構造生物　Vol.10 No.2
2004 年 11 月発行

自動精密化プログラム Lafire

北海道大学大学院理学研究科

姚　閔

An automatic refinement program: Lafire

Division of Biological Sciences, Graduate School of Science,
Hokkaido University

For the refinement of protein crystal structures, it is usually necessary to rebuild the model by computer graphics (fitting) intervening in rounds of the refinement. Moreover, manual building is required in the most cases for linking and extending the fragments of initial model, although automatic model building programs such as Solve/Resolve or ARP/wARP are now available for constructing the initial model of the protein structure. All these processes (linking, extending and fitting) are time-consuming, and the whole refinement process takes a long time (several months).

For realizing the manual-intervention-free refinement, we have developed a Local-correlation-coefficient-based Automatic FItting for REfinement program (Lafire). This system consists of evaluation of existing model, automatic model modification, partial model building, a graphic monitor system and an interface with existing refinement programs. In conjunction with the refinement programs such as CNS or REFMAC5, the system is already in the state that structure refinement without manual intervention can actually be realized in a few hours or days.

１．はじめに

遺伝子工学の発展，シンクロトロン放射光利用の普及，多波長異常分散法の開発などにより，蛋白質の構造解析は，近年，大幅なスピードアップが行われてきた．しかし, ゲノム解析プロジェクトの進展によるタンパク質の１次配列情報の蓄積に比べると，構造解析のスピードは遅く，さらに 1 桁以上アップすることは急務である．タンパク質の構造解析には, 全過程を通してさまざまな困難なプロセスがあるけれども, これまでに, 解析の様々なプロセスについて試行錯誤を軽減する自動化ソフトウェアが開発されてきている．例えば，データの処理段階の HKL2000，重原子サーチから位相の計算・改良，さらにモデル構築までの自動的に行われる Solve/Resolve，位相の改良から自動モデルの構築，部分的に精密化を含めた ARP/wARP などがよく使われている優れたソフトである．これらのソフトを使うことで，構造解析にかかる時間は

大幅に短縮されてきた．しかし，構造解析の最終ステップである構造の精密化は自動的に行われるプログラムがまだない状態で，最も時間を要し，且つ手作業で行うため，行う人の熟練度によって結果が変わる過程である（図 1）．私たちは構造解析をスピードアップするために，まず，この構造解析計算の全過程で人の手作業を最も頻繁に必要とする構造精密化の自動化に着手した．

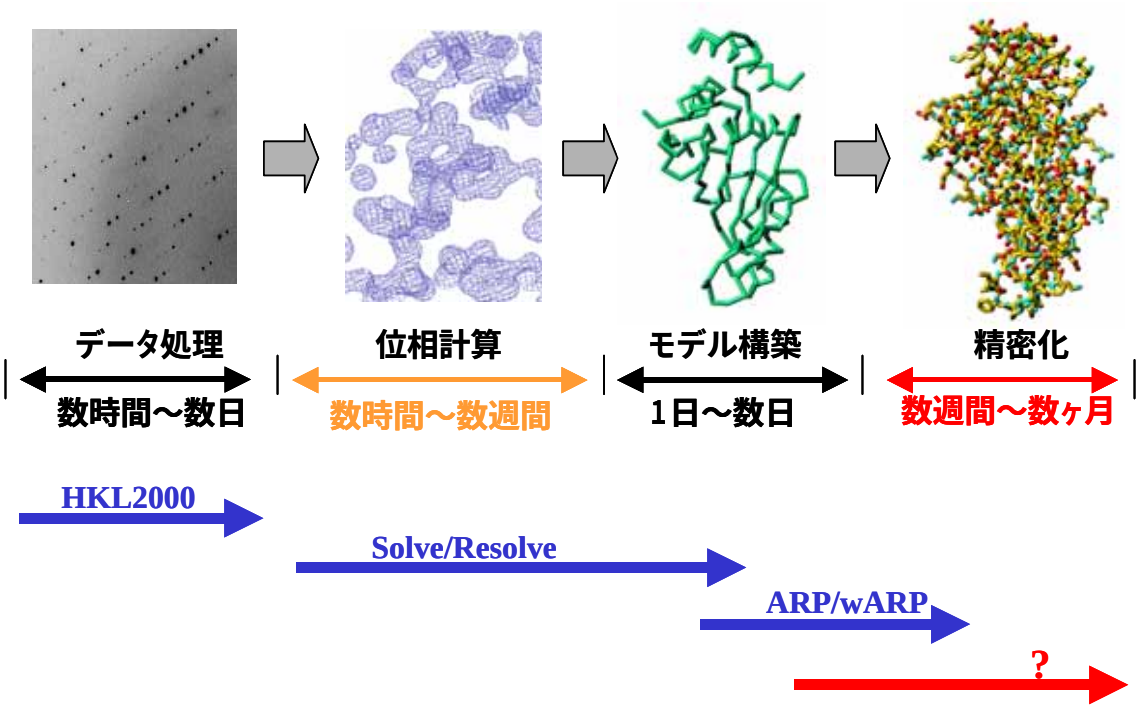


図１．構造解析の流れと各ステップに要する時間
HKL2000，Solve/Resolve，ARP/wARP は，汎用の自動化プログラム．
精密化ステップには，自動化したプログラムがない状態であった．

２．構造解析の精密化

X線結晶構造解析では,「位相」は直接測定されるのではなく，重原子同形置換法や多波長異常分散法，あるいは分子置換法などによって間接的に求められる．そのような手段によって求められた位相には，実験及び計算上の様々な誤差が含まれており，計算した電子密度図は，通常，モデル構築を行えるほど良質なものではない．そのため溶媒平滑化や分子平均化などの位相の改良を行う手法が開発されてきた．しかし，そのような努力により分子モデルが構築できるようになったとしても，位相の誤差やデータの分解能の限界などにより，得られた電子密度は不十分なものであり，初期モデルには大きな誤差が含まれている．したがって，初期モデルから計算した構造因子の振幅F_{cal}と測定した構造因子の振幅F_{obs}が一致するようにモデルを精密化しなければならない．

