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）

2023 年度 研究開発実施報告

概要

研究開発課題名	蛋白質構造データバンクのデータ駆動型研究基盤への拡張
開発対象データベースの名称（URL）	日本蛋白質構造データバンク（PDBj） (https://pdbj.org)
研究代表者氏名	栗栖 源嗣（90294131）
所属・役職	大阪大学 蛋白質研究所 教授（2024 年 3 月時点）

目次

§1. 研究実施体制

グループ名	研究代表者または主たる共同研究者氏名	所属機関・役職名	研究題目
研究代表者グループ（大阪大学グループ）	栗栖　源嗣	大阪大学・教授	蛋白質構造データバンクの構築・検証・公開、統合化および高度化と利用促進、国際協力
研究分担者グループ（蛋白質研究奨励会グループ）	栗栖　源嗣	(財)蛋白質研究奨励会・客員研究員	蛋白質構造データバンクの構築・検証・公開と安定運用

§2. 研究開発対象とするデータベース・ツール等

(1) データベース一覧

【主なデータベース】

No.	名称	別称・略称	URL
1	PDB Archive	PDB Archive	https://pdbj.org

【その他のデータベース】

No.	名称	別称・略称	URL
1	BMRB	BMRB	https://bmrbj.pdbj.org
2	eF-site	eF-site	https://pdbj.org/eF-site/
3	ProMode elastic	ProMod	https://pdbj.org/promode-elastic/
4	MoM	MoM	https://pdbj.org/mom/

(2) ツール等一覧

No.	名称	別称・略称	URL
1	EM Navigator	EM Navi	https://pdbj.org/emnavi/
2	DASH	DASH	https://sysimm.org/dash/
3	MolMil2	MolMil	https://pdbj.org/molmil2/
4	OneDep	OneDep	https://deposit-pdbj.wwpdb.org/deposition

§3. 実施内容

(1) 本年度の研究開発計画と達成目標

本課題は、2011 年度からの統合化推進プログラムにて構築・運営された PDBj および BMRBj(PDBj-BMRB を改称)をさらに発展させ、国内外の他のデータベースとの統合化を図り、研究者コミュニティや関連する製薬企業等の産業界が必須としている基盤的データを提供するものであり、完全な接続性を保つ。2023 年度の具体的な実施計画と達成目標を以下に示す。

i. 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

i-1) 国際基準での登録・編集・検証・公開

習熟した専従研究員(アノテータ)が、X 線結晶解析、クライオ電子顕微鏡構造解析、NMR 構造解析の実験研究者から登録された蛋白質構造データ(原子座標とメタデータ)と NMR の実験情報データ(化学シフトなど)を、世界共通のアノテーション手法により厳密な品質を管理しつつ登録・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実施する。wwPDB の他メンバー(RCSB-PDB, PDBe, BMRB, EMDB)や、アジア諸国の関連研究者との国際合意に基づいて、wwPDB で割り当てられた地域範囲を一つの漏れなく分担して完全にデータ処理を行う。i-2)で述べる登録・編集・検証システム(OneDep システム)の高度化によって、構造データの品質を検証する記載精度を高めつつ、アノテータの数を増やすことなく、増加し続ける構造データの処理に対応する。

本年度は、2022 年度に上海で活動を開始した PDB China 用に中国発のデータの一部を転送し、上海からリモートでのデータ編集・検証業務を監督することを継続するとともに、複雑なエントリーであってもwwPDB の既存メンバーと同等の品質でデータ処理できるよう、PDB China のメンターリングを継続する。PDB China が利用している編集・検証システム(登録システムは当面上海には設置しない)を PDB China のメンバーが自分でインストールし維持・更新できるよう、OneDep システムの汎用性拡充を進める(現在は全て PDBj のメンバーが行っている)。それにより、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。

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、一般公開されている。予測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測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測構造自身を評価しつつ、予測構造と実験構造が混合した場合に、構造データ全体をどう評価し活用していくのが最良であるのか、wwPDB の他メンバーとともに 2022 年度にタスクフォースを組織したので、予想構造を含めた統合的な検証レポートの検討を進める。

i-2) OneDep 登録システムの高度化

第3期までの活動により、X線結晶解析、クライオ電子顕微鏡構造解析、NMR 構造解析の3手法全てに対して、検証レポート(Validation Report)が提供できる状況となっている。構造解析の論文を投稿する際に、wwPDB の正式検証レポートを添付して投稿することを必須要件とするように、Nature や Science をはじめ、主要な学術誌に依頼済みで、多くの Journal で投稿必須要件としていただいている。それにより wwPDB の各サイトに登録されてくるエントリーの最初の段階での品質が著しく向上し、アノテータがエントリーの修正を依

頼する件数の減少に貢献している。しかし、今現在も、特にクライオ電顕の構造エントリーは、分子サイズが大きく構造が複雑で、かつ構造解析の分解能が低い場合が多いので、エントリーの処理に著しく時間を取られているのが現状である。配列データベースや CSD などの化合物データベースとの相互チェックや登録ユーザーインターフェースの高度化・自動化を進めるのみならず、より一層の厳しい検証項目を議論し実装することで、エントリー1つあたりのアノテーション処理にかかる時間を短縮する。コロナ禍前に設置したクライオ電子顕微鏡の Validation Task Force (VTF)での議論をベースに、クライオ電子顕微鏡の専門家が承認する新たな検証項目(実験データと原子座標の consistency)の追加を継続して実施する。2023 年度の追加支援により、NMR の構造エントリーを効率よく処理できるよう、OneDep/BMRBdep の処理システムでの NMR データ統一フォーマットへの自動変換と自動処理を進める。

ii. 他のデータベースとの統合化および高度化

ii-1) PDB 統合利用ポータルの構築(統合利用に向けたプログラム類、データベースの検討選定)

