



日本蛋白質構造データバンク

News Letter Vol. 5

2004 年 12 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。

1. ニュース

wwPDB 諮問委員会

国際蛋白質構造データバンク (wwPDB) の第1回諮問委員会 (wwPDBAC: world wide Protein Data Bank Advisory Committee) が、米国ワシントン DC において2004年11月21日に開催されました。Structural Genomics 社の Stephen K. Burley 博士が議長となり、その他に下記の方々が参加されました。(1) PDB 各サイトの代表諮問者: Wayne A. Hendrickson 教授 (RCSB, コロンビア大学), Gerard Kleywegt 博士 (MSD, ウプサラ大学), James H. Naismith 教授 (MSD, セント・アンドリュース大学), 西村 善文 教授 (PDBj, 横浜市立大学), 若槻 壮市 教授 (PDBj, フォトンファクトリー), (2) 国際学会の代表者: Edward N. Baker 教授 (IUCr, オークランド大学), R. Andrew Byrd 博士 (ICMRBS, 米国立癌研究所), (3) PDB 各サイトのリーダー: Helen M. Berman 教授 (RCSB, ラトガース大学), Kim Henrick 博士 (MSD, 欧州バイオインフォマティクス研究所), 中村 春木 教授 (PDBj, 大阪大学), (4) 運営開発基金の配分機関代表者: Christopher L. Greer 博士 (NSF), Michael Dunn 博士 (Wellcome Trust), Deborah Colson 博士 (Wellcome Trust), Josefina Enfedaque 博士 (European Commission), 小池俊行 氏 (JST), 高橋秀貴 氏 (JST)。

wwPDB の憲章にも書かれておりますように、wwPDB のミッションは、世界中の誰もが無料で利用できる唯一の生体高分子構造データ・アーカイブを維持することにあります。そこで、wwPDB 諮問委員会は、その役割として、「永続的に無料でアクセスできる公共財産として、構造生物学の実験データに対する、安全で唯一の国際的アーカイブを PDB が維持することを援助・保証すること」であると合意いたしました。

会議は、Helen Berman 教授による PDB と wwPDB に関するレビューから始まり、wwPDB の3つのサイトの代表者 (RCSB, MSD-EBI, PDBj) からそれぞれのサイトの活動状況について短い報告がなされました。次に、参加者全員が、上記した wwPDB 諮問委員会のミッションについて合意・確認をしました。そして、PDB ファイル・フォーマット、PDB エントリーの改善、X 線結晶解析・NMR スペクトル解析・電子顕微鏡解析の構造生物学実験データの登録問題、実験データの品質管理問題、wwPDB および wwPDB 諮問委員会メンバー間の今後のコミュニケーション等、様々な問題が議論されました。

wwPDB 諮問委員会は、全員が合意した議事内容と勧告を記述したレポートを作成しましたので、wwPDB メンバーは、これらの親切な勧告に感謝し、できるかぎり対応したいと思います。今回の第1回国際諮問委員会は、将来への大きな一歩となることを確信しております。

RCSB からの訪問者

PDBj で2ヶ月間、アノテーション法を共有し交流を図るため来日された RCSB の Suzanne Richman さんよりコメントを頂きました。

まず、今回の私の受け入れをしてくださった PDBj の方々にお礼を申し上げます。私がこちらに来て以来、礼儀正しく、とても親切に助けて頂いていることをありがたく思っています。

当初、地下鉄のシステムがとても難しく思いましたので、2ヶ月間、寮を出ずに滞在しようと思っていました。しかし、私は直ぐに勇気を出し、日本、主に関西地区を探索するという素晴らしい時間を持ちました。日本でマシュマロを買うことさえ簡単に出来るように、世界がグローバル化に向かうなか、日本文化の多くの側面が本当に独特であることを知ってホッとしました。2、3例を挙げてみます。特に食べ物（たこ焼き、おにぎり、納豆、お好み焼き、卵寿し、おたべ、梅干し）はどれも私にとっては初めての物ばかりでした。そして、車より自転車が駐車スペースを多く占め、数え切れないくらいの超高層ビル、背の高いアパートの建物が古い民家と隣接して並んでいます。更に、通りの各角には賑やかな照明のパチンコ店があります。また、京都では数え切れないくらいの美しい寺院もありました。

私は、日本語の語彙や言いまわしとほんの少し、「ひらがな」を楽しく覚えました。時々、言葉の壁でコミュニケーションをとるのが難しいと感じることもありますが、心からの笑顔や笑い、お辞儀、そして絵を描いたり、ジェスチャーをすることが、話すよりも効果的なこともあります。