蛋白質構造の精密化は大量のデータを用いて多くのパラメータを精密化するので，計算に膨大な時間のかかる過程である．また回折データと精密化のパラメータの比が低いため，精密化の収斂範囲が狭い．精密化の収斂範囲を広げるために，様々な優れたアルゴリズム（条件付き最小二乗法，エネルギー最小法，分子動力学法，最尤法など）とプログラム（SHELX, PROLSQ, TNT, X-PLOR/CNS, REFMAC など）が開発されてきた．また，90 年代からのコンピュータの発展と共に，精密化の 1 サイクル（原子座標，温度因子の精密化を含む大きな 1 サイクル）の計算時間は 1 週間から数日，さらに 1 日程度にまで短縮されてきた．しかし，初期モデルのずれが精密化収斂範囲を超えると, 計算だけでモデルを修正することが不可能になるので, 精密化の過程には，通常，精密化の計算後にコンピュータグラフィックスを利用して，原子座標を電子密度に合わせる過程，いわゆるマニュアルフィティングが必要である（図 2）．このマニュアルフィッティングの成否は人の熟練度にかなり依存する．図 3 には，我々の研究室における 2002 年度の構造解析例を示している．これらの構造解析では，長い場合２ヶ月，短くても半月間ぐらいかかっている．この中で，PH1161 と*pho*eIF-5Aを除いたタンパク質に対しては，構造解析の過程で最も時間がかかったのは精密化の段階である．この段階のマニュアルフィッティングを自動化すれば，構造解析のスピードアップと省力化ができることは間違いないであろう．このような観点から，我々は自動精密化プログラムLafire （Local-correlation-coefficient-based Automatic FItting for Refinement）の開発を進めてきた.

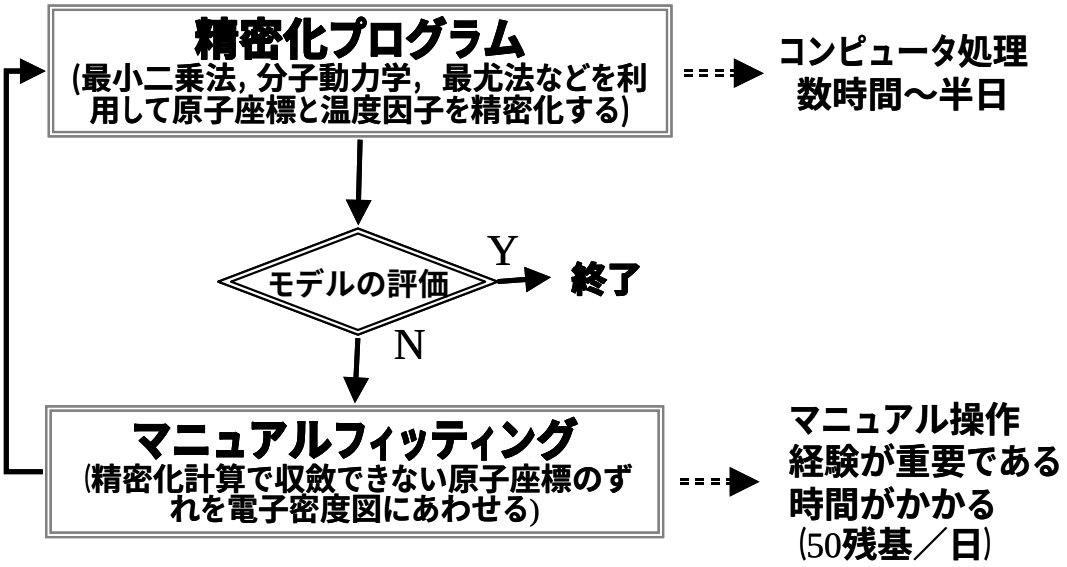


図２．従来の精密化過程

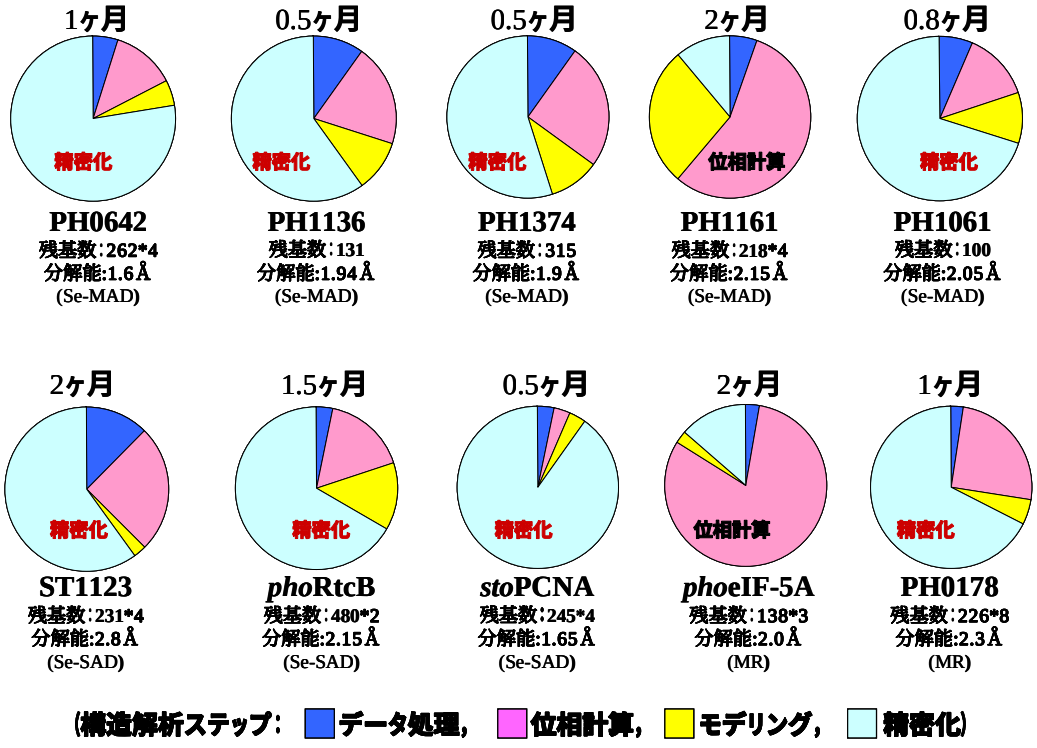


図３．実際の構造解析の全体と各ステップに要した時間の割合
ここに示したのは我々の研究室において 2002 年度に解析された構造の一部である．上の時間は解析全体にかかった時間である．

