



**PDBj**  
Protein Data Bank Japan

日本蛋白質構造データバンク

# News Letter Vol. 6

2005年8月

<http://www.pdbj.org/>

PDBj は大阪大学蛋白質研究所が  
(独) 科学技術振興機構の支援により運営しています。

PDBj





### 3. 配列検索と GASH 構造アラインメント・プログラムの改良

前回のニュースレターでは、私たちは PDB ID を入力することで配列や類似構造を割り出す「配列検索」、および「構造探索」するウェブツールを紹介しました。また、私たちは新しい構造アラインメント・プログラム、GASH も紹介しました。ここでは、主に“Structure Navigator”と“GASH”についてご紹介します。

“Structure Navigator”は私たちに配されているコンピューティングプラットフォーム：15PC クラスターの長所を活かして書き直されました。以前の“Structure Navigator”は、検索したい構造と同様の構造のリストを作り出すために多数の操作を行なわなければなりません。類似構造を簡単に検索することができるように、予めほとんどあらゆる PDB ID に対する結果を格納したため、現在、これらの操作を行う必要はありません。

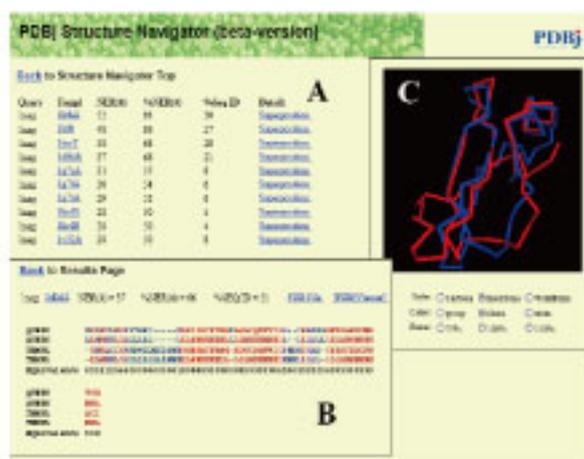
類似構造リストがあらかじめ計算される際に、相同性配列構造を含んでいます。ユーザは、相同性配列構造を取り除くかどうかという選択をすることができます。つまり、“Sequence Navigator”の機能性の大部分は“Structure Navigator”に組み入れてあるということです。1つの例外が、PDB ID からではなく、実際の配列からなる検索をしたときにあります。そのような場合が、“Structure Navigator”の現在の課題です。

新しい“Structure Navigator”と関連するデータバンクを構築する際の一つの課題は、容易に検索できるような方法で、ほぼ 1 テラバイトのデータを格納することでした。この問題は、PC クラスター上の 14 個のローカルディスクにデータを分けることによって、解決されました。コードの書き直しに加えたこれらの改良は、検索時間を最大 1 分からわずかに数秒までに減少させました。

基本的なコードとデータバンクをより効率的にするための作業と平行して、入力と出力データの新しい XML バージョンを開発してしています。Java ベースのアプリケーションか簡単なアクセスプロトコル (SOAP) サーバによって、XML データにアクセスすることができます。この実装は私たちの前の CGI インタフェースよりはるかに安定しています。

また、GASH 構造アラインメント・プログラムも改良されています。コードは、500 未満の残基での検索にかかる平均 CPU 時間がおおよそ 15 秒程度であるように調整され、100 未満の残基での検索にかかる時間は、おおよそ 5 秒とサイズに従って減少しています。これは一つのアラインメントを成すために最大 1 分を必要とした旧バージョンよりはるかに改良されています。GASH は 3,000 構造以上のデータセットの DaliLite と CE に対して評価基準をしています。また、一定の RMSD の下に配列された残基の数か相同残基 (NER) の数のどちらかに関して、GASH は一貫して DaliLite、或いは CE と同様、若しくはそれ以上に優れています。

“Structure Navigator”と同様に、GASH の入出力データは XML フォーマットで記述されています。GASH データの XML バージョンを利点を活かしたウェブベースの Java インタフェースとコマンドライン SOAP のサービスの両方が PDBj サイトを通して利用可能です。



類似構造の表示例：(A) 個々のアラインメント、(B) 構造の重ね合わせ、(C) 検索された構造モデル。

## 4. xPSSS における EDM の新たな進展

PDBj ニュースレターの第5巻、2004年12月号で、PDBjにおける電子密度マップ可視化 (EDM) サービスを紹介しました。今回は、この EDM の進展についてレポートします。

これまでの EDM サービスの対象は、2004 年半ばまでの PDB データに限定されていました。構造座標と構造因子のデータは急速に増加し、当時 12,310 件であったのが2005年7月1日の時点で16,857件となっています。