言葉や文化の違いがあるにもかかわらず、PDBj で働くことは、アメリカとほとんど変わりが無いように感じます。地球の反対側であっても、一つのプロジェクトの中で、同じ考えや気持ちを持って、私たち

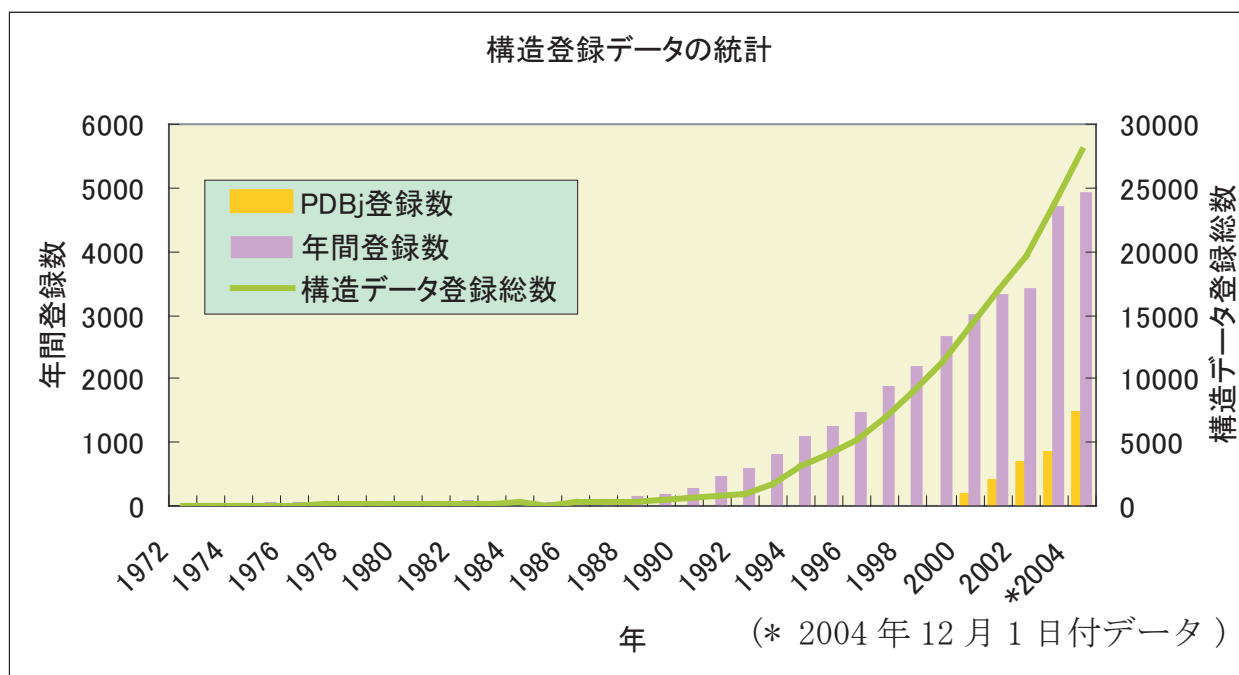


Suzanne Richman

は働いています。そして、アノテーション部屋は怖いほど静かで全ての作業を、集中してすることができます。

私は大阪で過ごした全ての時間がとても好きです。そして、このような素晴らしい体験が出来たことをとても感謝しています。

2. PDBj 構造登録データ統計

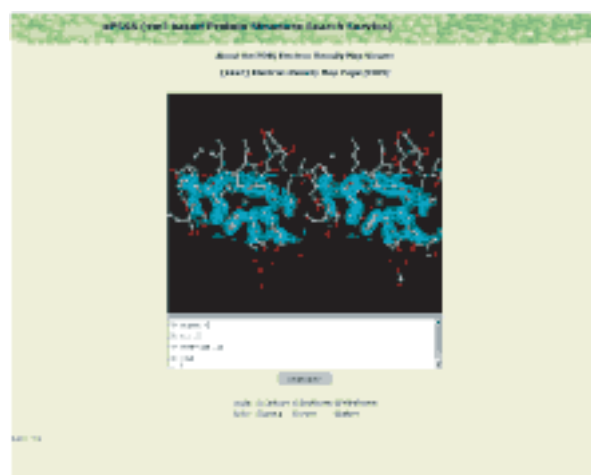


3. 電子密度マップ (Electron Density Map) と xPSSS

PDBj では PDB データのブラウザーとして xPSSS (xml-based Protein Structure Search Service) のサービスを提供しています。このサービスの一部として、分子構造モデルの画像を会話的に表示するビューアがあります。ウェブ上では Java アプレットとして使用されていますが、スタンドアローンとしても使えます。私たちは、最近このビューアを利用して、電子密度マップの表示機能を追加しました。