３．自動精密化プログラム Lafire

Lafire プログラムはモデルのアミノ酸残基のチェック・置換, モデルの評価･自動修正, 未構築部の自動構築，モニター及び精密化プログラムとのインタフェースから構成され，計算部分には主に C 言語を，モニター（Lafire_molview）にはグラフィクスラーブラリーOpenGL，インタフェースにシェルを使用した（図 4）．機能の増減及びプログラミングを簡単化，明瞭化するという点から，全プログラムの構造はモジュール化した．また，プログラムとして，精密化とマップの計算などには CCP4 プログラムパッケジーと CNS を利用している．

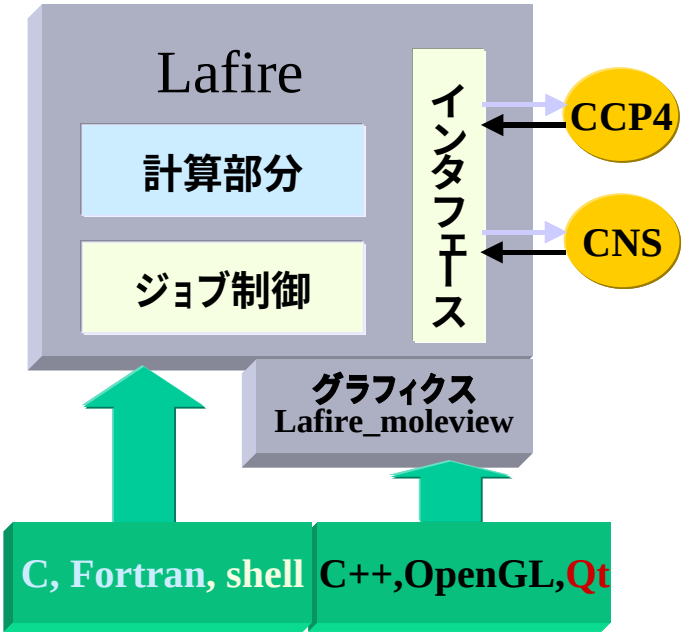


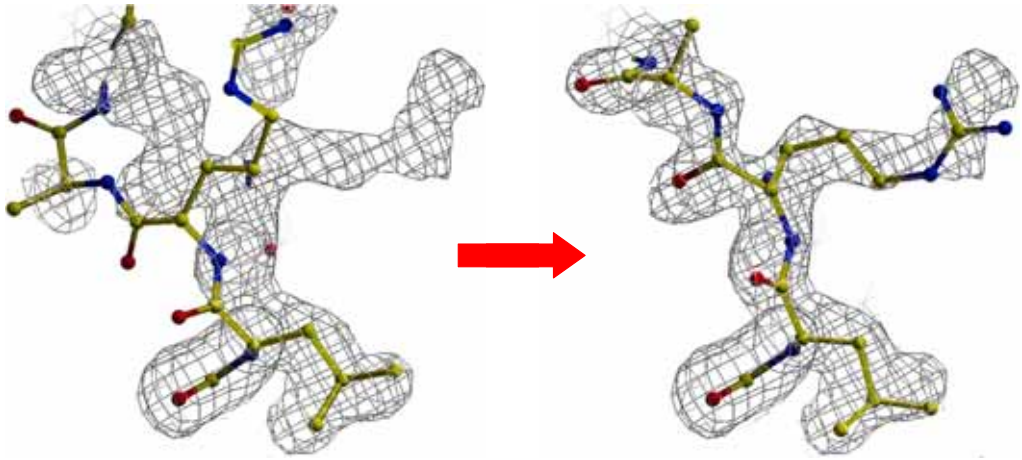
図 4．Lafire プログラムの構成

３－１．モデルの評価・自動修正

プログラムLafireの開発では，まず，自動フィッティングするためのモデルと電子密度図の一致を評価するには，電子密度マップの質を反映する重み（*g*）をつけたグループ化ローカル相関係数（*GLCC*）の数学モデル（式１）を構築し，その*GLCC*を用いて測定した構造因子の振幅（F_{obs}）とMAD位相などから計算したsigma-weight電子密度マップとモデルの一致を評価した．このローカル相関係数*GLCC*により検出した電子密度マップと一致しない残基の修正をCβの含む主鎖と側鎖を分けて行う部分修正アルゴリズム及びプログラムも完成した．図５に自動修正プログラムを

$$GLCC_i = g(\rho_{i,obs}) * \frac{< \rho_{i,obs} * \rho_{i,cal} >}{\left[(< \rho_{i,obs}^2 >) * (< \rho_{i,cal}^2 >)\right]^{1/2}} \qquad \text{式１}$$

ρ_{obs}：2Fo-Fc，Fo-Fc，或いは初期位相から計算した電子密度の観測値
ρ_{cal}：モデルから計算した電子密度の計算値
i：残基番号
< >：*i* 番目の残基に属す原子グループの電子密度に関する平均
　主鎖グループ：　N, Cα, Cβ, C, O
　側鎖グループ：　Cβを除いた側鎖の原子
g：電子密度の観測値$\rho_{i,obs}$により見積もられた電子密度図の質係数



(a) **(b)**

図５．Lafire による自動修正（タンパク質 MC１）
(a)．自動修正前.
(b)．自動修正後.

