

PDB登録システムの進展と蛋白質立体構造解析手法別の実験生データ収集活動

岩田武史⁽¹⁾、横地政志⁽¹⁾、池川恭代⁽¹⁾、多久和綾子⁽¹⁾、Gert-Jan Bekker⁽¹⁾、
常住規代⁽¹⁾、宮ノ入洋平⁽¹⁾、栗栖源嗣⁽¹⁾

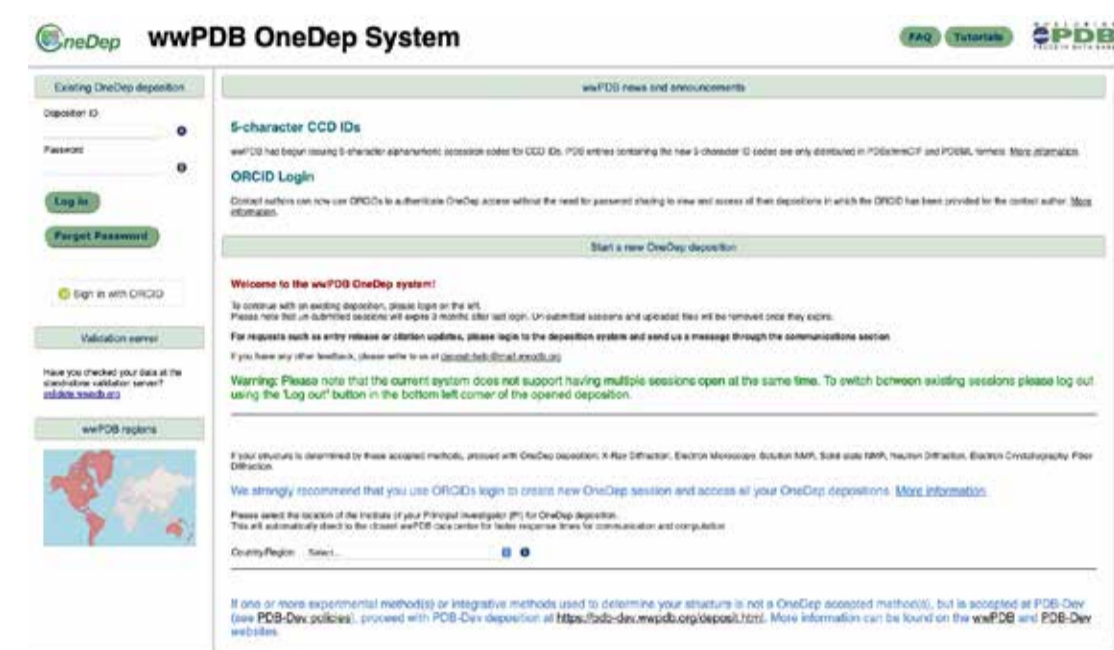
(1) 大阪大学蛋白質研究所

PDBjは、年間4000件以上の生体高分子の立体構造データを受け付けています。的確なデータキュレーションサービスに加えて、構造情報の品質評価を行う検証レポートの提供を通じて、wwPDB全体で計22万件を超える高品質なデータアーカイブを構築し、分子の立体構造を利用する様々なコミュニティへの貢献しています。wwPDBのデータ登録システムである国際共同開発のOneDepは、各諮問委員の提言をもとに機能を追加し、最新の実験手法に対応できるように改良し続けています。

また、将来の技術開発のための基盤データの整備や、第三者による解析結果の妥当性評価を可能にすることを目的として、PDBjは研究者の貴重な実験生データの収集を強化しています。X線回折データと電子顕微鏡データについては、それぞれXRDAとEMPIAR-PDBjという受付サービスを提供しています。さらに、核磁気共鳴に関して、従来のBMRBに加え、新たに柔軟性の高いBMRbigと呼ばれるデータ受付を開始し、蛋白質立体構造解析の主要な3手法を網羅する実験生データの受け入れ環境が整いました。

