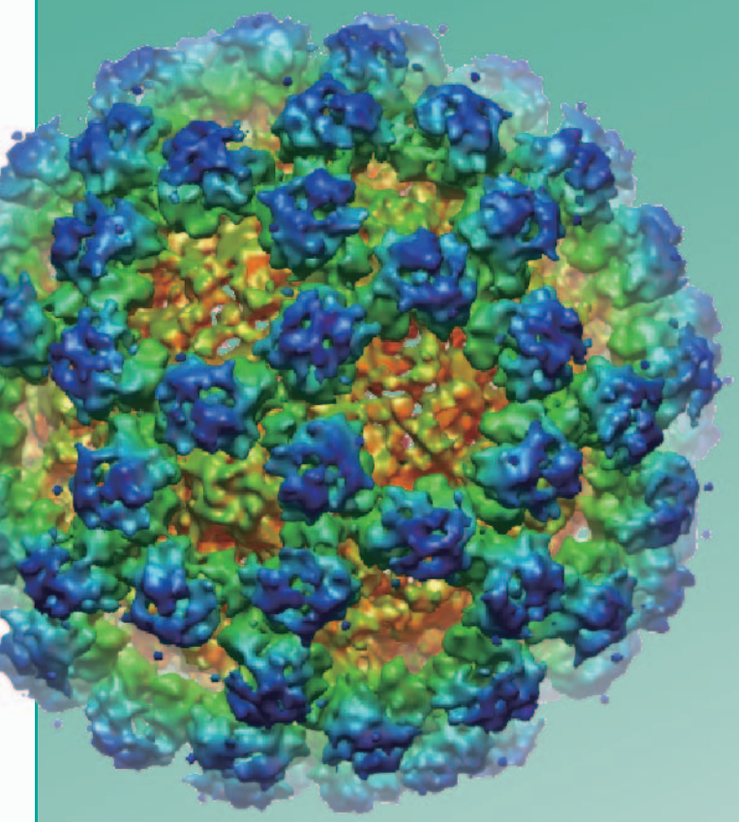
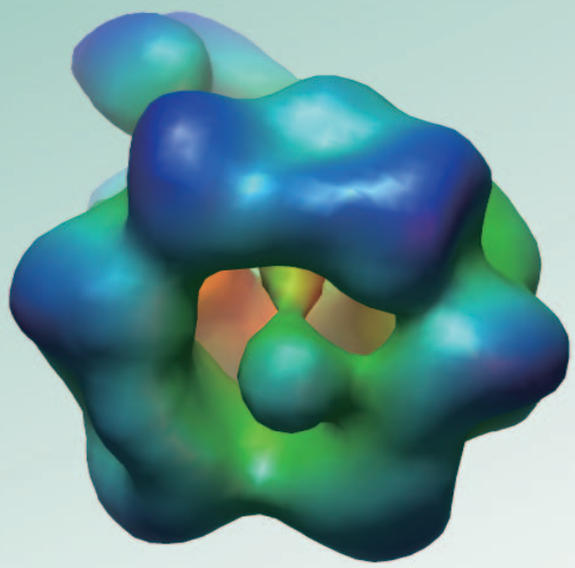
PDBj
Protein Data Bank Japan

http://pdbj.org/





Newsletter vol.15

2013 年 6 月

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って運営しています。

PDBj における PDB データ登録数が 2 万件を超えました

日本蛋白質構造データバンク（PDBj: Protein Data Bank Japan, http://pdbj.org/）では、国際蛋白質構造データバンク（wwPDB, http://wwpdb.org/）のメンバーの一員として、主にアジア・オセアニアの構造生物学研究者から登録された蛋白質や核酸、糖およびそれらの複合体である生体高分子の原子分解能の立体構造データについて品質を維持管理するための編集を行い、世界共通の無償で利用できるアーカイブ：PDB (Protein Data Bank) として、インターネットを通じて公開しております。

