^+(\text{obs})|} \quad (26)$$

がよく使われます。

2.3 多波長異常分散法 (MAD 法)

Okaya & Pepinskiにより提案された多波長異常分散法による位相決定¹²⁾は、放射光が汎用的に利用できるようになった1980年代後半から急速に普及してきました。特に遺伝子操作によりタンパク質のメチオニンの代わりにセレンメチオニンに置換したタンパク質を発現することが可能となったために、それまで重原子誘導体の調製に要していた試行錯誤を必要とせず位相決定を行うことができるようになりました。

Hendricksonらは、Karle & Karleの取り扱い¹³⁾を拡張し、結晶中に異常分散を示す原子が1種類しか存在しない場合に、複数の波長の回折強度と位相の関係を解析的に取り扱う方法を提案し、¹⁴⁾初めてのセレンメチオニン置換タンパク質の構造解析に成功する¹⁵⁾など、いくつかのタンパク質の構造解析に適用しました。

一方、Hendricksonらによる取り扱いと平行して、1つの波長のデータをネイティブとし、その他の波長のデータを同形置換体として取り扱うことにより重原子同形置換法と同じプログラムを使って位相決定を行う方法も提案され、¹⁶⁾重原子同形置換法と組み合わせたり、2つ以上の異常分散原子を取り扱えるなど有利な点が多いこともあり、現在ではこちらの方がよく使われています。

異常分散効果は、K吸収端の f'' で4電子程度、L_{III}吸収端 f'' で10電子程度と、重原子誘導体に使われる重原子の電子数が通常数10電子以上あるのに比べて小さい寄与しかありませんが、同形性のずれがほとんどないため、良好な結果が得られる場合が多いようです。

図12 MAD法の原理。(Harker diagram to illustrate phase determination by multiple wavelength anomalous diffraction (MAD).) 異常分散を示す複数の波長(図の場合2波長)を用いることにより正しい位相を求めることができる。

図13 SAD法の原理。(Harker diagram to illustrate phase determination by single wavelength anomalous diffraction (SAD).) 異常分散を示す1つの波長のデータからは正しい位相を一義的に求めることができない。SAD法では、電子密度改良法などと組み合わせることにより、一義的な正しい解を得る。

2.4 単一波長異常分散法 (SAD 法)

2.2.1に書いたようにネイティブと1つの重原子誘導体のみからは、一義的に位相決定を行うことができず、最良位相角 α_{best} は、重原子の位相 ϕ_a と同じになります。同様に1つの波長の回折強度データからは、Bijvoet 差を利用して正しい解を得ることはできません(図13)。ここで、正しい解を α_T 、間違った解を α_F とすると、 $\alpha_{\text{best}} (= \alpha_H = (\alpha_T + \alpha_F)/2)$ を位相としたフーリエ変換は、

$$\rho_{\text{SAD}}(x) = \frac{1}{2V} \sum_{h=0}^{\infty} \{ |F(\mathbf{k})| \cdot \exp(i\alpha_{\text{best}}) - |F(\mathbf{k})| \exp(2i\alpha_H) \exp(-i\alpha_T) \} \cdot \exp(-2\pi i \mathbf{k} \cdot \mathbf{r}) \quad (27)$$

となりますが、この式は、

$$\rho_{\text{SAD}}(x) = \{ \rho_{\text{protein}}(x) + (\text{noise}) \} / 2 \quad (28)$$

を意味しています。すなわち、1つの波長のデータしかない場合に Bijvoet 差のみから得られる電子密度は正しい構造を示す電子密度にノイズが乗ったようなものとなります。

ここで、溶媒領域平滑化法 (Solvent flattening) などの位相改良*4を用いて実空間上の電子密度を改良していくことによって、正しい位相情報を得ることが可能となります。この方法は、単一波長異常分散法 (SAD 法) と呼ばれて、1980年代半ばに Wang によって提案されていますが¹⁷⁾

(当時は Single anomalous scattering : SAS と呼ばれました)、電子密度改良法で正しい位相を得るためには、正しい位相確率分布が得られることが重要で、最尤法などが用いた最新のアルゴリズムの開発によって、その利用が現実になってきました。この方法を用いると、必ずしも波長可変の放射光を利用する必要がなく、仮に利用するにしても、少ない数の回折強度データセットから位相決定できるため、非常に有効な方法として、ここ数年急速に適用例が増えてきています。特に、タンパク質中のメチオニンやシステインなどのイオウ原子の異常分散効果のみで、しかも実験室系のX線発生装置を使って構造解析ができる可能性をもつ非常に魅力的な方法です。従来からの銅をターゲットとしたX線 ($\text{CuK}\alpha = 1.54 \text{ \AA}$; $f''(\text{S}) = 0.556$) や、より長波長のX線の得られるクロムをターゲットに使用したX線発生装置 ($\text{CrK}\alpha = 2.29 \text{ \AA}$; $f''(\text{S}) = 1.141$) を利用してネイティブ結晶中のイオウの異常分散のみで位相決定を行う、S-SAD法の適用例も増えてきています。