登録されたデータに対する精密化とうまく精密化された構造に対する電子密度マップの計算は、適当な CCP4 ソフトにインターフェースされた Python 言語で行われています。EDM サービスを常に最新のものに保つため、これらの記述を完全に自動化するよう発展させました。日本の PDBj において保守がなされ、毎週更新されているミラーサイト (<ftp://pdb.protein.osaka-u.ac.jp/pub/pdb/>) のデータを参照して作業がなされています。私どもの EDM サーバでは、PDB の座標データと構造因子データを、毎週水曜日にミラーサイトからコピーしてダウンロードし、新たなエントリーに対しては精密化と電子密度マップの計算を行い、古くなって PDB への登録が削除されたものに対してはこれらのデータも削除します。こうして、今日、EDM サービスでは、16,207 の精密化された構造と 15,250 の電子密度マップを提供しています。

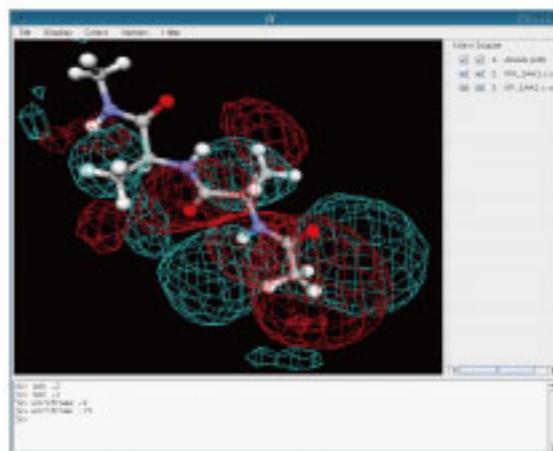
OpenGL を用い、スタンドアローンと Java アプレットとして利用できるグラフィックビューアの *jV* では、さらに *jogl* ライブラリを必要とします (<http://www.pdbj.org/PDBjViewer/>)。この Java アプレットとしての利用は、Linux 上での Opera 8.01 と Mozilla Firefox 1.0.4、および Windows 上での Internet Explorer での稼動が確認されています。推奨される Java run-time 環境は 1.4.2\_05 ですが、より新たなバージョンも有効であり、また 1.5.0\_03 は Linux 環境で有効です。

EDM サービスで利用される電子密度の等高線と等値面を発生するソフトウェアは、3次元のグリッドデータ上で稼動します。そのため、グリッドデータを用いるその他のデータにも簡単に応用できますので、私たちは、分子モデリングへの応用を行っております。

実際、分子の電荷や分子軌道を計算し、BMSML (BioMolecular Simulation Markup Language <http://www.biogrid.jp/BMSML/>) と称する XML 記述によるフォーマットによってこれらデータが蓄えられます。これらの BMSML ファイルを、C、C++ および Fortran プログラムから読み出すための応用プログラムライブラリ (API) が用意されており、私たちは Python との結合のために Fortran 用の API を用いています。これらの BMSML ファイルを可視化し、電子密度マップを描くのに使われた同一の XML のポリゴンファイルのフォーマットで等高線を描いたり等値面を発生するために、EDM の等高線を描くプログラムがわずかな書き換えによって利用されています。

上の図は、Ala-Ala のジペプチドにおける分子軌道の等高線を表示している例であり、青は正、赤は負の軌道を示します。この正と負の等高線は、それぞれ別々のファイルで作られているため、これらの表示は選択的に on-off することが可能です。

右記の表は、エントリー数の増加を示します。精密化されたエントリー数は、登録された全件数の約 50% です。PDB のこれまでの年数を考えると、この増加率はたいへん急激で、構造ゲノムプロジェクトの活動によるところが大きいといえましょう。



