

ライフサイエンスデータベース統合推進事業
(統合化推進プログラム 2017年採択課題)

研究開発終了報告書

「蛋白質構造データバンクのデータ検証高度化と統合化」

栗栖 源嗣 | 大阪大学 蛋白質研究所 教授

研究開発期間：2017年4月～2022年3月



©2022 栗栖 源嗣 (大阪大学) Licensed under CC BY 4.0

§1. 研究開発実施の概要

【研究開発の目的】

本提案で構築する主なデータベース PDB Archive (Protein Data Bank:蛋白質構造データバンク)と BMRB Archive (BioMagResBank:NMR 実験情報データバンク)は、生命科学を支える基盤データベースとしての役割を担っている。特に創薬の現場では、PDB や BMRB を活用した構造生物学をベースとする研究手法が、確立した創薬基盤技術のひとつになっている。PDBj(<https://pdbj.org>)は、2003 年に欧米のパートナーと一緒に国際蛋白質構造データバンク(worldwide Protein Data Bank: wwPDB, <https://wwpdb.org>)を設立し、その一翼を担ってデータを収集して、厳しい品質管理を行いつつデータベース構築と公開を継続してきた。PDB と BMRB のデータ登録・利用においては、個々のデータの質を利用者へ示す検証レポートの重要性が増している。本研究開発では、PDB と BMRB の利用者・登録者に、個別の品質管理データや独自アノテーション等を RDF 形式で提供して、統合データベースとしての発展的な利用を促進することを目的とする。

【実施内容】

以下の個別項目を開発し、PDB と BMRB の利用者・登録者に提供する方法をとった。

- i) 検証レポートを全ての実験手法で確立し、検証レポートを RDF 化して利用者に提供すること。
- ii) 蛋白質に結合する化合物の情報を効率よく収集・利用できるよう、既存の低分子化合物構造データベース(CSD)との連携・統合化を進めること。
- iii) 分子系統樹情報を利用した可視化により、配列データが大量であっても効率的な情報抽出を可能にするツールを開発すること。
- iv) 利用者が希望する構造情報(原子座標という形でなく、分子立体構造上のどの位置で、どのような性質を持ち、どのようなダイナミクスの特徴があるか?という情報)を PDBj 独自のアノテーション情報として提供すること。
- v) BMRB のデータをもれなく収集・利用できるよう、新しいツールを継続開発し、wwPDB パートナーはもとより一般ユーザーに提供すること。
- vi) アノテータやユーザーの人材育成に積極的に取り組み、基盤データベースとしての PDB と BMRB を継続的に維持・管理・利用できる体制を構築すること。

【得られた成果】

日米欧の国際協力により、進展著しい電子顕微鏡による構造解析にも完全に対応して、PDB と BMRB を継続的に構築・検証・公開した。wwPDB で分担しているアジア・中東地域からの全データを 100%処理し、各エントリーの論文発表に合わせて遅滞なく全世界に公開することができた。新型コロナウイルス感染症による運営への影響は最小限に抑えることに成功するとともに、新型コロナウイルスの蛋白質構造情報については、タイムリーかつ正確に情報発信することができた。PDBj オリジナルの活動として、特にデータ品質管理のための検証レポートの拡充と RDF 化を図る一方、位置情報の自動アノテーション、CSD との連携、NMR データ登録・検証の高度化を行った。また、ASH viewer の開発を進め、可読性の高いツールを構築し、プロジェクト終了までに実装できる見込みである。データベースの利用促進と人材育成・国際協力に関する特筆すべき成果として、PDB China と PDB India の開始に道筋をつけたことが挙げられる。

§2. 研究開発実施体制

1. 研究グループ

(1) 「栗栖」グループ(研究代表者グループ)

・担当項目

日米欧の国際協力により進展著しい電子顕微鏡による構造解析にも完全に対応して、PDB(蛋白質構造データバンク)とBMRB(NMR 実験情報データバンク)を継続的に構築・検証・公開する。その際、特にデータ品質管理のための検証レポートの拡充とセマンティック化を図る一方、低分子化合物のCSD(ケンブリッジ結晶構造データベース)との統合化や位置情報の自動アノテーション、NMR データの高度化を図る。データベースの利用促進と人材育成にも注力する。

(2) 「藤」グループ(主たる共同研究者グループ(1))

・担当項目

新規のアラインメントビューア ASH viewer を開発する。ASH viewer では分子系統解析情報に基づくクラスタ化に基づく可視化や立体構造グラフィクスと連動させることで、大量アラインメントであってもユーザインタラクティブな機能情報抽出を容易に行えるようにする。また、ネットを介してRDF 化されたデータベースから取得された各種情報をアラインメントにマッピングして機能解析を実施できるようにする。

(3) 「清水」グループ(主たる共同研究者グループ(2))

・担当項目

生物工学的に見いだされた有用物質作成のための変異情報を PDB の情報に反映させて総合的に最適化することで、更なる改変方針の創出を目指す。

2. 有識者会議等

(1) 会議概要

名称	蛋白質立体構造データベース専門部会
目的	PDBj は大阪大学蛋白質研究所の共同利用・共同研究拠点活動の一部と位置づけられており、研究所の「共同利用・共同研究委員会」の下に、外部有識者からなる「蛋白質立体構造データベース専門部会」を置いている。この会議は実質的な運営諮問会議(Advisory Committee)に相当し、PDBj の運営に対する外部有識者からの助言・意見・要望などの収集を行う。本会議は wwPDB からみると wwPDB 運営諮問委員会(wwPDB Advisory Committee)の地域委員会という位置付けで、専門部会委員のうち 2 名は、必ず wwPDB の Advisory Committee (wwPDB AC) 会議に毎年出席し、wwPDB AC においても議論と運営諮問に加わっていただいている。
委員数	9人

(2) 開催歴

年月日	場所	参加人数	主な議題
R3年10月19日 (予定)	オンライン会議システム Zoom (European Bioinformatics Institute 主催、英国)	23人	wwPDB-AC: wwPDB コアアーカイブの運営について (PDBjからは3人出席)
R3年3月17日	大阪大学蛋白質研究所 大会議室 (Zoom 併用)	12人	専門部会: 1. 蛋白質立体構造データベースの運営状況について、2. 次年度運営計画について、3. その他
R2年9月30日、 10月1日	オンライン会議システム Zoom (コネチカット大学 BMRB 主催、米国)	20人	wwPDB-AC: wwPDB コアアーカイブの運営について (PDBjからは3人出席)
R2年2月28日	大阪大学蛋白質研究所 4 階セミナー室 (Zoom 併用)	8人	専門部会: 1. 蛋白質立体構造データベースの運営状況について、2. 次年度運営計画について、3. その他
R1年10月18日	大阪大学蛋白質研究所 大会議室 (PDBj 主催)	22人	wwPDB-AC: wwPDB コアアーカイブの運営について (PDBjからは3人出席)
H31年2月5日	大阪大学蛋白質研究所 中会議室、大阪大学東京 オフィス (TV 会議システム)	10人	専門部会: 1. 蛋白質立体構造データベースの運営状況について、2. 次年度運営計画について、3. その他
H30年11月2日	Madingley Hall、Cam bridge、U.K.	20人	wwPDB-AC: wwPDB コアアーカイブの運営について (PDBjからは3人出席)
H30年2月28日	大阪大学蛋白質研究所 大会議室	8人	専門部会: 1. 蛋白質立体構造データベースの運営状況について、2. 次年度運営計画について、3. その他
H29年10月13日	Rutgers 大学 RCSB PDB 会議室 (米国)	20人	wwPDB-AC: wwPDB コアアーカイブの運営について (PDBjからは4人出席)

§3. 研究開発の目的、実施内容及び成果

1. 研究開発の背景

前研究開発課題において、構造生命科学の基盤的データベース開発として PDB と BMRB のデータを wwPDB の一翼を担って収集し、厳しい品質管理を行いつつデータベース構築と公開を継続して行った。JST-NBDC による統合化推進プログラムの活動結果として、いち早くデータの RDF 化を実施し、PDBj および PDBj-BMRB(現在は BMRBj と改名)から他のデータベースにリンクする環境の整備を構築することができていた。しかし一方で、PDB と BMRB のデータにおいては、そのデータの質を利用者へ示すための検証レポートの重要性が増している。X 線結晶構造に対しては検証方法がほぼ確立された一方で、NMR と電子顕微鏡(EM)に対しては、観測データと構造データの consistency に対する検証は未だにできていないのが実情であった。今期プロジェクト期間は、検証レポートを全ての実験手法で確立し、その検証レポートを RDF 化する必要性が増している状況にあると考えた。特に、EM データは急速にデータが増大し、その構造情報も生命科学にとって極めて重要なものになってくると推測されたが、EM データの検証は前研究開発課題で国際的にも積み残された喫緊の課題であった。また、RDF 化という DB 統合化における技術面は解決されたものの、利用者が特に希望する構造情報(原子座標、という形ではなく分子立体構造上のどのような位置で、どのような電気的性質を持ち、どのようなダイナミクスの特徴があるか?という情報)をアノテーション情報として追記することも必要であり、検証レポートの RDF 化とともに、さらなる DB 統合化技術として成熟させるべき課題であった。

2. 研究開発対象のデータベース・ツール

(1) データベース

・主要なもの

正式名称	略称	概要
PDB Archive	PDB	生体高分子の立体構造データベース、wwPDB と協力して構築、RDF を開発、公開
BMRB Archive	BMRB	生体高分子の化学シフト、緩和データ、相互作用データ等の NMR の実験データのデータバンク

・上記以外のもの

正式名称	略称	概要
eF-site	eF-site	蛋白質の分子表面の形状と物性(静電ポテンシャルと疎水性度)を機能部位情報と結合したデータベース。維持・更新のみ
ProMode elastic	ProMode	二面角を変数とする基準振動解析プログラムによって計算された蛋白質のダイナミクス・データベース。維持・更新のみ
MoM	MoM	社会で話題となっている内容に関わる分子を PDB から選び、機能と構造に関して解説。維持・更新のみ。

(2) ツール等

正式名称	略称	概要
EM Navigator	EM Navi	生体分子や生体組織の3次元電子顕微鏡データ(EMDB)閲覧用 web site
ASH viewer		PDBj、UniProtなどのデータベースから得られる配列、あるいは構造のアラインメントから、その機能や機能部位に関連する情報の抽出を支援する
MolMil	MolMil	インターネット上の web 環境で稼働する JavaScript による分子構造ビューア

※データベース、ツールの詳細は別紙参照。

3. 達成目標及び実施計画

(1) 当初の実実施計画・達成目標

1) 国際組織 wwPDB メンバーとしての蛋白質構造(PDB)アーカイブの構築・データ検証・公開【実施計画】

習熟した専従作業員が、X線結晶解析、電子顕微鏡解析、NMR解析の研究者からの蛋白質構造情報とNMR実験情報データを、世界共通のアノテーション手法により、厳密な品質を保ちつつ登録・編纂を行って、統合化に必要なキュレーション・アノテーションを継続的に2022年度まで実施する。wwPDBのメンバー(RCSB-PDB、PDBe、BMRB)やアジア諸国の関連研究者との国際連携に基づいて、開発中の登録・検証(validation)システム(OneDepシステム:X線結晶構造の登録については2016年に完成し運用中)によって、実験データの品質と検証記載の精度を高めて、恒久的に持続可能な蛋白質構造データバンクを実現する計画である。特にNMRと電子顕微鏡による立体構造情報の検証システムの組み込みを、wwPDBとして進める。