PDB は、現在 28,000 件以上のデータが登録されています。これらの登録数の 85% が X 線結晶解析により構造を決定したものであり、そのうちの 65% には、実験による構造因子の振幅データが登録者によって提供されています。このデータにより電子密度マップが計算されます。この電子密度マップは、結晶学を専門としない多くの PDB ユーザにとっては馴染みが薄いかも知れませんが、構造を決定するための基本的な実験データとして重要なものです。

電子密度マップの可視化は、ウプサラ電子密度サーバ (EDS: Uppsala Electron-Density Server) [<http://eds.bmc.uu.se/eds/>] において、かなり以前から行われています。このサーバは、G. J. Kleywegt らによって作られました¹。PDBj の Arno Paehler 博士によって開発された私たちのサーバも、基本的には同様の方法を採用していますが、EDS とは少なくとも次の 2 点の重要な違いがあります。



xPSSS での EDS 表示例。

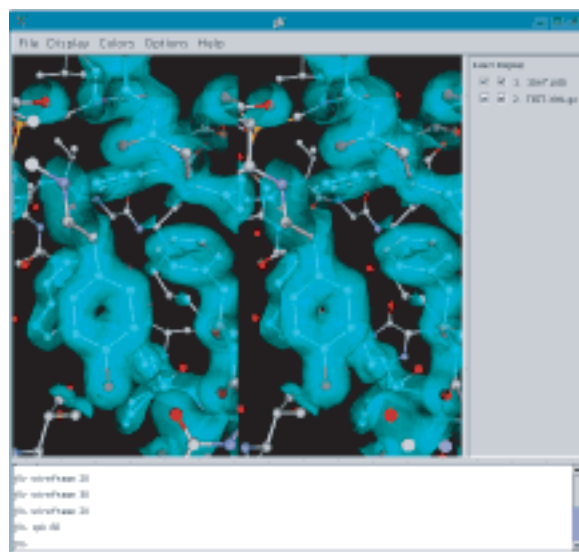
まず、EDSでは、CCP4 のプログラム:refmacを用いて、位相を単に計算するために、構造因子と分子モデルの構造情報を利用しているだけですが、私たちは同じrefmac プログラムを用いて、それらの情報を基に10サイクルの構造精密化を行っています。ある場合には、登録された分子モデルと構造因子データとの対応が異なっていることもあり、もし、この違いが大きければ、EDS はマップを作り出すことが出来ません。私たちは、そのような場合でも、引用されている構造因子とモデルとの一致を構造精密化計算によって再構築し、精密化を改善できることさえあることを見出しています。R 因子の値の分布の平均値および中央値は、登録された元の構造に比べ、私たちの精密化した構造では約1.5 %も低くなっています。

次に、EDS は電子密度の等高線だけを表示していますが、これは金網のような表示として知られています。私たちは同様の等電子密度線も使用していますが、それに加え等値表面によって電子密度が可視化されます。全ての必要とされる計算は、Python と、大きな計算資源を要するものに対しては Python と Fortran90 の組み合わせによって実施されています。

ユーザは電子密度マップの表示を操作するために、会話的にいくつかのパラメーターを選ぶことができます。電子密度は現在、PDB の2004年5月時点に基づき、登録数の11,284に対して利用することができます。それ以来、約2,000の構造因子データが追加されましたので、近いうちに、それらに従い私たちの電子密度マップをアップデートする予定です。電子密度マップが登録されているエントリーに対して利用される際には、ウェブページ上でxPSSSにより表示されます。等電子密度線または等電子密度面は、あらかじめ計算された電子密度マップから、ユーザのリクエストに応じて直ちに構築され、会話的にPDBj viewer上で表示されます。

会話的に利用されるアプレット版もスタンドアローン版の両方のプログラムとも、OpenGL を用いたJavaで記述されています。そのため、Java と、Java 用の OpenGL のインストールが必要です。構造精密化および電子密度マップの計算に使われる構造因子データと、電子密度マップそのもの、そして構造精密化された分子モデル構造が、それぞれダウンロードできます。私たちが使用しているマップのフォーマットは、歴史的なプログラム Frodo とその後継プログラム0において、マップを記述されるためにバイト値で表されるフォーマットに大変良く似ています。これらのファイル・フォーマットの詳細情報は、PDBj のウェブ上で入手できます。