用いて修正した実例の一つを示している.

３－２．部分構造の構築

精密化の過程に上記したように自動修正が必要であるが，実際の構造解析において，「初期モデル」は，初期電子密度図の質に応じて，千差万別である．そのため，Solve/Resolve，ARP/wARP などの自動モデル構築プログラムの出力は，ほとんど全構造を含んでいる場合もあるが，時にはかなりの欠失部があり，また時には，単なるフラグメントの集まりであることもある．したがって，全自動精密化を実現するには，自動フィッティング機能だけでは不十分であり，欠失している部分の構造構築プログラムの開発も必要であった.

Lafireの未構築部の自動構築機能では，既存の自動モデリングプログラム，例えばSolve/Resolve或いはARP/wARPなどで自動的にモデリングできなかった部分，例えばループや端末部分の構造を，電子密度図とアミノ酸配列に基づいて自動的に構築する．また，分子置換法の場合には，挿入残基の構築および電子密度から大きくずれた残基の再構築も行う．さらに，Lafireは，精密化の過程において，自動修正しても，電子密度図との一致度が改善されない残基，すなわち，Cβを含む主鎖グループのローカル相関係数$GLCC_i$値（式 1）が，しきい値（しきい値は分子全体の相関係から見積もられる.）より小さいままの残基を自動的にomitしながら，精密化計算を行い，そしてomitしたモデルの 2Fo-Fc あるいは初期位相で計算した電子密

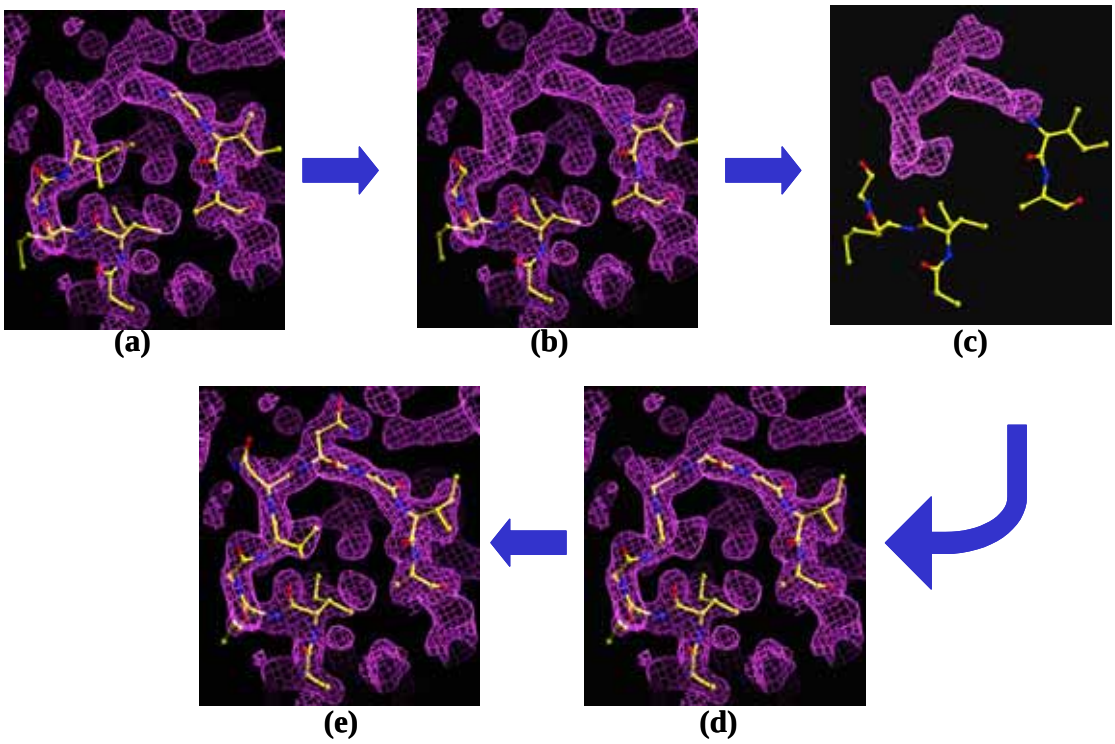


図６．欠失構造の構築
(a). 欠失している初期モデルと電子密度.
(b). 既構築部の両末端 1 残基ずつの削除.
(c). 取り出した欠失している部分の電子密度.
(d). 主鎖の構築.
(e). 側鎖の構築.

度図とomitしたモデルのFo-Fcの２つの電子密度図を組み合わせて使用するように，omitした部分を再構築していく．部分構造構築では，map-pruning法を利用した．図６では，実際の例を使って部分構造構築の流れを示している．

３－３．モニタープログラム（Lafire_molview）

自動修正プログラムを開発すると共に，精密化の過程をモニターするプログラム Lafire_molview を開発した．良好な視野を確保するために，図 7 に示しているようにモニタープログラムは N 端から 3，5 或いは 7 残基ずつを表示し，指定した原子の周りだけの電子密度図を表示する．また，表示された残基が全体構造のどの部分であるかをサブウィンドウに分かりやすいように表示している．フィッティングの様子を表すグループ化したローカル相関係数 *GLCC* 及び Ramachandran 図も表示する．