今後のデータサイエンスの推進において、基礎となるデータの検証と品質管理が極めて重要な課題として認識されている。wwPDBでは、いち早くその問題に着目し、データ生産を行う専門家によるデータ検証のための国際的なタスクフォースを組織し、タスクフォースからの提言に基づいて構造データに対する検証レポート(Validation Report)の自動生成と公開を行ってきた。この検証レポートでは、利用者が自身でデータの信頼性を判断できるように、全体の構造の精度だけでなく、各アミノ酸・塩基および結合している低分子リガンドやイオン毎に、その原子位置情報の精度が示されている。しかし、これまでの検証レポートは人間が眺めるためのPDF版とその元となっているXML版のみしか存在しておらず、セマンティックな活用ができていなかった。

一方、PDBjでは、以前からwwPDBの標準XML表記であるPDBMLを開発し、その利用を推進してきた。2011～2013年度および2014～2016年度の統合化推進プログラムにて、OWL準拠のRDF化したPDBデータ(wwPDB/RDF)とNMR実験データ(BMRB/RDF)を開発・拡充し、wwPDBにおける世界標準フォーマットとして採用されるに至っている。このように、wwPDBの他のメンバーおよびUniProtからも、PDB関連データのセマンティック化に関するPDBjの寄与が広く認識されている。そこで、2017年度から新たなPDBjの研究開発事業として、上記した検証レポートおよび今後拡充されるNMR、電子顕微鏡による構造の検証レポートのセマンティック化とその応用を実施し、利用者がゲノム等の他のデータベースとの統合的な利用の際に活用できるようにする。上記作業は、PDBjが統合化プログラムで実績を残してきた方向性であり、wwPDBからも今後の貢献

が強く期待されている。検証レポートの拡充については wwPDB と共同で実施し、構築後は他の PDB 版とともに公開される世界標準となる計画である。

【達成目標】

- ・ wwPDB の欧米メンバーと協力して、厳しい品質管理を行いつつ、さらに増加する立体構造情報を登録・検証・公開する。
- ・ 勃興する EM 構造を中心に検証レポートの高度化を進める。
- ・ 全 PDB データの検証レポートの XML 化、RDF 化を行うツールを wwPDB の国際的な協力関係においてプロトタイプを開発し、セマンティック化を進める。
- ・ 検証レポートを XML 形式あるいは RDF 形式に変換し、それらを WEB ツールへ実装する事でより高度で恒久的な構造品質評価ツールとして実装していく。

2) 他のデータベースとの統合化および高度化

2-1) CSD データベースとの統合化

【実施計画】

PDB には蛋白質ポリペプチド鎖の原子座標のみが含まれている訳ではない。PDB データの実に 75% 近くが非蛋白質分子、すなわち「リガンド化合物」を含んでいる。この PDB に含まれるリガンド化合物の原子座標は、構造化学的に精度が低いものがあり、実験的に得られた電子密度マップとの整合性の点でも信頼度の低い構造も含まれている。現在では、登録時に Ligand Validation を行い、データ提供者がリガンド化合物の化学組成や構造化学情報(立体配座やキラル位置)、結合情報(単結合や二重結合の位置)の確認と入力を行っている。これら PDB に含まれるリガンド化合物の情報は、Chemical Reference Data として wwPDB の ftp サイトに別ファイルとして収納されているが、一般ユーザーには決して利用しやすい形態とはなっていない。

PDB とは別に、低分子化合物の結晶構造データベースである CSD には、850,000 件を超える原子座標が metadata つきで収録されている。製薬企業での化合物スクリーニングの際にも PDB と併用される基盤的な化学データベースである。創薬候補化合物のターゲット蛋白質へのドッキングシミュレーションの精度を高めたり、補欠分子属や基質分子の結合様式を考察する酵素化学的知見を高めたりする上で、CSD との統合化と上述の問題点の改善を並行して進めることは必須と考える。本研究計画で、Chemical Reference Data を積極的に活用し、CSD データベースとの統合化を図る。CSD と緊密に連携をとることで CSD をベースとした検索、PDB をベースとした検索の両者で有効に情報を共有できる利用者の立場で活用しやすい環境を整備する。

【達成目標】

- ・ CSD データベースとの統合化を完了する

2-2) ASH の高度化

【実施計画】

アラインメントと分子系統樹を入力として、分子系統樹に基づいて 10~30 程度のクラスタを自動同定し、各クラスタに対応するアラインメントのコンセンサス配列を表示する。系統樹もクラスタ間の系統関係として粗視化して表示し、ノードのクリックにより系統樹の展開/圧縮ができ、それに伴ってコンセンサス配列も再構築されるようにする。また、クラスタに含まれる配列や構造について、由来する生物の情報や機能の情報を要約して表示できるようにする。さらにクラスタを指定することで、機能部位やクラスタ間の機能差に関連するサイトを推定し、その結果をコンセンサス配列上で表示して、グラフィクスを使って立体構造上にも反映する。タンパク質の情報については、注釈行

に含まれる情報に加えて RDF 化された種々のデータベースからネットを介して情報を集取できるようにする。

アラインメントは FASTA 形式、あるいは clustal 形式に対応するファイルを読み込めるようにする。また、分子系統樹は Newick 形式のファイルを読み込めるようにする。ビューアは Java で開発する。JavaFX と Java2D を使ってビューアを開発する。プログラム内部でのデータベースの取り扱いには、JDBC を使って SQLite で処理する。RDF データの検索には Jena を使って SPARQL エンドポイントにアクセスする。立体構造の表示には PDBj で開発された Molmil を使用する。機能解析には、これまで行ってきた保存度や変異度に基づく解析に加え、Kullback-Leibler 情報量や相対 von Neumann entropy を利用したパラログ間の機能差を決定する手法も導入する。

まずビューアの基本となるアラインメントと系統樹の入力部分と、その表示部分を作成する。順次、クラスタの情報要約、機能解析、立体構造へのマッピング、RDF 化されたデータベースの検索と得られた情報のアラインメントや構造へのマッピングなどの機能を付け加えていく。

【達成目標】

- ・アラインメントと系統樹の入力機能の実装
- ・クラスタ化に基づく可視化機能の実装
- ・クラスタの展開、圧縮、要約機能の実装
- ・アラインメントからの機能部位推測機能と、その結果の立体構造への反映機能の実装
- ・ネットを介して RDF 化されたデータベースからの関連情報の取得機能の実装

2-3) 各アミノ酸の蛋白質立体構造中における位置情報の自動アノテーションとその応用

【実施計画】

蛋白質立体構造データからしか得られない情報として、分子内及び分子間の原子レベルないし残基レベルの相互作用情報がある。これらは原子座標として既に PDB のデータに含まれているが、原子座標の値そのものは構造決定実験の座標系に依存するため意味がない。座標系に依存しない相互作用情報の記述法として、残基間ないし原子間コンタクトをリストアップすることが考えられる。そこでまず、分子内コンタクト及び分子間コンタクトをデータベース化し、既に運用されている PDBj Mine に統合・公開する。さらにこのコンタクトの情報を適切なオントロジーを付して RDF として公開することにより、核酸配列を含む様々なアノテーションリソースと統合できるようにする。これにより、例えばゲノム上の変異が蛋白質の安定性やリガンドとの相互作用に及ぼす影響などを解析する際に必要なリンク情報が容易に取得できるようになる。

【達成目標】

- ・各アミノ酸の蛋白質立体構造中における位置情報の自動アノテーションを進める。

2-4) 統合化された NMR データベースおよび関連ツールの活用

【実施計画】

NMR 研究においては特殊同位体ラベル、超高速マジック角試料回転、1.2GHz 超高磁場マグネット、DNP 感度増大法、細胞内 NMR、非線形サンプリング法による多次元 NMR などの測定解析技術、複数の構造解析法を組み合わせたハイブリッド解析法など多くの技術革新が芽生えつつある。本提案では、これら今後多様化が予想される登録データの科学的意味を喪失することなく machine readable な形式として高品位な情報として電子化し、データベースとして公開していく。NMR 実験データの登録の際に用いる UI の充実、特にデータ品質に関わるパラメータを可視化す

る WEB ツールの開発を行う。アクセスログの解析などからユーザーのトレンドを解析し、UI をより洗練され永続性の高いデータベース、WEB ツールとして最終年度までに開発を行う。

【達成目標】

- ・ wwPDB との国際的な協議により標準化を進め、OneDep 次期バージョンへの搭載を協力していく。
- ・ 標準化された NMR 構造および EM 構造の品質検証に関するツール群の更なる高度化と実装への国際的協力を進める。

3) データベースの利用促進・人材育成

3-1) 利用者・研究者コミュニティとの連携および講習会等の開催

【実施計画】

従来は、関連学会年会でのセミナーや講習会等、アカデミアのデータ提供者およびデータ利用者との連携を中心に実施してきたが、上に記載したように国内の製薬企業も創薬の現場において PDB データを多用している。この状況に対応するため、まず PDB の国内諮問委員会である大阪大学蛋白質研究所「蛋白質立体構造データベース専門部会」のメンバーに企業研究者に入っただき、企業からの要望等を集めて緊密な連携体制をとる。また、PDBj の代表が Vice President を務めている wwPDB Foundation (米国内の NPO 財団) の枠組みを利用して、企業と wwPDB との連携も図る。現在は、OpenEye、FEI、DECTRIS、Anton Paar 等の米国内の企業がスポンサーとなっているが、今後は日本企業にもスポンサーとして参加していただき、wwPDB の活動に対する支援を受ける。

利用者・研究者向けのセミナーや講習会については、他のデータベースとも協力し、2016 年までと同様に All-in-one 講習会や学会年会中あるいは前後での利用講習会を実施する。

一方、これまでも参加してきたサイエンスアゴラの活動等、一般社会人向けの生命科学におけるデータサイエンスの振興も積極的に実施する。

【達成目標】

- ・ 利用者・研究者コミュニティとの連携を進める。

3-2) アノテータの育成・国際協力

【実施計画】

PDB および BMRB のデータベース構築・検証にはコンピュータの利用だけでなく英語の語学力と特殊な知識および経験が必要であるが、アノテータに対する雇い止めや自己都合による退職が発生する恐れがある。問題点を先取りして対応するため、新たにアノテータ候補を非常勤として雇用し育成を図っておく必要がある。

2020 年頃には、現状の構造データの生産・利用状況から、中国の PDB China およびインドの PDB India が、wwPDB のフランチャイズに加わるのが想定されており、アジア圏を代表している我々 PDBj は、積極的にこれらの組織を支援し、それら新たな組織におけるアノテータの育成に協力して、グローバルなデータベース活動に拡充していく。

【達成目標】

- ・ アノテータを継続的に育成する
- ・ 中国、インドを中心に国際連携を進める

(2) 期間中に追加・削除・変更した実施計画・達成目標

・追加実施 1 (2017 年)

【実施計画】

システムバイオロジーへの展開: シグナル伝達系の解析においては、蛋白質や遺伝子の細胞内濃度や分子間の親和性などをもとにした数理モデルがよく用いられている。疾患の遺伝子情報を入力するには、ゲノム疾患医薬品のネットワークデータベース(KEGG MEDICUS)や疾患ヒトゲノム変異データベース(DBKERO)が用いられる。これらに加えて、系における蛋白質と薬剤の相互作用を PDB のデータを用いてドッキングシミュレーションする。

代謝工学への展開: 代謝工学モデルを基盤に実験的に選び出された変異部位に着目して、PDB の立体構造データと配列相同性解析 (ASH) とを緊密に連携させることで、より詳細な構造・機能の相関を解明する。

【達成目標】

システムバイオロジーへの展開:

- ・ 薬剤のシグナル伝達系全体に対する影響をより詳細に予測する。

代謝工学への展開:

- ・ コハク酸生産株の最適化を可能にする。

・追加実施 2 (2020 年)