第3期までの統合化プロジェクトにおいて、PDBj は 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。第 I、II 期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 における世界標準の正規フォーマットとして採用されるに至っている。しかし、SPARQL 検索を駆使したデータベース横断的な利用が広く普及している状況とは言えず、リッチなメタデータを十分に活用した検索や機械学習用データセットの生成ができていないと現状分析をしていた。そこで、今年度も引き続き、以下を実施する。

ii-1-1. PDB/RDF アーカイブの拡充

PDBj が主体となって開発した PDB の RDF 表現のアーカイブ(PDB/RDF)は、第 III 期プロジェクト期間中に拡張して、全ての PDB コアアーカイブ(PDB, CCD, BIRD, VRPT, SIFTS)を網羅した。一方、PDB/RDF アーカイブのポータルサイト (https://rdf.wwpdb.org)で閲覧可能なアーカイブは、PDB と CCD だけであり、エントリーID のみ検索可能な状態に留まっている。そこで、上記の全ての PDB コアアーカイブの RDF 表現を閲覧可能にするとともに、クエリ文の入力による検索を可能にする為の開発を継続する。

ii-1-2. 統合利用に向けたプログラム類、データベースの検討選定

すでに RDF 化されて公開されているデータベースに関しては、wwPDB と UniProt や Pfam、GO、SCOP、CATH などとの残基レベルでのマッピングを行っている SIFTS を通して、比較的容易に統合利用を検討することができると考えられる。例えば、Endpoint Browser (https://sparql-support.dbcls.jp/endpoint-browser.html)を用いて各 RDF データベースの中身を解析し、個々のデータベースのみからは得られない統合的な検索例を作成する。一方、生命科学分野のデータベースは RDF 化されていないものも数多く存在する。それらとの統合のために、統合データウェアハウスを構築するフレームワークの 1 つである、InterMine (http://intermine.org)の活用について検討を継続する。

ii-2) PDB 統合利用ポータルの構築(AI 開発の動向、生物学・化学の視点からの検索等に関する需要の調査と予備解析)

今期、上記の RDF 形式などによる(メタ)データを活用した高度な検索および部分データ取得システムを構築するため、「情報科学や結晶学に詳しくないユーザーが生物学・化学的興味から適切なデータを発見す

る」のと、「情報科学研究者が機械学習などに必要なデータセットをプログラム的に生成する」という両面を支援し、ゲノム等の他のデータベースとの統合的な利活用ができるよう PDB/RDF アーカイブの拡充と統合利用に向けたプログラム類、データベース検討選定を実施した。本年度、継続して各種学会等で細かな意向調査(アンケート)を継続実施し、需要調査と予備解析を実施する。

ii-3) 化合物情報に特化した機械学習用データセットの公開

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%近くのエントリーが非蛋白質分子、すなわち「リガンド化合物」を含んでいる。創薬ターゲットの蛋白質の場合には、このリガンド化合物の構造情報が極めて重要である。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、実験的に得られた密度マップとの整合性の点で信頼度の低い構造も含まれている。実際に、結晶学や NMR 分光学、電子顕微鏡学を専門としない一般の生命科学研究者がデータを利用する際に、「とりあえず2Åよりも高分解能」といった単一的な尺度で構造データをフィルタリングし、機械学習やデータ分析に用いられている例が多いが、2Å を超える高分解能でも相互作用が弱くリガンド化合物の占有率が低い場合などは実験データとの整合性は著しく低い。第 III 期の統合化プログラムにおいて、低分子化合物の結晶構造データベースである CSD との連携を強め、現在では、登録時に Ligand Validation を行って、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、CSD 構造情報とのリンクが完全に行われている。更に、リガンド化合物の実験データとの一致度を残基レベルで評価できる指標として「実空間信頼度因子(Real Space Reliability Factor: RSR)」や「マップファイルとの相関係数(CC)」、化学結合距離や結合角度の標準的な値からのずれを示す統計値などを、計算機で機械的に解釈できるように検証レポートを XML 化し更に RDF 化して、PDBj のサイトと wwPDB のサイトから公開している。そこで、PDBj では構造生物学実験の専門家がデータベースの構築に深く関与している環境を生かし、これら実験データとの整合性を専門的に評価する豊富な指標を駆使して、高精度・高品質のデータを事前に選抜した選抜データセットを公開するため、昨年度にさまざまな学会で選抜基準や利用動向についてアンケートを実施し、利用者の動向を丁寧に調査した。今年度は動向調査を継続するとともに、以下を実施する。

ii-3-1. 選抜基準の策定と必要とする検証パラメータの検討

創薬候補化合物のターゲット蛋白質へのドッキングシュミ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化学的知見を高めたりする上で、利用者が必要とする構造精度は目的によって異なってくる。例えば、量子化学計算を伴う QM/MM 法などでは、結合長や結合角まで構造解析の精度が大きく計算結果に影響するであろうし、MD をベースとした結合シミュレーションであっても実験的な根拠の低いノイズのような実験根拠を基に、いくら結合をシミュレーションしたとしても、途中の結合の様子を正確にトレースすることは困難であろう。実装済みの RDF 化した検証レポートの恩恵を最大限に活かし、化合物に特化した機械学習用のデータセットを選択する上で、目的別に実験データとの整合性をどの程度、どういった視点で検証し、フィルタリングするのが最適であるのかを、利用者の声を聞いた結果を検討して、目的ごとの機械学習用データセット選定のための基準作り、第一弾の選抜データセットの作成を行う。

ii-4) NMR 制限情報の標準化と検証結果の可視化

NMR を用いて実験的に決定された立体構造を PDB に登録する際、化学シフトと NMR 制限情報などの実験データの提出が義務付けられている。このうち化学シフトは、NMR-STAR フォーマットで標準化され、原