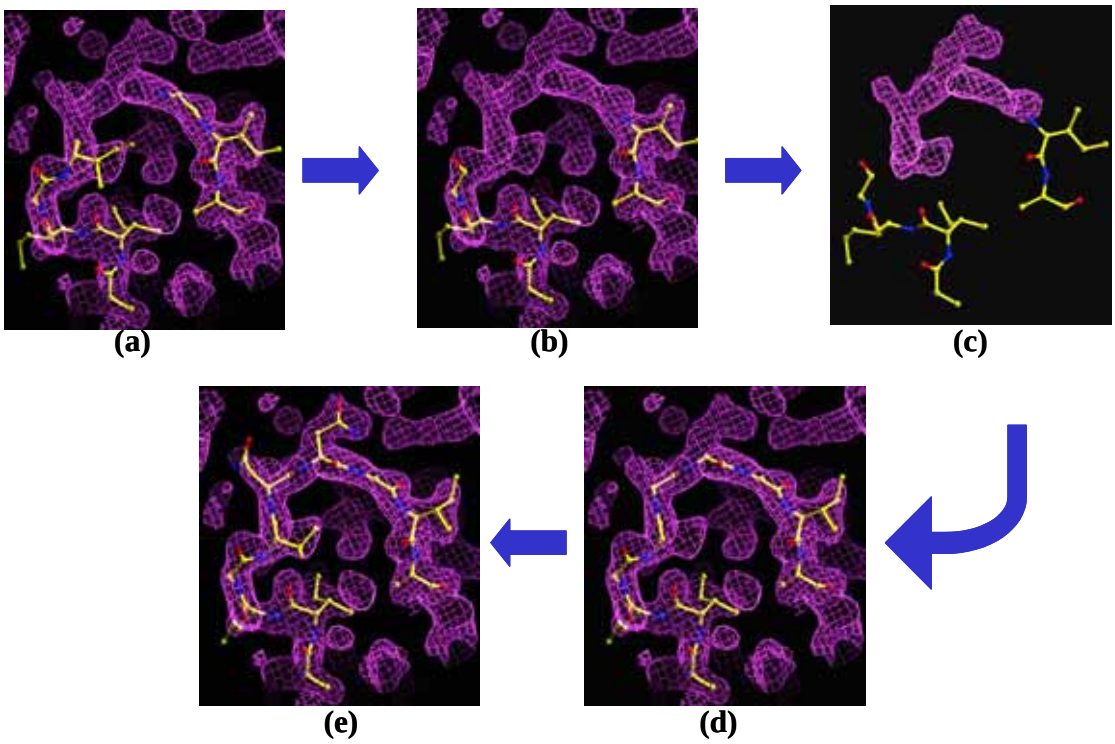


図６．欠失構造の構築
(a). 欠失している初期モデルと電子密度.
(b). 既構築部の両末端 1 残基ずつの削除.
(c). 取り出した欠失している部分の電子密度.
(d). 主鎖の構築.
(e). 側鎖の構築.

密度図と omit したモデルの Fo-Fc の２つの電子密度図を組み合わせて使用するように，omit した部分を再構築していく．部分構造構築では，map-pruning 法を利用した．図６では，実際の例を使って部分構造構築の流れを示している．

３－３．モニタープログラム（Lafire_molview）

自動修正プログラムを開発すると共に，精密化の過程をモニターするプログラム Lafire_molview を開発した．良好な視野を確保するために，図 7 に示しているようにモニタープログラムは N 端から 3，5 或いは 7 残基ずつを表示し，指定した原子の周りだけの電子密度図を表示する．また，表示された残基が全体構造のどの部分であるかをサブウィンドウに分かりやすいように表示している．フィッティングの様子を表すグループ化したローカル相関係数 *GLCC* 及び Ramachandran 図も表示する．

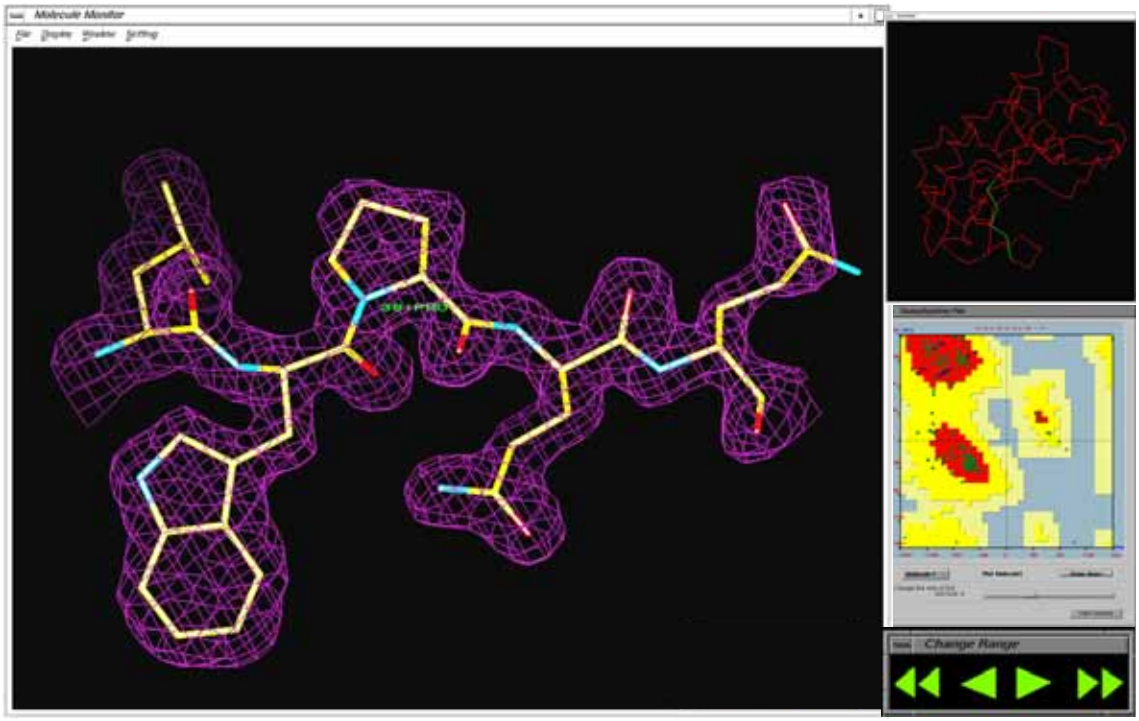


図７．モニタープログラム Lafire_molview
左　：メインウィンドウ.
右上：overview サブウィンドウ.
右中：Ramachandran 図サブウィンドウ.
右下：表示する残基の変更用ボタン.