【実施計画】

COVID-19 のような新興感染症の創薬研究では、これまで直接創薬に携わっていなかった研究者が新たに参入するケースも想定される。そこで、非専門家であっても関連する有用な PDB および BMRB エントリーに直感的に到達できる仕組みを構築する。検索対象が限られるコロナウイルス蛋白質に限定すれば、利用傾向を分析することも比較的容易であり、約半年間という限られた時間で成果をあげることが期待できる。

【達成目標】

- ・ 既存計画で整備してきた検証レポートを積極的に活用すると共に、コロナウイルス感染症に関するエントリー利用傾向を分析して、非専門家が利用しやすい新型コロナウイルス特集ページを整備する。

4. 実施内容

(1) 実施内容

1) 国際組織 wwPDB メンバーとしての PDB アーカイブの構築・データ検証・公開

新プロジェクト開始後も引き続き、wwPDB の欧米のメンバーと協力して、厳しい品質管理を行いつつ増加する一方の立体構造情報と NMR 実験情報のキュレーションをおこなった(見学、張、佐藤、金、岩田)。これは全世界の登録の約 24% に相当する^{A-8)}。wwPDB で分担しているアジア・中東地域からの全データを 100% 処理し、各エントリーの論文発表に合わせて遅滞なく全世界に公開することができている(Bekker、工藤、山下、岩田)。

データ検証レポートの活用に関しては、クライオ電子顕微鏡(Cryo-EM)および NMR により解析された構造データの品質を検証する手法を確立することに重点を置いた。本研究期間中に、新しく EM 構造と NMR 構造の検証レポートを作成し公開した。NMR については、データ検証システムを組み込んだ新しいデータ登録ツール(BMRBdep)の開発を進めた(横地、池川)。PDB データの

うちX線結晶解析のデータについて、品質管理のための検証レポートのセマンティック対応を進めた(横地)。検証ファイルの RDF 化においては、第 2 年次中に PDBj が独自に高度化した検証ファイル RDF 化を完成することができた。先行して X 線結晶構造解析の検証レポート^{A-12)}を RDF 化し、他の wwPDB メンバーと dictionary の整合を確認した。PDBj-Mine2 への実装が完了し、週次更新に合わせて RDF の更新を行うシステムの構築を行った(Bekker、山下)。wwPDB の標準的 RDF ファイルとして欧米の PDB サイトでも公開する方策を検討し、wwPDB 全体で検証ファイルの RDF が共有されている(横地)。PDB に含まれる糖鎖データの remediation を進める全ての糖タンパク質の記述法を統一した。野口研の山田博士(JST-NBDC 木下チーム)に協力をお願いし、PDB の新しい糖鎖記述ルールに従って PDB(RDF フォーマット)ファイルから GlyYouCan へ自動でリンクを組み込む準備を進めた(Bekker、横地)。

毎年 wwPDB 運営諮問会議を開催し、引き続き OneDep システムの高度化を進め^{A-9)}、新たに BMRBdep を導入して登録システムの更なる効率化を進めた。全世界の蛋白質構造データ処理のうち、PDBj で地域分担しているアジア・中東地域では、中国からの登録数がうなぎ登りに増えている。wwPDB ではデータ生産量に比例してその国で責任をもってデータ処理を行うという方針を立てて 2021 年 1 月に新しい wwPDB の定款(Charter)を更新した。

2) 他のデータベースとの統合化および高度化

2-1) CSD データベースとの統合化

低分子化合物データベース CSD との連携は CSD ID での検索を全ての PDB エントリーに対して継続して可能とし、データ検証においても CSD のデータを積極的に活用した(Bekker、池川)。データ登録者からの要望が多かった CSD の構造化学情報を最大限活用した X 線結晶構造解析の電子密度マップの計算係数を公開し、wwPDB で共通の電子密度を提供できるようになった(Bekker、池川)。これにより、登録者が事前に化合物と電子密度との一致度を正確に検証できるようになりデータ処理の効率が改善した。

2-2) PDB データを用いた 3D 構造同士の構造アラインメント・サービス

ASH ビューアの開発は、Java を使ってビューアの開発を行なった(藤)。アラインメントを読み込み、粗視化された系統樹とアラインメントをビューアのウィンドウに表示し、別ウィンドウに無根系統樹および保存度/変異度のプロファイルを、デフォルトのクラスタリングに従って表示できるようにした。当初計画では、系統樹は別途作成された Newick 形式のファイルを読み込む予定であったが、報告会での評価委員のコメントから、ビューア内部で近隣結合法により系統樹を作成する形に変更した。粗視化された系統樹は、ノードのクリックで展開/圧縮が可能であり、それに伴い粗視化アラインメント、無根系統樹のクラスタリング、保存度/変異度のプロファイルを再構築できるようにした。また、アラインメントに含まれる配列に関する GO 情報や由来する生物種の情報を Linked Open Data から収集し、無根系統樹、および保存度/変異度のプロファイルに表示できるようにした。Linked OpenData 収集のためには、東京都市大学大学院(開発当初は NBDC)の山口敦子教授により開発された PSurfer を利用した。当初、SPARQL を使って Linked Open Data を収集する予定であったが、上記の PSurfer を使用する形に変更した。構造データへの保存度/変異度の反映には Molmil を使用する予定であったが、そのために必要な機能がいくつか足りないことがわかり、PyMol を使用している。現在、プログラムの整理と新規機能の追加、また応用を並行して行なっている。

2-3) 各アミノ酸の蛋白質立体構造中における位置情報の自動アノテーションとその応用

蛋白質分子内コンタクトの情報は、すでに別のサービス(GIRAF)で内部的に利用しているプログラムを改変することにより、PDB アーカイブの mmCIF ファイルから抽出できる。そこで、このデータを単純なスキーマのテーブルで表現し、PDBj Mine2 の関係データベースに追加して、立体構造そのものから得られる情報と、その構造に関するメタデータが統合的に検索できるようにした(Becker、金城)。

2-4) 統合化された NMR データベースおよび関連ツールの活用

NMR データベースの登録業務の自動化と簡略化を推進した(岩田、横地)^{A-6, A-10}。米 Madison-Wisconsin 大学を中心とした wwPDB メンバーと共同で新しい登録システム(BMRBdep)を共同開発することにより、作業時間を現在より大幅に短縮できた(横地)。PDB アーカイブと同様に BMRB もデータの検証ファイルの作成と活用が非常に重要となってきた。NMR 構造座標、実験データとの整合性に関する検証ツールを wwPDB との協力を通じて開発し、データ検証レポート(validation report)を作成して、その内容の充実を進めた(横地)。wwPDB で共同開発を進めている OneDep システムで、PDBj が NMR-START または NEF 形式の単一ファイルで NMR 実験データのアップロードを受付できるようにした(横地、池川)。

3) データベースの利用促進・人材育成

3-1) 利用者・研究者コミュニティとの連携および講習会等の開催

PDB の国内諮問委員会である大阪大学蛋白質研究所「蛋白質立体構造データベース専門部会」を毎年開催し、企業を含めたユーザーからの要望や意見交換を行った(中村、中川、藤原、栗栖)(別掲§2.2(3))。第1年次に利用促進の呼び水となるよう行った研究成果を効果的に発信し、生物工学会に参加するなどして代謝工学やシステム生物学の研究者と意見交換を行った¹⁰。利用者・研究者向けのセミナーや講習会については、毎年テーマを変えて開催しており、ユーザーの利用促進や人材育成に役立っている。

3-2) アノテータの育成・国際協力

アノテータの育成・国際協力の観点では、引き続き更新されていく OneDep 登録システムを活用した世界で同一の品質管理によるデータ登録を実施する必要があり、継続して国際連携による人材育成を進めた。具体的には、定期的に行っている TV 会議による講習の他に、欧州 EBI-PDB e と米国 RCSB-PDB と合同で毎年 OneDep Summit を開催し、アノテータのスキルアップをはかった。wwPDB AC 会議では PDB China の開始時の運用形態について検討し、PDBj にサーバーを設置し、当初は全てのエントリーを PDBj がモニターしながら行うことに決定した。新型コロナウイルス感染症の状況次第ではあるが、状況が許せば具体的な PDB China トレーニングを開始する予定である。

(2) 実施内容のうちの特定項目の詳細

① 研究コミュニティを含むデータ提供者との連携・協業

データ提供者が多く参加する学会(日本結晶学会、日本蛋白質科学会、NMR 討論会)の年会中に、データ提供者・登録者に対するアナウンスや標準フォーマット PDBx/mCIF の利用方法を説明する目的で、ランチョンセミナーやブース展示、講習会を毎年実施した(別掲§9.2)。また、よくある質問についてはQ&Aのページ(<https://pdbj.org/help/faq>)を用意して、登録者の利便性を向上した。

一方、PDBj の活動やアナウンスをまとめて紹介するため、毎年英文と日本語のニュースレターを継続して発刊した(図1)。また、ランチョンセミナーでの発表内容は PDBj の web 上に公開して常に閲覧できるようにした。

新型コロナウイルス(SARS-CoV-2)は中国で最初に発生したこともあり、2020年1月26日に中国の研究者から最初の関連構造が PDBj に登録された。そこで wwPDB の各代表で相談し合意の下、新型コロナウイルスの構造情報が PDB に登録された場合、個別に登録者(研究者)と連絡をとり、論文発表を待たずに即時公開することを強く勧めることにした(アジア・中東地区のデータ登録者には PDBj が連絡し、今現在も随時継続して依頼中)。

規模の大きなデータ算出プロジェクトとして、AMED の創薬等先端技術基盤プラットフォーム事業(BINDS)がある。BINDS 事業と連携して PDBj の講習会を開催し、データ登録時のノウハウや注意事項を経験豊富な登録者を講師に迎えて説明していただいた。また、データ寄託時に PDBx/mCIF フォーマットでのみ登録を受け付けるという方針変更の際には、専門誌にその背景と理由などを詳細に解説し、研究コミュニティへの周知に勤めた(A-7, A-10)。

② データベース利用者への周知、利用者との連携・協業

データベース利用者が多く在籍する関連する学会(日本生物物理学会、日本分子生物学会、生命医薬情報学連合大会、CBI学会)の年会において、ほぼ毎年ランチョンセミナーやブース展示、利用者講習会においてアンケートをとって多数の意見を得た。特に、製薬企業研究者が多く参加される CBI 学会においては、低分子化合物のデータベースである CSD を企業向けに販売する化学情報協会と協力して、スポンサーセッションを企画し、CSD と PDB の統合利用を啓発した(別掲§9.2)。

2020年3月11日以降、創薬研究の加速を支援する目的で、新型コロナウイルス SARS-CoV-2 の構造情報データのみを正確に集約して、日英中韓の各言語で PDBj の特集ページから発信している(<https://pdbj.org/featured/covid-19>)。

アカデミアに限らず企業でのデータ利用者との協力関係を確かなものとするため、PDBj の有識者会議と位置付けている蛋白質立体構造データベース専門部会に、新しく協和発酵キリン(株)低分子医薬研究所から齋藤副所長に加わっていただき、創薬企業との連携を強めた。

③ 利用者にとって有用なデータ基盤の構築

PDB エントリー内の糖鎖に関する新たなデータ表現方法と、高分子構造中における糖鎖の検索性と相互運用性を向上させることを目的としたデータベースの高度化を実施した。PDB の分子表

