

# MIR-OAS : 統合型蛋白質構造解析ソフトウェア

濱田 賢作\*・冨 和彦†・田中 信夫††

\*島根大学総合理工学部数理・情報システム学科

†島根大学理学部情報科学科 (現在: 日立ソフトウェア株式会社)

††東京工業大学大学院生命理工学研究科

## MIR-OAS: A New Software for Protein Structure Determination by Multiple Isomorphous Replacement method

Kensaku HAMADA, Kazuhiko KAKOI, Nobuo TANAKA

The MIR-OAS is a user-friendly software to be used for macromolecular structure determination by multiple isomorphous replacement method with optimized anomalous scattering on X-ray crystallography. The software is designed to be flexible, allowing users to solve macromolecule structure without a good knowledge of X-ray crystallography. The graphical user interface is written by Tcl/Tk, and the programs to calculate for MIR-OAS are written mainly by standard Fortran 77. The MIR-OAS runs on wide range of architectures and operating systems.

### 1. はじめに

生命科学およびその応用分野の研究・開発において、蛋白質の原子レベルでの立体構造(三次構造)情報は益々重要になってきている。蛋白質分子の立体構造を決める方法に X 線結晶構造解析がある。X 線結晶構造解析では、蛋白質結晶からの少なくとも数万点の微弱な X 線回折強度から何らかの方法で位相情報を導き出し、フーリエ合成で得た結晶の電子密度から分子の立体構造を決める。位相情報を持たない X 線回折強度から初期位相情報を求める方法として、蛋白質結晶の場合、次の方法がある<sup>[1]</sup>。(a)蛋白質自身の結晶とそれに異なる重原子を付加させた二種類以上の結晶からの回折強度データで解析する重原子同型置換法(MIR法)<sup>[2]</sup>。(b)既知の蛋白質の立体構造を初期モデルとして解析をする分子置換法(MR法)。(c)異常分散効果が利用できるセレンなどを有する蛋白質に適用できる、放射光の特性の一つであるチューナビリティを利用した多波長異常分散法(MAD法)。このうち、MR法のようにモデル構造を必要とせず、またMAD法のように異常分散効果の大きな原子を必要もしないMIR法は少なくとも2種の重原子同型置換結晶を得るだけで、解析が可能となり、汎用性の高いX線結晶構造解析法である。遺伝子工学による蛋白質の大量発現による結晶化スクリーニングが容易となり、放射光とイメージングプレート<sup>[3]</sup>やCCD<sup>[4]</sup>の利用によりデータ収集までは大幅に迅速化されてきた。しかしながら、図1に示すように