３－４．Lafire による精密化

精密化を自動化するため，Lafire には，モデルのアミノ酸残基を，配列ファイルと比較しながら，置換する機能も増加した．また，上述した自動修正と構築のプログラムに加えて，精密化プログラム CNS，REFMAC5 と連動するインタフェースを作製し，さらに，私たちが今まで蓄積した精密化のノウハウを精密化のストラテジーとして組み込んで，R free 因子を指標として収斂するまで精密化サイクルを繰り返す Lafire プログラムを完成した．図 8 に示しているように，初期モデルから水分子を拾うまでの，熟練した研究者のマニュアル作業を必要とした精密化過程が Lafire によりコンピュータで自動的に行えるようになった.

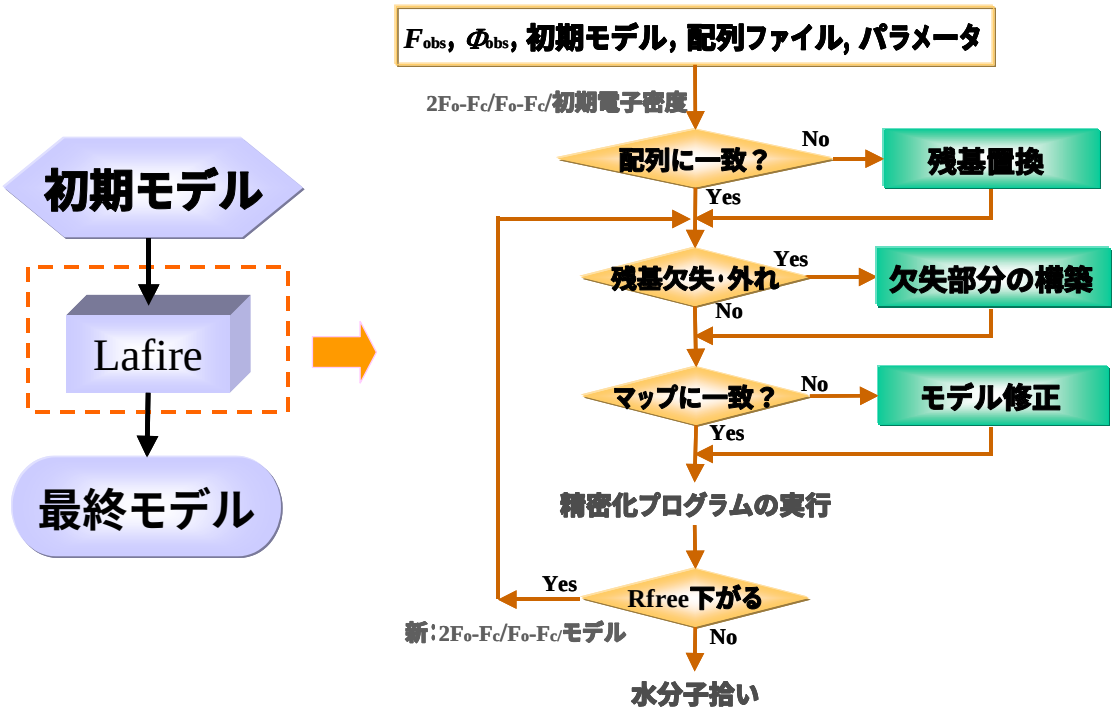


図８．Lafire による精密化のフローチャート

４．Lafire の応用及び結果

いくつかのテスト使用後，実際の構造解析への応用した．その結果，多くの場合，精密化にかかる時間を大幅に短縮することに成功し，また，精密化の条件を簡単に探索することに成功した例もある．

４－１．自動精密化

実際に，これまで数週間から数ヶ月かかっていた精密化を Lafire により，１日以内で，自動的に完成することができた例をいくつか紹介する．図 9 には，筆者らの研究室において構造解析された６つのタンパク質を例として，データ収集してからの構造解析にかかった時間を４つの段階（データ処理，位相計算，モデル構築，精密化）に分けて棒グラフで示した．この中で，MolA，MolB，MolC，MolD は既に，マニュアル精密化の終了した構造をテストしたサンプルであり，MolE と MolF は実際の構造解析への応用である．図 9 に示しているように，従来の精密化では，マニュアルフィッティングによる修正が必要であるので，精密化に要する時間は反射データと初期モデルの質だけではなく，人の経験にもかなり依存し，長い場合は数カ月，短くても 1 週間は必要とした．MolA と MolB は分子置換法により構造解析された例である．MolA の構造では，２つのループが位相計算に使われたモデルとかなり異なり，また非対称単位中に

