

分子置換法によるセミオートマチック蛋白質 構造解析システム

野澤 佳代*・濱田 賢作**

*島根大学大学院理学研究科情報科学専攻

**島根大学総合理工学部数理・情報システム学科

Semi-automatic Molecular Replacement On Web

Kayo NOZAWA, Kensaku HAMADA

SAMR-Web was designed as a user-friendly system for crystal structure determination of proteins by the method of molecular replacement (MR) running over World Wide Web. It was linked to Web sites of PDB and Swiss-Model, and was able to run *AMoRe* or EPMPR which were programs for MR. It is easy and efficient to prepare a search model and input data for running the programs *AMoRe* or EPMPR by using a favorite web browser.

1. はじめに

最近、インターネット技術の急速な発展と普及により、あらゆる分野で、様々なサイトが利用されており、研究に重要な役割を果たしている。生命科学分野では、DNAの塩基配列、蛋白質のアミノ酸配列そして蛋白質の立体構造などのデータベースが開発され、インターネットで利用できるようになってきている。さらに、これらデータベースをもとに情報解析する技術が急速に進展し、インターネットを利用することで、生命情報解析が可能となってきている^[1]。

遺伝子は蛋白質に翻訳され、蛋白質の機能発現には活性体としての3次元構造（原子レベルでの立体構造）をとる必要がある。そのため、蛋白質の立体構造を決定することは、益々重要となってきている。決定された蛋白質の立体構造はProtein Data Bank (PDB)^[2]に登録されている。

蛋白質の立体構造決定法の中心はX線結晶構造解析法である。それには、まったく未知な立体構造を決定できる重原子同型置換法 (MIR法)、多重異常分散法 (MAD法)そして、分子置換法 (MR法)とよばれている既知の蛋白質立体構造あるいはそれを修飾したモデル構造を利用して新規な蛋白質の立体構造を決定する方法がある^[3]。

重原子誘導体結晶の調整や遺伝子組換えによるセレノメチオニンの導入などを必要としない分子置換法は、PDBへの登録件数の爆発的な増大だけでなく、ホモロジーモデリングな

どコンピュータによる立体構造予測の確実性の向上とともに、迅速かつ容易に初期蛋白質立体構造が決定できることから、これまで以上に頻繁に利用されるようになる。モデル構造を作成するには、アミノ酸配列の相同性の解析、PDBから類似蛋白質の取得と修正が必要となる。さらにホモロジーモデリングによりモデル構造を作成することも必要となる。作成されたモデル構造と収集したX線回折データにより分子置換法プログラムを使って解析するには、いろいろなパラメータを設定する必要があるとともに、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、解析が煩雑となり、専門的知識も必要となってくる。しかしながら、これら一連の作業を一元的に統括して解析するソフトウェアは存在していない。そこで、インターネット利用技術を適用して、解析したい蛋白質のアミノ酸配列とX線回折データのみで、MR法により解析できるユーザーインターフェースを開発した。なお、既存の蛋白質X線構造解析のためのソフトウェアCCP4^[4]に組み込まれているAMoRe^[5]および進化的アルゴリズムを適用したEPMR^[6]を選択してMR法を実行できるようにした。

2. 構築上の問題点と基本的な考え方

MR法では、決定したい立体構造が未知である蛋白質結晶のモデル構造と結晶学的データ(格子定数、空間群など)と回折強度データが必要である。モデル構造は出来る限り真の立体構造に類似していなければならない。そのために、既知構造の蛋白質のアミノ酸配列と比較し、相同性の高い既知構造の蛋白質をもとにしてモデル構造は作成されるべきである。このために、蛋白質立体構造データベースの元締めであるPDBの検索、さらに相同性の検討が必要である。PDBはインターネットを通じて公開されており、PDBコードあるいは蛋白質名などで検索し、ダウンロードできる。しかしながら、アミノ酸配列から相同性の高い蛋白質を検索する機能を持っていない。アミノ酸配列に相同性があり、既知構造の蛋白質を検索できるインターネットサイトはいくつかあるが、SWISS-MODEL^[7]は既知構造を鋳型としたモデル構造をつくり、電子メールで返信することを目的とし、前処理としてアミノ酸配列の比較により相同性の既知構造の蛋白質を列挙する機能を持っている。