機能拡張を続けるPDB登録システム

登録・検証・アノテーションを行うパイプラインシステム

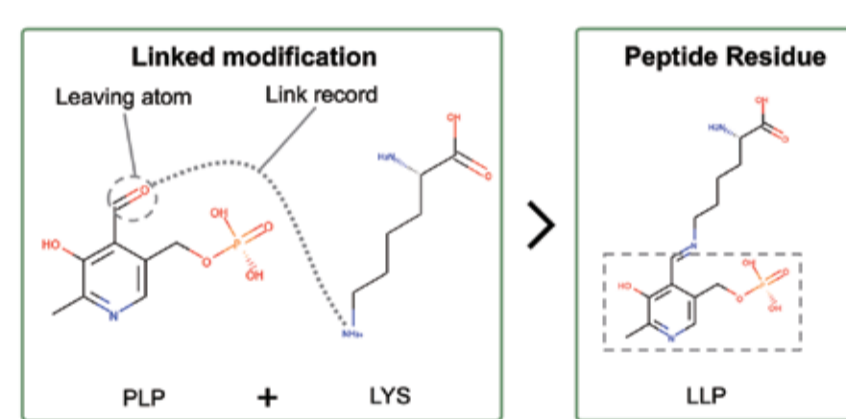


wwPDBが開発した、登録・検証・アノテーションを行うパイプライン・システムメンバーサイト(RCSB PDB, PDBe, PDBj)で共通のシステムを使用している。X線結晶解析、NMR、EM等の実験手法に対応。登録者PIの国・地域によって、担当処理サイト(RCSB PDB, PDBe, PDBj)へ自動的に振り分けられる。PDBx/mmCIFフォーマットを使用し、アーカイブ上のデータ品質の向上にも貢献するとともに、各諮問委員の提言をもとにシステム全体の機能拡張を続けている。

最近の機能拡張や変更内容

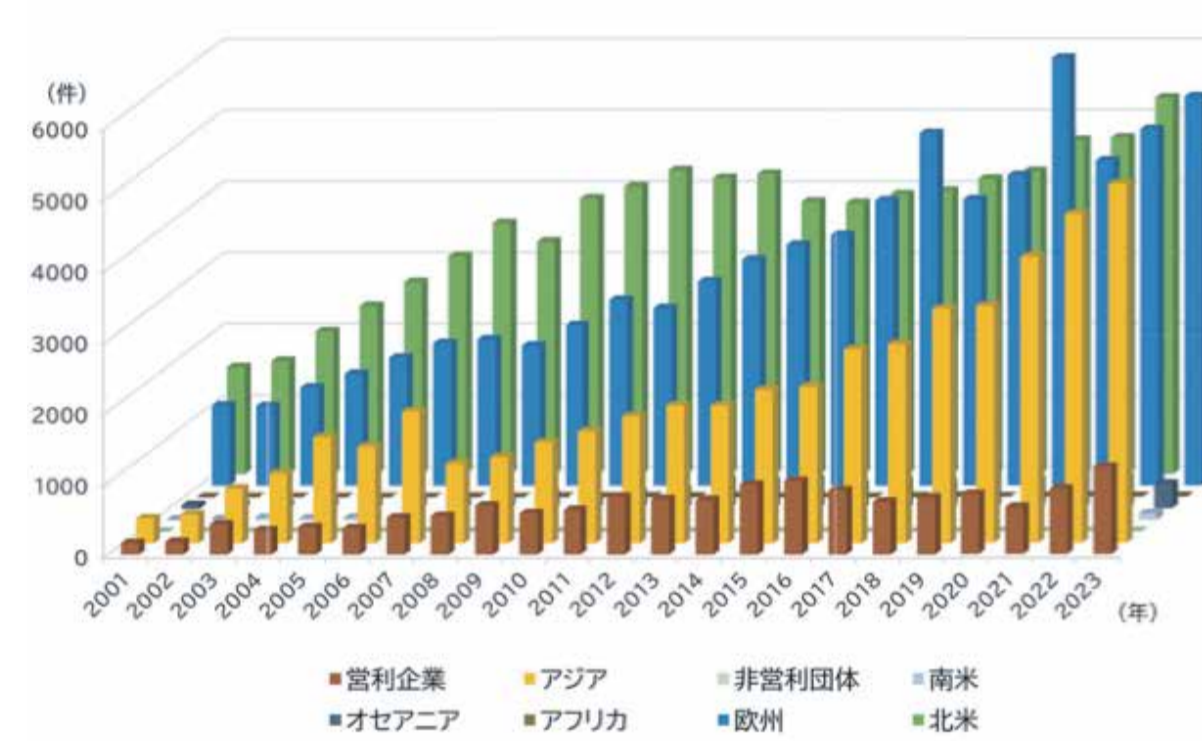
- PDBIDは8桁、CCDIDは5桁へ
- ORCID-IDを使用してOneDepシステムにアクセス可能に
- 修飾表現の標準化

