

単結晶X線構造解析 基礎講座

第9回 タンパク質構造解析と低分子構造解析

山野 昭人*

1. はじめに

前回までは低分子を試料とする単結晶X線構造解析について述べてきたが、今回はX線回折によるタンパク質の構造解析に話を進める。とくにタンパク質の構造解析と低分子の構造解析との違いに着目して解説する。

タンパク質構造解析でも低分子構造解析でも、単結晶X線構造解析の目的は分子構造の決定であり原理も同一だが、構造解析の各段階の実行方法がかなり異なっている。この違いは主に対象となる分子の違いに起因している。最大の違いは分子の大きさだ。低分子は通常数百から数千程度の分子量であるのに対して、タンパク質は数千から数百万Da程度の分子量を持っている。また分子の構成要素や化学的な自由度も大きく異なっている。低分子を構成する原子種や分子の形状は多様で原理上制限がないのに対して、タンパク質は20種類の構造既知のアミノ酸から成っている。しかも地球上の生物が持つタンパク質は、少数の例外を除きすべてL型アミノ酸のみから成り立っているため、絶対配置も既知である。また、タンパク質の特徴的な構造の1つである α -ヘリックスは、N末端からC末端方向にたどった際、必ず右巻き螺旋になっている。

このようにあらかじめ構造に関する情報が多いにもかかわらず、新規タンパク質の構造解析は、対象となる分子の発現精製段階を含めると、数年かかるのもまれではない。なぜこのような手間のかかる作業をするのか。それはタンパク質の構造が決定されると、いろいろな恩恵が得られるからである。

2. タンパク質構造解析でなにが分かるのか

いったい完成までにどのくらいの時間がかかるのかは分からないが、人類は生命現象の原子レベルでの完全理解を目指して進んでいるように見受けられる。生命現象の理解で重要なのが原子レベルでの代謝や生体反応の理解で、これは主役となるタンパク質の原子レ

ベルでの構造の解明なしには成し得ない。ひとたび特定の代謝経路が判明すると、関与するタンパク質に作用する化合物を作り出すことにより、反応を制御することが可能となる。

たとえばウィルスの生活環のカギとなるタンパク質の構造が解明されれば、抗ウィルス薬を作ることができる。X線構造解析の結果得られた標的タンパク質の3次元構造に基づいて薬を開発する方法を、Structure Based Drug Design (SBDD) という。SBDDの黎明期の顕著な成功例は、HIVが持つプロテアーゼの働きを阻害する薬剤だろう。HIVは自身を再生する最終段階で、宿主にまとめて作らせたタンパク質を切断し、部品を完成する。この反応をつかさどるのがHIVプロテアーゼであるため、ひとたびこのプロテアーゼがブロックされると、HIVは感染することはできても自身を再生することができなくなる。現在ではさらに進んで、HIVのようなレトロウィルスのみが持つDNA-RNA逆転写酵素や、HIVが宿主の遺伝子に融合する段階を触媒する酵素インテグラーゼを阻害する、より特異性の高い医薬品も登場している。

より身近な例では、頭痛薬の例がある。図1はアルコールを過剰摂取した際の頭痛の発生機序である。摂取されたアルコールはアルコール脱水素酵素によりアセトアルデヒドに分解される。アセトアルデヒドはアセトアルデヒド脱水素酵素により、アセチルCo-Aに変換され、TCAサイクルに入る。TCAサイクルは解糖系の中心となる代謝経路で、生体内での代表的なエネルギー担体であるATPを作り出す代謝経路である。さてアセトアルデヒドがアセトアルデヒド脱水素酵素によりすべて分解されれば問題はないのだが、分解が間に合わない場合、たとえば大量のアルコールを急激に摂取した場合には頭痛が起こる。アセトアルデヒドには血管の平滑筋の弛緩作用があり血管を拡張するが、血管が拡張すると炎症がおこり、ブラジキニンという物質が作られる。ブラジキニンはホスホリパーゼA₂を活性化する。活性化したホスホリパーゼA₂は、