モデル構造を結晶格子内で回転、並進し、収集した回折強度データと一致する初期位相を求めるMR法のプログラムはこれまでにいくつか開発されているが、比較検討した結果、高速化アルゴリズムをもつAMoReと計算時間はかかるが、進化的アルゴリズムを適用し比較的低い相同性をもつモデル構造でもかなりの能力を示すEPMRを利用することにした。これらMR法プログラムを利用するには、モデル構造の修正、入力データを必要とするが、それには分子置換法の原理やそれぞれのプログラムのアルゴリズムをある程度理解していることが要求される。

このようなことを踏まえて、インターネットを介して、モデル構造の作成からMR法の自動実行までを行える、優れたユーザーインターフェースを持つシステムを次の目標を立てて設計した。

- (1) PDBサイトやSWISS-MODELサイトにリンクし、或いは自己ファイルからモデル

構造データを入手する。

- (2) モデル構造データから、MR 法に必要な情報を取得し、最終的なモデル構造を自動生成する。
 - (3) 蛋白質結晶学及び MR 法プログラム使用法の知識の要求を可能な限り最小限にする。
 - (4) 各解析ステップ間のデータハンドリングの簡易化。
 - (5) OS に依存しない移植性，保守が容易。
 - (6) CGI (Common Gateway Interface) による優れたシンプルなユーザーインターフェイス。
- (1), (2)を実現するために、テキストデータを柔軟に処理でき、文字列処理に適した言語でなければならない。

また(3), (4)を実現するために、各解析ステップでの入力を最小限にし、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インターフェイスを実現する。

最近では種々の OS が利用されるようになってきたことから、これに依存しないソフトウェアが要求されている。これに対応するために CGI を導入する。CGI プログラムは、クライアントコンピュータではなくサーバーコンピュータ上で実行するという点で、ほかの Web プログラムと異なっている。このため、サーバーとのやり取りを必要とする作業に適していると考えられる。Perl 言語は優れたテキスト処理機能や使いやすさなどから CGI プログラムの主流となっているだけでなく、本システム開発に必要な全てのツールをもち、容易に開発できる言語であることから、Perl 言語により開発した^{[8],[9],[10]}。

3. システム構成

図 1 に示すように、本システムでの MR 法による蛋白質の立体構造の決定を Web ブラウ

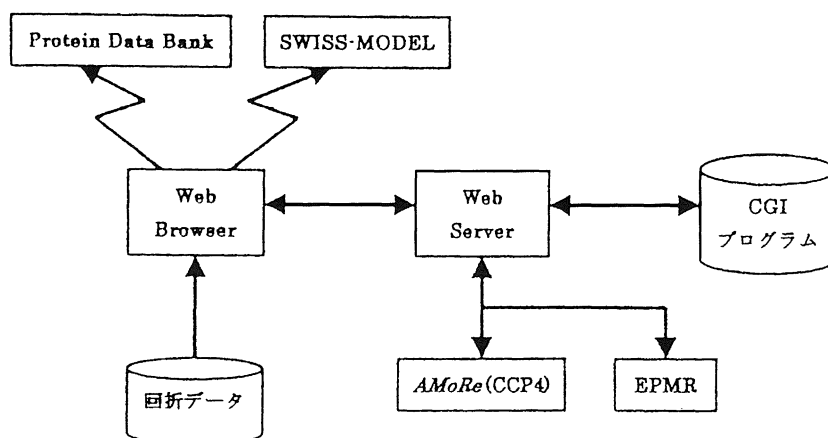


図 1 SAMR-Web ソフトウェアの全体構成

ザが要求したとき、CGI プログラムの実行、HTML (Hyper Text Markup Language) が生成され、そして Web ブラウザに提示される。提示された HTML ドキュメントの要求にユーザーが答えたデータをもとに、AMoRe あるいは EPMP の実行に必要なデータを作成し、実行する。PDB サイトあるいは SWISS-MODEL サイトでモデル構造を取得するために、これらのサイトにリンクされている。