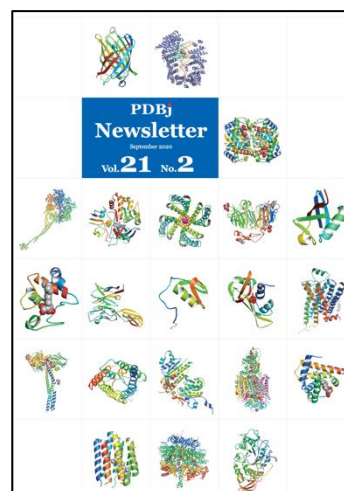


図1:PDBj Newsletter
vol.21. no.2 号、2020 の表紙

記に糖鎖科学のコミュニティで一般的に使われている線形記述法を採用し、メタデータの糖鎖修飾部位に関する記述を変更した。この過程で PDB の化合物辞書 (CCD) にある 1,000 を超える単糖の原子命名法を標準化し、8,000 を超える PDB エントリーにおいて分岐エンティティ表現をオリゴ糖に適用することとなった^{A-4)}。この高度化事業には、糖鎖コミュニティ、特に糖鎖データベースとの連携が必須であり、米国ジョージア大学の GML から短縮型 IUPAC を、NBDC 統合化プロジェクトの木下チームで開発された PDB2Glycan から WURCS を、そしてドイツの pdb-care から LINUCS の各ツールを wwPDB のシステムに提供していただき、wwPDB の OneDep システムに組み込んで運用している^{A-5)}。

本事業期間中に、クライオ電子顕微鏡による構造解析の新しい流れに対応するため、2019 年 1 月 13-15 日と 2020 年 1 月 23-24 日の 2 回、EM data-management workshop を開催し、クライオ電子顕微鏡による構造をどのように検証してデータの利活用を進めていくのか議論を行った (図 2)。この会議結果を基に 2020 年に新しい EM の検証レポートを作成して提供を開始した。



図 2 :EM Workshop
(左:2019 年, 右:2020 年)

④ 持続的なデータベース運用体制の構築に向けた取り組み

PDB バージョニングシステムを導入することにより、公開済み PDB エントリーに対して同じ PDB ID を保持したまま座標の更新ができるようにした。これにより、新しい構造精密化ソフトの導入等によって精密化が進み、同じ実験データから更に高精度の構造が得られた場合や解釈の間違いを発見した場合に、登録者自身がエントリーを更新することでデータベースの品質向上が期待される。2019 年に導入した当初は、2014 年以降に登録された比較的新しい PDB エントリーに限定していたが、2020 年からは全ての公開済み PDB エントリーの登録者 (PI) が、同じ PDB ID を保持したままモデル座標を更新できるようにした。これは、PDB ID と論文とのリンクを維持しつつデータベースの品質を向上する取り組みである。

データ寄託者が最初から高品質のデータを登録することがデータベースの品質管理と持続的な運用につながる。この考え方の下で、wwPDB として X 線結晶解析の検証レポート改良を行い、蛋白質に結合したリガンドの検証情報をより明確に記述することにした。改良した検証レポートには、リガンドの幾何学的な検証を強調して表示した図と、結晶解析構造の場合には電子密度の 3 次元の図も含まれる。wwPDB が検証レポートの作成時に使用した電子密度マップの係数も登録者に提供し、登録者自身がより注意深く信頼度の高い構造解析を進める助けになると期待している。

⑤ 人材の育成

アノテータについては日常的なメールや Zoom 会議による海外の他の wwPDB メンバーのアノテータとの交流 (毎週水曜日)、および OneDep Summit にアノテータが参加・訪問することによるアノテーション技術の習得を実施した。

OneDep の開発に参画している研究員については、Zoom で行っている OneDep 開発者会 (毎週木曜日) を通じて多岐に渡る IT 技術、NMR 関連の知識とノウハウを非常に高いレベルで獲得できた。それらの知識と経験を生かした今後のキャリアにおいても広範囲な分野における活躍が期待出来る。

⑥ 国際連携・国際貢献

wwPDB の既存拠点との間では、定期的な会議を通じてこれまで以上に緊密な国際連携を進めている。OneDep の共同開発においても PDBj のメンバーが特にユーザー登録システム (DepUI) の作成で貢献している。

2021 年の wwPDB 定款の更新により新設された wwPDB 准メンバーへの加入を目指し、中国とインドにおける新たなアーカイブ拠点が活動を開始した。「中国蛋白質構造データバンク」(Protein Data Bank China: PDBc) を設けるにあたっては、当拠点の統括責任者が公募された際に、国際蛋白質構造データバンク (wwPDB) の中心メンバーとして人事選考に意見を述べた。特に中国・上海での拠点立ち上げについては、大阪大学蛋白質研究所と国立蛋白質科学研究センター・上海との間で学術交流協定を締結し、緊密な交流をベースにした協力を準備中である。

⑦ その他

2021 年の wwPDB 定款 (Charter) の更新により、データアクセスライセンスをクリエイティブ・コモンズの 標準的なオープンソースライセンスである CC0 1.0 全世界 (CC0 1.0) に変更しました。また、この wwPDB 定款の更新に合わせて、PDB アーカイブのフォーマット規定を作り、PDBx/mmCIF フォーマットを Primary format、PDBML(XML) と RDF フォーマットを正式な Secondary format と定めた。さらに、wwPDB に准メンバーというカテゴリーを作り、新しい国や地域が wwPDB に参加するルールを明文化した。そのほか、一般向けアウトリーチ活動も積極的に行った。

§4. 主要なデータベースの利活用状況

1. アクセス数

(1) 実績

表 1 研究開発対象の主要なデータベースの利用状況(月間平均)

名称	種別	2017年度	2018年度	2019年度	2020年度	2021年度 (8月末時点)
PDB Archive	訪問者数	39,958	46,652	60,716	86,333	72,629
	訪問数	64,507	89,827	117,744	142,052	120,882
	ページ数 (Robot 込)	664,773 3,388,982	2,073,619 3,931,783	2,109,904 5,815,753	2,789,881 8,124,654	2,630,181 6,162,792
BMRB Archiv e	訪問者数	2,092	4,707	5,147	5,094	3,241
	訪問数	3,772	7,254	9,115	9,485	5,739
	ページ数	204,157	397,399	225,454	298,114	289,254

(2) 分析

- ・ PDB Archive として構築・維持している PDB データベースは、訪問者数、訪問数ともに 2020 年までは 16～42%の範囲で増加している。2021 年度の利用状況は 8 月末時点での値を元にした推計となるが、100%程度の著しい増加が見込まれる。本研究開発期間中、主要な Journal 各誌に対し、米国 RCSB PDB への LINK でなく wwPDB の DOI landing page へのリンクに変更を依頼し、利用者が landing page から各 PDB 拠点を選択できるようにしたことが、2020 年度後半から少しずつ実装されてきた影響であると思われる。ファイルのダウンロード数においても、2019 年度に 83%の急上昇があり、2021 年度にも 146%のダウンロード増が期待されている。これは、PDBj を含む各 wwPDB 地域拠点から全く同一のファイルをダウンロードできることを、関連学会への出展等でアピールしたことが浸透してきたからと考えている。データ寄託エントリー数も、引き続き 4～16%の範囲で右肩上がりに増え続けており、コロナ禍にもかかわらずアジア地区での研究活動が相変わらず活発であることを示している。
- ・ BMRB Archive は、2020 年度の訪問者数と 2019 年度のページ数を除き、訪問者数、訪問数、ページ数ともに、前後年度に比べて 4～125%の範囲で増加している。2020 年度の訪問者数と 2019 年度のページ数がそれぞれ 1%、32%減少している理由は思い当たらない。しかし一方で、エントリーのダウンロード数は、2019、2020 年度とも 14～96%の範囲で全て増加しており、利用は堅調に推移していると考えられる。2020 年度後半から各数値が上昇していることは、主要な Journal 各誌に米国 BMRB への LINK でなく、wwPDB の DOI landing page へのリンクを張り、利用者が landing page から地域拠点を選択できるように依頼した影響であると思われる。BMRB への登録数は、毎年度 100 件程度で横ばいであることを考えると、研究コミュニティは成熟化していると考えられる。
- ・ 分析のための付帯情報:wwPDB では各パートナーサイトが全く同一のデータを ftp サイトにて提供していることを鑑みて、ftp のダウンロード数については wwPDB 各パートナーサイトの合計数で発表する方針に変更した。また、各 Journal からのリンク先として、wwPDB ブランドの DOI landing page を共同で準備し 2019 年 10 月から提供を開始した。

2. データベースの利用状況を示すアクセス数以外の指標

一次データベースである PDB および BMRB においては、web page への訪問だけでなく、オリジナルデータのデータ寄託・登録件数、PDB 座標ファイルと BMRB エントリーファイルのダウンロードが重要な指標であり、両件数を以下に記載する。

PDBj にデータ寄託・登録したエントリー数

2017 年度	PDB:2,866 件、BMRB:107 件
2018 年度	PDB:3,065 件、BMRB:83 件
2019 年度	PDB:3,359 件、BMRB:109 件
2020 年度	PDB:3,500 件、BMRB:107 件
2021 年度(8 月末時点)	PDB:1,698 件、BMRB:46 件

PDB 座標ファイルのダウンロード数

2017 年度	70,187,534 件
2018 年度	69,732,015 件
2019 年度	128,135,982 件
2020 年度	124,982,039 件
2021 年度(8 月末時点)	128,139,570 件

BMRB エントリーファイルのダウンロード数

2017 年度	6,496,636 件
2018 年度	12,793,380 件
2019 年度	17,768,885 件
2020 年度	25,607,251 件
2021 年度(8 月末時点)	12,115,198 件

3. データベースの利活用によりもたらされた産業への波及効果や科学技術のイノベーション（産業や科学技術への波及効果）

1. DeepMind 社の John Jumper と Demis Hassabis らのグループは、PDB に登録されている全ての構造情報と UniProt の蛋白質配列情報をもとにした機械学習をベースに、完全に新しいニューラルネットワークを基にした蛋白質立体構造予想システムを構築することに成功した。既にこの手法は膨大な配列情報に対して適用されて、21 種の生物種がもつ全ての蛋白質の立体構造が予測され、AlphaFoldDB というデータベース化も進んでいる。
(本研究開発課題の研究参加者とは関係の無い研究グループによる研究成果)

関連: John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurno v, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenst ein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushm eet Kohli and Demis Hassabis, “Highly accurate protein structure prediction wit

h AlphaFold”, Nature, vol. 596, pp. 583-589, 2021 (DOI: 10.1038/s41586-021-03819-2)

§5. 今後の展開

【研究開発終了後、残された課題や取り組みたい研究開発等】

1) 国際組織 wwPDB メンバーとしての蛋白質構造 (PDB) アーカイブの構築・公開と AI で予想された構造との統合検証レポートの作成

習熟した専従研究員 (primary annotator) と X 線結晶解析、電子顕微鏡解析、NMR の研究者によって、wwPDB のメンバー (RCSB-PDB、PDBe、BMRB、EMDB) やアジア諸国の関連研究者と連携し、研究開発期間終了後も世界共通のアノテーション手法により蛋白質構造情報 (原子座標、密度マップ) と NMR 実験情報の登録・編纂を行っていきたい。日本からの 20 年以上にわたる PDB アーカイブ構築への貢献は信頼をもって世界から高く評価されており、統合的に必要なキュレーション・アノテーションを継続的に実施していきたい。世界中で増え続ける登録エントリーをより効率的に処理していくためには、国際協力により開発中の OneDep システムを継続して高度化し、恒久的に持続可能な蛋白質構造データバンクを実現させていくしか方法はないと考えている。