*株式会社リガク X線機器事業部応用技術センター

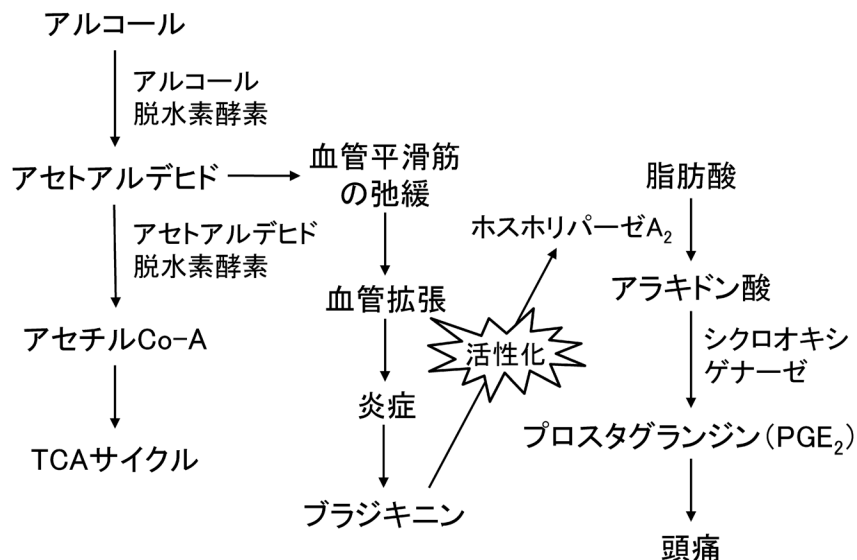


図1. アルコール過剰摂取時の頭痛の発生機序。ホスホリパーゼA₂やシクロオキシゲナーゼの活性を阻害すると、発痛物質が産出されないため鎮痛作用が期待できる。

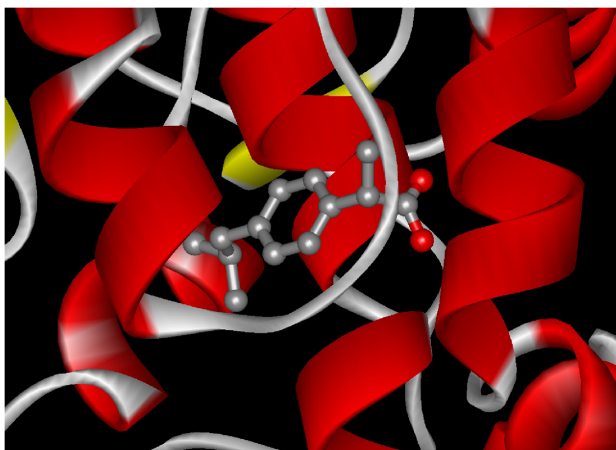


図2. シクロオキシゲナーゼ阻害剤であるイブプロフェンとシクロオキシゲナーゼの共結晶構造。イブプロフェンは市販の鎮痛薬の薬効成分として含有されている。

脂肪酸を分解し、アラキドン酸にする。アラキドン酸はさらにシクロオキシゲナーゼによりプロスタグランジン (PGE₂) に変換される。このPGE₂が発痛物質であり頭痛が発生する。このように特定の症状を引き起こす代謝経路が判明し、関与するタンパク質の分子構造が決定されると、阻害剤を設計することが可能となる。この例ではホスホリパーゼA₂やシクロオキシゲナーゼを阻害する分子を設計すれば、それが頭痛薬の出発点となる。薬剤の候補化合物を探す最初の段階では、化合物ライブラリーからコンピュータによるドッキング計算等で選択した化合物や、既存の医薬品との比較や薬草等に含まれる天然物などから出発する。これらを実際のタンパク質に結合させてX線構造解析を行い、より結合力の高い化合物になるように構造の最適化をしていく。図2はPDB (Protein Data Bank) にある、シクロオキシゲナーゼと鎮痛剤イブプロフェンの

共結晶構造である。イブプロフェンは風邪薬の鎮痛成分としてすでに市販されており、シクロオキシゲナーゼの活性部位に効率よく結合していることが分かる。

3. タンパク質構造解析と低分子構造解析はどこがちがうのか？

タンパク質構造解析と低分子構造解析は、目的・原理・解析の手順は同様だが、解析の各段階の内容が大きく異なる。まずは低分子構造解析の手順を復習する。

3.1. 低分子構造解析の手順

低分子化合物は通常化学合成により作製する。次に化合物を結晶化する。結晶化は主にバッチ法という手法で行われる。化合物を溶媒に溶解し、徐々に溶媒を蒸発させることにより過飽和状態を作り出す。結晶化の成否は、溶媒の種類や濃縮速度、温度によって決まる。

