

# 蛋白質立体構造データベースの構築と利用システムの開発

大阪大学蛋白質研究所 楠木 正巳

Research theme : Construction of the three-dimensional structure database of proteins and development of a software system to utilize the database.

Affiliation : Institute for Protein Research, Osaka University

Name : Masami Kusunoki

Abstract :

The Protein Data Bank is the sole archive of the three-dimensional structures of biological macromolecules and its entries has dramatically increased recently. We have started the deposition and processing of PDB entries from Japan and Asia-Oceania regions in July, 2000. We have developed a software system to solve the crystal structure of a protein automatically by molecular replacement using the Protein Data Bank on a WEB basis and now are improving its performance.

## 1. 緒言

プロテインデータバンク ( P D B ) は、蛋白質や核酸などの生体高分子の立体構造データベースであり、米国のRCSB (Research Collaboratory for Structural Bioinformatics) が中心となって運営している。P D BにはX線結晶解析、N M R解析など実験的方法で決定された生体高分子の原子座標、立体構造の記述、実験方法、参考文献などが記載され、さらにX線回折データも収録されている。P D Bは生体高分子の立体構造のアーカイブであり、エントリー数の急激な増加とともに構造生物学、医薬品開発などの分野での重要性が大きくなっている。

本研究課題の1番目の研究開発題目は、「蛋白質立体構造データベース編集システムの構築と運用」である。プロテインデータバンクを構築・運営する国際的な協力体制の一環として大阪大学蛋白質研究所においてアジア・オセアニア地区のP D Bエントリーの登録・編集することを企画し、現在、エントリーの登録・編集体制を構築し、順調に運用している。

本研究課題の2番目の研究開発題目は、「蛋白質立体構造データベースを利用した全自動分子置換解析システムの開発」である。この研究では、蛋白質のネイティブ結晶のX線回折データだけを用いる分子置換法に基づき、利用者はインターネット上のWEBサーバに必要なデータを入力することにより構造解析を実行するシステムを開発することである。現在、データの入力から構造解析のシステムの表示までのシステムが完成し、今後より性能をあげるための改良を行っている。

本研究課題の3番目の研究開発題目は、「蛋白質基本構造解明プロジェクトへの協力」では、2番目の研究開発題目がほぼ実用化する来年早々に協力ができるものと考えている。

## 2. これまでの研究の進展

### 2-1. 研究開発題目1「蛋白質立体構造データベース編集システムの構築と運用」

#### a. プロテインデータバンクの検索、登録サーバの設置

蛋白質立体構造データベース編集システムの構築と運用の時間経過を表1に示す。

表1 大阪大学蛋白質研究所のプロテインデータバンクの活動の経緯

1999年6月末	P D B検索サーバのミラーを設置し、運用を開始した。
2000年2月	P D Bエントリーの登録サーバ ( A D I T ) の試験的運用の開始
2000年7月1日	P D Bエントリーの編集業務を開始した。

1999年6月末に米国RCSBが開発したWEB上のPDB検索サーバのミラーを、大阪大学蛋白質研究所に設置し運用している。それまでは米国ブルックヘブン国立研究所で開発されたPDBの検索サーバを運用し、さらにPDBの登録システムAutoDepを設置し運用を予定していたが、1999年6月30日をもって米国でのPDBの運営がブルックヘブン国立研究所からRCSBに移ったため、RCSBのシステムを採用することになった。

PDBの登録と編集のシステムを図1に示す。

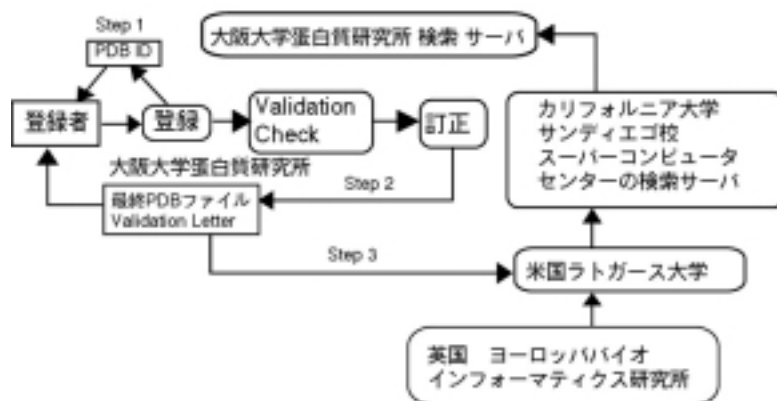


図1 PDBエントリーデータの登録・編集処理の流れ

#### b. 大阪大学蛋白質研究所における編集体制の整備

図1に示すようにPDBのエントリーは、蛋白質の立体構造を決定した研究者が、米国ではラトガース大学、ヨーロッパではEuropean Bioinformatics Institute、アジア・オセアニア地区では大阪大学蛋白質研究所において運用しているWEBサイトを通して登録する。2000年2月から蛋白質研究所で試験的にPDBエントリーの登録を開始した。2000年2月と3月にそれぞれ1ヶ月間、PDB編集員の1人が、ラトガース大学にて編集の訓練を受けた。そして、2000年6月21日から公式にPDBエントリーの登録を受け、7月1日から編集作業も蛋白質研究所で行っている。現在、蛋白質研究所にはPDB編集員が2人いるが、3人目を募集中である。登録する項目数は数百にのぼり、登録には1日または2日かかる。登録されたPDBエントリーは、PDBの編集員が半日から3日かけて、各種データの矛盾点が無いかどうかチェックし、形式を整える。この作業に0.5日から3日かかる。2000年10月末時点で、編集処理したPDBエントリー数は79個である。