モデル構造の作成、分子置換法解析プログラムへの入力データの作成を処理する CGI プログラムは図 2 に示す構成になっている。

(a) モデル構造作成

モデル構造作成は MAKE A MODEL が担っている。ここではモデル構造ファイルの有無を判断し、所持していないとき、PDB サイトに行き、モデル構造のもととなる蛋白質の立体構造データを取得する。あるいは、SWISS-MODEL サイトに行き、決定したい蛋白質のアミノ酸配列から相同性の高い蛋白質を取得する。ホモロジーモデリング法によるモデル構造の作成をする場合も SWISS-MODEL サイトに行き、作成する。取得したモデル構造データは、ユーザーとの対話をとおして、分子置換法に利用できるように加工される。このタスクウインドウを図 3 に示す。

(b) 分子置換法実行データの作成と実行処理

分子置換法は AMoRe あるいは EPMP のどちらかを選択して実行可能であり、それぞれについて、MOLECULAR REPLACEMENT METHOD: AMoRe と MOLECULAR REPLACEMENT METHOD: EPMP の 2 つから構成されている。なお、回折強度データ (MTZ ファイル) は拡張子を判断して取得される。MTZ ファイルは CCP4 で定められた回折強度データファイルであり、これに結晶学的データが記述してあることから、CCP4 プログラムを自動起動させ、結晶学的データの一部である格子定数、空間群を取得する。つぎにこれと、(a) で得たサーチモデルにより、AMoRe あるいは EPMP を実行するためのスクリプトの作成と実行を行う。これらのタスクウインドウを図 4, 5 に示す。

4. SAMR-Web の実行例

PC-LINUX (RedHat6.2, CPU Pentium III 600 MHz/133 MHz System Bus, メモリー 128 MB, ハードディスクシーク速度 8.9 ms) に本システムを導入した。Phormidium lapideum 由来のアスパラギン酸アミノ基転移酵素 (PIAspAT) の立体構造の決定を行った。なお、PIAspAT は 384 のアミノ酸残基で構成されるサブユニットからなるホモダイマー (分子量 84,000) である。結晶学的データは空間群: $P4_12_12$, $a=b=133.40$, $c=125.83$ Å であり、2.8 Å 分解能まで収集した 14,689 反射数の回折強度データを使い、Thermus thermophilus 由来の AspAT の立体構造を PDB サイトから検索し、ポリアラニンからなるダイマーをサーチモデルとし、分子置換法には、AMoRe と EPMP のいずれをも実行し、PIAspAT の初期構造を得た。ここで実行までの所要時間は約 2 分である、従来これに数時間要しており、解析の時間が大幅に短縮できたことを実証した。