子座標と適切に対応づけされ、最終的に BMRB コアアーカイブとして公開される。一方で NMR 制限情報については、OneDep 登録システムのサポートが不十分な状態が続いている。最近の一部の例外を除いて、NMR を実験手法とするほぼ全ての PDB エントリーの NMR 制限情報について、PDB 側は登録者の使用したソフトウェア固有のフォーマットをそのまま公開している。BMRB 側では、PDB エントリー公開のあと、適宜 NMR-STAR フォーマットに変換して公開しているが、未対応のフォーマットや（本来データ登録時に対処すべきな）データの不整合には対処できてない。過去に登録されたエントリーの NMR 制限情報は、化学シフトと異なりデータの再利用と検証が困難な状態が続いている。従って NMR 制限情報と原子座標を対応づけし直し、標準化されたデータファイルを作成して公開することが求められている。

先のプロジェクトの開発成果により、OneDep 登録システムは化学シフトと NMR 制限情報を一つのファイルにまとめた標準フォーマットを用いた NMR 構造データの登録が可能になった。このシステムを応用して、2023 年度も継続して過去の NMR エントリーの NMR 制限情報を標準化、再検証を行うとともに、検証結果の可視化に向けて必要なツール類の整備を進める。

iii. データベースの安定運用と利用促進、国際協力

iii-1) データベースの安定運用

世界的な生命科学の主要情報基盤であると自負している PDB および BMRB の安定運用が、電気、ガス、水道と同様に、1 日も途切れることなく求められていることは、新型コロナウイルス感染症関連のエントリーを取り扱った際に痛感した出来事であった。PDBj のメインサーバー群は 2018 年まで、大阪大学吹田キャンパスにある蛋白質研究所のサーバー室にバックアップ用を含め 2 セット設置されていたが、大学の基幹ネットワークの定期メンテナンスの際には、一時的なサービス停止を全世界に向けてアナウンスしなければならない状況であった。そこで、大阪大学蛋白質研究所とは地理的に6km ほど離れた位置にある（財）蛋白質研究奨励会に、蛋白質研究所の経費でサーバー室を借用し、2019 年から PDBj および BMRBj のバックアップサーバーを構築して、定期メンテナンスやメインサーバーの機種更新の際に活用を開始している。このバックアップサーバーの機能を維持・拡張し、2022 年より、アノテーション業務の一部も（財）蛋白質研究奨励会のデータベース研究支援部門において実施する体制を構築して、より安定的なデータベース運用を行っている。

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、wwPDB で採用している STAR 形式の PDBx/mmCIF フォーマットに対する深い知識、および何よりアノテーション業務に携わった長い経験がものをいうが、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と、データベースの安定運用に大きな懸念が生じる。問題点を先取りして対応するため、（財）蛋白質研究奨励会の研究員を兼務する栗栖が、経験抱負な専属アノテータ 4 名を蛋白質研究奨励会で雇用して安定かつ信頼性の高いデータベースとして運用を継続する。

iii-2) 利用者・研究者コミュニティとの連携

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、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、2017 年以降 PDBj 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ていただいている。2023 年度は Axcelead Drug Discovery Partners 株式会社の曽我部主任研究員に委員を交代していただく予定である。また、研究代表者である栗栖が Vice President（無報酬）を務めている wwPDB Foundation（米国内の NPO 財団）の枠組

みを利用して、継続して企業とwwPDBとの連携を図る。具体的には、今後は構造予測ソフトAlphaFold2で注目を集めているDeepMind社の意見を聞けるよう交渉した結果、現在は協力可能かどうかの返事待ちの状況である。国内諮問委員になっていただいていた企業研究者からの具体的な要望を受けて、日本の製薬企業からも欧米と同程度にデータ登録が行える環境整備を継続して進めていく。

iii-3) 国際協力

現状の構造データの生産・利用状況から、中国にPDB China (PDBc)およびインドにPDB India (PDBi)を設置し、wwPDB のフランチャイズを増やす可能性が長く議論されてきた。この議論をベースに、wwPDBの運営諮問委員会においてPDBcの早期実現が諮問され、アジア圏を代表している我々PDBjが、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貢献していくことが求められた。国立蛋白質研究センター上海（National Facility/Center for Protein Science in Shanghai）のWenqing Xu所長（兼PDBc代表）と、毎月定例のZoom会議を開催し、PDBcの立ち上げに必要な協力を継続して実施していく。2022年度にwwPDBの准メンバーとして迎えいれた PDB China に対し、継続してアノテータのトレーニングや、実際のデータベース運用のノウハウなどをwwPDBの他メンバーと分担しながら支援していく。