X 線結晶解析で確立されてきた検証レポートを、本研究開発期間中に NMR 法と電子顕微鏡法により決定した立体構造の検証に展開することができた。これは大きな進展であると自負するとともに、今後もデータベースの品質管理・品質向上の観点から検証レポートの充実をはかっていく必要性を強く認識している。時を同じくして、機械学習により配列情報から高精度に立体構造を予測する手法 (AlphaFold2 システムおよび RoseTTAFold システム) が開発され、本研究開発期間の最終年度に一般公開された。予想構造と低分解能の実験情報とを組み合わせた新しいタイプの構造解析が、急速に研究者コミュニティから発信され始めている。例えば、配列の 70% は実験により構造決定し、残り 30% の実験データは不十分であるものの予想構造を積極的に活用して 100% の全体構造として PDB に登録する事例などが該当する。ベイズ統計の手法を用いて予想構造自身を評価しつつ、予想構造と実験構造が混合した場合に、構造情報をどう評価し活用していくのが最良であるのか、wwPDB の喫緊の課題として議論を進めている。今後の展開の第一として、予想構造を含めた統合検証レポートを、計算生物学コミュニティと連携しながら積極的に構築していきたい。

2) 創薬候補化合物を含む低分子や生理活性物質と蛋白質との高精度な相互作用情報を提供する統合化

新型コロナウイルス感染症 (COVID-19) 治療薬の開発を例に取るまでもないが、研究者コミュニティは、より高精度な構造解析を求めている。PDBj では、低分子化合物のデータベースである CSD と連携し、リガンド化合物の立体化学を評価する仕組みを導入して CSD との統合化も進めた。検証レポートの高度化を含め、これまで解像度 (分解能) にのみ依存してきたデータ選別基準が、より具体的に目標化合物へ着目した基準に置き換わり、かつ、機械的に選別できるようになることは大きな進歩である。しかしながら、PDB や BMRB に含まれる低分子化合物の構造情報を、より豊富な生理活性情報とともにユーザーに提供することはできていない。wwPDB では Biologically Interesting molecule Reference Dictionary (BIRD) という辞書を作成し、抗生物質やペプチド阻害剤、その他の複雑なリガンドを独自に定義している。BMRB には BMRB-metabolomics という独自のメタボロミクス情報も持っている。BIRD や BMRB-metabolomics の情報も、PDB や BMRB に含まれる低分子化合物の構造情報と併せて、KEGG-DRUG 等の外部データベースの情報を取り込み、PDBj のサービスに構造ポータル機能を持たせる開発を進めて、創薬などの応用研究に貢献していきたい。

3) 専門家以外でも使いやすいセマンティック技術をコアにしたデータベースの応用

特に、今後のデータサイエンスの推進においては、基礎となるデータの検証と品質管理が極めて重要な課題であり、検証レポートを論文登録の必須要件とする対応が構造生物学の分野では広く浸透してきている。本研究課題により、検証レポートのセマンティック化が完了し、計算機を解して構造データを評価することが技術的に可能となった。しかし、実際には検索条件の設定時に、構造生物学の専門知識を必要とするため、一般の生命科学研究者がゲノム等の他のデータベースと統合的に利用する際に、自由自在に活用できるまでにはなっていない。wwPDBのPI会議においても、セマンティック技術を使った分解能に代わるデータ利用ガイドラインの設定・運用をPDBjが強く提案している状況にある。

データ利用ガイドラインの具体例を示す。特に電子顕微鏡による構造の場合、分解能が一様ではなくLocal Resolutionで表現される。例えば、RDF化した検証レポートをベースに、Grade A: 側鎖情報や化合物の官能基のレベルまで精度があるエントリー、Grade B: 主鎖構造のトレースには十分な精度があるエントリー、Grade C: ヘリックスの本数など二次構造のレベルまでの精度はあるエントリー、Grade D: 蛋白質分子のサイズや全体形状のレベルでのみ活用できるエントリー、といったガイドラインを付すことで、検索・利用が容易になる仕組みづくりを進めていきたい。

【開発したデータベースによって、将来的に想定される生命科学分野、その他の研究分野、産業に対する波及効果、科学技術イノベーションへの貢献等】

創薬研究を加速することを目的に、多くの新型コロナウイルスの蛋白質構造が報告されている。しかし、あまりにも急いで構造解析をおこなったため、中には不十分な精密化のものも多く含まれている。コロナ禍に起こったPDBに関連した大きな動きは、ボランティアベースでPDBの構造を再精密化しデータベースの品質を向上させようという動きである。X線の回折イメージを公的なアーカイブに寄託していれば、構造解析の専門家が最良の構造に再精密化するという、この取り組みは「The Coronavirus Structural Task Force」と呼ばれ、データ駆動型の新しい研究スタイルとして注目されている(図4: Nature Struct. Mol. Biol., vol. 28, pp.404-408, 2021)。PDBエントリーに紐づいたX線の回折イメージを、きちんとアノテーションし登録・公開することができれば、品質向上はコミュニティーが行うのである。この取り組みをサポートすることを目的に、大阪大学蛋白質研究所にもX線回折イメージのアーカイブを研究所費用にて開設した(<https://xrda.pdbj.org>)。今後、PDBデータベースの品質向上を目的に、積極的に連携していきたい。

クライオ電子顕微鏡の急速な発展により、オルガネラレベル、細胞レベルの構造情報の分解能が著しく向上している。クライオトモ

comment [Check for updates](#)

Making the invisible enemy visible

Structural biology plays a crucial role in the fight against COVID-19, permitting us to 'see' and understand SARS-CoV-2. However, the macromolecular structures of SARS-CoV-2 proteins that were solved with great speed and urgency can contain errors that may hinder drug design. The Coronavirus Structural Task Force has been working behind the scenes to evaluate and improve these structures, making the results freely available at <https://insidecorona.net/>.

Tristan I. Croll, Kay Diederichs, Florens Fischer, Cameron D. Fyfe, Yunyun Gao, Sam Horrell, Agnès Præven Joseph, Luise Kandler, Oliver Kippes, Ferdinand Kirsten, Konstantin Müller, Kristopher Nolte, Alexander M. Payne, Matthew Reeves, Jane S. Richardson, Gianluca Santoni, Sabrina Ståb, Dale E. Tronrud, Lea C. von Soosten, Christopher J. Williams and Andrea Thorn

When the COVID-19 pandemic hit in early 2020, the structural biology community quickly swung into action to determine the atomic structures of the 28 viral proteins encoded by SARS-CoV-2 (ref. 1). A total of 1,146 structures covering 18 SARS-CoV-2 and SARS-CoV-2 proteins have been released over the course of just 12 months. They are freely and publicly available in the Worldwide Protein Data Bank (wwPDB), which celebrates its 50th anniversary this month. These models serve as the basis for structure-based drug design and vaccine development. They are also essential for understanding how the virus hijacks human cells and causes disease. However, errors occur in even the most carefully determined structures and are probably more common in structures solved quickly and under immense pressure. Even small errors can have severe consequences for structure-based drug discovery, structural bioinformatics and computational chemistry because they can be misinterpreted as biologically and pharmaceutically relevant.

While the wwPDB is an invaluable tool, setting a structural biologist's archive of records, it is also largely static. Released structures can only be updated by the original depositors, and there is often little motivation to make corrections once associated papers are published. 90% of PDB structure downloads are not conducted by experimental structural biologists but by scientists who use the structural data and who may lack the training to identify and correct erroneous sites in the molecular model.

In the global crisis, it is vital to ensure that the available structural data are the best they can be, which requires us to push our methods to the limit. The Coronavirus Structural Task Force, a diverse international team of structural biologists involved in

Table 1 | Examples of quality indicators pointing to potential problems in PDB entries, calculated using our automatic evaluation pipeline

Key indicators in evaluation	Number of depositions (Percentage of total)
Key (crystallography (999 depositions))	34 (3.4%)
Completeness <90%	2 (0.2%)
R _{int} > 35%	52 (5.2%)
Potential twinning	93 (9.3%)
Contaminated by ice diffraction	86 (8.6%)
Isolated image	46 (4.6%)
Single-particle cryo-EM (260 depositions)	46 (19%)
Average model-map FSC <0.4	56 (10%)
MI score <0.4	64 (10%)
SUOC <10%	122 (8.8%)
Other indicators (1392 depositions, including 31 NMR and 2 neutron diffraction)	338 (22%)
CAR-AM outlier conformations >22%	122 (8.8%)
CAR-AM seven C _α outliers >20%	23 (1.7%)
Sequence mismatches	23 (1.7%)

Partial twinning was identified by Croll¹ and an alternative and correct model was identified by a user structure of 6S29K (ref. 2). The three outliers for FSC and MI score structures were all reported between the map and model. SUOC: lower MI score indicates that more than 10% of the residues of a structure fit poorly with the map and could potentially be improved.

methods development, responded to this challenge by rapidly categorizing, evaluating and reviewing all experimental protein structures of SARS-CoV-1 and SARS-CoV-2, which comprise the subgenus Sarbecovirus. We do a weekly automatic post-analysis as well as a manual reprocessing and reuploading of representative structures from each of the 18 structurally characterized Sarbecovirus proteins. Every Wednesday, when new PDB structures are released, our automated pipeline identifies new conformational structures and assesses the quality of the models and experimental data. This assessment, along with the original structures, is immediately made available in our online repository at <https://insidecorona.net/>. There we also supply a summary and

SQL database of key statistics and quality indicators, and individual results. After our validation effort began, researchers involved in *in silico* drug screening from Folding@Home 'OpenPandemic' and the EU Joint European Therapeutic Initiative (JETI) expressed great interest. These groups aim to simulate the conformational flexibility of coronavirus proteins and their interactions with each other and with host cell proteins and to design small-molecule inhibitors against key SARS-CoV-2 targets via high-throughput computational modelling, a task that is especially sensitive to the quality of the input model.

In addition to structure evaluation and improvement, <https://insidecorona.net/>

図 4: Nature Struct. Mol. Biol.に掲載の記事

グラフィックとよばれる手法で集められた細胞やオルガネラの 3D イメージは蛋白質粒子を判定できるレベルになり、PDB がもつ原子レベルの構造情報と組み合わせることで、非常に有用な構造情報となりうる可能性を秘めている。急速に発展する高分解能イメージング技術とうまく連携することで、分子細胞生物学などの研究分野や、製薬業界(創薬科学分野)に大きなインパクトを与えると確信している。欧米を中心に、細胞イメージングセンターや、それに準ずるデータセンターの設置が進むが、PDBj も日本国内で同様のセンターの設置・運用に積極的に協力していきたいと考えている。

§6. 自己評価

【目標の達成状況や実施内容に対する研究代表者自身の評価】

申請時のミッションは基本的に全て完了できる見込みである。国際的なデータベースの構築拠点が国内にあり、国際的に広く信頼されている利点を最大限に生かして、正確で使いやすいデータを日本語(中国語、韓国語)で迅速に提供できた。特に本研究課題期間の後半には、新型コロナウイルス感染症の急拡大が社会問題となった。水道や都市ガスのように、使えて当たり前だと思われる PDB のような情報基盤は、緊急事態になればなるほどスピード感をもって、いつも以上の精度で提供することが重要であることを痛感している。

【本プログラムの趣旨に対する、本研究開発課題への貢献の程度】

『わが国の生命科学研究等によって産出された研究データを広く収集するデータベースを対象とし、より多くの多様な研究者にとってより価値のあるものへと発展させる研究開発を推進』という点では、PDBj は 100% の貢献ができていると自負している。日米欧の 3 拠点の中で、PDBj へのデータ登録数の伸びが依然として著しく大きく、PDBj のデータ処理能力は限界に近い。OneDep 登録・評価システムのより一層の効率化と、余裕のある人員での運用が、安定で信頼性のあるデータベース運用の根幹であることを実感している。

『他に開発されているデータベースとの連携・統合化とそれに必要な技術開発』の観点では、RDF 化を推進し、データサイエンスを駆動する基盤的データベースとしての役割を果たした。wwPDB の各パートナーサイトとの連携や、他の統合化プロジェクトとの連携(木下チーム)も行っており、100 点満点とはいかないが合格点で PDBj を運用できていると自己評価している。