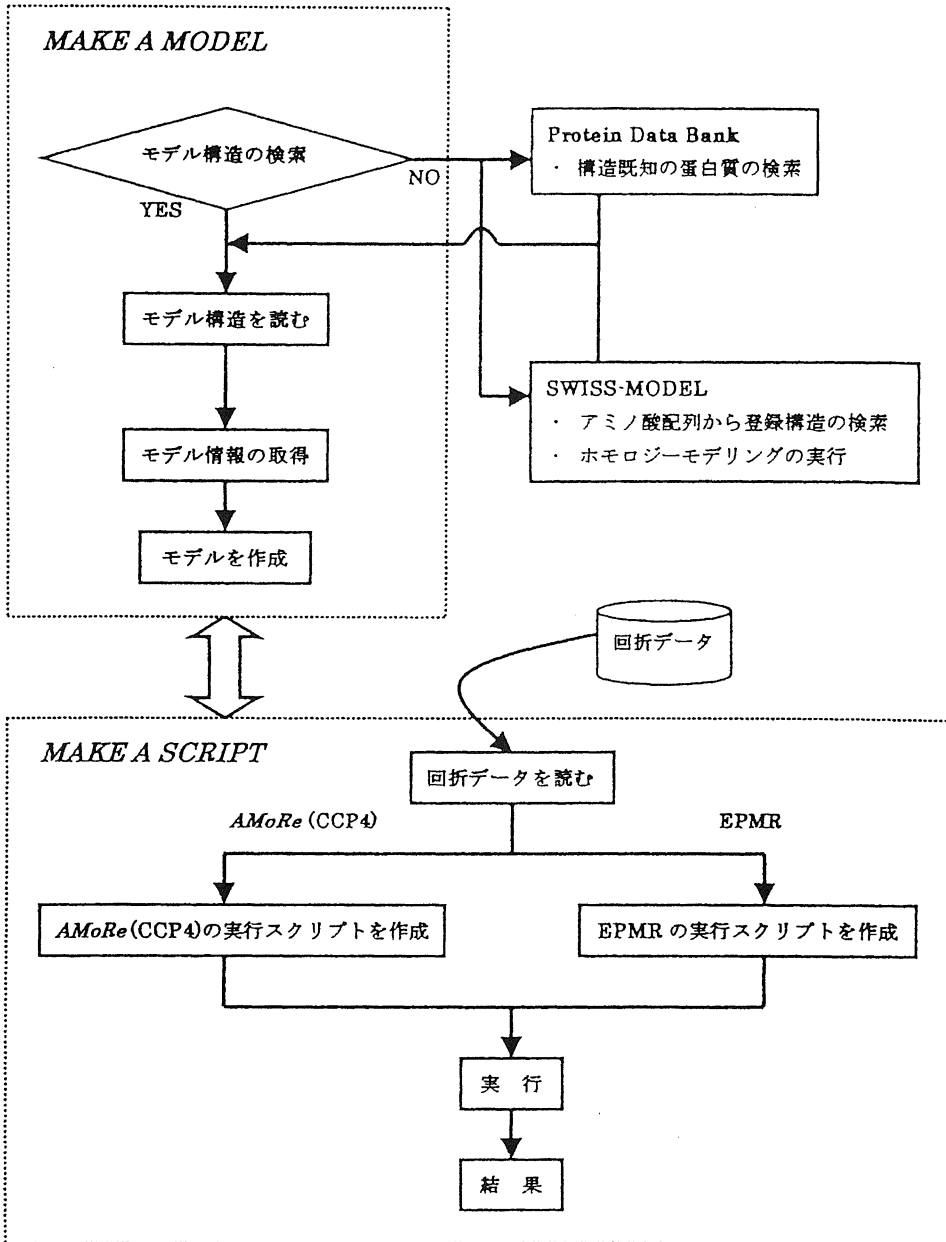
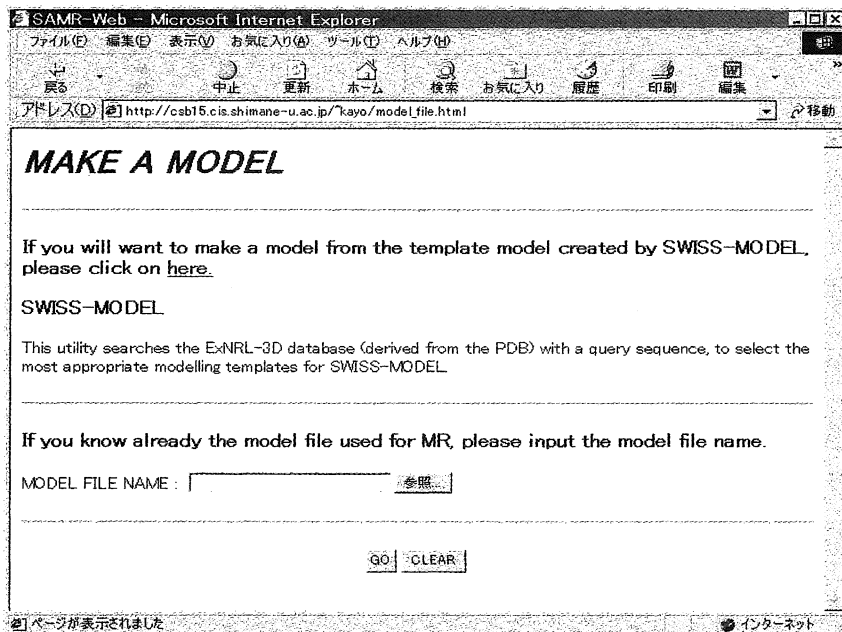
