

研究開発実施報告

□概要

研究開発課題名	蛋白質構造データベースのデータ検証高度化と統合化
開発対象データベースの名称(URL)	日本蛋白質構造データベース(PDBj) (https://pdbj.org)
研究代表者氏名	栗栖 源嗣
所属・役職	大阪大学 蛋白質研究所 教授

□目次

§1. 研究実施体制	2	① 概要	8
§2. 研究開発対象とするデータベース・ツール等	2	② 招待講演	8
(1) データベース一覧	2	③ 口頭講演	8
(2) ツール等一覧	2	④ ポスター発表	8
§3. 実施内容	3	(4) 知的財産権の出願	9
(1) 本年度の研究開発計画と達成目標	3	(5) 受賞・報道等	9
(2) 進捗状況	5	§5. 研究開発期間中に主催した活動(ワークショップ等)	
§4. 成果発表等	7	9
(1) 原著論文発表	7	1. 進捗ミーティング	9
① 論文数概要	7	2. 主催したワークショップ、シンポジウム、アウトリー	
② 論文詳細情報	7	チ活動等	9
(2) その他の著作物(総説、書籍など)	8		
(3) 国際学会および国内学会発表	8		

§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
栗栖グループ	栗栖 源嗣	大阪大学・教授	PDBおよびBMRBアーカイブの構築および高度化・統合化
藤グループ	藤 博幸	関西学院大学・教授	ASHビューアの開発

§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://pdj.org/

【その他のデータベース】

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

(2) ツール等一覧

No.	名称	別称・略称	URL
1	EM Navigator		https://pdj.org/emnavi/
2	ASH		https://sysimm.ifrec.osaka-u.ac.jp/ash_service/
3	MolMil		https://pdj.org/help/molmil
4	ASHviewer		これから開発するため、まだない

§3. 実施内容

(1) 本年度の研究開発計画と達成目標

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

- 引き続き, wwPDB の欧米のメンバーと協力し, PDBj (PDB Japan) および PDBj-BMRB が厳しい品質管理を行いつつ, 増加する一方の蛋白質立体構造情報 (Protein Data Bank: <https://pdbj.org>) と NMR 実験情報 (BMRB: <https://bmrj.pdbj.org>) に対応する. ベテランアノテータが引退し新人アノテータが 2 名加わった新体制で, 継続的に世界標準のデータベースを構築し公開することを目標とした. 第 2 年次には, 引き続き 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 wwPDB の欧米メンバーと協力して確立することに重点を置いて対応する計画であった.
- PDB 全データの品質管理のための検証レポートのセマンティック化を進める. 特に, 検証レポートの構築法がほぼ確立している X 線結晶構造解析のデータを先行して RDF 化し, PDBj-Mine への実装を進める.

2) 他のデータベースとの統合化および高度化

2-1) 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: Alignment of Structural Homologs) の高度化

本年度は、ビューア上での「クラスタ圧縮展開機能の作成」と「SPARQL 検索機能の作成」を行った。

クラスタ圧縮展開機能は、ユーザインタラクティブに、アラインメントの表示を変化させることで、アラインメントの可読性を向上させる機能である。ビューア上で粗視化された系統樹とそれに対応するアラインメントのコンセンサス配列を隣接させる形で表示させる。粗視化とは、系統樹に従いアラインメントをいくつかのサブクラスタに分割し、上記の表示を行うことを意味する。デフォルトでは 3 つのサブクラスタに分割するように設定されているが、分割数は変更できる。ビューアの左パネルに粗視化系統樹を、右パネルに対応するアラインメントのコンセンサス配列が表示される形になっている。系統樹のクラスタリングと対応させる形になっているので、単純なコンセンサス配列よりも可読性が高く、解釈が容易にできるようになっている。左パネルの系統樹の内部節をクリックすると、その配下のサブツリーが圧縮され、外部節に変更される。それに伴い、右パネルのアラインメントのコンセンサス配列の表示も圧縮された形に変更される。左パネルの系統樹の外部節(葉)をクリックすると、そこに圧縮されていたサブツリーが展開される。展開の単位はサブクラスタの分割と同じで、デフォルトは 3 個のサブツリーに分割された形で展開される。右パネルのコンセンサス配列も展開された形で再構築・表示される。