PDBc および PDBi が設立され稼働を始めた場合には、対抗するのではなく同一のデータベースを構築する国際協力機関として協力していく。実際、大阪大学蛋白質研究所と学術交流協定も締結し、継続してより一層の国際交流を進めていく。

(2) 進捗状況

i. 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

i-1) 国際基準での登録・編集・検証・公開

本年度も引き続き、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。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。具体的には、2023 年度に PDB 全体では 21,786 件エントリーが増加し、そのうち PDBj は 6,738 件を処理した。PDBj が処理したエントリー数のうち、1,555 件は PDB China 用のサーバーにデータを転送し PDBj の指導の下で PDBc のアノテータが登録処理した件数となる。BMRB は全体で 775 件増加し、そのうち 102 件を PDBj から新規登録した。

予測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に対応するため、ベイズ統計の手法を用いて予測構造自身を評価し、予測構造と実験構造が混合した場合に構造データ全体をどう評価し活用していくのを検討するタスクフォースを組織した。このプロジェクトへの経済的支援を求めるべく、AlphaFold2 を開発した Deep Mind 社に経済的支援（主にタスクフォースメンバーの旅費、滞在費）を要請し、提案書を Deep Mind 社で検討していただいたが、残念ながら支援には至らなかった。そこでまず初期モデルに AlphaFold の構造を用いたのか完全に実験で決定したのかを厳密にする目的で、今後のタスクフォースでの議論を見越し、新しい PDBx/mmCIF カテゴリ_pdbx_initial_refinement_model を導入して、X 線、電子顕微鏡、NMR 法の初期モデルに関する情報収集を改善した。これにより、実験的に得られた初

期モデルと計算で得られたモデルが区別される。初期モデルが得られたリソースの出どころ（例えば、PDB、AlphaFoldDB、RoseTTAFold など）と、そのアクセッションコード又は識別子が公開されている場合は、その情報を取得できるようにした。

i-2） OneDep 登録システムの高度化

本年度は、OneDep 登録システムの高度化案件として、NMR 構造を登録する際に問題となる NMR データファイルの統一システムの高度化を実施した（追加支援）。NMR の実験データには様々なフォーマットが存在するが、wwPDB では NEF および NMR-STAR フォーマットを標準としている。OneDep 登録システムの高度化により PDB エントリーの NMR データを 1 つの NMR-STAR/NEF ファイルに統合して提供できるようにした（図1）。 wwPDB では、NMR データ（距離制限、化学シフト、場合によってはピークリスト）の NMR-STAR/NEF 形式での単一ファイルのアップロードを進めており、将来的には OneDep でのソフトウェア固有の形式でのアップロードを廃止する予定である。この改善により、NMR 構造の1エントリーの処理にかかる時間を削減することが期待される。

コロナ禍前に設置したクライオ電子顕微鏡の Validation Task Force （VTF）での議論をベースに、クライオ電子顕微鏡の専門家が承認する新たな検証項目（実験データと原子座標の consistency）を追加した。この内容は専門誌 IUCr J に論文発表した。

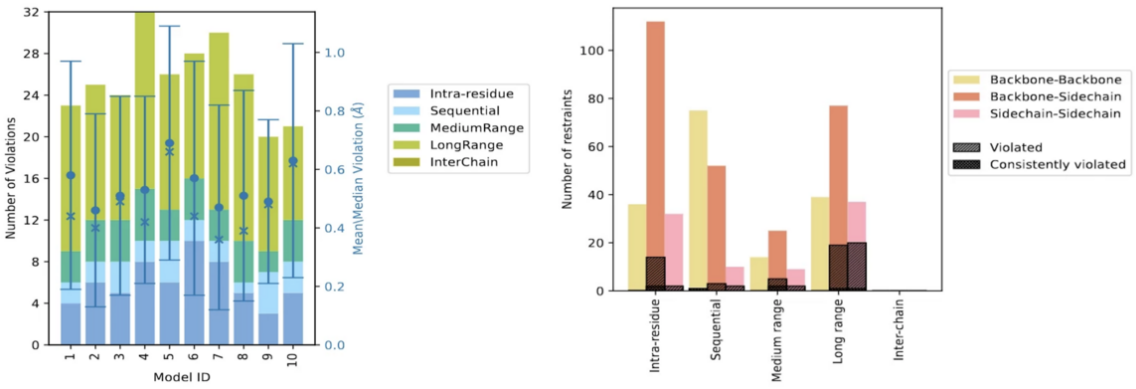

図 1. 統一 NMR データで拡張して計算された検証レポート

ii. 他のデータベースとの統合化および高度化

ii-1-1. PDB/RDF アーカイブの拡充

PDBj が主体となって開発した PDB の RDF 表現のアーカイブ（PDB/RDF）は、第 III 期プロジェクト期間中に拡張して、全ての PDB コアアーカイブ（PDB, CCD, BIRD, VRPT, SIFTS）を網羅した。一方、PDB/RDF アーカイブのポータルサイト（https://rdf.wwpdb.org）で閲覧可能なアーカイブは、PDB と CCD だけであり、エントリーID のみ検索可能な状態に留まっている。そこで、上記の全ての PDB コアアーカイブの RDF 表現を閲覧可能にするとともに、クエリ文の入力による検索を可能にする計画であった。しかし、担当者