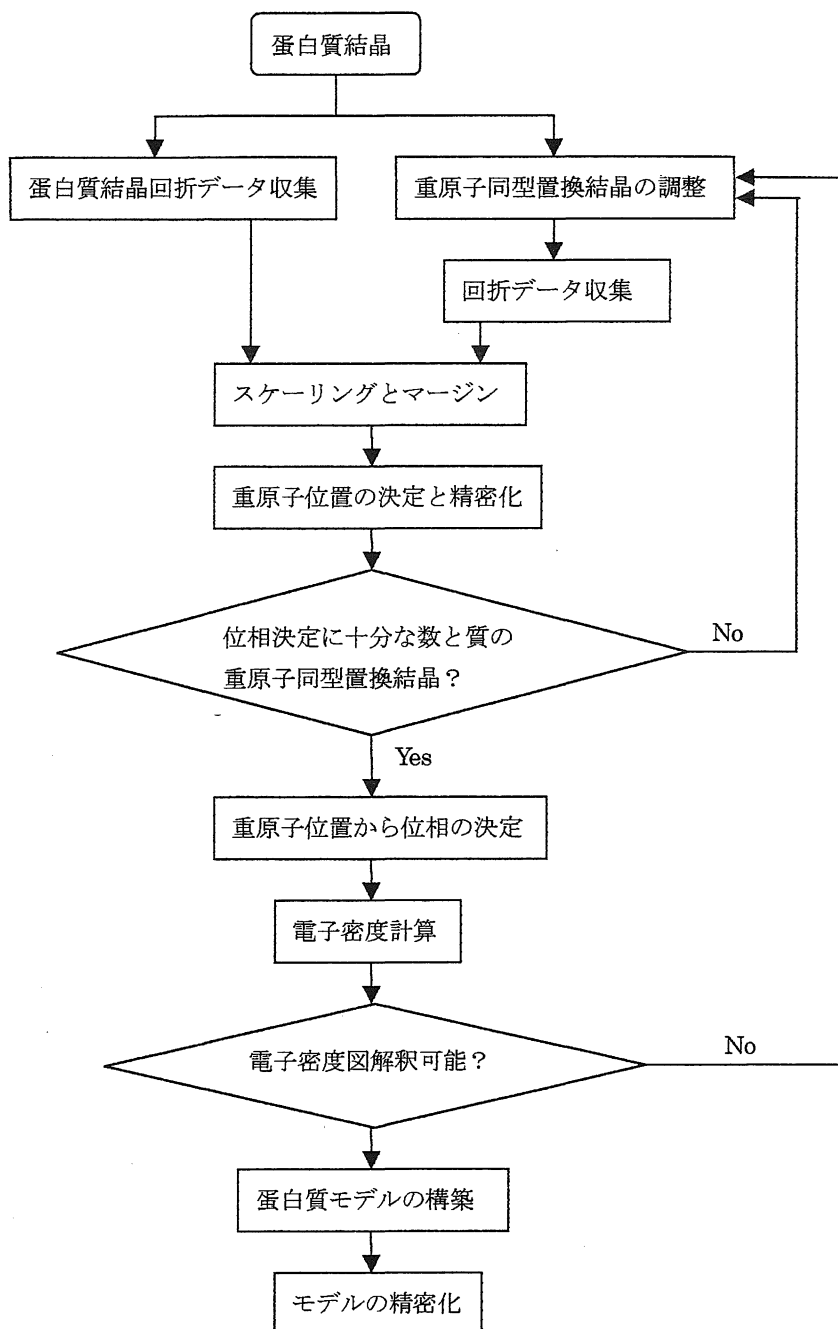


図1 重原子同型置換法による蛋白質の X 線結晶構造解析の手順

MIR 法での解析の各ステップにおいて蛋白質結晶学の高度な知識を必要としている。また、膨大な回折強度データを取扱い、解析過程でも膨大なデータが生じてくるため、解析は煩雑となるだけでなく、計算量も膨大となり、解析にはコンピュータの取扱いに習熟していることが要求され、時間も長くかかっていた。このような問題を克服し、蛋白質結晶学に習熟していない生化学・分子生物学者でも容易に蛋白質の構造解析を行える異常分散も考慮した MIR 法によるソフトウェアを開発した。

## 2. 構築上の問題点と基本的な考え方

MIR 法で X 線構造解析を行うには、原理的には格子定数、空間群とネイティブ結晶（もとの蛋白質そのものの結晶）の回折データ、そして 2 種類の重原子同型置換結晶（ネイティブ結晶と格子定数、分子の立体構造と結晶での分子の相対配置が変わらないで水銀などの重原子導入した同型な結晶）の回折データを必要とする。蛋白質の位相情報が得られるのに有効な重原子の寄与率を持つ重原子同型置換結晶を調整しなければならないが、重原子位置を決定するまで判定できない。重原子同型置換結晶の調整は試行錯誤的に行われ、解析に適した重原子同型置換結晶を得るまでに多数の重原子同型置換結晶のチェックが必要となる。重原子位置の決定には蛋白質結晶学の知識を必要とする。このため、ユーザーに結晶学的知識を基にした判断を要求しないで、かつ迅速に解析できるソフトウェアでなくてはならない。

