タンパク質結晶解析のためのソフトウェア紹介

理化学研究所播磨研究所 熊坂 崇

Takashi KUMASAKA: Introduction to Software for Protein Crystallography

Crystal structure determination of proteins is not so straight forward, since the direct method for small molecules would not be applied to them. To extract the phase information from experimental measurement, methodological and technical developments have been proposed and examined. The qualification of crystals is also extended by the recent development of related techniques such as molecular biology, computation and synchrotron radiation. We summarize here the recent trend of computer software for protein crystallography.

1. はじめに

タンパク質結晶の構造解析は、結晶の非対称単位に含まれる原子数の多さと回折分解能の低さが主な原因となって、位相問題を解決する直接法の適用が難しい。この位相問題を実験的に解決しなくてはならないという現実を前に、重原子同型置換法をはじめ多くの試みがなされて来た。また、分子生物学、計算機や放射光など関連分野の進歩に伴って、解析が効率化されるとともに、適用できる結晶の質的制約が少なくなった。ここでは、特にソフトウェアに関する最近の話題を整理してみた。

2. 回折データ処理

現在、タンパク質結晶の回折強度データはほとんど例外なく、二次元検出器を利用した振動写真法によって得た回折像を処理して得られる。今この処理に良く使われているのは、米国 HKL 社の製品である DENZO¹⁾とその開発版である HKL2000 であろうか。1 枚の回折像からでも、高い確率で結晶格子や方位についてのパラメータを決定することができる。また、たくさんのパラメータ(結晶や検出器の方位、カメラ距離、波長等)を精密化するために、さまざまな数値解析技術を詰め込み、頑健かつ容易に行えるようにした。最新版では GUI の導入や対応ハードウェアの拡大などのほか、高速な検出器の開発によって可能になったファインスライス(微小振動角振動写真法)にも対応している。ひとつ難点を言えば、国内での導入コストがやや高いことだろうか。続いて良く使われているのは MOSFLM/DPS²⁾であろう。MOSFLM は英国 MRC で開発が進められていたが、これを中心に、使いやすいグラフィカルユーザーインターフェイス(GUI)

と自動指数決定などの新たなプログラムを加えて統合したものが DPS である。こちらはソースコードが入手でき、非商用での利用であればコストはかからない。その他、国産で以前から実績のある PROCESS³⁾や 3D プロファイルフィッティングをいち早く導入した XDS⁴⁾ 自動処理を念頭に置いて開発が進められている d*TREK⁵⁾ XGEN などがある。

また、白色光を利用した Laue 写真処理用のソフトとしては、LaueGen、LaueView⁶⁾ があげられる。後者で適用されている反射形状の関数近似やその関数によるオーバーラップ反射の分離などの技術は振動写真の処理ソフトウェアにも応用されていくかもしれない。

3. 位相決定 I — 同型置換法

多重同型置換法(Multiple Isomorphous Replacement, MIR)や多波長異常分散法(Multiwavelength Anomalous Diffraction, MAD)など、同型結晶からの回折強度変化に基づく位相決定法において、重原子の部分構造を決定するのに、一般に差パターン図のピーク位置から逆算する方法が取られる。これを対話的に計算で行うのが RSPS (CCP4⁷⁾)である。このほか、低分子の構造決定で使われている最小関数法や直接法なども原理的に適用可能で、実際 SHELXS⁸⁾はこの用途にも用いられる。また、難解な構造を決定するために実空間と逆空間でそれぞれ最適化を行いながら解を探していくソフト SnB⁹⁾も多数の重原子を含む場合に威力を発揮している。これに関連して、重原子を含む高分解能データについては、そのパターンピークのサーチによって得られる部分構造からスタートして、位相を決定しようという試みもなされている(SHELXS, ARP/wARP¹⁰⁾)。