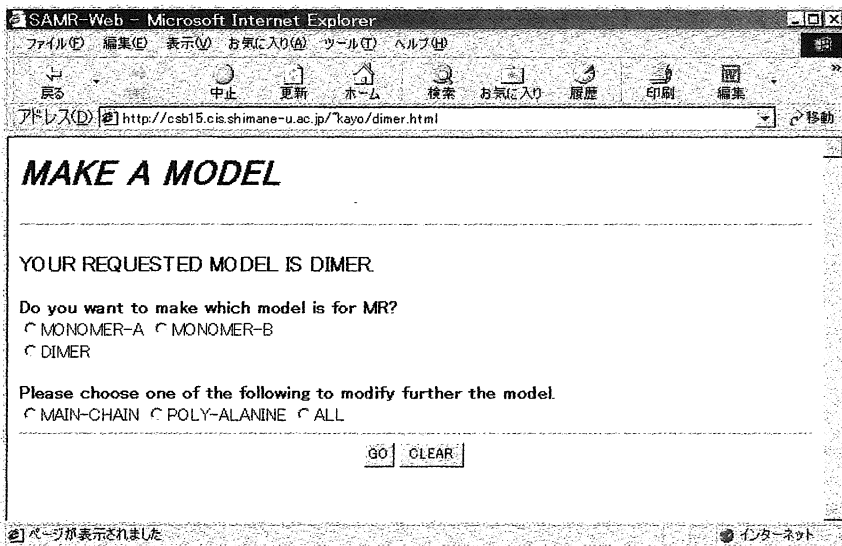