『研究者コミュニティ、食品業界、製薬業界などの産業コミュニティとの緊密な連携』の観点では、製薬メーカーとの連携はあるも他産業界との連携がやや手薄であり、もっと緊密な共同研究ができるのではないかと考えている。実際に、PDBj で行っている構造データの厳しい品質管理は、産業界がフォローしにくい内容ではあるが、例えば業界ごとに PDBj-mammal や PDBj-plant といった選択と集中を行うことによって管理基準をチューニングし、より緊密な連携が図れるかもしれない。

§7. 外部発表等

1. 原著論文発表

(1) 論文数概要

種別	国内外	件数
発行済論文	国内(和文)	5 件
	国際(欧文)	12 件
未発行論文	国内(和文)	0 件
	国際(欧文)	1 件

(2) 論文詳細情報

- A-1. Bekker GJ, Yokochi M, Suzuki H, Ikegawa Y, Iwata T, Kudo T, Yura K, Fujiwara T, Kawabata T, and Kurisu G. “Protein Data Bank Japan: Celebrating our 20th anniversary during a global pandemic as the Asian hub of 3D macromolecular structural data”, *Protein Sci.*, under review
- A-2. Sammer V, Burley SK, Kurisu G, Hoch JC, and Markley JL. “The Protein Data Bank Archive”, *Methods Mol. Biol.* Vol. 2305, pp. 3-21, 2021 (DOI: 10.1007/978-1-0716-1406-8_1.)
- A-3. Young JY, Berrisford J, and Chen M. “wwPDB biocuration: on the front line of structural biology”, *Nature methods*, vol. 18, pp. 431-432, 2021 (DOI: 10.1038/s41592-021-01137-z)
- A-4. Shao C, Feng Z, Westbrook JD, Peisach E, Berrisford J, Ikegawa Y, Kurisu G, Sammer V, Burley SK, Young JY. “Modernized uniform representation of carbohydrate molecules in the Protein Data Bank”, *Glycobiology*, vol. 31, pp. 1204-1218, 2021 (DOI: 10.1093/glycob/cwab039)
- A-5. Feng Z, Westbrook JD, Sala R, Smart OS, Bricogne G, Matsubara M, Yamada I, Tsuchiya S, Aoki-Kinoshita KF, Hoch JC, Kurisu G, Velankar S, Burley SK, Young JY. “Enhanced validation of small-molecule ligands and carbohydrates in the Protein Data Bank”, *Structure*, vol. 29, pp. 393-400, 2021 (DOI: 10.1016/j.str.2021.02.004)
- A-6. Romero PR, Kobayashi N, Wendell JR, Baskaran K, Iwata T, Yokochi M, Maziuk D, Yao H, Fujiwara T, Kurisu G, Ulrich EL, Hoch JC, and Markley JL. “BioMagResBank (BMRB) as a Resource for Structural Biology”, *Methods Mol. Biol.*, vol. 2112, pp. 187-218, 2020 (DOI: 10.1007/978-1-0716-0270-6_14.)
- A-7. Adams PD, Afonine PV, Baskaran K, Berman HM, Berrisford J, Bricogne G, Brown DG, Burley SK, Chen M, Feng Z, Flensburg C, Gutmanas A, Hoch JC, Ikegawa Y, Kengaku Y, Krissinel E, Kurisu G, Liang Y, Liebschner D, Mak L, Markley JL, Moriarty NW, Murshudov GN, Noble M, Peisach E, Persikova I, Poon BK, Sobolev OV, Ulrich EL, Velankar S, Vonnrhein C, Westbrook J, Wojdyr M, Yokochi M, and Young JY. “Announcing mandatory submission of PDBx/mmCIF format files for crystallographic depositions to the Protein Data Bank (PDB).” *Acta Crystallograph., Sec. D* vol. 75, pp. 451-454, 2019 (DOI: 10.1107/S2059798319004522)
- A-8. wwPDB consortium, “Protein Data Bank: the single global archive for 3D macromolecular structure data”, *Nucleic Acids Research*, vol. 47, pp.520-528, 2018 (DOI: 10.1093/nar/gky949.)

概要:PDB archive の成り立ちから最近の高度化まで含めたデータベースの全てについて詳細を記載した文献

- A-9. Young JY, Westbrook JD, Feng Z, Peisach E, Persikova I, Sala R, Sen S, Berrisford JM, Swaminathan GJ, Oldfield TJ, Gutmanas A, Igarashi R, Armstrong DR, Baskaran K, Chen L, Chen M, Clark AR, Di Castanzo L, Dimitropoulos D, Gao G, Ghosh S, Gore S, Guranovic V, Hendrickx PMS, Hudson BP, Ikegawa Y, Kengaku Y, Lawson C L, Liang Y, Mak L, Mukhopadhyay A, Narayanan B, Nishiyama K, Patwardhan A, Sahni G, Sanz-García E, Sato J, Sekharan MR, Shao C, Smart OS, Tan L, van Ginkel G, Yang H, Zhuravleva MA, Markley JL, Nakamura H, Kurisu G, Kleywegt GJ, Velankar S, Berman HM, Burley SK. "Worldwide Protein Data Bank biocuration supporting open access to high-quality 3D structural biology data", Database (Oxford), Jan 1, 2018 (DOI:10.1093/database/bay002.)

概要:worldwide Protein Data Bank で行っているバイオキュレーションのプロセスを紹介し、新規化合物のアサイン方法などの詳細を解説した論文

- A-10. Schober D, Jacob D, Wilson M, Cruz JA, Marcu A, Grant JR, Moing A, Deborde C, de Figueiredo LF, Haug K, Rocca-Serra P, Easton J, Ebbels TMD, Hao J, Ludwig C, Gunther UL, Rosato A, Klein MS, Lewis IA, Luchinat C, Jones AR, Grauslys A, Larralde M, Yokochi M, Kobayashi N, Porzel A, Griffin JL, Viant MR, Wishart DS, Steinbeck C, Salek RM, Neumann S. "nmrML: A Community Supported Open Data Standard for the Description, Storage, and Exchange of NMR Data", Anal Chem. Vol.90, No.1, pp.649-656, 2017 (DOI: 10.1021/acs.analchem.7b02795.)
- A-11. Kinjo AR, Bekker GJ, Wako H, Endo S, Tsuchiya Y, Sato H, Nishi H, Kinoshita K, Suzuki H, Kawabata T, Yokochi M, Iwata T, Kobayashi N, Fujiwara T, Kurisu G, and Nakamura H. "New tools and functions in data-out activities at Protein Data Bank Japan (PDBj)". Protein Science. Vol.27, No.1, pp.95-102, 2018 (DOI:10.1002/pro.3273)
- A-12. Gore S, García ES, Hendrickx PMS, Gutmanas A, Westbrook JD, Yang H, Feng Z, Baskaran K, Berrisford JM, Conroy M, Hudson BP, Ikegawa Y, Kobayashi N, Lawson CL, Mading S, Mak L, Mukhopadhyay A, Oldfield TJ, Patwardhan A, Peisach E, Sahni GS, Sekharan MR, Sen S, Shao C, Smart OS, Ulrich EL, Yamashita R, Quesada M, Young JY, Nakamura H, Markley JL, Berman HM, Burley SK, Velankar S, and Kleywegt GJ. "Validation of the Structures in the Protein Data Bank", Structure, Vol. 25, pp.1916-1927, 2017 (DOI: 10.1016/j.str.2017.10.009.)
- A-13. Burley SK, Kurisu G, Markley JL, Nakamura H, Velankar S, Berman HM, Sali A, Schwede T, and Trewthella J. PDB-Dev: a Prototype System for Depositing Integrative/Hybrid Structural Models. Structure, vol. 25, pp. 1317-1318, 2017 (DOI: 10.1016/j.str.2017.08.001.)

2. その他の著作物(総説、書籍など)

1. 栗栖源嗣, 新型コロナウイルスに関する蛋白質構造データベースの開発と応用, ファームステージ, vol. 20,, pp.67-70, 2020
2. 栗栖源嗣, 生体高分子の構造データ検索と解析なら PDBj, 実験医学, Vol. 38, pp889-896, 2020
3. Bekker GJ, 工藤高裕, 池川恭代, 山下鈴子, 栗栖源嗣, Protein Data Bankで利用するPDBx/mmCIF形式について, 日本結晶学会誌, Vol. 61, pp. 159-160, 2019 (DOI: 10.5940/jcrsj.61.159)

4. 中村春木, 栗栖源嗣, データサイエンスと日本蛋白質構造データバンク PDBj の活動, 生物物理, vol. 58, pp. 71-77, 2018 (DOI: 10.2142/biophys.58.071)
5. 小林直宏, 生体系 NMR の解析ツール(機械学習と自動化), 日本核磁気共鳴学会, Vol8, pp. 94-98, 2017

3. 国際学会発表及び主要な国内学会発表

(1) 概要

種別	国内外	件数
招待講演	国内	3 件
	国際	5 件
口頭発表	国内	4 件
	国際	4 件
ポスター発表	国内	9 件
	国際	5 件

(2) 招待講演

〈国内〉

1. 栗栖源嗣, 新型コロナウイルス感染症と立体構造データベース蛋白質科学の視点から, 日本学術会議主催公開 WEB シンポジウム, オンライン, 11月29日
2. 栗栖源嗣, Protein Data Bank における低分子/薬剤化合物の表示と検証レポート, 第 41 1 回 CBI 学会講演会(兼第 21 回 FMO 研究会), 東京工業大学キャンパスイノベーションセンター(東京), 2019 年 12 月 11 日
3. 栗栖源嗣, 蛋白質の構造解析と Protein Data Bank, 関西学院大学特別セミナー, 関西学院大学, 2018/7/6

〈国際〉

1. Genji Kurisu, Protein Data Bank Japan: 20 years and more as the Asian hub for 3D structure and the founding member of the wwPDB, 25th International Union of Crystallography Congress, Prague (Hybrid), Czech Republic, 2021 年 8 月 16 日
2. Genji Kurisu, Protein Data Bank Japan (PDBj): A member of the worldwide Protein Data Bank, Bioinformatics Institute, Singapore, 2019 年 12 月 18 日
3. Kurisu G, Structural atlas of Ferredoxin-dependent proteins in the Protein data Bank, Japan Science Day at RUB, Ruhr University Bochum, Germany, 2018/7/6
4. Kurisu G, "Ligand representation and validation in the Protein Data Bank", International Life Science Integration Workshop, Tokyo, 2018/3/6
5. Kurisu G, "Small molecule ligand/drug representation and validation in the Protein Data Bank", 24th International Union of Crystallography Congress, Hyderabad, India, 2017/8/22

(3) 口頭講演

〈国内〉

1. 工藤高裕, Protein Data Bank Japan (PDBj), 第 42 回 日本分子生物学会年会フォーラム、福岡国際会議場(福岡県福岡市), 2019 年 12 月 3 日
2. 栗栖源嗣, Protein Data Bank とデータの品質管理、CBI 学会, タワーホール船堀(東京), 2019 年 10 月 23 日
3. 栗栖源嗣, 蛋白質構造データバンクのデータ検証高度化と統合化, トーゴの日シンポジウム 2018, 2018/10/5
4. 栗栖源嗣, 蛋白質構造データバンクのデータ検証高度化と統合化, トーゴの日シンポジウム 2017, 東京, 2017 年 10 月 4 日