解析の色々な過程について独立に作ったプログラムを集め、必要なプログラムを呼び出して使用する従来の蛋白質 X 線構造解析ソフトウェア (program suite) では、解析の各段階で使用するプログラムは何かを知る必要があった。また、各プログラムの使用方法や入力データの作成が複雑あるだけでなく、各プログラム間のデータのハンドリングも複雑であった。このような問題を解消する優れたユーザーインターフェイスを開発する。

本システムの設計にあたっては、上記の大きな問題を考慮して、次のような目標をたてた。

- (a) 重原子位置はパターン関数から最小関数法<sup>[5]</sup>で決定する。
- (b) 蛋白質結晶学的知識の要求は可能な限り最小限にする。
- (c) 各解析ステップ間のデータハンドリングの簡易化。
- (d) OS に依存しない移植性、保守が容易
- (e) 優れたシンプルなユーザーインターフェイス。

(b), (c) を実現するために、各解析ステップでの入力を最小限に留め、必要な場合でも、マニュアル等が無くても理解して入力できるようにユーザーインターフェイスを実現する。煩雑さの要因の 1 つである解析過程で生成される多くのデータファイルをファイル名と拡張子で分類し、ユーザーインターフェイスでコントロールする。

慣れていないコンピュータ (OS も含む) の利用は、多くのユーザーにとって、障壁になることが多い、また、最近では種々のコンピュータが利用されるようになってきたことから、OS に依存しないソフトウェアが要求されている。そこで (d) を実現するために、それに対応したユーザーインターフェイスを導入する。(b)~(d) を実現するためには優れたユーザーイン

ターフェイスを必要とする。ユーザーインターフェイスを実現する言語は種々あるが、必要とする全てのツールをもち、容易に開発できる Tcl/Tk でユーザーインターフェイスを実現する。Tcl はスクリプト言語であり、Tk はグラフィックスツールボックスである<sup>[6]</sup>。

### 3. システム構成

本ソフトウェアの構成は図2のように、ユーザーインターフェイスと実行プログラムからなっている。すべてのデータファイルは、蛋白質名をディレクトリとするディレクトリに格納される。ユーザーインターフェイスは、X-ウィンドウ・システム、MS-Windows、マッキントッシュのいずれのグラフィックユーザーインターフェイスのもとで動作する Tcl/Tk スクリプト言語で記述し、蛋白質 X 線構造解析実行プログラム（実行プログラム）は標準的な Fortran 77 言語で記述し（一部を C 言語）、そして UNIX オペレーティングシステムのもとで開発した。1つのタスクウィンドウは解析の各ステップの実行プログラムの1つに対応している。しかし、重原子同型置換結晶の重原子位置を見つける SMIN や MINF などのタスクウィンドウは、自動的にパターンソート、ピークサーチを行い、最小関数で重原子位置を見つけられるように構成されている。