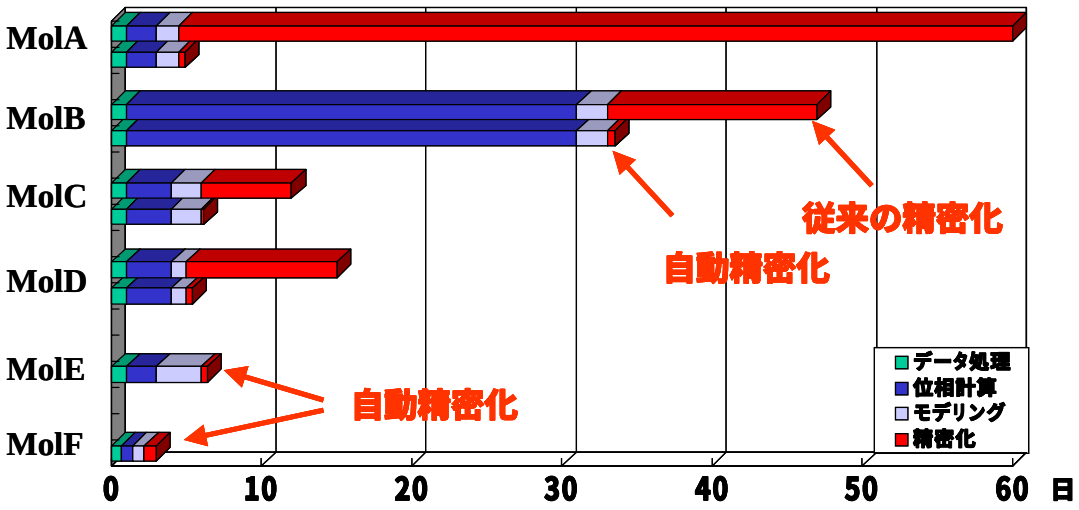


図８．構造解析にかかった時間の比較
上が従来の精密化，下が Lafire による精密化である．
MolA：191*2 残基，1.65Å 分解能での構造解析．
MolB：134*3 残基，2.0Å 分解能での構造解析．
MolC：131 残基，1.95Å 分解能での構造解析．
MolD：100 残基，2.05Å 分解能での構造解析．
MollE：192 残基，2.5Å 分解能での構造解析．
MolF：177 残基，1.92Å 分解能での構造解析

ある 2 分子の一つの N 末端にある長いループ（約 20 残基）の電子密度が貧弱であるため，精密化とモデルの修正に苦労した．MolB の構造解析では分子置換法に使われたモデルのアミノ酸配列の相同性は 42%に満たなかったが，モデルの修正と精密化を行ったのはかなり経験豊かな者であったので，比較的短い時間で終了した．一方，タンパク質 MolC と MolD は Se-MAD 法により解析された新規構造であり，良質な回折データが得られ，また非対称単位中のアミノ酸残基数も少ないため，極めて迅速に，半自動的に構造解析が行われた例である．但し，このような順調な場合でも，プログラム Solve/Resolve 或いは SHARP + DM + ARP/wARP を利用して，注意深く計算された位相から得られた電子密度図を使ってのモデルの自動構築では全体の 8，9 割しか構築できず，やはり人の手作業が必要であった．この４つタンパク質を例にとって，Lafire を用いて水分子を拾う前までの精密化を自動的に行った結果，数時間から 1 日の内に精密化が終了した．また，MolE と MolF は実際の構造解析への応用例である．両方とも Se-MAD 法により解析された．そのうち，MolF は 1.92Å 分解能で自動構造解析した例である．Lafire により自動精密化した結果，61 個の水分子を含んで R 因子

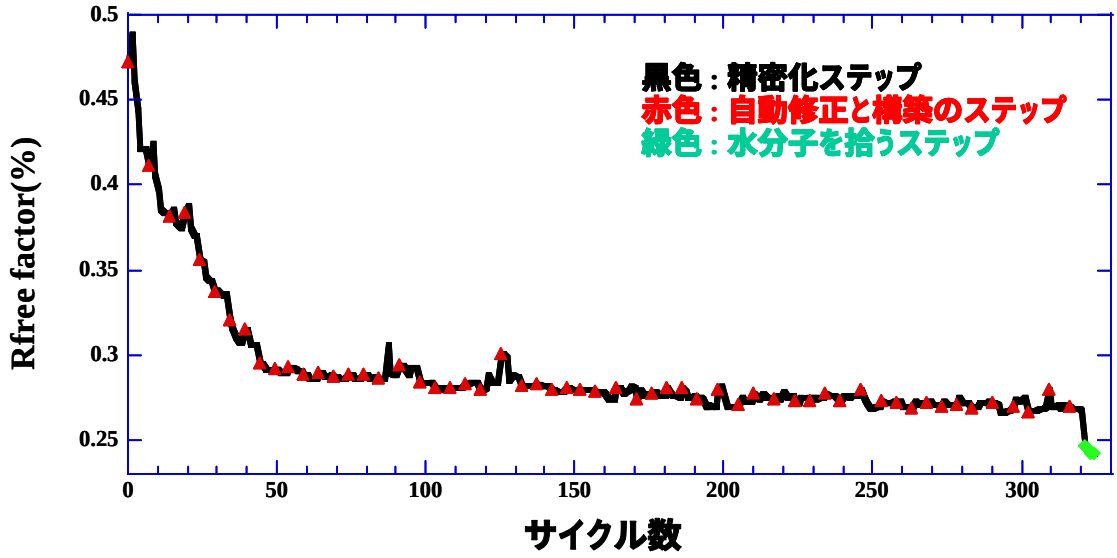


図９．MolF の精密化過程の R free 因子
精密化は Lafire により，CNS と連動して自動的に行なった．

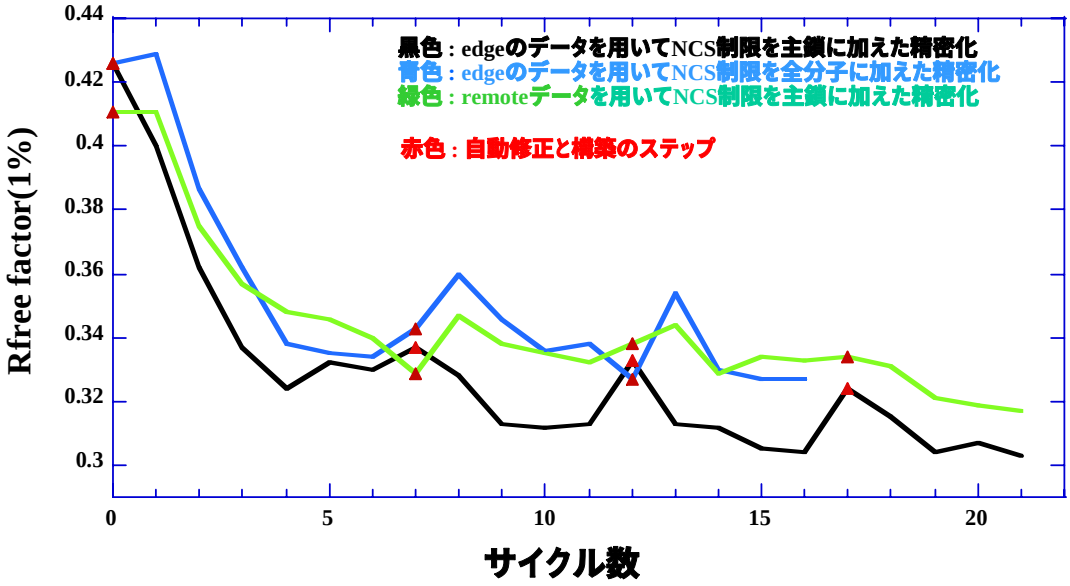


図 10．精密化条件の探索の例
精密化は Lafire により，CNS と連動して自動的に行なった．

と R free 因子がそれぞれ 25.1%と 26.2 %までに下がった．解析はデータを測定してから 3 日間で終了した．精密化過程における R free 因子の変化が図 9 に表示されている．