こうして得た重原子位置の精密化には、以前は最小二乗法をベースにしたソフトウェアが使われていた。しかし、最近ではほとんどが最尤法 (Maximum Likelihood, ML) を採用し、質の悪いデータを加えても、それに引き擦られずにより確度の高い位相を与えるようになった。MLPHARE (CCP4) は ML による最適化を行うようにした最初のソフトである。最近最も良く使われているのが SHARP¹¹⁾ で、完全に ML をサポートしてより良い位相情報を与えるものとして評価を受けている。また、Web ベースの GUI を用意してユーザーにも使いやすいものにした。このほか、位相改良において草分け的な役割を果たした Phases¹²⁾ パッケージをはじめ、多くのソフトウェアが位相計算のために作成、利用されている。

これらの手法を組み合わせ、自動で部分構造決定と位相精密化を行おうという試みもなされている。SOLVE¹³⁾ は逐次差パターンおよびフーリエ合成を行って部分構造を決定し、これを位相精密化に組み入れ、独自のスコア関数を導入して最適解を自動で探そうというソフトである。入力するパラメータ数もそれほど多くなく、経験の少ないユーザーにも使いやすい。また、Chart¹⁴⁾ や Elves¹⁵⁾ といった、既存のソフトを流用したエキスパートシステムも開発されている。

4. 位相決定 II — 分子置換法

分子置換法 (Molecular Replacement, MR) は類似の既知構造から位相を得ようというものである。モデル分子を適当な方位と位置で結晶格子中に置き、それによって計算で得られる構造因子と実験データの一致をみる。一つの分子を適切に格子中に置くためには、6つのパラメータが必要である。この計算量は膨大になるため、このパラメータを回転成分3つと並進成分3つに分け、それぞれ別ステップで求めるのが、依然として現在の主流になっている。そのなかで、現在良く使われているのは CCP4 にも含まれている AMoRe¹⁶⁾ と X-PLOR/CNS¹⁷⁾ であろう。前者は分子構造因子の概念を導入し、計算時間を短縮しているほか、複数の相関関数をサポートしている。適当なスクリプトを用意すれば、分子モデルの方位決定とその並進成分の決定が半自動で効率良く行える。後者は実空間でのサーチを行うことに特徴があり、PC-Refinement と呼ばれる分子内可動成分のパターン相関精密化により並進サーチをより正確に行える。

計算機の性能が向上してきていることや計算アルゴリズムの進展もあって、6つのパラメータを一気に決めてしまおうという試みも始まっている。epmr¹⁸⁾ はその一例で、進化的アルゴリズムに基づいて解を探すソフトである。入力ファイルが3つしかなく、コマンドライン一発で実行できるので、操作が非常に単純である。

しかし、いずれにしても、非対称単位中に複数の分子

を含む場合、いわゆる「多体問題」の解決は現在この手法の課題の一つになっている。自己回転関数や自己パターン関数から分子の相対配置を決定する試みや、6次元サーチ法の改良などが行われている。

5. 位相改良

上で述べた方法で位相が得られて、すぐに分子モデルを構築することができればそれに越したことはないが、実際は曖昧な位相情報しか得られないことがよくある。こういった場合、あるいはそうでなくても、位相改良法を適用することでより明瞭な分子の形を得られることも多い。特にモデルを一切仮定していない同型置換法の場合、その効果が顕著な場合がある。現在用いられている手法の多くは、実空間で電子密度を操作してそれを逆空間に戻し、実験位相と結合させる手法 (電子密度修飾法) に基づいている。結晶中の溶媒領域を均質な電子密度に置き換える溶媒領域平滑化法、電子密度の強度分布を仮定するヒストグラムマッチング法、非対称単位に含まれる複数の分子間で電子密度を平均化する方法などを組み合わせて、タンパク質結晶の電子密度として統計的により正しいかたちに加工する。

これらを複数組み合わせ処理できるのが DM¹⁹⁾ (CCP4) である。Squash²⁰⁾ から派生したこのソフトは上に述べた方法のほか、電子密度骨格化や Sayer 式による逆空間での位相改良などの機能を持つ。Squash も 2次元ヒストグラムマッチングの機能などを追加して改良が行われている。SOLOMON²¹⁾ (CCP4) は溶媒の構造をより精細に引き出す工夫がなされ、SHARP のインターフェイスからも容易に利用できる。