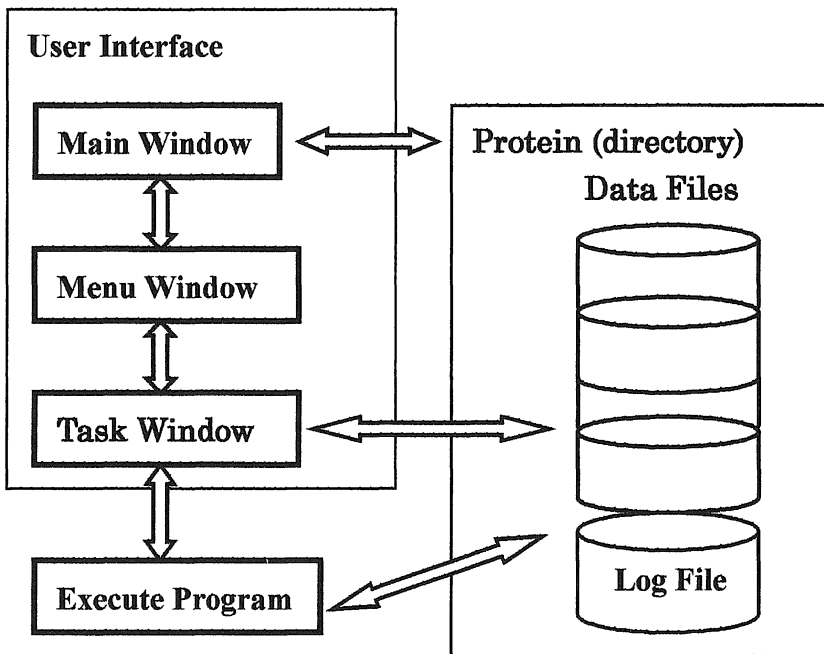


図2 MIR-OAS ソフトウェアの構成

#### 4. 解析実行プログラム

本システムでは、MIR法により、蛋白質の位相を決定し、電子密度を得るまでの各段階のプログラムから構成されている。それらプログラムは以下のとおりである。

- ◎**Start**: 結晶学データの入力、ネイティブ結晶データへ重原子同型置換結晶データをスケールリングする。
- ◎**Restart**: 新たに測定された回折強度データの追加し、スケールリングを行う。
- ◎**Patterson**: 各種パターン関数の計算。
- ◎**Min-f** (minimum function): パターン図をもとに求像法の1つである最小関数 (minimum function) の計算を行い、重原子同型置換結晶の単位格子内での重原子位置を求める。Min-fは3次元パターン空間で行われる。
- ◎**S-min** (symmetrical minimum function): 最小関数であるが、空間群 (結晶の持つ対称) により、パターン関数の結果である原子間ベクトルの集中する面あるいは線上であるハーカー断面での最小関数。これで、重原子位置 (x, y, z) の2あるいは1成分を求め、Min-fで重原子位置を最終的に求める。
- ◎**LSH**: 重原子パラメータ (位置, 温度因子, 占有率) をRossmannの方法による最小二乗法で精密化する<sup>[7]</sup>。
- ◎**LS-3D**: すべての重原子同型置換結晶データを使い、重原子パラメータを最小二乗法により精密化する。
- ◎**Phase**: 重原子パラメータより蛋白質の位相情報を求める。
- ◎**Peak**: Patterson, Min-f, s-min, D-four, Fourier, pat で得られたマップより、ピークを探索する。
- ◎**D-four**: 位相情報をもとに重原子同型置換結晶とネイティブ結晶の構造因子の差の電子密度図を計算する。通常、最小関数で見つけられなかったマイナー重原子位置を探すために利用する。
- ◎**Fourier**: 蛋白質の電子密度図を求める。
- ◎**MEM**: マキシマムエントロピー法で蛋白質の位相を精密化する。
- ◎**t-mapping**: 蛋白質分子モデル作成ソフトウェア Frodo, Turbo-Frodo のために、電子密度図ファイルの書式を変換する。
- ◎**t-phase**: ソフトウェア Phases へのファイル変換をする。

#### 5. ユーザーインターフェイス

ユーザーインターフェイスは、メインウィンドウとメニューウィンドウから呼び出すタスクウィンドウから構成されている。このユーザーインターフェイス部分において、解析手順の管理、ファイル管理、そして、各プログラムの実行に必要な最小限の入力データの作成を行う。

## (1) メインウインドウ

MIR-OAS を実行すると、メインウインドウが表示される (図 3)。このウインドウのディレクトリ名にネイティブ及び重原子同型置換結晶の回折強度データが保存されているディレクトリが表示される。但し、最初の起動時に MIR-OAS を実行したディレクトリが表示されるので、Protein name にその回折強度データが保存されているディレクトリ名を入力するか、▼ボタンをクリックして、目的のディレクトリを探す。この Protein name のディレクトリのもとに、すべての入出力ファイルが保存されるようになり、蛋白質を知っているだけで、ファイル管理の煩わしさからユーザーを開放している。