PDBj が 2000 年に最初の編集作業を始めてから 13 年が経ち、2012 年末には、総計 2 万件を超えるデータが PDBj にて編集されました。このデータ量は、PDBj が編集作業を始めた 2000 年における世界全体の登録量よりも多いものになっています。wwPDB 全体では、2013 年 5 月に 90,000 件を超え、NMR の化学シフトのデータベース（BMRB）や電子顕微鏡の画像情報のデータベース（EMDB, http://emdatabank.org/）とも連携し、構造生物学に関する幅広いデータが集められています。PDBj でも、NMR 実験情報の登録・編集を以前から実施しております(http://bmrb.protein.osaka-u.ac.jp/)。

一方、日本国内のライフサイエンス分野のデータベースの統合化を進めている独立行政法人科学技術振興機構バイオサイエンスデータベースセンター（JST-NBDC, http://biosciencedbc.jp/）に協力し、セマンティック・ウェブ技術を取り入れたデータベース統合化・高度化を行っています。具体的には、wwPDB で従来から用いられている PDBx/mmCIF 辞書に基づき、それと互換な OWL オントロジーを構築しました。さらに、このオントロジーに基づいて PDB の各エントリーの正式な XML 形式で記述された PDBML ファイルから、対応する RDF ファイルを自動生成する仕組みを開発して、毎週のエントリー更新に自動的に対応できるようにいたしました。各エントリーの各カテゴリーの RDF データをユニークな URL で参照してウェブ上で取得する仕組みを簡易検索機能とともに実装し、既に公開しております。同様な仕組みを wwPDB の低分子化合物データライブラリについても開発し、公開しています。この PDB/RDF（http://rdf.wwpdb.org/）は、wwPDB の正式フォーマットのひとつとして wwPDB ドメインで公開されております。また、これを受けて UniProt（http://uniprot.org/）が PDB/RDF のデータへのリンクを追加しておりますので、セマンティック・ウェブにご興味がある方は、是非、ご覧になってください。また、後述しますように、新しいウェブ技術を取り入れたポータルの開発と公開も、独自に進めています。

PDBj では、さらに安全で安定なデータベース運営のため、NBDC と協力して新たなポータルサイトを関西圏以外にも設置し、サービスに不具合が起きないように努力をしております。しかしながら、ウェブ・ブラウザの更新が頻繁に行われることもあり、様々なウェブ・ブラウザの全ての版に対応すべきというポリシーに、必ずしも対応できない事態も起き始めており、苦慮していることもあります。

利用者の皆様には、お気づきの点があれば、PDBj のウェブページ（http://pdbj.org/pdbj_contact_j.html）を通じて、ご遠慮なくお知らせ下さい。また、利用者講習会や学会でのセミナー、ブース展示発表も随時実施しておりますので、それらの機会を使って、ご意見を賜りたくお願いいたします。

PDBj の新しいウェブインターフェース

図 1：新ウェブインターフェース版トップページ

この度 PDBj はウェブインターフェースを刷新し、新しいインターフェース技術を使って PDB データを利用者へ提供することにいたしました。新しいウェブインターフェースは、単にウェブ画像を新たにしたというだけではなく、機能やアノテーションの情報、さらに関連性のあるデータを追加可能にする新しい仕様を備えた改良が加えられています。新しいインターフェースは、データを動的に統合して表示できる特長を備えています。すなわち、カスタム化の機能を用いて、利用者は自分の見たい情報とその表示場所を自身で選択できます。

新インターフェースにおいては、以前のインターフェースに比べ、様々なサービスがより緊密に統合されています。PDB データの検索エンジンである「PDBj Mine」、ホモログ蛋白質を探索する「Sequence Navigator」、類似構造を探索する「Structure Navigator」、弾性ネットワークモデルでダイナミクスを解析した「ProMode Elastic」、詳細な解説の「今月の分子」、「ヘルプページ」など、PDBj が提供する全てのサービスが統合され、内部リンクによる単一のウェブインターフェースを構築しています。他のサイトにおけるサービスに関連性がある場合には、その情報源へのリンクが上記の様々なサービスから提供されています。

新インターフェースによるもう一つの便利な特長は、ウェブ画面のヘッダーに多機能検索のためのバーが設置されたことです。この検索バーは、デフォルトでは「PDBj Mine」だけでなく、ニュースやヘルプ、今月の分子、