が i-2) OneDep 登録システムの高度化において NMR 構造を登録する際に問題となる NMR データファイルの統一システムの高度化・自動化を進める開発項目に予定以上の時間を取られてしまい、rdf.wwpdb.org の拡充に時間を割くことができなかった。令和 6 年度に優先的に対応する予定である。

ii-1-2. 統合利用に向けたプログラム類、データベースの検討選定

wwPDB と UniProt や Pfam、GO、SCOP、CATH などとの残基レベルでのマッピングを行っている SIFTS に含まれていないデータベースとの PDB との統合利用に関して、NBDC/DBCLS で開発されている、RDF 化されたデータベースを統合的に検索するためのフレームワーク TogoDX（https://togodx.dbcls.jp/human/）について調査した。PDB と Reactiome のパスウェイデータや医薬品類似化合物のデータベース ChEMBL に含まれる活性データと薬効データなどを組み合わせて検索できることを確認した。また、2021 年、22 年と日本蛋白質科学会、生命医薬情報学連合大会、日本生物物理学会、CBI 学会、日本結晶学会、ライフインテリジェンスコンソーシアムなどで、利用者がこれまでにどのようなデータベースと関連付けて PDB を利用してきたか、また今後したいと考えているかについて詳細な調査を実施し意見を収集した。この意見をもとに、ゲノム情報との統合利用の需要が大きいと判断し、まずは日本発のコホート調査のデータを基に作成された jMorp（日本人多層オミクス参照パネル）との連携を進めた。jMorp のデータを PDB データと統合利用するためには、まず Uniprot ベースのサマリーページを用意し、どの PDB エントリーを選ぶのが最適であるかを利用者が判断する必要がある。エントリー毎の Region explorer や Chain Topology の表示などの新機能を開発し、PDBj への実装準備を進めている（図 2）。

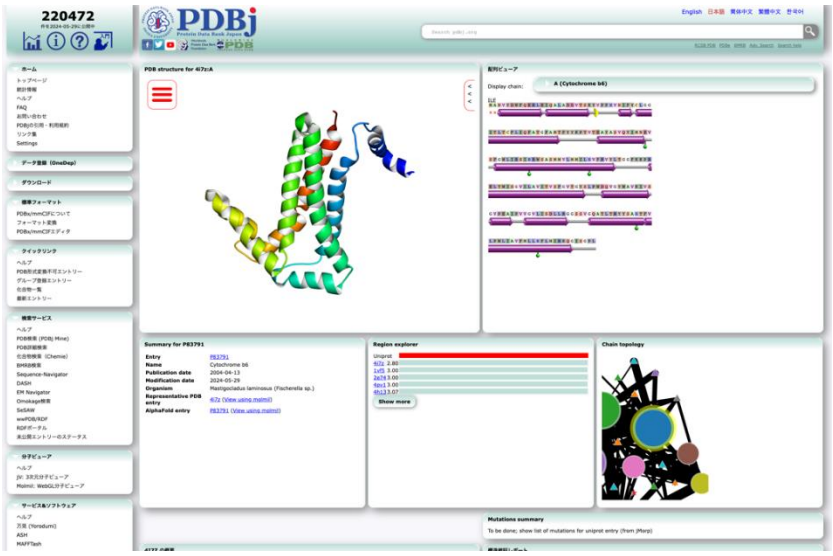

図 2. 開発途中の UniProt サマリーページ

ii-3) 化合物情報に特化した機械学習用データセットの公開

ii-3-1. 選抜基準の策定と必要とする検証パラメータの検討

創薬候補化合物のターゲット蛋白質へのドッキングシュミ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化学的知見を高めたりする上で、利用者が必要とする構造精度は目的によって異なってくる。化合物に特化した機械学習用のデータセットを選択する上で、目的別に実験データとの整合性

をどの程度、どういった視点で検証し、フィルタリングするのが最適であるのか、利用者の声を聞くために、2022 年度、23 年度に日本蛋白質科学会、生命医薬情報学連合大会、日本生物物理学会、CBI 学会、日本結晶学会で詳細な利用動向調査を実施した。目的ごとの機械学習用データセット選定のための基準作りを進めており、検証方法の確立している X 線結晶解析については概ね方法を固定した。最近エントリー数が急増しているクライオ電子顕微鏡構造については、検証方法が確立していないが Purdue 大学の木原教授らが開発している DAQ-score を用いることとした。

ii-4) NMR 制限情報の標準化と検証結果の可視化

NMR を用いて実験的に決定された立体構造を PDB に登録する際、化学シフトと NMR 制限情報などの実験データの提出が義務付けられている。先のプロジェクトの開発成果により、OneDep 登録システムで化学シフトと NMR 制限情報を一つのファイルにまとめた標準フォーマットを使えるようになった。2022 年、このシステムを応用して、過去の NMR エントリーの NMR 制限情報を標準化し再検証を行う Remediation の準備を進め、2023 年度に実装した。予定より 1 年近く遅れたが、令和 6 年度の早い段階で過去の全 NMR エントリーに対してデータ標準化の Remediation を実施する。

iii. データベースの安定運用と利用促進、国際協力

iii-1) データベースの安定運用

前年度と同様に、PDBj のメインサービスは、計画停電や学内ネットワークの定期メンテナンスの際にも、(財)蛋白質研究奨励会に設置したバックアップサーバーを活用することで、365 日サービスを提供することができた。2022 年度には、アノテーション業務の一部も(財)蛋白質研究奨励会のデータベース研究支援部門にて実施する体制を構築し、経験抱負な専属アノテータと OneDep のディベロッパーを新たに蛋白質研究奨励会で雇用して安定かつ信頼性の高いデータベース運用を行った。懸案であった(財)蛋白質研究奨励会のネットワーク環境の整備にも道筋をつけることができ、2024 年度中に 10Gbps で SINET6 に接続するメドがたった。