## (2) タスクウインドウ

メニューウインドウから呼び出されるタスクウインドウは、実行に必要な入力データの作成、編集を行う。各実行プログラムに必要なファイルは、名前と拡張子を判別して自動的にあるは、選択候補が表示される。いずれのタスクウインドウの下に、exec, view the result, end の 3つのボタンが配置されている。exec ボタンをクリックすると、タスクファイルを生成し、実行プログラムを実行する。実行が終了すると、そのボタンの色が変わり、実行の終了を知らせる。実行結果は、view the result ボタンのクリックで表示される。end ボタンでメニューウインドウに戻る。タスクウインドウは機能から、(a)回折実験で得た回折強度データファイル情報機能を持つ Start, Restart, (b)構造解析のステップのみを担っている Pat, S-Min, Min-f, LS-3D, D-fourier, Fourier, MEM, それに(c)構造解析結果を他のプログラム

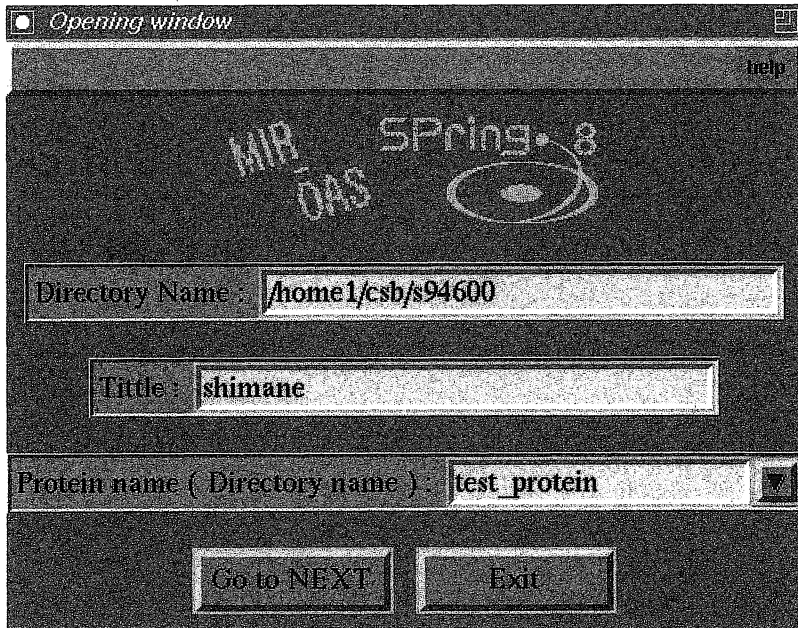


図 3 メインウインドウ

で利用するためのインターフェイスである t-mappage, t-phase の 3 つに分類される。

(a)では、図4に示すように、提示された蛋白質名のディレクトリに存在する回折強度データファイルから選択された重原子同型置換結晶の回折データのみが、データファイルのリスト表 (datafile.lst) に登録され、回折強度のスケールが行われる。また、結晶学的データファイル (crystal.dat) が作成される。したがって、解析開始時は Start を、新たに回

