



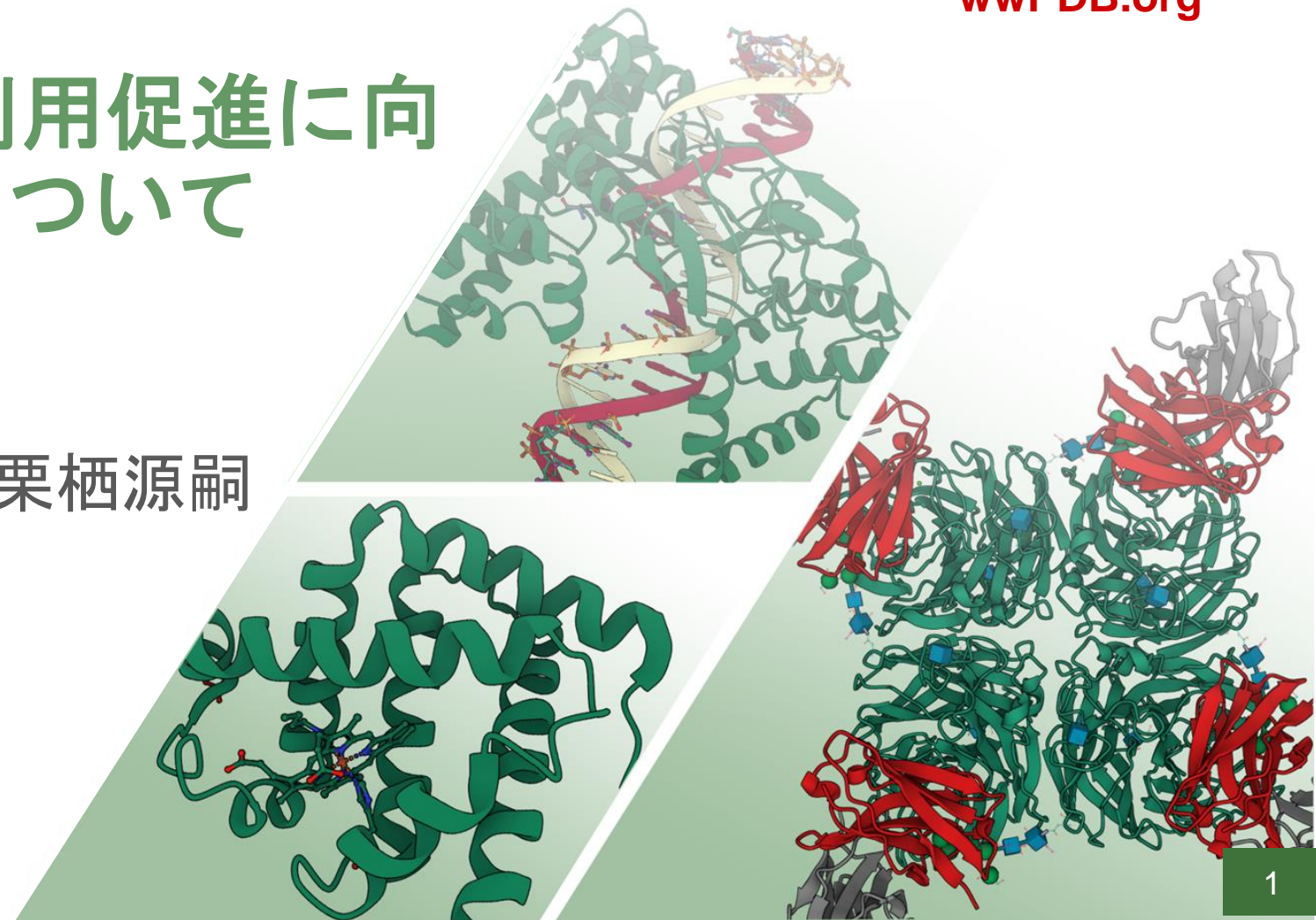
PDBj
Protein Data Bank Japan

WORLDWIDE
wwPDB
PROTEIN DATA BANK

wwPDB.org

PDBjの最近の活動と利用促進に向けた今後の活動方針について

大阪大学蛋白質研究所 栗栖源嗣



PDBは老舗デジタルデータリソース

CRYSTALLOGRAPHY

Protein Data Bank

A repository system for protein crystallographic data will be operated jointly by the Crystallographic Data Centre, Cambridge, and the Brookhaven National Laboratory.