PDBにおいて、タンパク質鎖の修飾に関する表現方法を標準化するプロジェクトを進めています。これにより検索性の向上が期待できる。

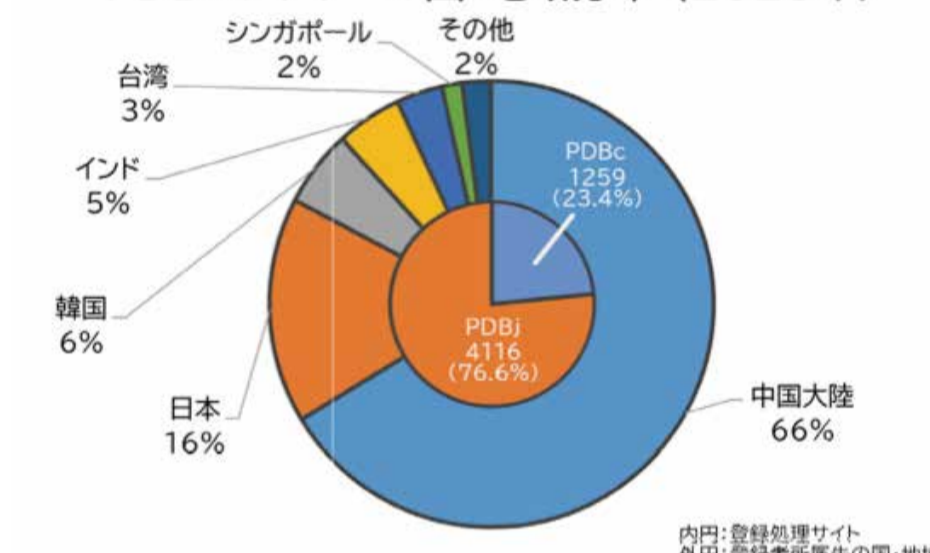


登録・処理状況と解析手法の内訳

wwPDB全体での地域別PDB登録数



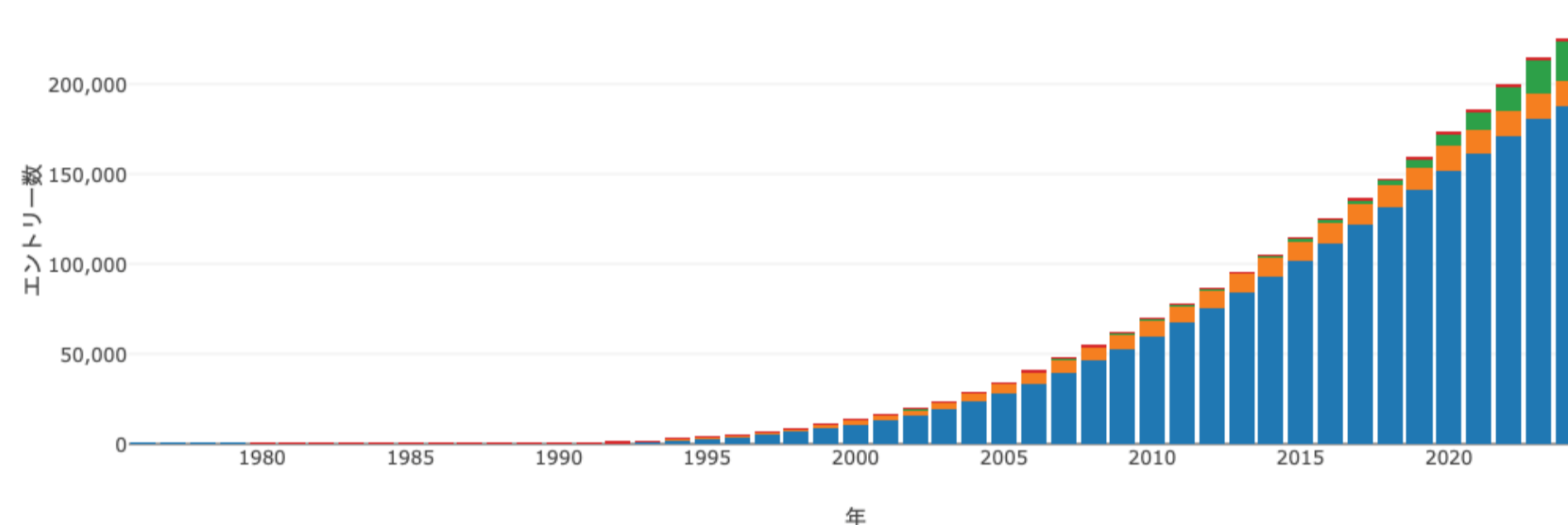
PDBj+PDBcが登録処理したPDBエントリの国・地域分布 (2023年)



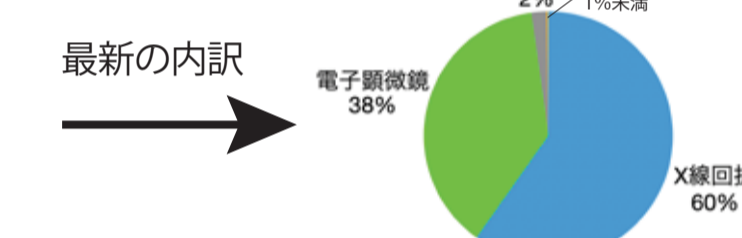
アジア発のデータ登録数は欧米に並ぶまで急速に伸びている。2023年はアジア地区の増加率は他地域より多く、PDBjの登録・処理割合は約32%に増加。

2022年よりPDBcがAssociate memberとして中国大陸のデータの処理を担当。(登録は全てPDBjにて受付)