４－２．精密化条件の検討

プログラム Lafire を用いた精密化は経験に依存せず，自動的に行われるので，解析にかかる時間は大幅に短縮するが，さらに，精密化条件の検討，例えば，最適なデータセットの選択，精密化プログラムの選択，精密化の制限条件の選択なども簡単にできる．図 10 は実際の応用例を示している．このタンパク質は 154 残基であり，非対称単位中に 2 分子が存在しており，Se-MAD 法により，2.62Å の分解能での構造解析に成功した．初期モデルの構築は Solve/Resolve 或いは ARP/wARP を用いて自動的に行うことができず，コンピュータグラフィクスを利用してマニュアルで行った．つまり，分解能のリミットと初期位相の誤差のために，初期モデルの構築と精密化は極めて難しい状況であった．マニュアルで初期モデルを構築した後に，Lafire による精密化は，NCS 制限を加えない，全分子に NCS 制限を加える，主鎖だけに加える，の三つの条件で行われた．さらに，Se 原子の吸収端の波長（edge）と参照波長（remote）で測定した 2 つのデータセットを使って精密化して見た．その結果，NCS 制限を主鎖だけに加える条件が最適で，最後に測定した remote のデータは結晶の X 線損傷のため良質ではなく，精密化に不適当ということが判明した．

５．おわりに

現在，Lafire は CNS 或いは REFMAC5 と連動して，数個のテスト及び 25 個以上の実際の構造解析への応用を経て，既に SGI（IRIS6.5 以上）と Linux（RadHat9.0）の 2 種類バージョンを一般公開している（図 11）．これから，Lafire は大量の構造解析例に適用しながら，改良を進めていく．現在は，特に分子置換法で得られた構造の omit 法による修正，未構築部分の構造構築法の改良，電子密度図が貧弱な場合での成功率の向上などについての改良が行われている．2Fo-Fc，Fo-Fc と初期位相から計算した電子密度の最適な取り扱いについての検討も行っている．

ヒトゲノムの配列解析が終了した現在，研究の焦点は塩基配列解読から遺伝子の機能解析，遺伝子産物の相互作用解析を中心としたポストゲノム科学へ移ってきた．多数のタンパク質の相互作用も含む細胞内での複雑な機能を総合的に理解するのに，個々のタンパク質の構造解析や複合体の構造解析は，今後増々必要とされよう．立体構造解析の迅速化，自動化が要求される所以である．今回，結晶構造解析で最も時間を要する「構造精密化過程」を自動化することに成功した．解析のハイスループット化を実現し，全自動構造解析に一歩近づいたと思っている．

http://altair.sci.hokudai.ac.jp/g6　　http://altair.sci.hokudai.ac.jp/g6/Lafire.html

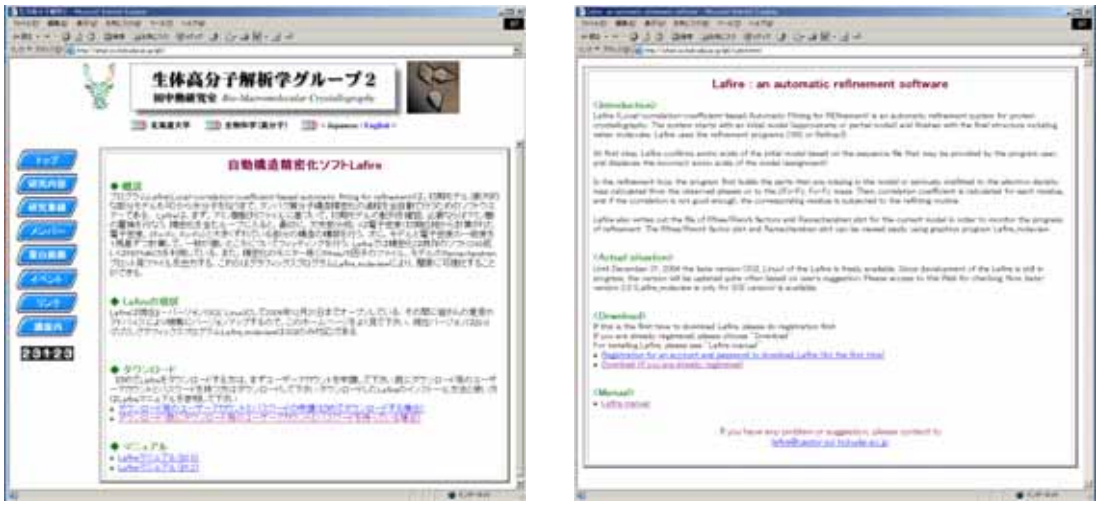


図１１．Lafire 公開の Web
左：日本語の公開ページ.
右：英語の公開ページ.

６．謝辞

この研究をアドバイス，サポートして頂いた田中勲教授，Lafire における計算法の開発及びプログラミングを担当した本研究室の博士学生である周勇さん，テスト及び実際の構造解析のサンプルを提供して頂いた日本のタンパク質結晶学者の皆様に感謝します．また，モニタープログラム Lafire_molview の開発を協力して頂いた株式会社システムｋに感謝します.