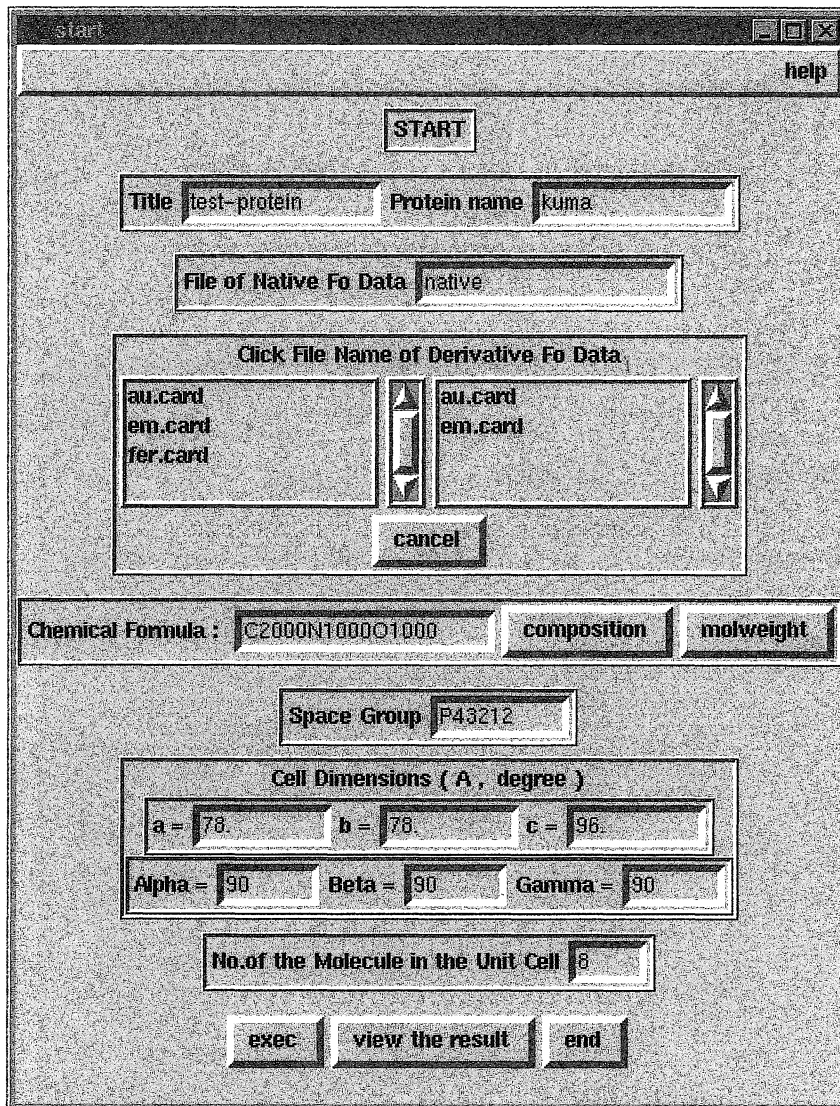


図4 Start のタスクウインドウ

折強度データを得たときは Restart を必ず行わなければならない。

(b)では、図5に示すように重原子同型置換結晶 (derivative) を選択し、入力データを画面の指示にしたがい入力する。実行後、view the result ボタンをクリックすると、図6のように結果を表示する。

## 6. 終わりに

コンピュータ依存性のなく、使用が比較的簡単な Tcl/Tk でユーザーインターフェイスを開発した。一方、X線構造解析計算プログラムは解析科学技術計算に優れた Fortran 77 (一部を C) で開発した。これにより、UNIX, LINUX, Winndows 95/NT, Macintosh で、操作性に優れ、高速で計算でき、また保守管理の容易なソフトウェアを実現できた。現在、各地で SUN (OS5.51), LINUX (Red Hat 5.2) のもとでの利用されているが、Winndows 95/NT, Macintosh で動作確認を必要とする。