首尾よく結晶が得られたら、X線回折装置で回折強度を求める。PILATUS 200Kに代表される半導体検出器では、通常数分から数時間程度で測定が完了する。PILATUS 200Kは読み出しが7 msと早くシャッターレス測定が可能のため、高精度なデータを短時間で取得することができる。

電子密度を計算するためには、初期位相を推定することが必須である。低分子構造解析では直接法を用いる。直接法の理論は難解だが、プログラム化されているため、実行は極めて簡単だ。反射強度と結晶学的な情報を与えるだけで、初期位相が算出される。多彩なプログラムが出回っており、無料もしくは比較的低価格で入手可能となっている。位相が決まると電子密度を計算することができる。直接法では規格化構造因子Eを用いるため、終了時にはE-mapのピーク位置のり

ストが出力される。E-map のピーク位置に原子が存在するとして、分子モデルを構築する。一般的には初期位相は精度が低いので、最初は金属原子や5員環や6員環など、実効的な電子密度が高い部分が観測される。モデルの一部が観測されない場合でも、そのまま最小二乗法を実行し、残差 $|F_o - F_c|$ を係数とする差フーリエ変換を実行する。この時点ではすでに初期位相よりも位相の精度が上がっているため、初期段階では観測できなかった残りの構造が現れる。単結晶 X 線構造解析の各段階は、より良い位相、真の位相に近づくこととするステップであると見ることもできる。

3.2. タンパク質構造解析

タンパク質構造解析でも、構造解析の基本的な段階は低分子構造解析と同一である。ただし各段階の実行方法が大きく異なっている。

3.2.1. 試料の取得

低分子は主に化学合成で作製するが、タンパク質は遺伝子工学的手法を用いて大量発現を行い、固定化金属イオンアフィニティークロマトグラフィーや、抗原抗体反応を利用したアフィニティークロマトグラフィー、分子の大きさで分離するゲル濾過クロマトグラフィーなどで精製する。これらの液体クロマトグラフィーで分離精製しやすいように、あらかじめ別のタンパク質をつなげたキメラタンパク質や、タンパク質の N 末端や C 末端にヒスチジン鎖のタグを付けたタンパク質として発現させることが主流となっている。大量発現を実行するのに必要な宿主も大腸菌や昆虫細胞、人の細胞など、目的や状況に応じて使い分ける。

3.2.2. 結晶化

タンパク質の結晶化については古典的な方法から、

電子素子や磁場を利用した最新手法までさまざまな方法が考案されている。結晶化の手法そのものについては、十分すぎるほどの研究成果が報告されている。しかしながらタンパク質の結晶化で最も重要なのは、結晶化そのものの手法ではなく、結晶化しやすいタンパク質を作製することである。タンパク質の大量発現を行う段階で、結晶化に重要であると推定される部位の改変や、ドメイン間の相対的な動きを制限するための改変を施したさまざまなコンストラクトを多数作成し、結晶化に最も適したタンパク質を選定する。

タンパク質の結晶化の主流は、蒸気拡散法と呼ばれる方法である (図 3(a))。タンパク質溶液を溶解している緩衝液と同じ緩衝液に、沈澱剤を入れたものを用意する。結晶化容器に十分な量のリザーブ溶液を入れます。タンパク質溶液に等量のリザーブ溶液を加え、逆さまにしてふたをする。結晶化ドロップの沈澱剤はおおよそ半分の濃度になっているため、リザーブ溶液と同様な濃度になるまで水が移動する。この過程で濃縮がおこり結晶が得られる。

現在ではデータ測定や位相決定に必要なハードウェアやソフトウェアが進歩しており、低分子構造解析と同様、良い結晶さえ得られれば以前とは比較にならないほど高い確率で、最終構造にたどり着くことができる。従って、タンパク質構造解析の成否の大半は、生化学的な実験如何にかかっていると言っても過言ではない。