登録データの公開状況も対象として検索を行います。さらに、 電子顕微鏡画像データベース EMDB や、NMR 実験情報データベース BMRB へも、データベースを検索するクエリを送ることができます。検索バーを直接使って行う高度な検索（アドバンスド・サーチ）では、どのフィールドを検索するか、またどの（特定のキーワードを含む）検索結果を除外するかを指定する一連のコマンドを使って、検索が行われます。以前のウェブインターフェースと同様に、新インターフェースにも、グラフィカル・インターフェースを使ってパラメータを指定するアドバンスド・サーチのオプションが備わっています。

また多言語でのサポートも、英語、日本語、中国語（繁体字と簡体字）、韓国語で、以前と同様に継続されます。メニューやデータ登録法の紹介等の基本的なインターフェースは各言語に翻訳されていますが、その他のサービスや情報は英語と日本語での紹介となっています。新インターフェースでは、例えばユーザーが検索エンジン経由でウェブにたどり着いた場合であっても、それぞれのページで言語を変更できるため、より使いやすくなります。

一方、プログラミング技術の新たな特徴として、新インターフェースは REST（REpresentational State Transfer）デザインによって動くため、そこから得られるデータは全て URL によってデータを要求するプログラミング言語でアクセス可能であり、また JSON や XML による記載を読むこともできます。これらの REST サービスを通して、利用者は PDBj のデータとサービスを利用者自身のソフトウェアに統合することもできます。

新インターフェースによって PDBj はウェブ技術開発の新たな一歩を歩み始めました。インターフェースのカスタム化のおかげで、PDBj が提供する情報を利用者個人の好みを最大限に活かしたかたちに変えて利用する、いわゆる「パーソナル化」を可能にしています。

いくつかの古い版のブラウザ、特に IE8 よりも古い版の Internet Explorer では、新インターフェースがうまく稼働しないことがわかっております。他のブラウザや Internet Explorer の新しい版を用いることをお勧めしますが、これまでのブラウザも引き続き別のサイト（http://legacy.pdbj.org/）から提供しておりますので、どうぞご利用ください。

新インターフェースに関するご質問、ご提案、コメントなどを歓迎しております。私どものウェブページ（http://pdbj.org/pdbj_contact_j.html/）から、お気づきの点をお知らせください。

wwPDB の新フォーマット　PDBx/mmCIF の紹介

「PDB フォーマット」といえば、固定長コラム・行指向の、あのフラットファイル形式を思い浮かべる人がほとんどかもしれません。この 1970 年代から使われてきたフォーマットは、人の目に読みやすく、計算機処理も一見やさしそうに見えるため、今日に至るまで広く使われてきました。しかし最近の巨大複合体の構造を記述したり、PDB 以外のデータベースとの連携を図ることを考えたとき、従来の PDB フォーマットの記述能力ではもはや太刀打ちできないことは明らかでしょう。例えば、Vault の構造データは、一つの PDB ファイルに 10 万原子以上を記述できないために、現状では 3 つのエントリー（2zuo, 2zv4, 2zv5）に分割されています。また、最近の構造生物学研究では、構造データ(原子座標）そのものに加えて、機能や系統、および PDB 以外のデータベースとの対応に関する情報が重要性を増してきています。これらのデータも従来の PDB ファイルに記載されてはいるのですが、それらを自動処理しようとすると、数多くの「例外処理」が必要になるという難点があります。この問題は固定長コラム・行指向形式に由来する弱点です。

以上に挙げた PDB ファイルの問題点を解決するために、wwPDB では mmCIF と PDBML という二つのフォーマットを提供してきました。mmCIF はもともと結晶学の分野で使われてきた CIF というファイルをもとに米国の RCSB PDB が拡張を施した形式で、PDBML はそれを XML に直訳したものです。2011 年の wwPDB 諮問委員会において、mmCIF を発展させた PDBx フォーマットを wwPDB のデータの登録および公開の公式フォーマットとすることが決定されました。これに伴い、従来の PDB フォーマットは「できる限り」のサポートは継続されますが、PDBx フォーマットがデータ交換の主要な形式として使われることになります。