SPARQL 検索機能とは、アラインメントに含まれる配列に関して、SPARQL 検索によって、RDF 化されたデータベースからインターネットを介して収集し、それを可読性の高い形で表示するための機能である。今回は、0 の molecular function, cellular component, biological process, また source organisms, および active sites の情報を収集し、それを表示する機能を開発する。SPARQL 検索については、ROIS の山口敦子准教授に協力いただき、山口准教授のグループで開発されていた LODSurfer をベースに、タンパク質に特化した PSurfer を作成してもらい、それをビューアに組み込む形とした。0 の molecular function, cellular component, biological process, また source organisms, については、入力されたアラインメント全体の無根系統樹を表示するパネルを作成し、その周辺に同心円状に 0 terms や生物種の情報を表示できるようにする。無根系統樹は、上記の粗視化処理に合わせた色分けを行い、粗視化アラインメントとの対応がとれる形にする。図中に表示された 0 や source organisms の詳細については、同心円をクリックすると、上記の左右パネル下部に置かれているテキストフィールドに、対応する 0 term や生物種の分類情報が表示される。また、active site については、粗視化処理によって得られたクラスタごとに、保存どのプロファイルを作成し、その上に得られた active site の位置を表示するようにする。

2-2) 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を、すでに別のサービス(IRAF) 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 ファイルから抽出する。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、これら一連の作業をPDBの毎週の更新に合わせて自動的に行う仕組みを整えることを目標とした。続いて同様のことを分子間コンタクト(蛋白質間相互作用、蛋白質-リガンド間相互作用など)に対しても行う。これにより、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する。

2-3) 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する。セキュリティおよびサーバ稼働状況の監視システムのプロトタイプを試験運用を開始する。具体的には、現在、1 件辺りの登録業務平均時間は数時間であり、登録者との通信業務、更新作業にも同様の時間を要している。これらを米 Madison-Wisconsin 大学とのツール共同開発により、作業時間として現在の半分程度になるよう簡略化・自動化を目指す。

- BMRB データベースの XML 化、RDF 化や更なる多彩なデータフォーマットへの変換ツールを公開し、検索 WEB サイトへ応用する。これによりデータの詳細を視覚的にも高度な検索、グラフィカルな表示を実行可能にしていく。
- 引き続き NMR 構造座標、実験データとの整合性に関する評価ツールを wwPDB との協力を通じて開発し、データの品質維持を客観評価可能な次世代技術について議論し、開発およびツールとしての公開を進める。
- BMRB データベースを活用した自動解析支援ツールの試用を開始する。創薬研究へ利用できるパイプラインの構築と実際に応用可能であることを示していく。

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」を定期的で開催し、企業を含めたユーザからの要望等を集める。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、代謝工学やシステム生物学の研究者の声を聞く機会を設けるよう努力する。
- 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、第 2 年次も PDBj 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。
- 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。

3-2) アノテータの育成・国際協力

- 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教育を進める。具体的には、TV 会議による情報共有と、米国 RCSB-PDB あるいは欧州 EBI-PDBe での合同講習会へ参加する。中国、韓国出身のアノテータが揃う予定なので、中国語と韓国語の PDBj ホームページの多言語化にもより積極的に取り組む。

(2)進捗状況

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

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。第 2 年次も、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、平成 30 年度中に PDB と BMRB を合わせて 3204 件を PDBj から新規登録した [参考:平成 29 年度は PDB と BMRB 合わせて 3087 件]。

本年度は、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 wwPDB の欧米メンバーと協力して確立することに重点を置いた。具体的には、EM validation task force を組織して EM コミュニティーからの意見収集に注力した。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。平成 30 年度中に wwPDB 運営諮問会議を英国 Cambridge において開催し、引き続き OneDep システムの高度化を進め、新たに BMRBdep を導入し登録システムの更なる効率化を進めていくこととした (図 1)。

PDB 全データの品質管理のための検証レポートのセマンティック化を進めた。特に、検証レポートの構築法がほぼ確立している X 線結晶構造解析のデータを先行して RDF 化し、他の wwPDB メンバーと dictionary の整合を確認した。PDBj-Mine への実装を目指し、先行して PDBj-BMRB の HP で β 版を公開し、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。

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増え、2015 年に日本からの登録数を上回った。その後も中国からの登録は増え続けているため、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の下、wwPDB の枠組みで中国に PDB の拠点を受け入れる準備を進めた⁽¹⁾。



図 1. 2018 年 wwPDB 運営諮問会議

2) 他のデータベースとの統合化および高度化

2-1) 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: Alignment of Structural Homologs) の高度化
(非公開)