3.2.3. データ測定

低分子構造解析では、結晶溶媒の離脱を防ぐ、原子の熱振動を抑える、結晶へのダメージを抑えるなどの理由で、低温で測定することも多くなっている。一方

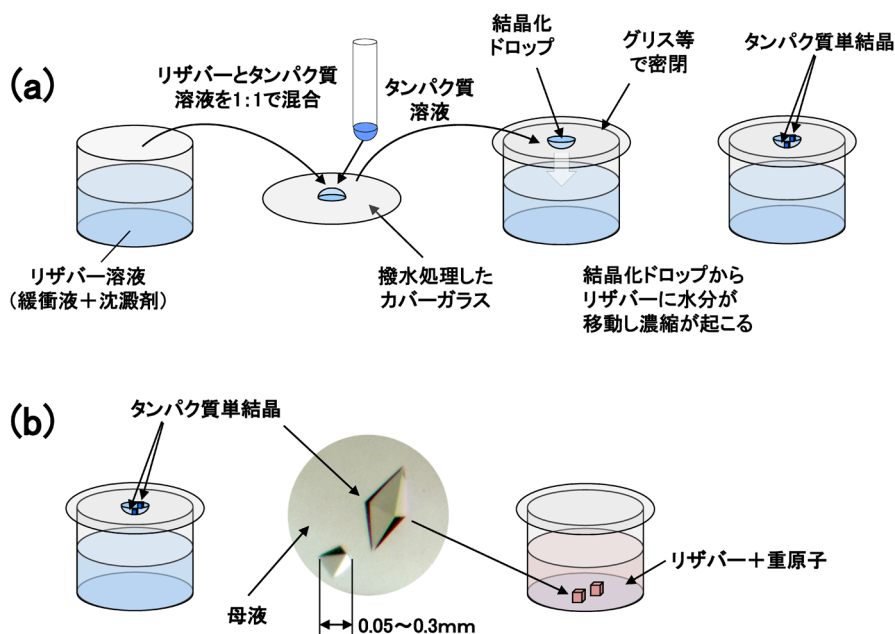


図3. (a) ハンギングドロップ蒸気拡散法による結晶化のセットアップ. (b) 重原子のソーキング方法.

タンパク質構造解析では、ほとんど例外なく常に 100 K 程度の低温で測定する。タンパク質結晶は 50% 程度の結晶水を含み、加えて X 線によるダメージをより受けやすいため、低温測定を行うには、抗凍結剤の選択が必要となる。問題なのは結晶化の母液である。母液は水を主成分とした緩衝液であるため、極めて急速な凍結を行わない限りそのまま凍結させると氷が生成し、データ測定の大きな障害となる。当初はタンパク質結晶の中にある水が凍結して結晶が生成するために、凍結時に結晶がダメージを受けると考えられていたが、結晶中の水はタンパク質自身により安定化されているため、凍結による氷の生成は限られていると考えられる。

通常の凍結は低温装置の窒素ガス流で行う。低温ガス流をカード等でブロックし、結晶をセットしたのちカードを素早く抜き去ることにより、凍結速度を上げる。この方法は急速冷凍法と呼ばれている。結晶をすくったループを、そのままデューワーに入れた液体窒素などの低温の液体に素早く入れる方法もある。気体と比較して液体媒体は熱伝導が良いとされているが、注意深く凍結しないとかえって逆効果になることがある。失敗の原因の多くは、液体窒素の上面に存在する窒素ガス層である。ガス層は比較的温度勾配が緩やかであるため、結晶の急冷の障害となる。これを避けるためには、液体窒素をデューワーの縁ぎりぎりまで満たす、もしくは何らかの方法でガス層を吹き飛ばすなどが考えられる⁽¹⁾。

3.2.4. 位相決定

低分子構造解析ではほとんど例外なく、計算により初期位相を決定する。一方タンパク質構造解析では、

直接法による位相決定が困難であるため、実験的に位相を決定する。代表的な手法が重原子多重同型置換法 (Multiple Isomorphous Replacement method: MIR 法) である。MIR 法は重原子を導入することにより生じた、わずかな位相のずれを利用して元のタンパク質 (Native) の位相を導出する方法である。理論上は Native 結晶と最低 2 種類の異なる重原子誘導体結晶の、合計 3 セットのデータがあると Native 結晶の位相を計算することができる。