Ala-Ala 分子軌道の表示例。

	2004年6月	2005年6月	増加率(%)
精密化エントリー数	11,971	16,207	35.4
電子密度マップ数	11,284	15,259	35.1

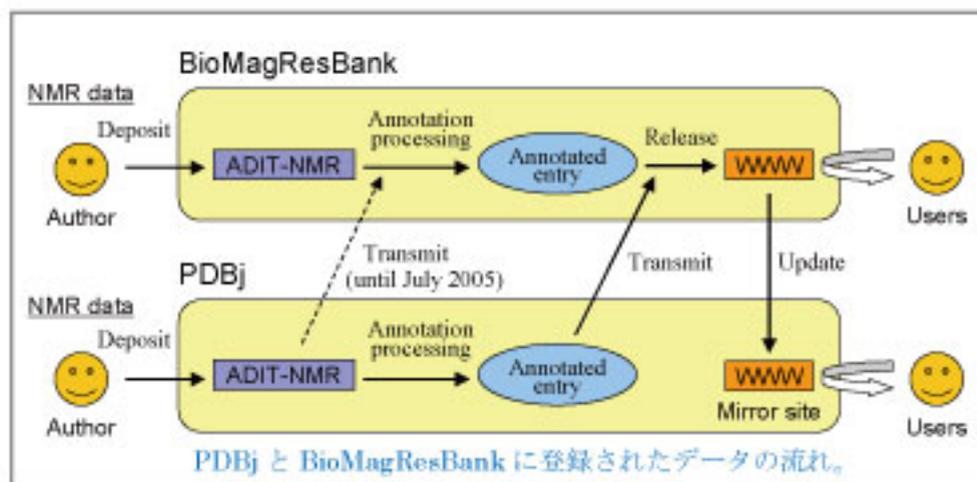
## 5. BMRB データベースのミラーと登録の開始

BMRB (NMR) データベースグループ (共同研究者: 阿久津秀雄 教授) は BMRB (BioMagResBank; ウィスコンシン大学マディソン、PI, John. L. Markley) のミラーサイトの運営を 2002 年 3 月に開始しました。2004 年 12 月に BMRB の新しいデータ登録インターフェイス「ADIT-NMR」を PDBj に移植したのを機会に、私たちは BMRB の NMR データ登録の受け入れを開始しました。



ADIT-NMR のトップページ (左) とデータ入力画面 (右)。

URL: <http://bmrbadit.protein.osaka-u.ac.jp/bmrbadit/>



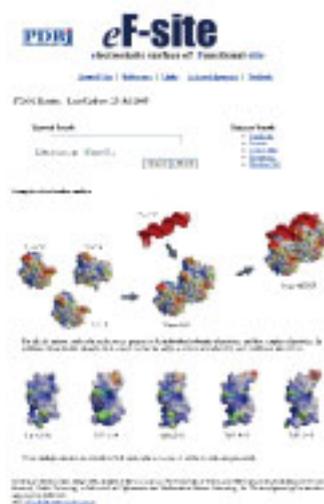
PDBj と BioMagResBank に登録されたデータの流れ。

さらに、BMRB グループでは PDBj に登録されたデータのデータベース化処理を 2005 年 8 月より開始しました。登録は PDBj でデータベース化処理が行われた後に BioMagResBank に転送されます。BioMagResBank による一連の手続きが完了された後、登録は BMRB データベース WWW サーバに追加されて公開されます。

## 6. 新しい eF-site データベースの公開

分子表面のデータベース eF-site が、内部構成を一新しました。新しい eF-site では、一部大きすぎて計算が難しいエントリーを除いて、ほぼすべての PDB エントリーが既に登録・公開されています。また、今後は PDB の日々の更新とも連携して更新が行われる予定です。

これまで eF-site は低分子の結合部位の予測法の開発のための基礎データベースであることを目標に開発を行ってきましたが、今後は、これからますます重要になってくるタンパク質間相互作用の研究にも資することを目指して、それぞれの PDB エントリーは、サブユニット毎と全体構造の両方で分子表面及び静電ポテンシャルの計算を行い登録してあります。また、NMR で決定された構造で複数の構造が登録されているエントリーに関しても、すべてのモデルで同様の計算を行った結果を登録してあります。その結果、2005 年 7 月 6 日現在、170,733 エントリーが登録されています。



eF-site のトップページ。

## 7. PDBj スタッフ (大阪大学蛋白質研究所)



後列左から

楠木正巳、池川恭代、山下鈴子、松木陽、Paehler, Arno、吉原淳郎、  
見学有美子、五十嵐令子、清水有希子、中谷英一、高橋亜紀、Standley, M. Daron、  
鎌田知佐、佐伯忍、中村春木、小佐田高史 (敬省略)

### 統括責任者

中村春木 (大阪大学蛋白質研究所・教授)

### PDBj データベース管理運営グループ

楠木正巳 (大阪大学蛋白質研究所・助教授)

小佐田高史 (大阪大学蛋白質研究所)

五十嵐令子 (JST-BIRD)

見学有美子 (JST-BIRD)

池川恭代 (JST-BIRD)

佐伯忍 (JST-BIRD)

### 新規蛋白質立体構造データベース構築グループ

伊藤暢聡 (東京医科歯科大学・教授)

Paehler, Arno (JST-BIRD)

Standley, Daron M. (JST-BIRD)

山下鈴子 (JST-BIRD)

吉原淳郎 (JST-BIRD)

清水有希子 (JST-BIRD)

高橋亜紀 (JST-BIRD)

### 解析システム開発と二次データベース構築グループ

大川剛直 (神戸大学・教授)

木下賢吾 (東京大学・助教授)

藤博幸 (九州大学・教授)

輪湖博 (早稲田大学・教授)

### BMRB データベース管理運営グループ

阿久津秀雄 (大阪大学蛋白質研究所・教授)

松木陽 (JST-BIRD)

中谷英一 (JST-BIRD)

### 研究チーム事務員

鎌田知佐 (JST-BIRD)

### PDBj

〒565-0871

大阪府吹田市山田丘3-2

大阪大学蛋白質研究所・附属プロテオミクス総合研究センター内

PDBj事務局：TEL(06)6879-4311、FAX(06)6879-8636

PDBjデータベース登録業務：TEL(06)6879-8638、FAX(06)6879-8636

E-mail:pdbjADMIN@protein.osaka-u.ac.jp