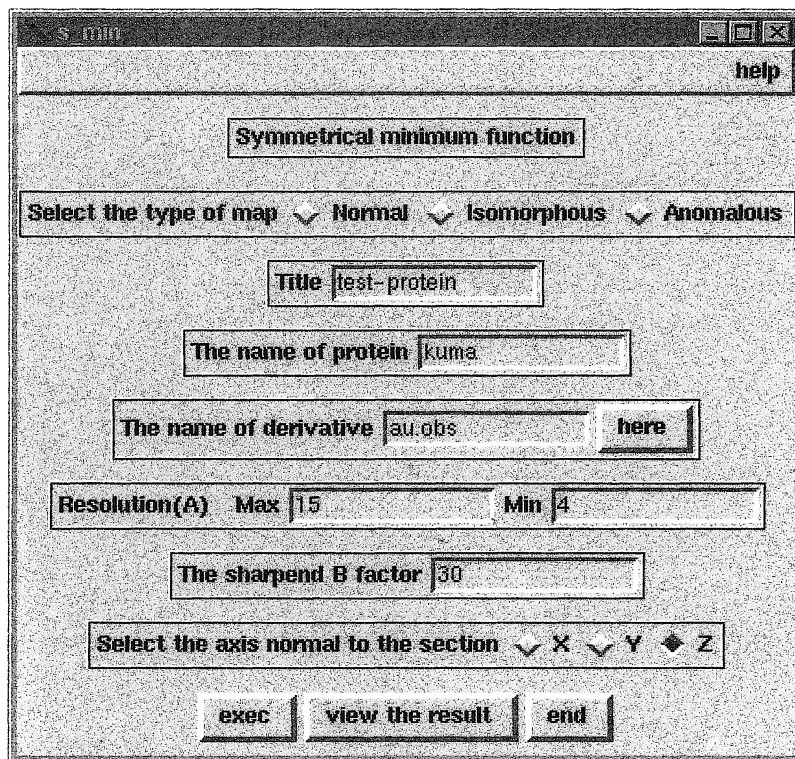


図5 S-min のタスクウインドウ



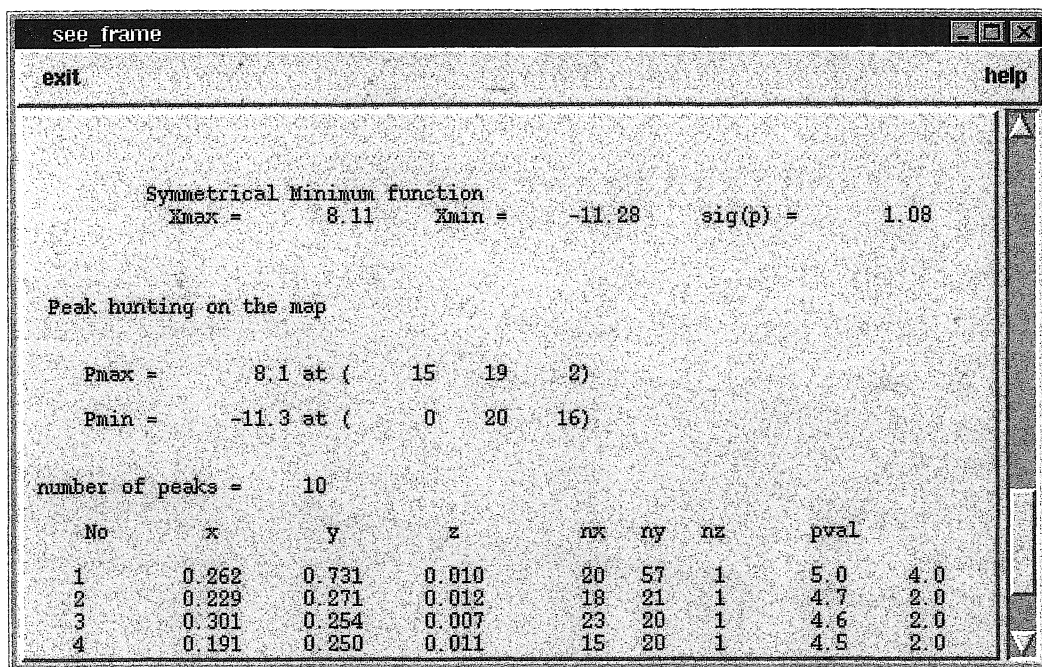


図6 S-minの結果

## 参 考 文 献

1. Jan Drenth (1994) *Principles of Protein X-ray Crystallography*, Springer, New York.
2. Hengming Ke (1997) *Methods Enzymol.* **276** Part B, 448-461.
3. Yoshiuki Amemiya (1997) *Methods Enzymol.* **276** Part A, 233-243.
4. Edwin M. Westbrook and Istvan Nady (1997) *Methods Enzymol.* **276** Part A, 244-268.
5. Martin J. Buerger (1959) *Vector space*, Jhon Wiley & SONS INC, New York.
6. John K. Ousterhout (著), 西仲芳幸, 石曾根信 (訳) (1995) Tcl & Tk ツールキット, ソフトバンク.
7. Michael G. Rossmann (1960) *Acta Crystallogr.* **13**, 221-226.