6. 構造構築

タンパク質の結晶解析はふつう原子分解能には至らないので、分子モデルを組み立てなくてはならない。コンピュータグラフィックスを使って、電子密度にアミノ酸残基をひとつずつ置いていくこの作業は、分子サイズの増加に伴って膨大なものになる。位相決定に要する時間が十分に迅速化した今、モデル構築を高速化・自動化することが急務になっている。

O²²⁾ はモデル構築のワークベンチとして、統合された機能を持っている。初期バージョンではコマンドベースの操作が、それまで広く使われていた FRODO とその派生ソフト (TURBO-FRODO など) のユーザーにとって違和感を覚えたが、その後のメニューの充実と、なんといっても強力なデータベースからの構造構築支援ありがたい。

近年 PC とグラフィックボードの性能が高まり、また 3D ライブラリが PC で使えるようになってきて、高価なワークステーションを使わなくても、これらの操作ができるようになってきた。Xfit は XtalView²³⁾ という GUI

ベースの統合パッケージに含まれるモデル作成ソフトである。早くから PC UNIX で軽快に動き、最近のバージョンではハードウェアステレオをサポートするようになった。メニューからの直観的な操作が可能で、モデルを詳しく見たい場合にも適している。前述の O も Linux 版や NT 版が開発されている。

これらの分子モデル構築作業を支援するプログラムに、電子密度の骨格化によって構造を線で抽出する Bones がある。この流れとして、自動解釈や位相改良のソフトウェアが派生している。自動解釈は、上記の方法の他、電子密度の局所的な勾配などを指標にして稜線をたどれるようなアルゴリズムや、ヘリックスやシート構造などの二次構造をパターン認識する方法が開発されつつある。

X-POWERFIT/X-AUTOFIT/X-BUILD は米国 MSI 社がリリースしている市販の自動モデリングソフトである。96 残基からなるタンパク質の 2.4 Å の電子密度図に対する C α トレース (90 残基分) をたった 1 分で完了してしまったそうである。さらに高分解能であれば数秒で完了する能力をもっている。人手で行う構造構築にはアーティファクトが多く含まれてしまい、構造精密化にてこずることもあるため、電子密度解釈の自動化による恩恵は構造精密化の迅速化にも繋がっていくと思われる。

7. 構造精密化

構造精密化は、タンパク質結晶の分解能的制限もあって、多くの場合、立体化学的束縛あるいは制約下での最小二乗問題になる。PROLSQ²⁴⁾ は、この種の制限を加えて R 因子を最小化するプログラムの最初であった。しかし、非線形最小二乗法は局所解に陥ると、そこから脱出できず、手動でモデルの組み換えをしなくてはならない。この問題を解決するために分子動力学 (Molecular Dynamics, MD) の手法を導入したソフトウェアが X-PLOR である。現在はその後継である CNS がこの目的に最も良く使われている。CNS になってからエネルギー項の荷重の計算の自動化や Web ベースでの入力ファイルの作成などが追加され、ユーザーの使いやすさも向上してきている。

PROLSQ の方法は改良が続けられて、現在では ML による最適化を行う REFMAC²⁵⁾ (CCP4) や TNT,²⁶⁾ SHELXL などに反映されている。さらに、分解能の制限がない場合の Full Matrix 法や異方性温度因子の精密化のような低分子結晶学で使われる手法については、SHELXL がタンパク質への応用を進めている。構造精密化は、局所解トラップという問題を抱えているが、これは現状 MD だけでは解決できない。ARP/wARP は位相改良—構造構築—構造精密化を自動で行おうとする試みである (ただし構造精密化計算には REFMAC を使う)。比較的高い分解能のデータなら、電子密度解釈から精密化まで自動で行っ

てしまう能力がある。

8. 各種ツール

得られた構造がこれまでのタンパク質構造と比較して、その立体化学的なパラメータにおいて適当かどうかは、分解能の低い構造解析の場合には特に注意を払わなくてはならないポイントの一つである。この目的に限らず、分子構造の妥当性を評価検定するソフトウェアには PROCHECK²⁷⁾ (CCP4) などがある。いわゆる Ramachandran Plot 以外にも、既知構造から抽出したパラメータとの比較評価を行う。