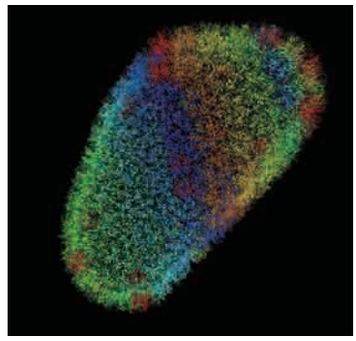


図 2：5 月 29 日にリリースされた HIV カプシド(3J34, 3J3Q, 3J3Y)の構造は 200 万以上の原子と 1000 以上のペプチド鎖を含み、25 の PDB エントリに分割されています(PDB ID 1VU4 など)

では、PDBx/mmCIF を使うことの利点は何でしょうか？まず、PDBx はその文法が厳密に定義されていることが挙げられます。このおかげで、PDB ファイルの悩みの種だった「例外処理」が不要になります。次に、PDBx はそのデータの項目が PDBx 辞書によって厳密に定義されています。これによって、どのようなデータがどの項目に記述されているかが容易に参照できるばかりでなく、項目間の関係までも把握できます。PDBx 辞書自体の形式も厳密に定義されているので、それを機械的に処理することによって、データベースやプログラムの作成も容易

になります。現在約 300 の項目(カテゴリ)が定義されていますが、その中には低分子化合物や外部データベースへの参照も含まれているため、PDB のデータと他のデータベースとの連携も容易になります。第三の利点として、人間と計算機の両方に取って読みやすいということが挙げられます。文法は厳密なのですが、見かけは従来のフラットファイルに近い感覚で扱えます。特に PDBML にみられる XML のタグに相当するものが非常に簡潔になっているので、メタデータに煩わされることなく、データを読み解くことが比較的容易です。

では、PDBx(mmCIF)の欠点は何でしょうか？やはり、現在使われている多くのプログラムが対応していないことが一番に挙げられるでしょう。しかし、いくつかのメジャーなプログラムの中にはすでに PDBx(mmCIF)への対応が完了しているものもあります (Jmol、OpenRasMol、Chimera、CCP4、Phenix など)。プログラマにとっては、これまで書き溜めたプログラムを多少書き直す必要があることも面倒な問題です。しかし、メジャーなプログラミング言語(C/C++、Python、Java、Perl など)では、すでに PDBx(mmCIF)のパーサライブラリが用意されていますので、これらを使えばかなりの手間は軽減できるでしょう。

これから PDBx/mmCIF を勉強される方のためにいくつかのリソースを紹介します。

(1) http://mmcif.pdb.org PDBx 辞書、mmCIF に関するあらゆる情報が得られます。
(2) http://pdbj.org/workshop/20120207/ito.pdf PDBj 講習会の資料。PDB、mmCIF、PDBML の違い、読み解き方を詳しく解説しています。

EVENT

■ サイエンスアゴラ 2012

(2012 年 11 月 10 日（土)･11 日（日)、日本科学未来館（東京))

毎年お台場で開催されている科学技術振興機構主催の一般向け科学イベント『サイエンスアゴラ』に初参加しました。来場者の皆さまに分子の形を 3D 画像で見ていただき、形（構造）の情報が何の役に立つのかを分かりやすく紹介しました。

＊サイエンスアゴラで紹介したコンテンツ
「万見プライム」　URL: http://pdbj.org/prime

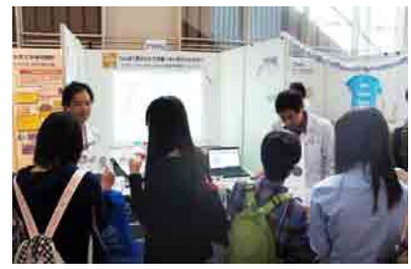


図 3：サイエンスアゴラでの風景

図 4：分子生物学会展示会場

■ 第 35 回日本分子生物学会年会･特別企画展示 ーバイオデータベースコーナーー

(2012 年 12 月 11 日（火）- 14 日（金)、マリンメッセ福岡)

第 35 回日本分子生物学会年会にて、JST-NBDC 主催の特別企画『バイオデータベースコーナー』に参加しました。展示ブースでは、PDBj のウェブサービスの利用法や、新しい開発の取組みを紹介しました。