重原子誘導体結晶の作成方法にはいくつかあるが、古典的な方法は Native 結晶を重原子試薬を含む溶液に浸す方法で、ソーキング法と呼ばれている (図 3 (b))。一見簡単そうだが、重原子の種類、濃度、ソーキング時間など、変更すべきパラメータは多数ある。条件が整うと結晶性が保存されたまま、重原子がタンパク質分子の特定の場所に結合する。重原子が導入されたか否かは、実際にデータを測定して評価するほかなく、したがって 1 つの誘導体の取得にも多大な労力が必要となる。

図 4 に MIR 法による位相決定の原理を示す。構造因子は波として考えることができるので、複素平面上で考えると便利である。青色で表したのが Native 結晶からの情報である。Native 結晶の反射強度はデータ測定により測定することができるが、位相は測定することができない。この状況は本来ベクトルである構造因子の矢印の長さは分かるが、その終点の位置が分からないということに相当する。ここで重原子誘導体の出番となる。

まずは重原子のみの構造因子を求める必要がある。重原子誘導体の同型性が十分高い場合、重原子誘導体

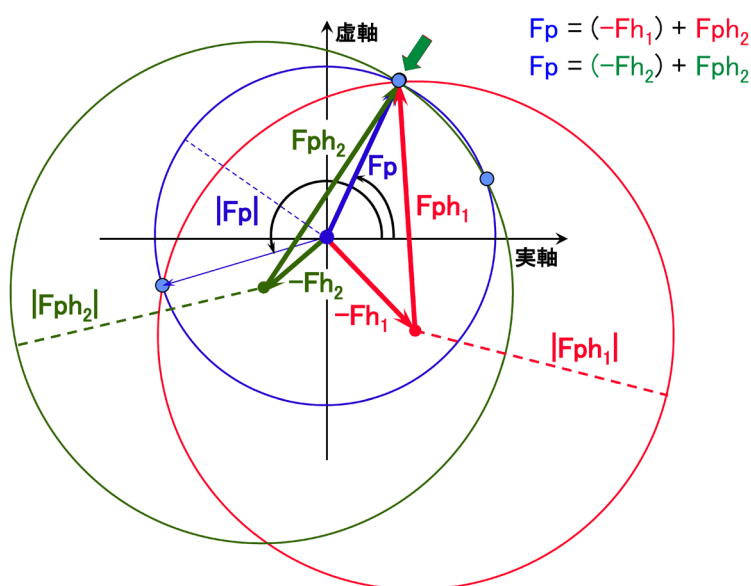


図 4. 重原子多重同型置換法による位相決定の原理。青・赤・緑はそれぞれ Native 結晶、1 つ目の重原子誘導体、2 つ目の重原子誘導体から得られる情報。

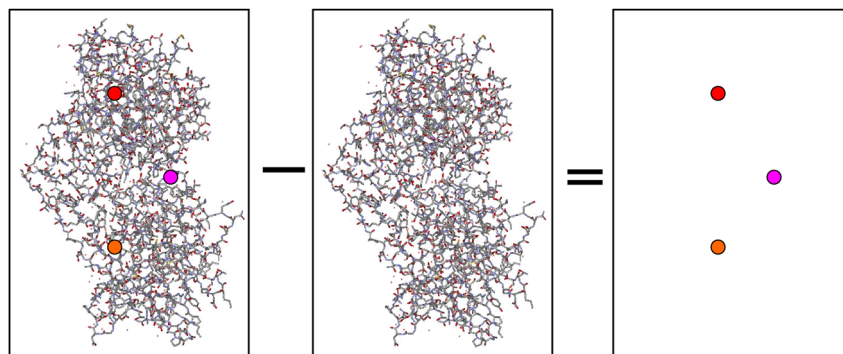


図5. 直接法による重原子のみの構造の決定原理. 重原子誘導体からNativeを差し引くと, 重原子のみの構造が残る. 重原子のみの構造は比較的単純であり, 直接法により重原子位置を特定することができる.