2-2) 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 (IRAF) 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する計画であった。ほぼ実装準備が整っている状況で、担当者である金城准教授 (大学承継ポスト) の異動までに、現有メンバーの横地と Bekker とを中心に実務内容を引き継ぎ開発完了できると想定していた。しかし、他メンバーも担当開発テーマを抱えた状況であり、本テーマの主担当者が不在の状況となり本年度中の開発を一時停止せざるを得ない状況となった。2019 年度に体制を立て直して対応する予定である。

2-3) 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。セキュリティおよびサーバ稼働状況の監視システムのプロトタイプの実験運用を開始した。現在、1 件辺りの登録業務平均時間は数時間であり、登録者との通信業務、更新作業にも同様の時間を要している。これらを米 Madison-Wisconsin 大学を中心とした wwPDB メンバーで新しい登録システム(BMRBdep)を共同開発することにより、作業時間として現在より効率よく処理できる目処がたった。
- BMRB データベースの XML 化、RDF 化や更なる多彩なデータフォーマットへの変換ツールを公開し、検索 WEB サイトを一新した。これに併せてデータの詳細を高度に表示できるシステムを実装した。
- 引き続き 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。
- BMRB データベースを活用した自動解析支援ツール(NMR Tool Box)の最新版を公開した。PDBj 主催の講習会や関連学会でのランチョンセミナーを開催して、創薬研究へ利用できるパイプラインの構築と実際に応用可能であることを示した。

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 2019 年 2 月 5 日に開催し、企業を含めたユーザからの要望や意見交換を行った。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。
- 利用者・研究者向けのセミナーや講習会については、JST-NBDC 主催の All-in-one 講習会(2019 年 1 月 28 日)に参加・協力した。データ登録者の学術団体である日本結晶学会と日本蛋白質科学会の年會中に、ランチョンセミナーを開催して、データ登録への継続的な協力と、データ受付方針の変更点などを説明し理解していただいた。また、データ利用者の学術団体である日本生物物理学会では DDBJ と協力してランチョンセミナーを開催し、生命医薬情報学連合大会においてもランチョンセミナーを開催して、利用者の声をアンケートの形で集めた。特に、2019 年 2 月 19 日(火)には、以下の様な NMR 講習会を開催し、BMRB データベースの積極的な利用を図った。

2018 年度 PDBj 講習会概要

日程:2019 年 2 月 19 日(火) 9:30 - 17:30

定員: 10 名

会場(部屋名) 大阪大学 蛋白質研究所 構造解析研究棟の 4 階セミナー室

プログラム:

<講義>

9:30 - 10:00 日本蛋白質構造データバンク PDBj (栗栖 源嗣)

10:00 - 10:30 生体高分子 NMR データバンク BMRB (横地 政志)

11:00 - 12:00 NMR 分光法の基礎 (宮ノ入 洋平)[講義資料]

12:00 - 13:00 休憩

<計算機実習>

13:00 - 17:30 NMR の構造解析の情報処理の計算機演習 (小林 直宏)

• これまでも参加してきたサイエンス・アゴラ 2018(11月10-11日)(図3)に加えて「ららぽーと EXPOCITY」で開催された大阪大学共創 DAY@EXPOCITY に参加し、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のため、積極的な啓蒙活動を行った。2018年12月2-5日にニュージーランド・オークランド大学で開催されたアジア結晶学会において、PDBj ポスター賞をスポンサーして、アジア地区からのベストポスターをPDBjが懸賞することを開始した(図4)。今後も毎年PDBjポスター賞の表彰をおこなっていく予定である。



図3. サイエンスアゴラ 2018 の PDBj コーナー



図4. アジア結晶学会での PDBj ポスター賞表彰風景

3-2) アノテータの育成・国際協力

Ver3.5 まで更新された OneDep 登録システムを活用して世界で同一の品質管理によるデータ登録を実施した。継続して国際連携による教育を進めるために、隔週金曜日におこなっている TV 会議による情報共有と、米国 RCSB-PDB で 2019 年 2 月に開催した OneDep Summit へ出席して人材育成につとめた。本年度より、中国、韓国出身のアノテータが揃ったため、中国語と韓国語の PDBj ホームページの多言語化により積極的に取り組んだ。具体的には、OneDep 登録用サイトがより正しく利用されるように、OneDep の News 欄に日、英、中、韓国語のニュースを掲載して、アナウンス効果の向上を図った。