■PDBj 講習会 @ 東京農工大

(2013 年 3 月 7 日（木)、東京農工大小金井キャンパス)

PDBj では､毎年違った地域で､web サービスの紹介やその使い方を、PC を使いながら具体的に学んで頂くワークショップを主催しています。2013 年 3 月 7 日、東京農工大小金井キャンパスにて、初心者から中級者レベルのユーザーを対象とした講習会を開催しました。

※講習会テキストやセミナー資料は、PDBj web の『過去の講習会』ページよりダウンロード出来ます。

図 5：PDBj 講習会の様子

Data Growth

統計データは wwPDB の web ページ（http://wwpdb.org/stats.html）でもご覧になれます。

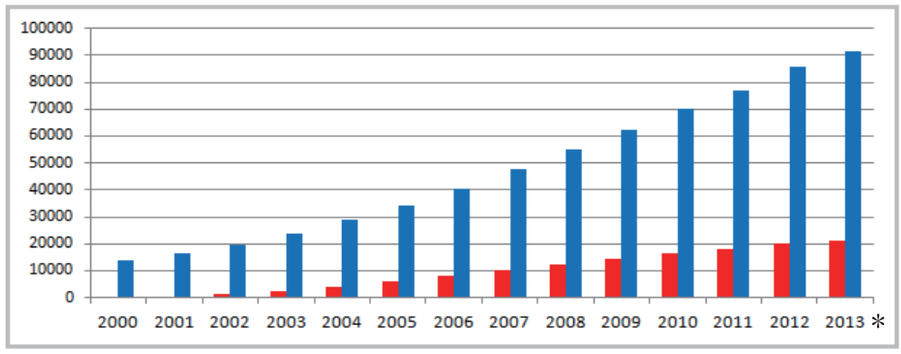


図 6：wwPDB と PDBj における構造データの登録数　　*2013 年 6 月 5 日付データ

NMR 実験情報データベース：BMRB

PDBj-BMRB グループは米国 BMRB および wwPDB と連携して、NMR データベースである BioMagResBank の開発や、拠点としてサイトの構築運営を行っています。現在 PDBj グループの一員として国内ライフサイエンスデータベース統合事業に参画しています。米国 Wisconsin 大学と共同でサービスを提供している BMRB ミラーサーバーではデザインの一新を行い、データの可視化ツールを充実させることで、登録者の NMR 実験データを分かりやすく確認できる機能を提供しています。また、公開データベース内の高度な SQL 検索をわかりやすいインターフェースで提供するアドバンスド･サーチ の公開も開始しました。登録した研究者から寄せられた意見を反映した新 PDBj-BMRB ポータルサイトを本年度より公開し、日本語、英語での新たな登録解説マニュアルに加え、我々が独自に開発した支援ツールである NMR データの解析支援を行うツール、MagRO、また今春からは NMR-STAR 形式から XML 形式へとファイルの変換を行う BMRBx Tool の提供を開始しました。実験・登録を行う研究者のみでなく、データベースを活用する研究者の方々の利便性も大きく向上させました。