また、日本におけるPDB編集活動の一般登録者へのメリットとして、登録作業の日本語のチュートリアルなどを作成して公開している。

### 2-2 研究開発題目2「蛋白質立体構造データベースを利用した全自動分子置換解析システムの開発」

#### a. 全自動分子置換解析システムの概略

蛋白質の結晶解析の方法には、重原子同型置換法、多波長異常分散法(MAD)、分子置換法などがある。分子置換法では、重原子同型置換法やMAD法と違って、ネイティブの蛋白質のX線回折データだけで構造解析できる利点がある。しかし、前二者とは違い、立体構造がある程度以上類似している蛋白質の原子座標が必要である。結晶構造解析された蛋白質の構造がすでに1万3千を超えたこと、蛋白質の立体構造の基本型の数が有限であることから、分子置換法で十分構造解析できる蛋白質結晶の割合が今後増えていくことが予想され、有力な方法となりつつある。

我々の分子置換法の自動解析システムの開発では、プログラムはすべて独自開発を心がけ、外国のプログラムは入っていない。当初、外国製のプログラムを組み込み、データの入力から出力までのシステム全体を構築した

後、外国製のプログラム部分を独自開発のプログラムに置き換えた。図2に我々の分子置換プログラムの概略を示す。

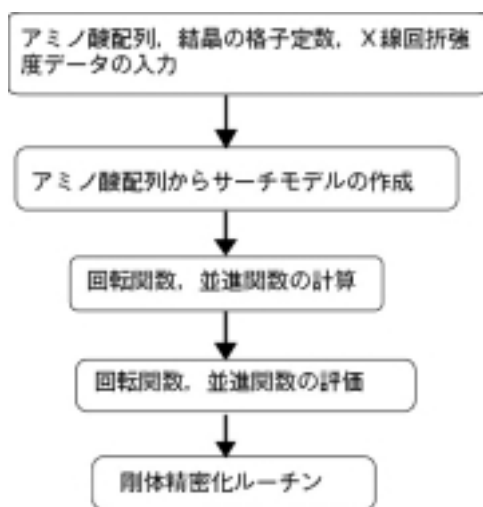


図2. 分子置換法自動結晶解析システムの処理の流れ図

#### b. 分子置換サーチモデル構築プログラムと回転・並進関数プログラムの開発

立体構造未知の蛋白質のアミノ酸配列から、分子置換のためのサーチモデルを構築するアルゴリズムを2種類開発し、プログラミングした。一つは、アミノ酸配列の類似性を手がかりにプロテインデータバンクのエントリーをサーチし、複数のエントリーが見つかった場合、その分子置換法に最適の平均構造を構築する方法である。もう一つは、その蛋白質のリガンドや基質が既知である場合、リガンドまたは基質の結合サイトを手がかりに分子置換法のサーチモデルを構築するアルゴリズムである。

回転関数・並進関数は東京工業大学の田中信夫教授のグループによって開発された。この方法では、最近の計算機の速度の進歩を考え、回転関数と並進関数の近似を従来の方法より減らし、より感度の高いアルゴリズムを開発し、プログラミングした。

#### c. 全自動分子置換解析システムのセキュリティと運用

全自動分子置換解析システムを実際に運用するためには、計算が正常に実行すること以外に、入力された貴重なX線回折データが、不正な手段で第三者に漏洩しないこと、システムの稼働記録が自動的に採取され、エラーフリーで安定に稼働することが必要である。これらのためのプログラミングはソフトウェアハウスに外注しており、我々は、アルゴリズム等の研究開発に専念している。

### 3. ネットワークの活用について

研究開発題目1「蛋白質立体構造データベース編集システムの構築と運用」では、アジア地区からPDBのデータ登録者からインターネットを通して、蛋白質研究所にデータファイルの転送がある。蛋白質研究所のPDB検索サーバへの一般利用者のアクセスは頻繁にある。蛋白質研究所で編集済みのPDBエントリーは米国ラトガース大学に送られる。PDBの検索サーバは、週1回米国カリフォルニア大学サイディエゴ校スーパーコンピュータセンタから更新ファイルが蛋白質研究所に送られる。

研究開発題目2「蛋白質立体構造データベースを利用した全自動分子置換解析システムの開発」では、まだ一般に公開されていないので、ネットワーク利用はまだない。

#### 4．まとめ及び今後の予定

研究開発題目1「蛋白質立体構造データベース編集システムの構築と運用」は、予定されていたPDBのデータ構築体制はすべてセットアップされ順調に推移している。しかし、今後データの登録数が増加したときに備えた体制整備が必要であり、データベース編集に必要な計算機の更新などが必要である。

研究開発題目2「蛋白質立体構造データベースを利用した全自動分子置換解析システムの開発」では、データの入力から構造解析の結果の出力まで実行するシステムは完了したが、パフォーマンスをあげることに、一般の人の研究中のデータを取り扱うのでセキュリティ確保の対策が必要である。

#### 5．研究実施体制

研究開発題目1「蛋白質立体構造データベース編集システムの構築と運用」では、データの登録・編集体制についてはラトガス大学と緊密な連絡と協力体制を取って、PDBの編集作業を軌道に乗せた。今後も、増え続けるデータに対応できるように、より効率的な協力関係構築にむけて努力が必要である。RCSBのグループとは少なくとも1年に1回会ってPDB活動の維持・発展のための議論を行っている。

研究開発題目2「蛋白質立体構造データベースを利用した全自動分子置換解析システムの開発」では、東京工業大学の田中信夫教授にグループと緊密な共同開発体制を今後も維持していく予定である。