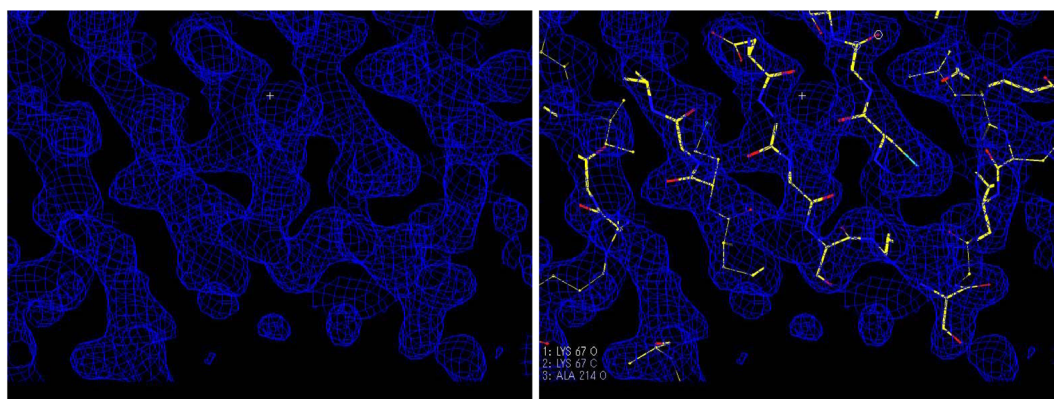


図6. 左: 溶媒の平滑化後の電子密度図. 右: モデルの自動構築プログラムを用いて作成した分子モデルと電子密度.

結晶の反射強度 $|F_{ph}|$ から Native 結晶の反射強度 $|F_p|$ を引いたものは, 重原子のみに起因する強度さと考えることができる. 実空間では重原子のみの構造を決定すればよいことを意味する (図5). 重原子のみの構造は単純であり, これであれば何とかして構造を決めることができそう. 以前はパターンソン法により重原子の位置を決定していたが, 現在は直接法を用いるのが一般的になっている.

重原子の位置が分かると, 重原子分の構造因子を計算することができる. 構造因子はベクトルとしてあらわされるので, 図4の Fh_1 のように書き込むことができる. ただし重原子誘導体の重原子以外の部分については, その反射強度しか分からないので, 重原子の構造因子の終点を中心とする構造因子の振幅を半径とする円を描く (図4の赤円). 誘導体とNative結晶に由来する赤円と青円の交点は, 重原子誘導体とNative結晶, 重原子のみの構造因子のベクトル演算が満たされている. しかしながら交点は2つ存在あり, 位相角の可能性は2つあることになる. そこで2つ目の重原子誘導体を用意し, 同様な作業を繰り返す. 2つめの誘導体についても, 重原子の位置を決め重原子分の構造因子を計算する (図4の Fh_2). Fh_2 の終点を中心とする, 重原子誘導体の振幅を半径とする円を描くと, 今

回も2つの交点ができるが, 1つめの誘導体で得られた2つの交点と矛盾しないのは1つの交点のみとなる (図4の太い矢印). これで2つの位相角の可能性が1つに決定できたことになり, 従ってタンパク質結晶の位相が決まったことになる.

3.2.5. 分子モデルの構築

位相を計算することができると, 電子密度を計算することができる. 初期位相で計算された電子密度は, マイナスの領域を持つなど現実にはありえない電子密度マップとなっている. このような不合理を除くために, 溶媒領域の平滑化という作業を行う. 溶媒領域の平滑化により実空間で修正した電子密度マップをフーリエ変換することにより, 位相を改良することができる.

図6に示す電子密度マップは, 溶媒領域の平滑化を行った後の電子密度である. このような質の高い電子密度マップが得られたら, 電子密度の形を見ながらアミノ酸の鎖を当てはめていく. トリプトファンやチロシン, フェニルアラニンなどの特徴的な形状を持つアミノ酸側鎖は良い手掛かりとなる. 新規タンパク質といえども構造解析開始以前にアミノ酸配列は決定しているのが通例であるため, 一部確実に合致する部位があれば, 残りはアミノ酸配列にしたがって当てはめて

表 1. 低分子構造解析とタンパク質構造解析のまとめ.