実験手法別の公開PDBエントリー数



2023年の解析手法別登録数

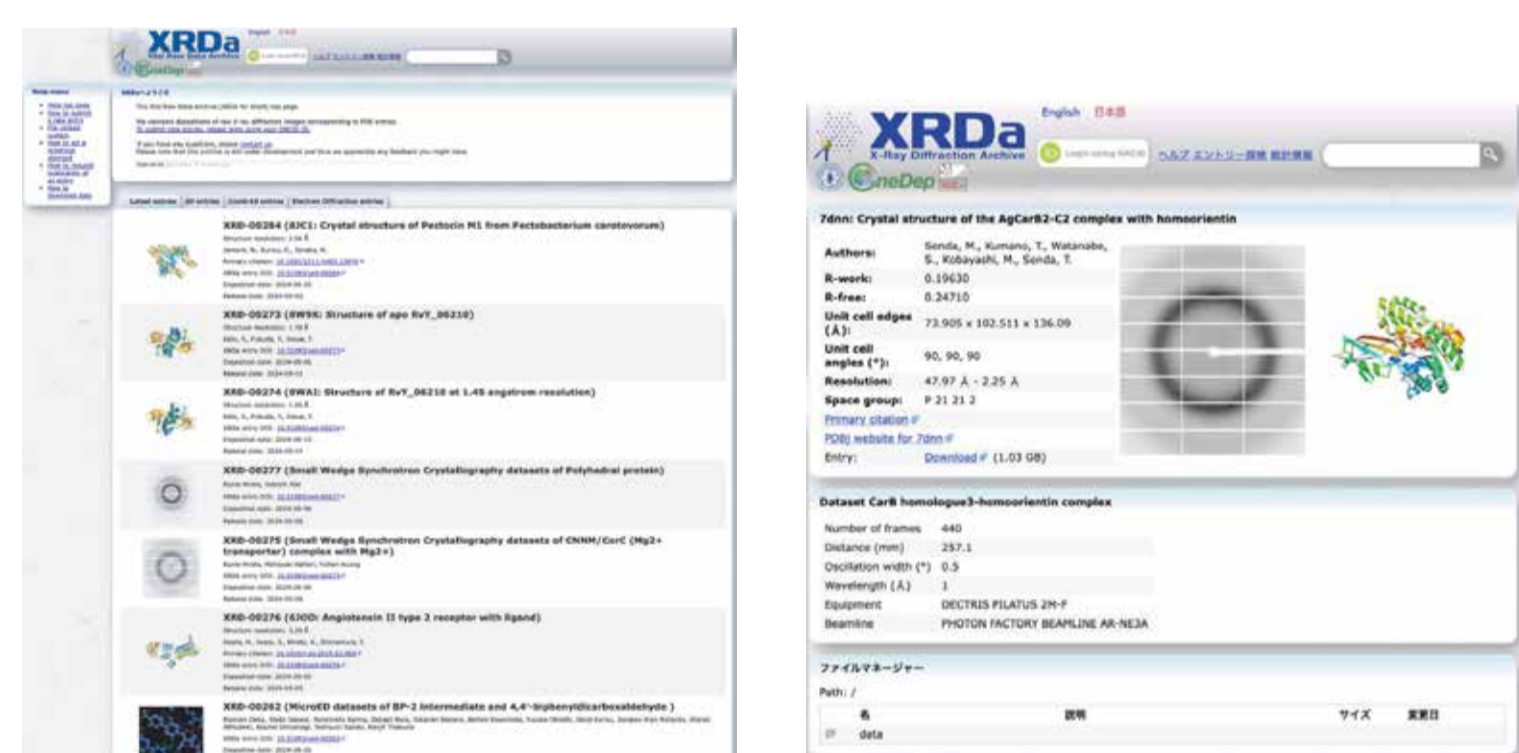


PDBへの登録数は年々増加、解析手法はX線回折や核磁気共鳴に加え、近年は電子顕微鏡によって構造が決定された登録が大幅に増加している。

2021年に28%程度だった全登録に占める電子顕微鏡による構造決定の割合が2023年には40%近くを占めるまで増加した。主要3手法による構造決定が99%を超えている状況には変化はない。