iii-2) 利用者・研究者コミュニティとの連携

研究代表者である栗栖が Vice President(無報酬)を務めている wwPDB Foundation(米国内の NPO 財団)の枠組みを利用して、構造予測ソフト AlphaFold2 を開発した DeepMind 社と意見交換をおこなった。i-1)でも触れた通り、ベイズ統計の手法を用いて予測構造自身を評価し、予測構造と実験構造とが混合したエントリーにおいて、構造データの評価法を検討するタスクフォースを財政的に支援していただく案を提出していた。DeepMind 社内で支援の可否と規模を検討していただいたが、大変残念なことに、最終的に支援は行えないという回答を得た。

並行して、Cell 誌の Editor から「PDB のリリース前データを査読の際に査読者が見られる仕組みを作ってもらえないか？」という問い合わせが来ていた。他の国際誌を出版する編集社からも同様の相談を受けていた。そこで、各出版社からデータベースを査読者に公開するシステムを開発する費用を負担していただけないか、現在問い合わせ中である。

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 2024

年3月13日にオンラインで開催した。Zoomにて韓国、台湾のユーザー代表にも出席していただいた（NBDCからも陪席していただいた）。PDBjの活動報告と次年度計画について報告を行い、委員の皆様から運営について専門的な諮問をいただいた。特に、今後のデータ量の増加に対する対応や、配列データベースとの統合利用の方向性について意見交換をおこなった。

iii-3) 国際協力

2022 年度に立ち上がった PDB China（PDBc：国立蛋白質研究センター上海）のメンターとして、PDB China の Wenqing Xu 所長（兼 PDBc 代表）と、毎月定例の Zoom 会議を開催し必要なアドバイスを行なった。2023 年 9 月 21 日に上海で開催された PDB China の設立記念シンポには、PDBj 代表の栗栖がオンラインで参加して記念講演を行った。PDBc が完全に立ち上がるまでは、アジア地区からの登録エントリー受付は全て PDBj で行い、PDBc のアノテータの習熟度に応じて、適切な数および種類のエントリーを PDBc 用に PDBj が準備したサーバーに転送することで、上海からリモートでデータ処理を進めている。PDBc でのデータ処理量が増えて、PDBj のアノテータに余裕ができてきた場合には、オーストラリアとニュージーランドからのエントリーを PDBj で受け付ける予定となっている。PDBc と PDBj とは対抗するのではなく、引き続き同一のデータベースを構築する仲間として、国際協力しながらデータ科学の発展に寄与していきたい。

§4. 成果発表等

(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)	0 件
	国際(欧文)	7 件
未発行論文 (accepted, in press 等)	国内(和文)	0 件
	国際(欧文)	0 件

② 論文詳細情報

1. Vallat B, Webb BM, Westbrook JD, Goddard TD, Hanke CA, Graziadei A, Peisach E, Zalevsky A, Sagendorf J, Tangmunarunkit H, Voinea S, Sekharan M, Yu J, Bonvin A AMJJ, DiMaio F, Hummer G, Meiler J, Tajkhorshid E, Ferrin TE, Lawson CL, Leitne r A, Rappsilber J, Seidel CAM, Jeffries CM, Burley SK, Hoch JC, Kurisu G, Morris K, Patwardhan A, Velankar S, Schwede T, Trewhella J, Kesselman C, Berman HM, Sali A.. IHMCIF: An Extension of the PDBx/mmCIF Data Standard for Integrative Structu re Determination Methods, J Mol Biol, 2024 (DOI: 10.1016/j.jmb.2024.168546).

2. Baskaran K, Ploskon E, Tejero R, Yokochi M, Harrus D, Liang Y, Peisach E, Persikov a I, Ramelot TA, Sekharan M, Tolchard J, Westbrook JD, Bardiaux B, Schwieters CD, Patwardhan A, Velankar S, Burley SK, Kurisu G, Hoch JC, Montelione GT, Vuister GW, Young JY.. Restraint validation of biomolecular structures determined by NMR in the Protein Data Bank, Structure, 2024 (DOI: 10.1016/j.str.2024.02.011).

3. Kleywegt GJ, Adams PD, Butcher SJ, Lawson CL, Rohou A, Rosenthal PB, Subraman iam S, Topf M, Abbott S, Baldwin PR, Berrisford JM, Bricogne G, Choudhary P, Croll TI, Danev R, Ganesan SJ, Grant T, Gutmanas A, Henderson R, Heymann JB, Huisko nen JT, Istrate A, Kato T, Lander GC, Lok SM, Ludtke SJ, Murshudov GN, Pye R, P intilie GD, Richardson JS, Sachse C, Salih O, Scheres SHW, Schroeder GF, Sorzano C OS, Stagg SM, Wang Z, Warshamanage R, Westbrook JD, Winn MD, Young JY, Burle y SK, Hoch JC, Kurisu G, Morris K, Patwardhan A, Velankar S.. Community recomm endations on cryoEM data archiving and validation, IUCrJ, 140-151, 2024 (DOI: 10.11 07/S2052252524001246).

4. wwPDB Consortium. EMDB-the Electron Microscopy Data Bank, Nucleic Acids Res, D 456-D465, 2024 (DOI: 10.1093/nar/gkad1019).