〈国際〉

1. Kawabata T, Suzuki H, and Kurisu G. Databases and Web services from PDBj for Electron Microscopy, Japan-Korea Bilateral Symposium on Multi-Scale Structural Life Science and the Advanced Technologies, 2019/03/15
2. Kawabata T, Suzuki H, and Kurisu G. Databases and Web services from PDBj for Electron Microscopy, Asian Crystallographic Association, 2018/12/2-5
3. Yamaguchi A, Toh H. Implementing LOD Surfer as a search system for the annotation of multiple protein sequence alignment. JIST 2018: The 8th Joint International Semantic Technology Conference, Awaji, Japan. 2018/11/27
4. Kurisu G. "The Protein Data Bank Japan(PDBj): A Partner of the wwPDB Accepting Structure Depositions through One Dep System", International Symposium on Establishment of Structural Biology Network in Asia and Oceania, 新竹・台湾, 2017/12/6-7

(4) ポスター発表

〈国内〉

1. 横地政志, 岩田武史, 池川恭代, Gert-Jan Bekker, 宮ノ入洋平, 児嶋長次郎, 藤原敏道, 栗栖源嗣, PDB/BMRB のデータ登録システム改善および PDB エントリーの糖鎖データ表現方法の改訂, トーゴの日シンポジウム, オンライン, 2020 年 10 月 5 日
2. 工藤高裕, Gert-Jan Bekker, 山下鈴子, 鈴木博文, 横地政志, 由良敬, 栗栖源嗣, 蛋白質構造データバンク(PDB)の高度化と統合的運用, トーゴの日シンポジウム, オンライン, 2020 年 10 月 5 日
3. 岩田武史, 横地政志, 見学有美子, 張羽澄, 池川恭代, 佐藤純子, 金宙妍, 中川敦史, 宮ノ入洋平, 児嶋長次郎, 藤原敏道, 栗栖源嗣, PDB 検証レポートの高度化と NMR 統合ファイル検証機能の開発, トーゴの日シンポジウム, 日本未来館(東京都江東区), 2019 年 10 月 5 日
4. 工藤高裕, Gert-Jan Bekker, 山下鈴子, 鈴木博文, 横地政志, 由良敬, 栗栖源嗣, 蛋白質構造データバンク(PDBj)の高度化と統合的運用, トーゴの日シンポジウム, 日本未来館(東京都江東区), 2019 年 10 月 5 日
5. 藤博幸, 山口敦子, 工藤高祐, 山下鈴子, Linked open data を利用するアラインメントビューアの開発, 第 18 回蛋白質科学会年会, 新潟市, 6 月 26 日
6. 栗栖源嗣, 蛋白質構造データバンクのデータ検証高度化と統合化, トーゴの日シンポジウム 2017, 2017/10/4-5
7. 小林直宏, PDBj-BMRB における統合化データベースの取り組み: 高度な WEB 検索表示機

能への応用, トーゴの日シンポジウム 2017, 2017/10/4-5

8. 藤 博幸, 工藤高祐, 山下鈴子, ASHViewer の開発, トーゴの日シンポジウム 2017, 東京大学弥生講堂, 2017/10/4-5
9. 鈴木博文, 川端猛, 栗栖源嗣, 中村春木, クライオ EM マップ及び多階層構造を対象とした状態構造検索サービス「Omokage 検索」の改良, 第 55 回日本生物物理学会, 2017/9/19-21

〈国際〉

1. Yokochi M, Iwata T, Miyanoiri Y, Kojima C, Kurisu G, Fujiwara T. BMRBj: Regional NMR Data Repository Site of PDBj. ISMAR-APNMR-NMRSJ-SEST2021, Osaka (Hybrid), 2021 年 8 月 22-27 日
2. Toh H, Yamaguchi A. Alignment Viewer with a Function to Collect Annotation of Proteins from LOD. InCob 2019 (2019) Jakarta. 9 月 11 日
3. Yamaguchi A, Toh H. Annotation of Proteins from LOD for a Viewer of Multiple Protein Sequence Alignment. 11th International Semantic Web Application and Tools for Health Care and Life Sciences Conference, Antwerp, Belgium. 12 月 4 日
4. Kobayashi N, Nagata T, Kojima C, Fujiwara T. “Highly automated NMR signal assignments using peak-noise recognition assisted by Deep Neural Networks”, Gordon Research Conference, Sunday River USA, 2017/7/16
5. Kinjo AR, Bekker GJ, Suzuki H, Tsuchiya Y, Kawabata T, Ikegawa Y, Kurisu G, and Nakamura H. “Protein Data Bank Japan (PDBj): updated semantic web services and tools for large structures”, 19th IUPAB congress, Edinburgh, 2017/7/16

4. 知財出願

(1) 出願件数

種別	件数
特許出願	国内 0 件
	国外 0 件
その他の知的財産出願	0 件

(2) 一覧

- ① 国内出願
該当なし
- ② 海外出願
該当なし
- ③ その他の知的財産権
該当なし

5. 受賞・報道等

(1) 受賞

1. * 文部科学大臣表彰科学技術賞、栗栖源嗣・中村春木, 2020 年 4 月 7 日
2. 大阪科学賞, 栗栖源嗣, 2019 年 11 月 13 日

(2) メディア報道

1. 毎日放送 ミント, 「タンパク質」のデータバンクが研究成果を先駆け公開...新型コロナ治療薬開発には「タンパク質」がカギ, 2020年3月17日
2. NHK 関西のニュース, ウイルス「たんぱく質構造」データベース化, 2020年3月17日
3. 朝日新聞 朝日新聞 DIGITAL、新型コロナの構造データ公開 大阪大、論文発表前に, 2020年3月12日
(プレスリリース”新型コロナウイルスの蛋白質構造情報を集約して3/11から公開“, https://resou.osaka-u.ac.jp/ja/research/2020/20200310_1)

(3) その他

1. NHK スペシャル「人体2」制作協力, 2017年11月5日放送
2. 東京書籍 高校生物, 生物基礎, スタンダード生物へ資料提供, 2019年度使用分

§8. 研究開発期間中の活動

1. 進捗ミーティング

年月日	名称	場所	参加人数	目的・概要
毎週金曜日 10:00-11:00	PDBj システム管理・運営グループミーティング(非公開)	大阪大学蛋白質研究所(Zoom 併用)	13 人	システム開発・運営・進捗についての情報交換
隔週金曜日 13:30-14:30	BMRBj システム管理。運営グループミーティング(非公開)	大阪大学蛋白質研究所(Zoom 併用)	5 人	システム開発・運営・進捗についての情報交換
隔週金曜日 12:30-13:30	PDBj および BMRBj 合同アノテータ・ミーティング	大阪大学蛋白質研究所(Zoom 併用)	11 人	アノテーション処理・運営についての情報交換
毎月不定期 22:00-23:00	wwPDB PI 会議(非公開)	Zoom	5 人	wwPDB の 5 つのメンバーの PI による種々の問題や今後の計画について議論
3ヶ月毎 22:00-23:00	wwPDB PI/OneDep 会議(非公開)	Zoom	18~15 人	wwPDB で共同開発している登録・アノテーションシステムの開発方針決定会議
毎週木曜日 22:00-23:00	OneDep Leader Team 会議(非公開)	Zoom	13~15 人	wwPDB で共同開発している登録・アノテーションシステムの開発進捗管理会議
毎週水曜日 22:00-23:00	Lead Annotator 会議(非公開)	Zoom	5~6 人	wwPDB のアノテータが行うアノテーションに関する毎週の情報交換
2021 年 4 月 13、14、15 日	OneDep Developer Summit(非公開)	Zoom	16 人	wwPDB で共同開発している登録・アノテーションシステムの開発会議
2020 年 3 月 30、31 日	OneDep Developer Summit(非公開)	Zoom	16 人	wwPDB で共同開発している登録・アノテーションシステムの開発会議
2019 年 2 月 26、27 日	OneDep Developer Summit(非公開)	Rutgers 大学、米国ニュージャージー州	16 人	wwPDB で共同開発している登録・アノテーションシステムの開発会議
2018 年 6 月 25、26 日	OneDep Developer Summit(非公開)	EMBL-EBI、英国ヒンクストン	16 人	wwPDB で共同開発している登録・アノテーションシステムの開発会議
2017 年 5 月 8-10 日	wwPDB OneDep Meeting(非公開)	EMBL-EBI、英国ヒンクストン	21 人	wwPDB で共同開発している登録・アノテーションシステムの開発会議、PI も一緒に議論

2. 主催したワークショップ、シンポジウム、アウトリーチ活動等

年月日	名称	場所	参加人数	目的・概要
2021年 12月1日～ 2日(予定)	第44回日本分子生物学会年会(MBSJ2021)内フォーラム	オンライン	未定	PDBのRDFグラフデータの拡張とwwPDB正規アーカイブ化についての紹介
2021年 11月26日 (予定)	第59回日本生物物理学会年会にてバイオフィジックスセミナーを開催	オンライン	100人 (予定)	PDBの最新情報やPDBjの開発ツールについて紹介する
2021年 11月24日 (予定)	PDBアジア地区50周年記念シンポジウム「アジア地区構造生物学の最先端とProtein Data Bank 50年の歩み」	オンライン	300人 (予定)	PDB設立50周年を祝い、アジア地区でのPDB関連活動の歩みと今後の展開について議論する
2021年 11月19日 (予定)	令和3年(2021年)度日本結晶学会年会にてセミナーを開催	オンライン	未定	PDBの最新情報やPDBjの開発ツールについて紹介する。
2021年 8月22-27日	第22回国際磁気共鳴会議、第60回NMR討論会(2021)合同会議	オンライン	200人	PDBjの活動紹介と最新動向紹介を行う
2021年 8月21日	第30回青少年のための科学の祭典「大阪大会 2021 Webサイエンスフェスタ 2021」	オンライン	169人	分子閲覧ソフトによる三次元構造の観察を通じて、新型コロナウイルスの感染について解説をおこなった。
2021年 6月16日	第21回日本蛋白質科学会年会 PDBjランチタイムセミナー	オンライン	75人	PDBの最新情報やPDBjの開発ツールについて紹介した。
2021年 5月4、5日	wwPDB主催PDB50周年記念特別シンポジウム	オンライン	880人	最新の研究成果についてポスターセッションで議論するとともに、PDBの歩みと今後の展開について議論した
2021年 1月20日	PDBj講習会「データベース登録ノウハウ講習会」	オンライン	79人	PDBjで管理・運営するデータベースへの登録方法について、オンラインで勘所を説明した。
2020年 11月22日	サイエンスアゴラ 2020「ようこそ、タンパク質ワンダーランド」	オンライン	134人	小学生高学年を対象に、身近で不思議な「タンパク質」のはたらきと形について解説した。
2020年 11月29日	日本学術会議主催公開WEBシンポジウム「COVID-19パンデミックを契機として考える日本の結晶学の現状と今後」	オンライン	200人	「新型コロナウイルス感染症と立体構造データベース」と題した講演を行い、COVID-19治療薬開発とデータベース