異なる種類の分子についての座標の重ね合わせは、意外に面倒な操作が必要である。topp/top3d (CCP4) は一次構造の相同性や二次構造の空間配置などに基づいて重ね合わせを行う。

得られた構造を表示し作画するツールは、近年の PC のグラフィック性能の向上もあり、紹介しきれないほど多く開発されてきている。雑誌の表紙も飾る図の多くは、専用の作図ソフトウェアによって作られている。Molscript²⁸⁾ はこの用途に以前からよく使われている。スクリプトを通すことで図が得られるため、直観的でない面もあるが、3次元のレンダリングを行えるパッケージ Raster3D²⁹⁾ の入力ファイルも作成できる。Setor は SGI 社製ワークステーション専用であるが、メニュー形式で作画ができるソフトである。また、簡便な表示用として Windows や Mac でも使える RasMol (ProteinExplorer) などがある。

9. 統合パッケージ

なんといっても、CCP4 (Colaborative Computational Project, 4) によって作成、管理が行われているパッケージが筆頭にあげられる。非商用には無料で取得が可能で、ソースコードが公開されている。また、事実上、回折強度データ、電子密度図などの標準ファイルフォーマットを規定しているといってもいい。制御用のスクリプトファイルを作成してバッチ形式で使用するのが基本であるが、最近 Tcl/Tk ベースの GUI (CCP4I) が付け加えられ、使いやすさの向上も目指している。

一方、CNS は解析用インタープリタ言語としての機能を拡張して、現在では MIR/MAD などによる位相計算もサポートするようになった。MD 計算の機能が強調されるが、NMR による解析の機能も持ち、生体高分子構造解析の統合パッケージとして CCP4 と双壁をなしている。

XtalView は GUI ベースの統合環境として異彩を放っている。上記の 2 つにはない構造構築用のグラフィック機能を有しているのが特徴である。このほかにも、Phases, Protein, MAIN などのパッケージが開発されている。

商用のパッケージは、タンパク質のモデリングや理論計算の方から発展してきたものである。タンパク質分子の構造予測や、分子力学や分子動力学による構造最適化

図1 タンパク質構造解析ソフトウェア概観。(Overview of protein crystallography softwares.). ただし、配布形態はアカデミックの場合を示した。

などがその主たる目的であった。これらは、近年注目を浴びつつある構造に基づく新薬開発とも密接な関係を持っている。米国 MSI 社はこの種のソフトを開発している企業である。数年前の合併によって生じた同機能のソフトを二重に持つ問題は、徐々に解消されているようだ。最近では Quanta/CHARMM を中心として X 線解析部分を統合したソフトウェア Xsight をリリースしている。これらは高価ではあるが、電子密度の自動解釈の機能を新たに加えるなど、タンパク質分子を計算で取り扱う巨大なパッケージとして充実してきている。これは、近年現実味を帯びてきたハイスループット、構造ゲノミクスなどの大量構造解析での需要を見込んだものだろう。

本文中で紹介したソフトウェアと解析の各段階との対応を図1にまとめた。ソフトウェアに関する情報の多くは IUCr のソフトウェアデータベース SINCRIS (<http://www.jp.iucr.org/sincris-top/>) で検索可能なほか、CCP4 のホームページ (<http://www.dl.ac.uk/CCP/CCP4/>) や PDB のリンクページ (<http://pdb.osaka-u.ac.jp/pdb/software-list.html>) などにまとめられている。ウェブサーチエンジンなども活用されたい。また、著者の能力から、多くの有用なソフトウェアを紹介できなかったことをお詫いしたい。

謝 辞

本文執筆の機会を与えていただき、拙文を検討していただいた姫路工業大学森本幸生先生に感謝致します。また、ここに紹介したものをはじめとするソフトウェアを開発した方々に敬意を表するとともに、ソフトの利用を含めた解析において多くの議論を交わしてきた方々に、この場を借りて感謝したいと思います。