5. Xu W, Velankar S, Patwardhan A, Hoch JC, Burley SK, Kurisu G. Announcing the la unch of Protein Data Bank China as an Associate Member of the Worldwide Protein Data Bank Partnership, Acta Crystallogr D Struct Biol, 792-795, 2023 (DOI: 10.1107/S 2059798323006381).

6. Vallat B, Tauriello G, Bienert S, Haas J, Webb BM, Žídek A, Zheng W, Peisach E, Pi ehl DW, Anischanka I, Sillitoe I, Tolchard J, Varadi M, Baker D, Orengo C, Zhang Y, Hoch JC, Kurisu G, Patwardhan A, Velankar S, Burley SK, Sali A, Schwede T, Berm an HM, Westbrook JD. ModelCIF: An Extension of PDBx/mmCIF Data Representation for Computed Structure Models, J Mol Biol, 2023 (DOI: 10.1016/j.jmb.2023.168021).

7. Choudhary P, Feng Z, Berrisford J, Chao H, Ikegawa Y, Peisach E, Piehl DW, Smith J, Tanweer A, Varadi M, Westbrook JD, Young JY, Patwardhan A, Morris KL, Hoch J C, Kurisu G, Velankar S, Burley SK. PDB NextGen Archive: centralizing access to int egrated annotations and enriched structural information by the Worldwide Protein Dat a Bank, Database (Oxford), 2024 (DOI: 10.1093/database/baae041).

(2) その他の著作物(総説、書籍など)

該当なし

(3) 国際学会および国内学会発表

① 概要

種別	国内外	件数
招待講演	国内	0 件
	国際	1 件
口頭発表	国内	1 件
	国際	1 件
ポスター発表	国内	2 件
	国際	0 件

② 招待講演

〈国内〉

該当なし

〈国際〉

1. Genji Kurisu, Gert-Jan Bekker, X-tal Raw Data Archive (XRDa): A crystallographic raw diffraction image archive in Asia、26th Congress and General Assembly of the International Union of Crystallography, Melbourne, Australia, Aug 22, 2023

③ 口頭講演

〈国内〉

1. 栗栖源嗣、AlphaFold 時代の Protein Data Bank、トーゴーの日シンポジウム、日本未来館、10 月 5 日

〈国際〉

1. Genji Kurisu, Protein Data Bank Japan: the Asian Hub of 3D macromolecular structural data, 26th Congress and General Assembly of the International Union of Crystallography, Melbourne, Australia, Aug 29, 2023

④ ポスター発表

〈国内〉

1. Gert-Jan Bekker、 Querying PDB data using the Mine 2 RDB service、トーゴーの日シンポジウム、日本未来館、10 月 5 日
2. 横地政志、NMR 実験データ標準化へのあゆみ、トーゴーの日シンポジウム、日本未来館、10 月 5 日

〈国際〉

該当なし

(4) 知的財産権の出願 （国内の出願件数のみ公開）

出願件数

種別		件数
特許出願	国内	0 件

(5) 受賞・報道等

① 受賞

該当なし

② メディア報道

1. （プレスリリース「大阪大学が世界の蛋白質構造データバンク(PDB)を運営して 20 年- 世界の PDB データの 4 分の1相当 5 万件に到達！」、 http://www.protein.osaka-u.ac.jp/achievements/20230511/）

③ その他の成果発表

該当なし

§5. 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（月平均）

DB 名	種別	2023 年度
PDB Archive	訪問者数	3,589
	訪問数	5,562
	閲覧ページ数	820,017
BMRB	訪問者数	3,151
	訪問数	8,993
	閲覧ページ数	245.726
eF-site	訪問者数	363
	訪問数	15,705
	閲覧ページ数	60,325
ProMode elastic	訪問者数	2,973
	訪問数	11,392
	閲覧ページ数	68,523
MoM	訪問者数	26,683
	訪問数	39,764
	閲覧ページ数	67,716

② 分析

上記数値は, 大阪大学に設置しているメインサーバーpdbjlvh1 とバックアップサーバーpdbjbk1、 それに(財)蛋白質研究奨励会に設置しているバックアップサーバーpdbjpw1 の3つのサーバーへの各アクセスログから該当サービスの URL へのアクセス分だけを抽出し、それぞれを AWStats で集計した月ごとの値を 2023 年度分 1 年間で合計し 12 ヶ月で割った値になります。

PDB Archive、 BMRB は rsync 等で機械的アクセスされるものを集計しているため、訪問者週は IP あぢレスで辿れる範囲になります。https://data.pdbj.org などのプロトコールを用いて Web ベースでダウンロードされたものは含みません。基本的に rsync 等で週次更新分をまとめて毎週ダウンロードされている利用者が多いと判断しております。アクセス数の増減で判断すると、コロナ禍で急増しておりましたので、ここのところ落ち着いてきていると考察しております。PDBe が rsync によるサービスを停止して全て Globus に移行中ですので、利用者が Globus を志向しているのかどうか動向に注目しております。

(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

2023 年に世界中で wwPDB に寄託された構造データは合計 17,064 件であり、そのうち実に31.5%に相

当する 5,376 件が PDBj へ登録されている。(PDBe は 29%の 4,990 件)

(3) データベースの利活用により得られた研究成果(生命科学研究への波及効果)

非公開

(4) 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)

非公開

§6. 研究開発期間中に主催した活動(ワークショップ等)