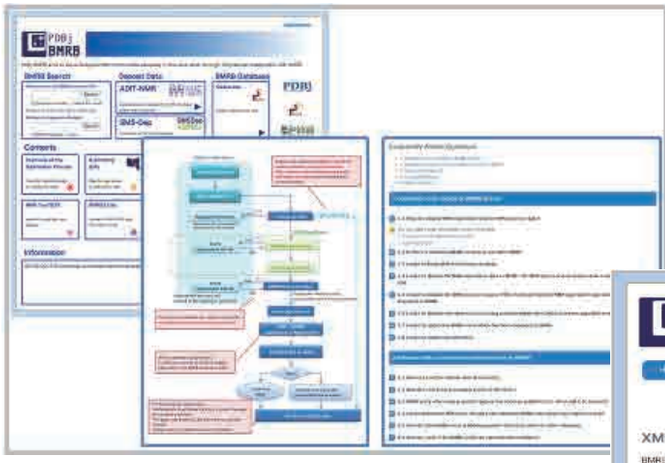


図 7：登録を行った研究者の方々のご意見を元に作成した解説ページ（左上）。「データ登録の流れ」（中央）と「よくあるお問合せ」（右）

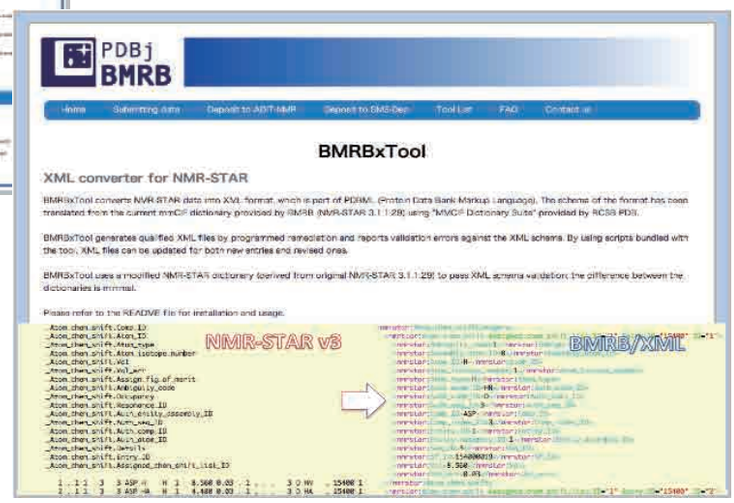


図 8：NMR-STAR ファイル形式から XML ファイル形式へと変換を行う BMRBxTool の提供ページと変換例

STAFF

統括責任者

中村 春木 (大阪大学蛋白質研究所・教授)

PDBj データベース構築グループ

中川 敦史 (大阪大学蛋白質研究所・教授)
松田 真 (大阪大学蛋白質研究所・特任研究員)
五十嵐 令子 (大阪大学蛋白質研究所・特任研究員)
見学 有美子 (大阪大学蛋白質研究所・特任研究員)
張 羽澄 (大阪大学蛋白質研究所・特任研究員)
池川 恭代 (大阪大学蛋白質研究所・特任研究員)
佐藤 純子 (大阪大学蛋白質研究所・特任研究員)

PDBj 国際的な運営高度化グループ

金城 玲 (大阪大学蛋白質研究所・准教授)
岩崎 憲治 (大阪大学蛋白質研究所・准教授)
鈴木 博文 (大阪大学蛋白質研究所・特任研究員)
山下 鈴子 (大阪大学蛋白質研究所・特任技術専門職員)
工藤 高裕 (大阪大学蛋白質研究所・特任研究員)
西川 建 (大阪大学蛋白質研究所・客員教授)
Bekker, Gert-Jan (大阪大学蛋白質研究所・特任研究員)

BMRB データベース管理運営グループ

藤原 敏道 (大阪大学蛋白質研究所・教授)
阿久津 秀雄 (大阪大学蛋白質研究所・招聘教授)
児嶋 長次郎（大阪大学蛋白質研究所・准教授）
小林 直宏 (大阪大学蛋白質研究所・特任研究員)
岩田 武史 (大阪大学蛋白質研究所・特任研究員)
高橋 あみ (大阪大学蛋白質研究所・特任研究員)
横地 政志 (大阪大学蛋白質研究所・特任研究員)

研究協力者

輪湖 博 (早稲田大学社会科学総合学術院・教授) **(for *Pro*Mode)**
伊藤 暢聡 (東京医科歯科大学大学院・教授)
木下 賢吾 (東北大学大学院情報科学研究科・教授) **(for *e*F-site)**
Standley, M.Daron (大阪大学免疫学フロンティア研究センター・准教授)
(for SeqNavi, StructNavi, SeSAW and ASH)
加藤 和貴 (大阪大学免疫学フロンティア研究センター・特任准教授) **(for MAFFTash)**

事務職員

晴氣 菜穂子 (大阪大学蛋白質研究所・事務職員)

連絡先

Protein Data Bank Japan (URL：http://pdbj.org/)
〒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蛋白質解析先端研究センター内
PDBj 事務局：TEL(06)6879-4311, FAX(06)6879-8636
PDBj データベース登録事務局：TEL(06)6879-8634、FAX(06)6879-8636