文 献

- 1) Z. Otwinowski and W. Minor: *Methods Enzymol.* **276**, 307 (1999).
- 2) M.G. Rossmann and C. G. van Beek: *Acta Cryst.* **D55**, 1631 (1999).
- 3) T. Higashi: *J. Appl. Cryst.* **23**, 253 (1990).
- 4) W. Kabsch: *J. Appl. Cryst.* **26**, 795 (1993).
- 5) J. Pflugrath: *Methods Enzymol.* **276**, 286 (1999).
- 6) I. J. Clifton, E. M. H. Duke, S. Wakatsuki and Z. Ren: *Methods Enzymol.* **277**, 448 (1999).
- 7) Collaborative Computational Project, Number 4: *Acta Cryst.* **D50**, 760 (1994).
- 8) G. M. Sheldrick: *Methods Enzymol.* **276**, 628 (1999).
- 9) C. M. Weeks and R. Miller: *J. Appl. Cryst.* **32**, 120 (1999).
- 10) A. Perrakis, R. Morris and V. S. Lamzin: *Nature Struct. Biol.* **6**, 458 (1999).
- 11) E. de La Fortelle and G. Bricogne: *Methods Enzymol.* **276**, 472 (1999).
- 12) W. Furey and S. Swaminathan: *Methods Enzymol.* **277**, 590 (1999).
- 13) T. C. Terwilliger and J. Brennden: *J. Appl. Cryst.* **55**, 849 (1999).
- 14) P. Emsley: <http://crick.chem.gla.ac.uk/~paule/chart/>
- 15) J. Holton: <http://ucxray.berkeley.edu/~jamesh/elves/>
- 16) J. Navaza and P. Saludjian: *Methods Enzymol.* **276**, 581 (1999).
- 17) A. T. Brunger, P. D. Adams *et al.*: *Acta Cryst.* **D54**, 905 (1998).
- 18) C. R. Kissinger, D. K. Gehlhaar, D. B. Fogel: *Acta Cryst.* **D55**, 484 (1999).
- 19) K. D. Cowtan and P. Main: *Acta Cryst.* **D52**, 43 (1996).
- 20) Y. P. Nieh and K. Y. T. Zhang: *Acta Cryst.* **D55**, 1893 (1999).
- 21) J. P. Abrahams and A. G. W. Leslie: *Acta Cryst.* **D52**, 30 (1996).
- 22) T. A. Jones, J. Y. Zou, S. W. Cowan and M. Kjeldgaard: *Acta Cryst.* **A47**, 110 (1991).
- 23) D. E. McRee: *Practical Protein Crystallography* (2nd ed.), Academic Press (1999).
- 24) W. A. Hendrickson: *Methods Enzymol.* **115**, 252 (1985).
- 25) G. N. Murshudov, A. A. Vagin and E. J. Dodson: *Acta Cryst.* **D53**, 240 (1997).
- 26) D. E. Tronrud: *Methods Enzymol.* **277**, 306 (1999).
- 27) R. A. Laskowski, M. W. MacArthur, D. S. Moss and J. M. Thornton: *J. Appl. Cryst.* **26**, 283 (1993).
- 28) P. Kraulis: *J. Appl. Cryst.* **24**, 946 (1991).
- 29) E. A. Merritt and M. E. P. Murphy: *Acta Cryst.* **D50**, 869 (1994).

プロフィール

熊坂 崇 Takashi KUMASAKA

理化学研究所播磨研究所構造生物物理研究室
RIKEN Harima Institute

〒679-5148 兵庫県佐用郡三日月町光都 1-1-1

1-1-1 Kouto, Mikazuki, Sayo, Hyogo 679-5148, Japan

Tel. 0791-58-2815, Fax. 0791-58-2816

e-mail: kumasaka@postman.riken.go.jp

最終学歴：東京工業大学大学院生命理工学研究科
博士後期課程

専門分野：蛋白質結晶学

現在の研究テーマ：放射光を利用した蛋白質結晶
構造解析