(1) 進捗ミーティング

年月日	名称	場所	参加人数	目的・概要
2023 年 4 月1日～2024 年 3 月 31 日(毎週開催)	PDBj開発者会議(非公開)	Zoom	10 人	研究進捗報告のためのミーティング
2023 年4月 1 日～2024 年 3 月 31 日(隔週開催)	PDBj Primary Annotator's meeting(非公開)	Zoom	14 人	同上
2023 年 4 月1日～2024 年 3 月 31 日(毎週開催)	BMRBj テクニカルスタッフミーテイング	大阪大学蛋白質研究所セミナー室	5 人	同上
2023 年 4 月 1 日～2024 年 3 月 31 日(毎月不定期)	wwPDB PI ミーティング	Zoom	4 人	wwPDB を構成するデータベースの各 PI による方針決定会議
2023 年 4 月 1 日～2024 年 3 月 31 日(毎週)	OneDep リーダー会議	Zoom	10 人	wwPDB を構成するデータベースのリードアノテータが出席して, OneDep による処理の方針を相談する会議
2023 年 4 月 1 日～2024 年 3 月 31 日(毎週)	OneDep 開発者会議	Zoom	8 人	OneDep の開発者が開発状況をシェアし, 開発項目を整理する会議
2023 年 4 月～2024 年 3 月 31 日(毎月)	PDBc-PDBj PI 会議	Zoom	3 人	PDB China の立ち上げに協力するため, 問題点や技術支援を進める相談をする会議
2023 年 10 月 27 日	wwPDB 運営諮問会議	Zoom ＋ EMBL EBI からの現地参加	22 人	wwPDB の運営に関して今後の方針や問題点を議論し, 運営方針を諮問する会議
2023 年 10 月 28-30 日	wwPDB サミット	EMBL EBI	30 人	wwPDB の Primary Annotator と OneDep 開発者, および PI が集まって 1 年間の開発方針を決める会議

(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
2023 年 7 月 5 日	第 23 回日本蛋白質科学会年会 PDBjランチョンセミナー	名古屋国際会議場(愛知県名古屋市)	100 人	学会参加者に向けた PDBjとサービスの紹介と, 利用動向調査を行なった
2023 年 7 月 29 日, 8 月 1 日	高校生のための蛋白研セミナー	蛋白研講堂	計 100 人	蛋白質に興味のある高校生に, PDBj の一般向けページを用いて模擬講義を実施
2023 年 7 月 8 日	大阪大学共創 DAY @EXPOCITY2023 出展	ららぽーと EXPCITY	参加者多数でカウントできず	体験イベント、ミニレクチャー、展示などを通して、大阪大学のさまざまな研究成果を紹介する取り組みに参加
2023 年 8 月 22-29 日	第 26 回国際結晶学連合会議ブース展示	メルボルン国際会議場(オ	参加者多数	wwPDB メンバーとしてブースを費用折半(日米欧で等分)にて出

年月日	名称	場所	参加人数	目的･概要
		ーストラリア）		展し，PDB の活動を登録者に周知
2023 年 9 月 8 日	第 12 回生命医薬情報学連合大会ランチョンセミナー	柏の葉カンファレンスセンター（千葉県柏市）	100 人	学会参加者に向けて PDBjの活動説明と，高度利用について講演し，利用動向調査を行なった
2023 年 10 月 29 日	令和 5 年度日本結晶学会年会ランチョンセミナー	山口大学（山口県宇部市）	85 人	学会参加者に向けた PDBjとサービスの紹介し，利用動向調査を行った
2023 年 11 月 15 日	第 61 回日本生物物理学会年会ランチョンセミナー	名古屋国際会議場（愛知県名古屋市）	100 人	学会参加者に向けた PDBjとサービスの紹介し，RCSB PDB の Burley 教授の講演を実施した

以上

別紙1　既公開のデータベース・ウェブツール等

No.	正式名称	別称・略称	概要	URL	公開日	状態	分類	関連論文
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	維持・発展	データベース等	Sameer Velankar, Stephen K. Burley, Genji Kurisu, Jeferry C. Hoch, Joh L. Markley, "The Protein Data Bank Archive", Methods Mol Biol., 2305, 3-21, 2021 (DOI: 10.1007/978-1-0716-1406-8_1)
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://bmrbj.pdbj.org	2011/4/1	維持・発展	データベース等	Jeffery C. Hoch, Kumaran Baskaran, Harrison Burr, Joh Chin, Hamid R. Eghbalnia, Toshimichi Fujiwara, Michael R. Gryk, Takeshi Iwata, Chojiro Kojima, Genji Kurisu, Dmitri Maziuk, Yohei Miyanoiri, Jonathan R. Wedell, Colin Wilburn, Hongyang Yao, Masashi Yokochi, "Biological Magnetic Resonance Data Bank", Nucleic Acids Res., 51:D368-D376, 2023 (DOI: 10.1093/nar/gkac1050)
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	維持・発展	データベース等	維持・更新のみ
4	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．	https://pdbj.org/promode-elastic/	2003/4/1	維持・発展	データベース等	維持・更新のみ
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	維持・発展	データベース等	維持・更新のみ
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	維持・発展	ツール等	維持・更新のみ
7	DASH	同左	PDBデータを基にした構造アラインメント（旧ASH）	https://sysimm.org/dash/		維持・発展	ツール等	維持・更新のみ
8	MolMil2	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	https://pdbj.org/help/molmil	2014/9/1	維持・発展	ツール等	維持・更新のみ