Kleywegt, G. J., Harris, M. R., Zou, J., Taylor, T. C., Wahlby, A. and Jones, T. A.
The Uppsala Electron-Density Server.
Acta Crystallographica (2004) D60, 2240-2249



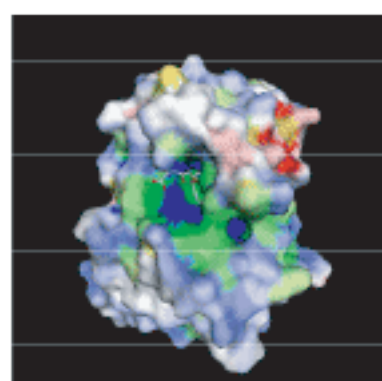
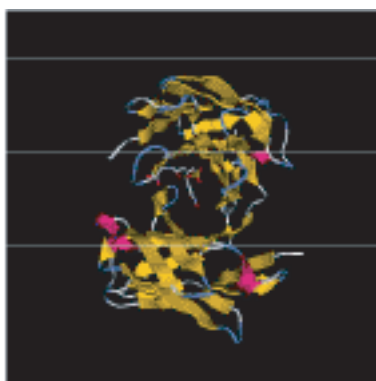
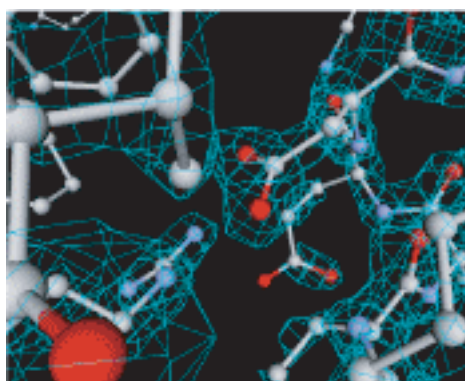
jV3 スタンドアローン版による EDM の表示例。

4. jV: PDBj ビューアー

jV version 3

PDBj viewer の最新版である jV3 が公開されました。実行形式とソースコードがフリーにダウンロード可能です。このバージョンは Java と OpenGL を利用することにより、windows 2000/XP、マッキントッシュ OS-X、Linux 環境で、applet や stand alone での利用が可能です。この最新版では、Rasmol や Chime などで利用可能な機能がほぼすべて利用可能な他に、複数分子の表示や、一般的なポリゴンの表示、アニメーションの表示、他のサーバとの http を介した情報のやりとりが可能になっています。

jV3 は既に、eF-site (<http://ef-site.protein.osaka-u.ac.jp/eF-site/>), DNA 結合部位予測サーバ (<http://pre-s.protein.osaka-u.ac.jp/~preds/>), PDBj 検索 (<http://pdbjs3.protein.osaka-u.ac.jp/xPSSS/>), Active site 予測サーバ (<http://bioinfo.tsurumi.yokohama-cu.ac.jp/p-cats/>), タンパク質構造百科辞典 (<http://eprints.protein.osaka-u.ac.jp/eProtS/>) など、いくつかのサーバでのビューアーとして利用されています。



JV3による様々な表示例。

5. 配列検索と構造検索

PDBjにおけるテキストベース領域を超えた新検索法：配列探索と構造探索

2004年12月、PDBjはウェブサイトにも三つの新しいサービスを加えました：

- SequenceNavigator (http://www.pdbj.org/cgi-bin/run_seq_hom.cgi),
- Structure Navigator (http://www.pdbj.org/cgi-bin/run_algn_struc.cgi),
- GASH (http://www.pdbj.org/cgi-bin/run_gash.cgi).