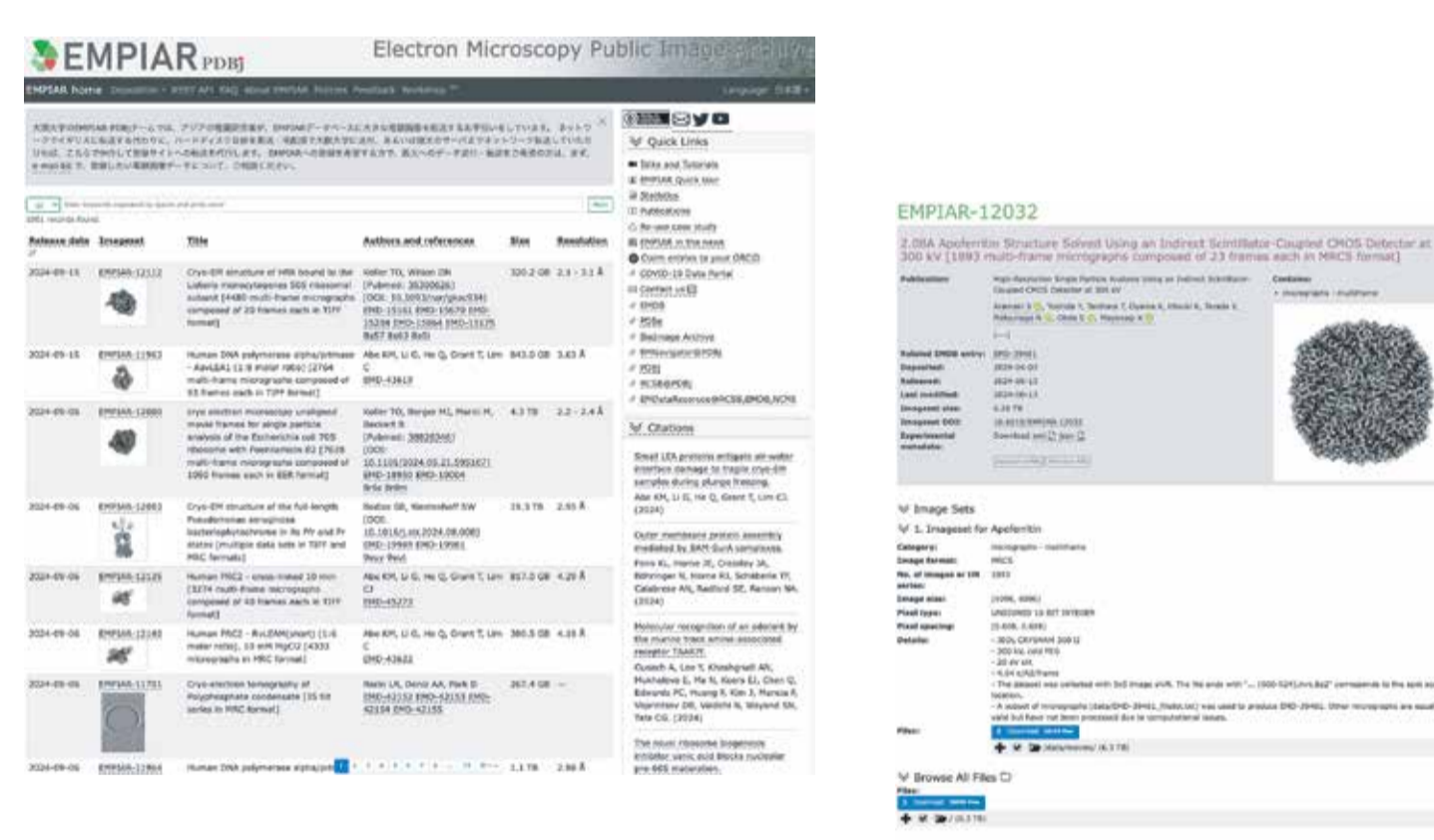
蛋白質立体構造解析手法別の生データ受付サービス

X線回折



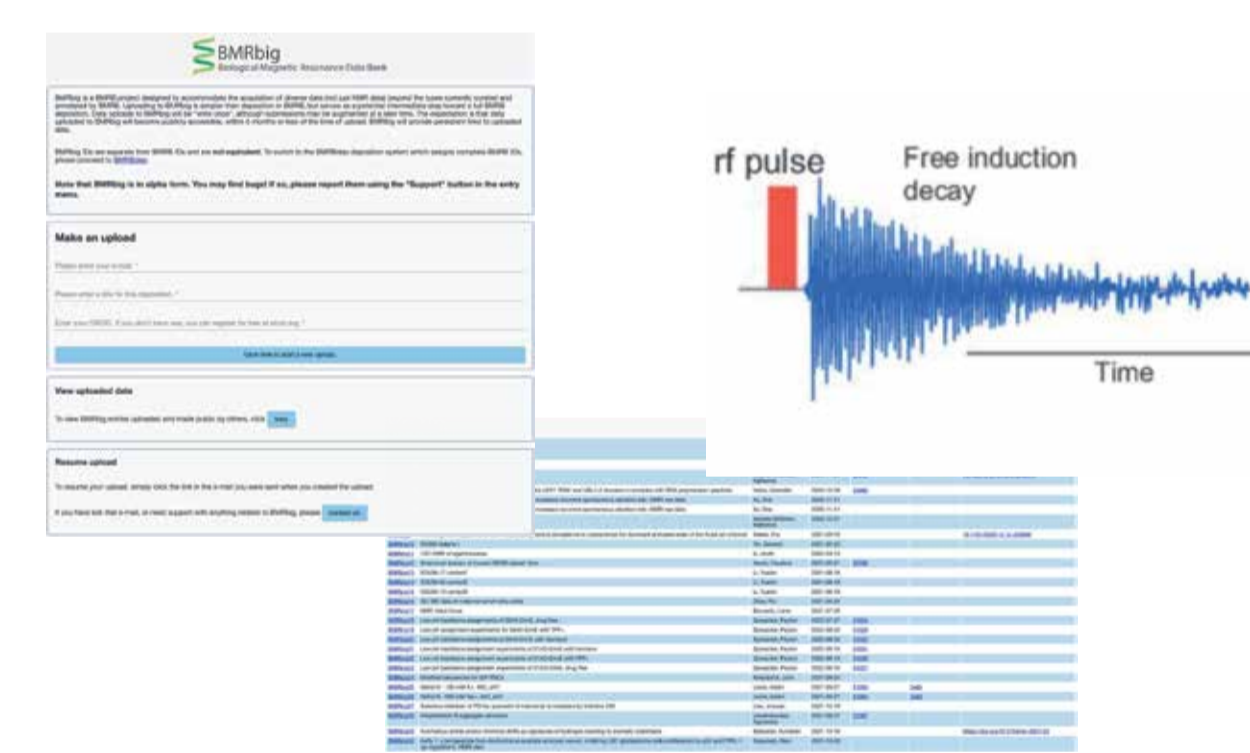
XRDa (Xtal Raw Data Archive) は、PDBj独自開発によるX線回折画像のデータベース。wwPDBメンバーの強みを活かしてOneDep登録後、構造エントリ情報がXRDa内のORCID-IDに自動的にリンクされる。OneDepへ登録したPDBID情報ともリンクして表示されるため、登録者と利用者どちらもエントリ検索しやすいシステムである。構造データ登録前に生データを先行登録することも可能で、その場合も後からリンクすることが可能である。また、低分子量の回折データなど、PDBの登録基準に満たない生データも登録が可能となっている。

電子顕微鏡



EMPIAR (Electron Microscopy Public Image Archive) は、3Dバイオイメージング実験から得られた生の2次元画像データおよび特定の3Dバイオイメージングデータセットのデータベース。公開された生データを使って三次元構造を再構成することにより、解析ツールの開発や解析手法の研究に用いられ、分解能向上に役立つことが想定されている。Webインターフェースに加えて、JSONファイルによるエントリ作成にも対応しているため、複数の登録やバージョン管理システムによる制御を行いたい場合であっても登録者側にとって最適な方法で登録を行える。