年月日	名称	場所	参加人数	目的・概要
				の役割について講演
2020年 10月27日	配列解析から学ぶバイオインフォマティクス入門	東京大学 医学研究所ヒトゲノム解析センター(オンライン)	不明	データベース検索からアライメントまでの流れを説明。その中で、ビューアの必要性を述べ、具体例として ASHViewer を紹介 (zoom 開催)
2020年 9月16-18日	第58回日本生物物理学会年会で PDBj オンライン展示を開催	オンライン	158人	PDBx/mmCIF フォーマットの説明と PDBj で開発している分子ビューアについて紹介した。
2020年 1月29日	ケンブリッジ構造データベース(CSD)利用セミナー開催	大阪大学蛋白質研究所、大阪府	52人	連携利用を進めている CSD の利用説明会を開催した
2020年 1月23-24日	第2回 wwPDB EM data-management Workshop	EMBL-EBI、英国 Hinxton	45人	電子顕微鏡構造の検証と検証ソフトの開発に関する会議を共催した
2020年 1月10日	韓国生体分子科学連合学会でランチョンセミナーを開催	光州科学技術院、韓国・光州市	200人	PDBx/mmCIF フォーマットの説明と wwPDB で開発している OneDep 登録システムについて紹介した。
2019年 12月3-6日	第42回日本分子生物学会年会・特別企画「使ってみようバイオデータベースー広がるデータ、つながる世界」に出展	マリンメッセ福岡、福岡県	7,129人/年会	wwPDB と PDBj の活動や開発ツールを紹介するブース展示を行った。
2019年 11月30日	2019 大阪大学共創 DAY@EXPOCITY にブース出展	ららぽーと EXPOCITY、大阪府	500人	分子閲覧ソフトによる三次元構造の観察や、3D プリンタやペーパークラフトでの蛋白質構造制作を行い、を分かりやすく一般へ紹介した。
2019年 11月19日	日本結晶学会令和元年度年會にてランチョンセミナー開催	金沢市文化ホール、石川県	80人	PDB の最新情報や PDBj で開発した mmCIF editor について紹介した。
2019年 10月5日	トーゴの日シンポジウム 2019 ワークショップ	日本科学未来館、東京都江東区	120人	生命科学分野のデータ統合にまつわる問題とともに考え、議論を深めるシンポジウムで PDBj の活動紹介を行った。
2019年 10月26日	さくらサイエンス	関西学院大学神戸三田キャン	20人	インドネシアのパジャジャラン大学、ウダヤナ大学の学生を招聘するプログラムの中でタ

年月日	名称	場所	参加人数	目的・概要
		パス		ンパク質の立体構造解析を molmil を使って説明
2019年 9月24日	第57回日本生物物理学会年 会で PDBj ランチョンセミナーを 開催	シーガイア コンベンシ ョンセンタ ー、宮崎県	100人	PDB の最新情報や PDBj の 開発ツールについて紹介し た。
2019年 9月10日	生命医薬情報学連合大会 (IIB MP2019) にて PDBj ランチョン セミナー開催	東京工業 大学大岡 山キャン パス、東京都	70人	PDB の最新情報や wwPDB で配布している検証レポート の最新情報について紹介し た。
2019年 8月18日	第29回青少年のための科学 の祭典(サイエンスフェスタ)大 阪大会 2019 にブース出展	ハービスホ ール、大阪 府	350人	分子閲覧ソフトによる三次元 構造の観察を通じて、新型コ ロナウイルスの感染について 解説をおこなった。
2019年 6月26日	第19回日本蛋白質科学会年 会にて PDBj ランチョンセミナー を開催	神戸国際 会議場、兵 庫県	110人	PDB の最新情報や wwPDB で配布している検証レポート の最新情報について紹介し た。
2019年 6月6日	統合データベース講習会: AJA CS 徳島	大塚製薬株 式会社 徳 島研究所、 徳島研	80人	PDBj の活動紹介を行った
2019年 5月23日	配列解析から学ぶバイオインフ ォマティクス入門	東京大学 医科学研究 所ヒトゲ ノム解析セ ンター	不明	データベース検索からアライ メントまでの流れを説明。 その中で、ビューアの必要性 を述べ、具体例として ASHVi ewer を紹介
2019年 3月1日	米国生物物理学会サテライト Workshop「Working towards federating structural models and data」	ボルチモア 国際会議 場、米国メ リーランド 州	38人	複数の実験手法と計算を組 み合わせた Integrated/Hyb rid 法による構造をどう集め、 検証するかを議論した
2019年 2月19日	PDBj 講習会「NMR による構造 計算入門」を開催	大阪大学 蛋白質研 究所、大阪 府	12人	BMRB を活用した NMR によ る構造計算について講義と 演習を行い、PDBj で開発し たツールの先端利用法を説 明。
2019年 1月28日	All-in-one 合同講習会 2018/ 統合データベース講習会に PD Bj から講演	JST 東京本 部別館、東 京都	不明	蛋白質構造データバンク PD B の使い方について講演を 行った

年月日	名称	場所	参加人数	目的・概要
2019年 1月13-15日	第1回 wwPDB EM data-management Workshop	EMBL-EBI、英国 Hinxton	49人	電子顕微鏡構造の検証と検証ソフトの開発に関する会議を共催した
2018年 11月28-30日	第41回日本分子生物学会年会にてブース出展	パシフィコ横浜、神奈川県	7,519人/年会	wwPDB と PDBj の活動や開発ツールを紹介するブース展示を行った。
2018年 11月17日	2018 大阪大学共創 DAY@EXPOCITY にブース出展	ららぽーと EXPOCITY、大阪府	500人/資料配布数	分子閲覧ソフトによる三次元構造の観察や、3D プリンタやペーパークラフトでの蛋白質構造制作を行い、を分かりやすく一般へ紹介した。
2018年 11月11日	日本結晶学会 2018 年度年会にてランチョンセミナー開催	東京工業大学大岡山キャンパス、東京都	70人	PDB の最新情報や PDBj で開発したサービスについて紹介した。
2018年 11月10、11日	サイエンスアゴラ 2018 -in 東京-に出展	テレコムセンタービル、東京都	480人/資料配布数	分子閲覧ソフトによる三次元構造の観察や、3D プリンタやペーパークラフトでの蛋白質構造制作を行い、を分かりやすく一般へ紹介した。
2018年 9月20日	生命医薬情報学連合大会 (IIBMP2018) にて PDBj ランチョンセミナー開催	荘銀タクト鶴岡、山形県	60人	PDB の最新情報や PDBj の開発ツールについて紹介した。
2018年 9月16日	第56回日本生物物理学会年会で PDBj ランチョンセミナーを開催	岡山大学津島キャンパス、岡山県	100人	DDBJ と共同で PDB の検証レポートの紹介と DDBJ のサービスについて紹介した。
2018年 7月10日	統合データベース講習会: AJACS 筑波4にて PDBj から講演	物質・材料研究機構、茨城県	30人	蛋白質構造データベース PDB の使い方について講演を行った
2018年 6月28日	第18回日本蛋白質科学会年会にて PDBj ランチョンセミナーを開催	朱鷺メッセ、新潟県	100人	PDB と BMRB の最新情報の提供と CSD との連携利用について紹介した。
2018年 3月27日	第138回日本薬学会年会にて PDBj ランチョンセミナーを開催	ホテル金沢、石川県	150人	PDB の最新情報や PDBj を活用したデータベース構築について紹介した。
2018年 2月20日	PDBj 講習会「生命科学のための立体構造データの使い方入門」を開催	大阪大学銀杏会館、大阪府	35人	電子顕微鏡により決定された構造と密度マップの利用の仕方を中心に講習した。
2017年	ConBio2017 生命科学系学会	神戸ポート	10,201	wwPDB と PDBj の活動や開

年月日	名称	場所	参加人数	目的・概要
12月6-9日	合同年次大会にてブース出展	ピアアイランド、兵庫県	人 /年会	発ツールを紹介するブース展示を行った。
2017年 11月24-26日	サイエンスアゴラ 2017 -in 東京-に出展	テレコムセンタービル、東京都	3,328 人/来場	分子閲覧ソフトによる三次元構造の観察や、3D プリンタやペーパークラフトでの蛋白質構造制作を行い、を分かりやすく一般へ紹介した。
2017年 11月24日	日本結晶学会 2017 年度年会にてランチョンセミナー開催	アステールプラザ 広島、広島県	100 人	PDB の最新情報や、PDB へのデータ登録の新機能等について紹介した。
2017年 9月27日	生命医薬情報学連合大会 (IIB MP2017)にて PDBj ランチョンセミナー開催	北海道大学 情報科学研究科、北海道	100 人	PDB の最新情報や PDBj を活用したデータベース構築について紹介した。
2017年 9月21日	第 55 回日本生物物理学会年会で PDBj ランチョンセミナーを開催	熊本大学 黒髪キャンパス、熊本県	100 人	PDB の最新情報や PDBj で開発したサービスについて紹介した。
2017年 8月21-28日	国際結晶学連合会議 (IUCr2017) でブース展示	ハイデラバード国際会議場、インド・ハイデラバード市	1,748 人/会議	wwPDB の活動や OneDep システムを用いた登録方法について紹介
2017年 6月22日	第 17 回日本蛋白質科学会年會にて PDBj ランチョンセミナーを開催	仙台国際センター、宮城県	100 人	PDB と BMRB の最新情報の提供と PDBj で開発した分子ビューア MolMil について紹介した。
2017年 5月8日	All-in-one 合同講習会 2017/生命科学の最先端に触れてみようにて講演	三島市民文化会館、静岡県	50 人	NBDC、DBCLS、DDBJ、PDBj 合同で、一般向けにゲノム情報や蛋白質構造などの生命科学のビッグデータを紹介するデータベース講習会を行った。
2017年 4月20日	大阪大学いちよう祭・蛋白研展示『タンパク質分子の『立体構造にさわる』体験をしてみよう』に出展	大阪大学 蛋白質研究所、大阪府	70 人	閲覧ソフトによる三次元構造の観察や、3D プリンタやペーパークラフトでの蛋白質構造制作を行い、を分かりやすく一般へ紹介した。

以上

別紙 研究開発対象のデータベース等

No.	正式名称	別称	概要	URL	公開日	状態	分類	生命科学系データベースアーカイブ	NBDCヒトデータベース	NBDC RDFポータル	関連文献 (論文リストに記載があれば、その番号でも可)
1	Protein Data Bank	PDB Archive	生体高分子の立体構造データベース, wwPDBと協力して構築, RDFを開発, 公開	https://pdbj.org	2002/4/1	継続・発展	データベース等	公開済	対象外	公開済	A-2, A-3, A-8
2	Biological Magnetic Resonance Data Bank	BMRB	生体高分子の化学シフト, 緩和データ, 相互作用データ等のNMRの実験データのデータバンク	http://bmrdep.pdbj.org	2011/4/1	継続・発展	データベース等	公開済	対象外	公開済	A-2, A-6, A-10
3	eF-site	同左	蛋白質の分子表面の形状と物性(静電ポテンシャルと疎水性度)を機能部位情報と結合したデータベース. 維持・更新のみ	https://pdbj.org/eF-site/	2002/3/1	継続・発展	データベース等	公開済	対象外	対象外	A-11
4	ProMode elastic	同左	二面角を変数とする基準振動解析プログラムによって計算された蛋白質のダイナミクス・データベース. 維持・更新のみ.	https://pdbj.org/promode-elastic/	2003/4/1	継続・発展	データベース等	公開済	対象外	対象外	A-11
5	Molecule of the Month	MoM	RCSB-PDBより毎月提供されている分子解説記事「Molecule of the Month」を日本語に訳したもの. 社会で話題となっている内容に関わる分子をPDBから選び, 機能と構造に関して解説. 維持・更新のみ.	https://pdbj.org/mom/	2008/4/1	継続・発展	データベース等	公開済	対象外	対象外	
6	EM Navigator	同左	生体分子や生体組織の3次元電子顕微鏡データ(EMDB)閲覧用web site	https://pdbj.org/emnavi/	2007/5/1	継続・発展	ツール等	公開済	対象外	対象外	
7	ASH	同左	PDBデータを基にした構造アライメント	https://sysimm.ifre.c.osaka-u.ac.jp/ash_service/		継続・発展	ツール等	公開済	対象外	対象外	
8	Molmil	同左	インターネット上のweb環境で稼働するJavaScriptによる分子構造ビューア	https://pdbj.org/help/molmil	2014/9/1	継続・発展	ツール等	公開済	対象外	対象外	A-11