	低分子構造解析	タンパク質構造解析
試料の作製	化学合成.	遺伝子工学により生化学的に発現・精製.
結晶化	バッチ法. 有機溶媒に溶解して濃縮する方法や, 貧溶媒を拡散させる.	蒸気拡散法. 個々のタンパク質に最適な pH を持つ緩衝液に溶解しているタンパク質に, 沈殿剤を加え濃縮する.
データ収集	結晶の吸収や絶対構造の決定の有無により Mo/Cu を使い分ける. 測定の際の振動角は, 狭い場合も広い場合もある.	絶対構造の決定不要. 主に軽元素からなる巨大分子であり, 格子が大きいため, 実験室では通常 Cu を用いる. 振動角は通常 1°以下.
位相決定	統計的な手法を用い, コンピュータで計算. 直接法.	一般的には直接法が使えないため, 実験的に決定. MIR 法など.
モデルの構築	予想構造や化学的な知識, 結合距離や電子密度の高さを頼りに構築. 形状や原子種が多彩.	構成要素が 20 種類のアミノ酸に限られるため, 基本的にはアミノ酸鎖を当てはめる. 自動構築プログラム多数.
構造の精密化	現在は専ら最小二乗法による.	SA 法等が主流. 最小二乗法も使われる. Refmac, Shelxl など.
結果の評価・利用	分子構造や結晶構造, 絶対構造の利用など.	専ら分子構造を利用し, 生体反応や創薬に利用する.

いくことになる. 現在ではモデルの構築ソフトウェアが複数出回っており (Buccaneer⁽²⁾, ARP/wARP⁽³⁾, Solve/Resolve^{(4),(5)}), アミノ酸のシークエンスを与えると自動的に分子モデルを構築してくれる. 分子モデルの構築プログラムの出現により, タンパク質構造解析に必要な時間は大幅に短縮された. 構成単位が限られているタンパク質は, 自動化に適しているのかもしれない.

3.2.6. 構造の精密化

分子モデルが完成すると, 次は精密化の段階となる. タンパク質構造解析に特徴的な精密化のアルゴリズムは, Simulated Annealing 法 (SA 法) だ. SA 法はエネルギーの最小化を行った後, 仮想的に分子の温度を上昇させ, 熱浴を切り離し徐々に冷却する. 最小二乗法と比較して収束半径が大きく, より高いエネルギー障壁を超えて正しい構造に収束するとされている. SA 法の代表的なプログラムは CNS/CNX である. タンパク質構造の精密化には, 最小二乗法も利用されている. タンパク質構造解析の基本ソフトウェアである CCP4 には, Refmac⁽⁶⁾ という最小二乗法のソフトウェアが搭載されている. また低分子構造解析の精密化プログラムの標準となっている SHELXL にも, タンパク質構造の精密化の機能が備わっている. SA 法でも最小二乗法でも, タンパク質構造解析では反射数すなわちデータ点数とパラメータ数の比が低いので, アミノ酸の構造が崩れないような束縛をかけて精密化

を実行する.

タンパク質構造解析では, タンパク質の一部として振る舞う水分子が散見される. また分子表面で水の水素結合ネットワークを作って, 可溶化に貢献する水分子も存在する. 通常第 1 層目の水は局在化しており, これらも分子モデルに取り込んでいく. 最終的な結晶学的 R 値は一般的に 10~30% 程度になる.

4. まとめ

表 1 にタンパク質構造解析と低分子構造解析の各段階をまとめる. 手順は同一だが, 内容が著しく異なるところもある. 低分子構造解析の困難さは, 結晶学的な問題に起因している場合がほとんどである. 最も頻度が高いのは, 双晶と気づかずに解析している場合だ. 一方タンパク質構造解析の難しさは, 主に実験の困難さである.

参考文献

- (1) M. Warkentin, V. Berejnov, N. S. Husseini and R. E. Thorne: *J. Appl. Crystallogr.*, **39** (2006), 805–811.
- (2) M. D. Winn *et al.*: *Acta. Cryst.*, **D67** (2011), 235–242.
- (3) G. Langer, S. X. Cohen, V. S. Lamzin and A. Perrakis: *Nature Protocols*, **3** (2008), 1171–1179.
- (4) T. C. Terwilliger: *Acta Cryst.*, **D59** (2003), 38–44.
- (5) T. C. Terwilliger: *Acta Cryst.*, **D59** (2003), 45–49.
- (6) G. N. Murshudov, A. A. Vagin and E. J. Dodson: *Acta Cryst.*, **D53** (1997), 240–255.