図2 SAMR-WebソフトウェアのCGI部分の構成



(a)



(b)

図3 (a)MAKE A MODEL のタスクウインドウ
(b)MAKE A MODEL 実行後のタスクウインドウ

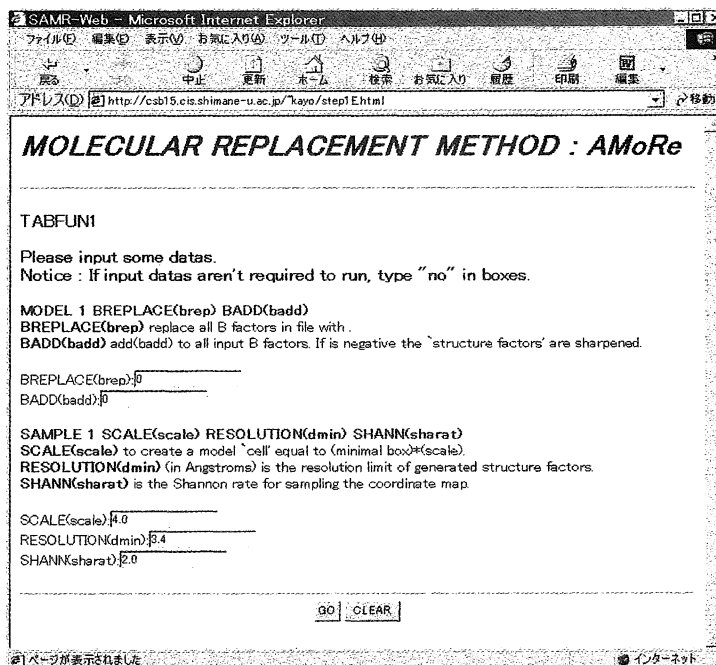


図 4 MOLECULAR REPLACEMENT METHOD: AMoRe のタスクウインドウ

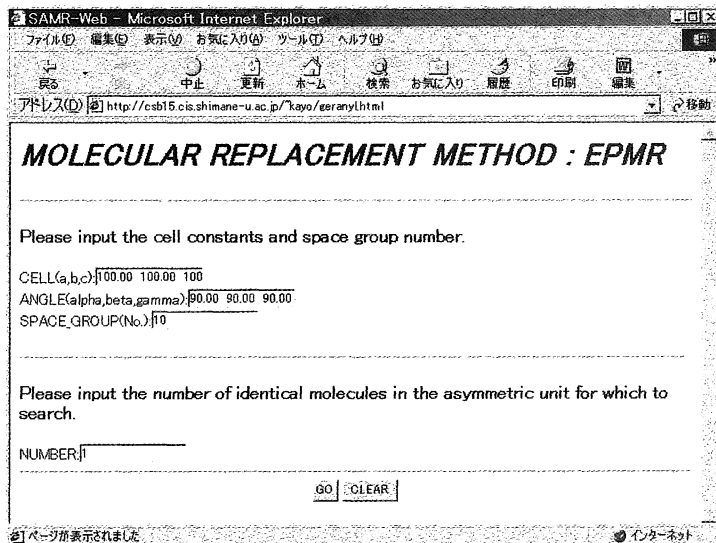


図 5 MOLECULAR REPLACEMENT METHOD: EPMR のタスクウインドウ

5. ま と め

分子置換法での解析の成否のカギを握っているモデル構造を、ユーザーサイトのブラウザで一貫してほぼ自動的に作成できるようにしたことで、これまでの煩雑で時間を要していたモデル構造の作成を短時間で容易にできるようにした。分子置換法プログラムを実行するには、測定した回折データとモデル構造の他に、種々のパラメータを入力する必要があり、蛋白質結晶学の知識を必要としていたが、本システムでは、それらを自動生成できるようにし、蛋白質結晶学の深い知識を必要としない。

本システムでは、SWISS-MODEL サイトで構築されたモデル構造をメールで受信する。しかし、これには時間がかかるだけでなく、ユーザーに煩雑な処理を強いている。これは、モデリングのプログラムを Web サーバーにインストールすることで改善される。利用して頂いた人々たちから、いくつかの改善点を指摘されている。今後これら問題点を検討し、さらに利用しやすいシステムに改善していく予定である。

参 考 文 献

- [1] A. D. Baxevanis (2000) The Molecular Biology Database Collection: an online compilation of relevant database resources, *Nucleic Acids Res.* **28**, 1-7.
- [2] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne (2000) Protein Data Bank, *Nucleic Acids Res.* **28**, 235-242.
- [3] J. Drenth (1994) *Principles of protein X-ray Crystallography*, Springer.
- [4] Collaborative Computational Project (1994) The CCP4 Suite: Programs for Protein Crystallography, *Acta Cryst.* **D50**, 760-763.
- [5] J. Navaza (1994) *AMoRe*: an automated package for molecular replacement, *Acta Cryst.* **A50**, 157-163 (1994).
- [6] C. R. Kissinger, D. K. Gehlhaar and D. B. Fogel (1999) Rapid automated molecular replacement by evolutionary search, *Acta Cryst.* **D55**, 484-491.
- [7] N. Guex and M. C. Peitsch (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling, *Electrophoresis* **18**, 2714-2723.
- [8] S. Asbury, M. Glover, A. Humphereys, E. Weiss, J. Methews and S. Sol (著), アクロバイト (訳) (1998) Perl 5 パワフルテクニック大全集, インプレス.
- [9] 三島俊司 (1998) CGI のための実践入門 Perl, 技術評論社.
- [10] M. McMillan (著), トップスタジオ (竹内里佳) (訳) (1998) Perl 5 マスターブック, 翔泳社, 1998.