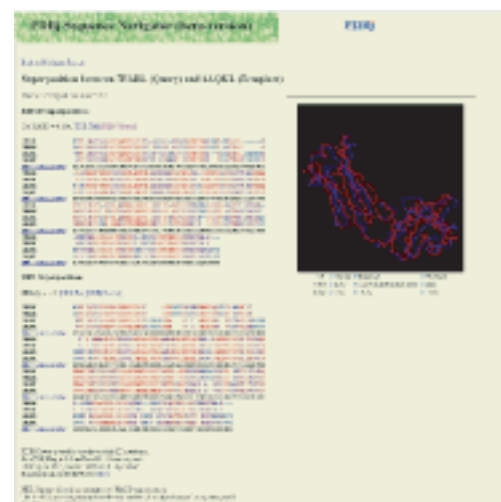
PDBjのDaron Standley博士によって開発されたSequence Navigator とStructure Navigatorは、配列の相同性と構造の類似性をそれぞれ利用してProtein Data Bank (PDB)のデータベース検索を行うツールです。また、Daron Standley博士 (PDBj)と藤 博幸 教授 (京大・化研)によって開発されたGASHは、藤教授によって開発されたASHプログラムを拡張した構造アラインメント・プログラムです。

World Wide PDB (Newsletter vol. 3参照) の誕生により、PDBjは単なる蛋白質立体構造のアーカイブだけではなく、蛋白質の構造や機能の検索と解析を行う公的データベースとして展開しています。現在のPDBの登録数は28,000件を越えており、その構造の数は毎年数千単位で増加しています。従来は、このデータを検索するため、蓄積された生データファイルからキーワードに関連する領域を検索しておりました。XML技術を利用して、PDB中に文献や他のデータベースからの情報を追加しまとめることにより、一連の検索可能な領域が拡張されてきました。この目的のため、PDBjではnative XML データベースを利用し、複合検索が可能なXPath検索エンジンを検索ブラウザに加えています (Newsletter vol. 4参照)。

これらの高度なテキストベースの検索ツールとは異なり、今回私たちは、PDBのヘッダー部分にあるようなテキスト情報の検索ではなく、複雑な入力を要しないで、実験に基づくPDBデータを検索するツールを開発しました。すなわち、アミノ酸配列と立体構造を利用してPDBに登録されたエントリー間の類似関係を定めれば、テキストとして記述される分類やキーワードの制限を受けることがなくなります。更に、専門的な知識が無いユーザーでも、PDBデータの検索ができるようにいたしました。

どのPDBエントリー (検索) から始めても、Sequence Navigatorは相同性のある配列 (テンプレート) のリストを作成します。そして、どれかのテンプレートをクリックすると、BLAST配列アラインメントに基づき、検索した構造の上にテンプレートの構造を重ね描きします。構造の類似性は、偏差の二乗平均平方根 (RMSD) と、新しい指標である等価残基の個数 (NER: Number of Equivalent Residue) ²によって定量化されます。リストにあるそれぞれのテンプレートは、さらにクリックすることによって、新しい検索として使用できます。

Structure Navigatorも同じような働きをしますが、構造の類似性検索は配列の類似性検索よりもより



Sequence Navigator の表示例。

曖昧であり、対応する構造アラインメントの検索は、より計算に負担をかけます。このため、代表的な構造ペア間のアラインメントが、あらかじめASHプログラムを使用して計算され、私たちのサーバーに蓄積されています。ある検索がなされると、Structure Navigatorはまず、その代表構造のリスト中から最も適合するものを選び、配列アラインメントを利用して、実際に検索された蛋白質にその蓄積されている代表構造をマップします。数多くのエントリーを考慮するため、この処理には若干時間を要します。

いくつかの場合、アラインメントがたいへん複雑となるため、正確な結果はASHからは得られません。また、有名なDaliやCEなどを使用しても得ることはできません。このため、私たちはASHのアルゴリズムを拡張したGASHという新しいプログラムを開発しました。このプログラムではただ一つのアラインメントのみを考慮するのではなく、多数の可能なアラインメントを考慮します。GASHプログラムは、二つの構造から一つのアラインメントを計算するのに若干時間がかかりますが、複雑な構造が比較される場合でも、アラインメントはたいへん正確に行われます。

²Standley DM, Toh H, Nakamura H.
Detecting local structural similarity
in proteins by maximizing number of equivalent residues.
Proteins 2004;57(2):381-391

統括責任者

中村 春木（阪大蛋白研、教授）

PDB データベース管理運営グループ

楠木 正巳（阪大蛋白研、助教授）

小佐田 高史（阪大蛋白研）

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、教授）

Paehler Arno（JST-BIRD）

Standley Daron M.（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（阪大情報科学研究科、助教授）

木下 賢吾（東大医科研、助教授）

藤 博幸（京大化研、教授）

輪湖 博（早大社会学、教授）

BMRB データベース管理運営グループ

阿久津 秀雄（阪大蛋白研、教授）

松木 陽（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