ベテランアノテータが 1 名引退し、新人アノテータを 2 名加えた新体制でスタートしたが、産休明けの新人 1 名が自己都合退職された。遅滞無くデータ処理を進めることができたのは、新人 2 名のうち 1 名が非常に高い適応能力を示されたためである。

§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)	1 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. wwPDB consortium, “Protein Data Bank: the single global archive for 3D macromolecular structure data”, Nucleic Acids Research, vol. 47, No. 12, pp.520-528, 2018

(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	1 件
	国際	1 件
口頭発表	国内	1 件
	国際	3 件
ポスター発表	国内	1 件
	国際	1 件

② 招待講演

〈国内〉

1. 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, 2018/11/20

〈国際〉

1. Kurisu . , Structural atlas of Ferredoxin-dependent proteins in the Protein data Bank, Japan Science Day at RUB, Ruhr University Bochum, Germany, 2018/7/6

③ 口頭講演

〈国内〉

1. 栗栖源嗣, 蛋白質構造データベース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

〈国際〉

1. 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
2. Kawabata, T., Suzuki, H., Kurisu, . Databases and Web services from PDBj for Electron Microscopy, Asian Crystallographic Association, 2018/12/2-5
3. Kawabata T, Suzuki H, Kurisu . 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

④ ポスター発表

〈国内〉

1. 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

〈国際〉

1. 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) 知的財産権の出願

該当なし

(5) 受賞・報道等

該当なし

§5. 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
2018年4月1日～ 2019年3月31日 (毎週開催)	PDBj開発者会議 (非公開)	大阪大学蛋白質研究所構造解析研究棟4階セミナー室	10人	研究進捗報告のためのミーティング
2018年4月1日～ 2019年3月31日 (隔週開催)	PDBj Primary Annotator's meeting (非公開)	大阪大学蛋白質研究所構造解析研究棟4階セミナー室	14人	同上
2018年4月1日～ 2019年3月31日 (隔月開催)	wwPDB PI ミーティング	Skype	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
2018年4月1日～ 2019年3月31日 (隔月開催)	OneDep リーダー会議	Skype	10人	wwPDBを構成するデータベースのリードアナテータが出席して、OneDepによる処理の方針を相談する会議

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
2018年 6月28日	第18回日本蛋白質科学会年会 PDBjランチョンセミナー	朱鷺メッセ	100人	学会参加者に向けた PDBjとサービスの紹介
2018年 9月16日	第56回日本生物物理学会年会 PDBjランチョンセミナー	岡山大学津島キャンパス	100人	学会参加者に向けた PDBjとサービスの紹介
2018年 9月20日	第7回生命医薬情報学連合大会 PDBjランチョン	荘銀タクト鶴岡	80人	学会参加者に向けた PDBjとサービスの紹介

年月日	名称	場所	参加人数	目的・概要
	セミナー			
2018年 10月3日	CCP4 School and Workshop	SPring-8	40人	蛋白質構造解析者むけ講習会で登録方法の説明
2018年 11月9-11日	サイエンスアゴラ2018	テレコムセンター (東京都)	750人	子供、中高生を含む一般参加者に向けた蛋白質立体構造の展示、解説等
2018年 11月11日	日本結晶学会 H30 年度年会 PDBjランチョンセミナー	東京工業大学大岡山キャンパス	80人	学会参加者に向けた PDBjとサービスの紹介
2018年 11月17日	大阪大学共創DAY@EXPOCITY	ららぽーと EXPOCITY (大阪府)	800人	子供、中高生を含む一般参加者に向けた蛋白質立体構造の展示、解説等
2018年 11月28-30日	日本分子生物学会 H30 年度年会データベース展示企画 ブース出展及びフォーラム講演	パシフィコ横浜	約100人	学会参加者に向けた PDBjとサービスの紹介
2019年 1月28日	All-in-one 合同講習会2018	JST-東京	45人	学生及び研究者を対象とした PDBjサービスの紹介

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	
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ベース	http://bmrdep.pdbj.org	2011/4/1	維持・発展	データベース等	
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	
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	維持・発展	データベース等	
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	
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	
7	ASH	同左	PDBデータを基にした構造アラインメント	https://sysimm.ifreec.osaka-u.ac.jp/ash_servic		維持・発展	ツール等	
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	維持・発展	ツール等	