付録 最小自乗法と最尤法

観測した量から最も確からしいパラメータを得るにはどうすればよいでしょうか。ある測定値 x_i が得られたとき、真の値 x^0 はその近傍にあると考えられます。真の値が x^0 であるとき、実際に得られた値が観測される確率を $P(x_i|x^0)$ と表します。実際には真の値 x^0 を知ることはできないので、代わりにその推定値 $\hat{x}$ を使って $P(x_i|\hat{x})$ を考えます。ここでデータ R とあるモデルが与えられたときの仮定 H の尤度 (Likelihood) を $L(H|R)$ とすると $L(H|R)$ は

*4 詳細は、本入門講座「(3.2) 位相改良 (溶媒平滑化など) と手動モデリング」を参考にしてください。

$P(R|H)$ に比例する値になると定義することができます。

$$L(H|R) \propto R(R|H) \equiv P(H;R) \quad (29)$$

$L(H|R)$ と $P(R|H)$ の違いは、前者は R が定数で H が変数であるのに対し、後者はその逆であるということです。つまり、 $L(H|R)$ は、データ R が与えられたときに、仮定 H がどれくらい「もっともらしい」かを表すことになります。

先の例で、 N 回測定を繰り返すことを考えると、この測定から得られる推定値 $\hat{x}$ の尤度 $L(\hat{x})$ はそれぞれの値の積で表すことができます。

$$L(\hat{x}) = \prod_{i=1}^N L(\hat{x} | x_i) \quad (30)$$

この尤度関数が最大となる値を真値 x^0 に対して最も確からしい値 (最尤推定値; Maximum likelihood estimate) とする方法は最尤法 (maximum likelihood estimation) と呼ばれています。実際の計算では、 $L(\hat{x}|x_i)$ の積を求める代わりにその対数 $\ln(L(\hat{x}|x_i))$ (対数尤度; log-likelihood) の和を最大にする方法がよく使われます。

$$LL(\hat{x}) = \sum_{i=1}^N \ln\{L(\hat{x} | x_i)\} \quad (31)$$

観測値から最適なパラメータを求める方法としては、「最小自乗法 (最小二乗法)」がよく利用されます。これは、ある関数 ($G(x)$) を考えたときに、観測値から得られる値 G^{obs} と関数から求められる値 G^{calc} の差が最も小さくなるように、

$$\phi = \sum (G^{\text{obs}} - G^{\text{calc}})^2 \quad (32)$$

を最小にするようなパラメータを求める方法です。多くの方が、大学の学生実験などで、この方法でパラメータを求めたことがあるでしょう。実は、最小自乗法は、観測値がモデルに対して正規分布 (Gauss 分布) するときに最尤値を求めることと同じです。言い換えると、観測値から最も確からしい情報を得ることができるのは、観測値のば

らつきが正規分布するときのみになります。

重原子位置の精密化や構造の精密化では、観測値がモデルに対して正規分布するという仮定は成り立たないので、最小自乗法ではなく最尤法を適用したほうがより確かな結果を得ることができるため、最近のプログラムではよく利用されています。

文 献

- 1) D. W. Green, V. M. Ingram and M. F. Perutz: *Proc. Roy. Soc. A* **225**, 287 (1954).
- 2) B. W. Matthews: *J. Molec. Biol.* **33**, 491 (1968).
- 3) G. M. Sheldrick: *Crystallographic Computing*, 5 p.145, IUCr and Oxford University Press (1991).
- 4) T. C. Terwilliger and S. -H. Kim: *Acta Cryst.* **A43**, 1 (1987).
- 5) G. M. Sheldrick: *Acta Cryst.* **A46**, 467 (1990).
- 6) J. -X. Yao: *Acta Cryst.* **A39**, 35 (1983).
- 7) C. M. Weeks and R. Miller: *J. Appl. Cryst.* **32**, 120 (1999).
- 8) T. R. Shneider and G. M. Sheldrick: *Acta Cryst.* **D58**, 1772 (2002).
- 9) T. L. Blundell and L. N. Johnson: *Protein Crystallography*, Academic Press (1976).
- 10) S. Sasaki: KEK Report, 83-22 (1984).
- 11) Collaborative Computing Project Number 4: *Acta Cryst.* **D50**, 760 (1994).
- 12) Y. Okaya and R. Pepinski: *Phys. Rev.* **103**, 1645 (1956).
- 13) J. Karle: *Int. J. Quantum Chem. Symp.* **7**, 357 (1980).
- 14) W. A. Hendrickson: *Trans. Am. Cryst. Assoc.* **21**, 11 (1985).
- 15) W. Yang, W. A. Hendrickson, R. J. Crouch and Y. Satow: *Science* **249**, 1398 (1990).
- 16) V. Ramakrishnan and V. Biou: *Method Enzymol.* **276**, 538 (1997).
- 17) B. C. Wnag: *Method Enzymol.* **115**, 90 (1985).

プロフィール

中川敦史 Atsushi NAKAGAWA

大阪大学蛋白質研究所

Institute for Protein Research, Osaka University

〒565-0871 大阪府吹田市山田丘3-2

3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

e-mail: atsushi@protein.osaka-u.ac.jp

最終学歴: 大阪大学大学院理学研究科博士課程
中退

専門分野: 構造生物学, 蛋白質結晶学