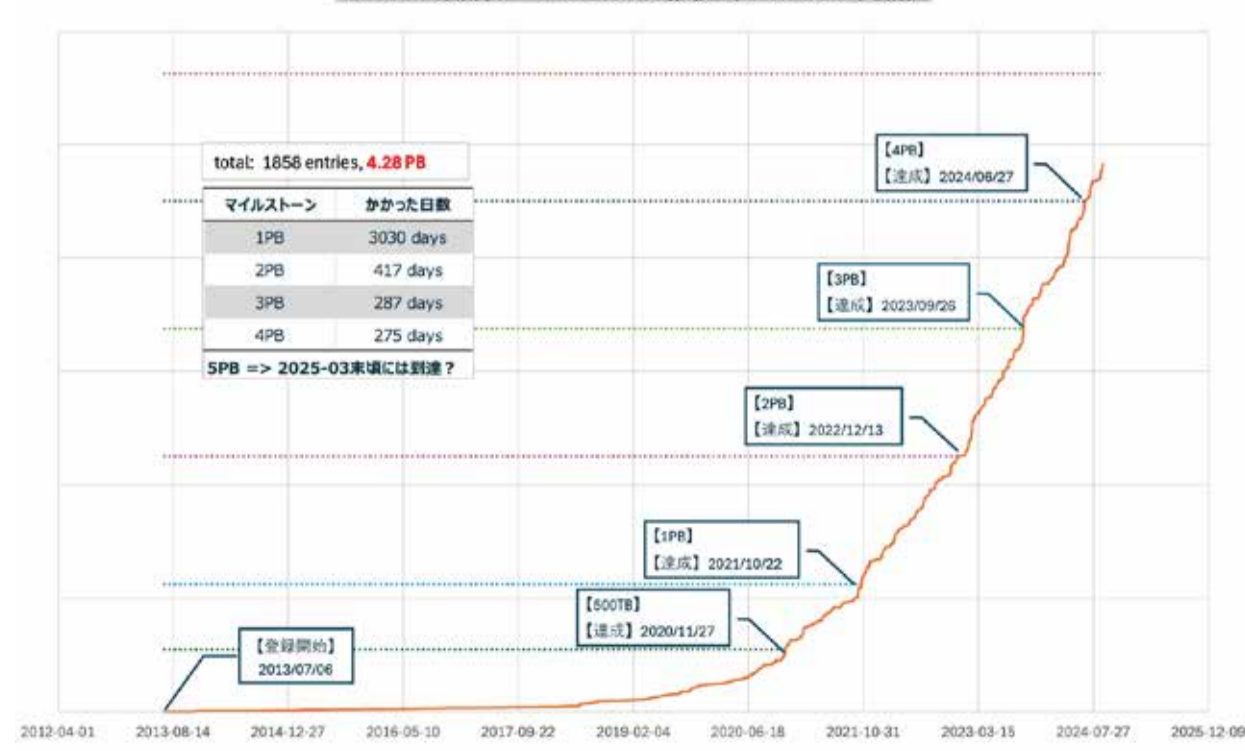
核磁気共鳴



BMRBigはFID (Free Induction Decay) やピークリストなどNMRによる測定で得られるすべて情報をアーカイブするデータベース。多数のサンプルによる実験を一つのエントリとして扱えるなど柔軟なデータ受け入れを行っている。大規模な研究プロジェクトにおいて、NMR分光計からBMRbigに直接送信して、実験データを逐次整理しておくことで、プロジェクト公開時のデータ共有の迅速化を図ることが想定されている。BMRbigでは、全てのデータファイルがバージョン管理され、公開後もデータファイルの追加、更新が可能である。

増大するRAWデータサイズへの対応

EMPIAR公開済みデータサイズの推移 (2024-09-18更新)



電子顕微鏡で構造解析に使用した大量の生の2D画像をアーカイブするEMPIARデータベースでは、すでに公開データサイズが4PBに達しました。2025年の第1四半期には5PBまで増えることが予測されています。PDBjでは肥大化するデータに対応するため、非圧縮のデータを独自で圧縮してデータ容量の削減する機能の実装を行っています。また、数TBサイズの大容量のデータの転送ツールの導入やネットワーク環境に制約がある研究者への登録支援として、オンラインの他、HDDの直接送付による受付にも対応している。

日本蛋白質構造データバンク (PDBj)

〒565-0871 大阪府吹田市山田丘3-2
大阪大学蛋白質研究所・構造解析研究棟 3F
蛋白質構造データバンク構築研究室内
TEL: 06-6879-4311 (全般)
06-6879-8634 (PDB登録)
Web問い合わせ: <https://pdbj.org/contact>



日本蛋白質構造データバンク (PDBj: Protein Data Bank Japan) は、「JST-NBDC」「大阪大学蛋白質研究所に措置された共同利用・共同研究拠点経費 (文部科学省)」の支援を受け、米国 RCSB、BMRB、および欧州 PDBe、EMDB と協力して、生体高分子の立体構造データベースを国際的に統一化された PDB アーカイブとして運営するとともに、様々な解析ツールを提供しています。