The system will be responsible for storing atomic coordinates, structure factors and electron density maps and will make these data available on request. Distribution will be on magnetic tape in machine-readable form whenever possible. There will be no charge for the service other than handling costs. Files

Nature New Biology **233**, page 223 (1971)



**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

2021.11.24 [Wed.] 10:00-17:05 (日本時間)
(17:00-17:06 閉会挨拶)

ポスターセッション
PDBjポスター賞
学生・ポスドクによる優れた研究発表に対しPDBjポスター賞を贈呈します。

記念シンポジウム

13:30-14:15 Session 1	アジアにおけるPDB50年の歴史 講師 栗橋 謙嗣 大阪大学 蛋白質研究所 中村 善本 大阪大学
14:25-15:10 Session 2	Structure of the sodium-dependent phosphate transporter reveals insights into human solute carrier SLC20 講師 Yuh-ju Sun 国立清華大学 (台湾) 中川 敏史 大阪大学 蛋白質研究所
15:20-16:05 Session 3	Molecular movies and beyond 講師 若田 想 京都大学大学院数科学研究科
16:15-17:00 Session 4	NMRによる膜タンパク質の機能解明 講師 横田 一夫 理化学研究所 生命機能科学研究センター 高遠 謙 大阪大学 蛋白質研究所
	高速撮影と高分解能を両立したクライオ電子顕微鏡による構造生物学 講師 難波 啓一 大阪大学大学院生命機能研究科、理研放射光科学研究センター、生命機能科学研究センター
	Structural and mechanistic dissection of the Wnt signaling pathway 講師 Wenqing Xu 国立蛋白質科学研究所、上海科学技術大学 (中国)
	A new era in structural biology: the impact of structure prediction using AI methods 講師 Sameer Velankar 欧州生物情報学研究所 (英国)

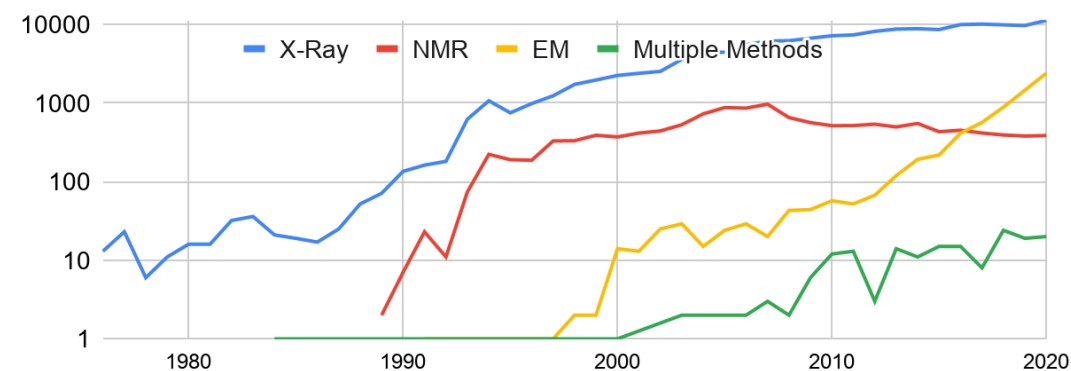
参加登録費用 無料
締切
●ポスター発表登録締切
2021年10月17日 (日)
日本時間 23:59 まで
※学生発表枠 (100件) に達した場合は、その時点で締め切ります。
●参加登録締切
2021年11月8日 (月)
日本時間 23:59 まで

PDB Archiveの推移 (1)

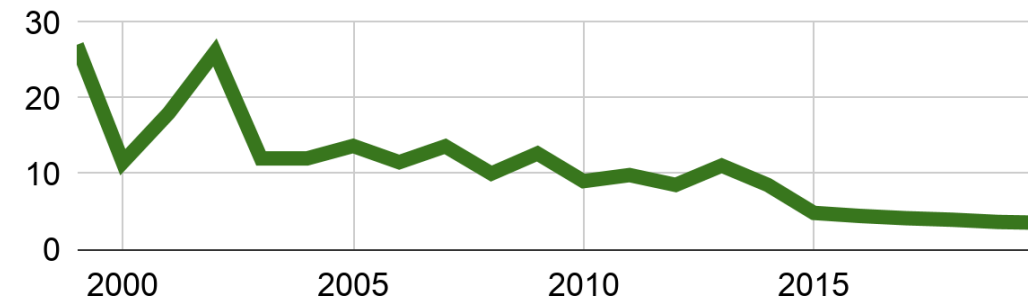


- 190,000件を超える専門家が編集処理したデータをCC0 1.0として自由に利用可能
 - ❖ 年率 ~8%でデータ量が増加
- 全世界で >400 を超える外部データベースで活用されている
- 2020年集計でクライオ電顕のエントリーが
 - ❖ 前年比60%増
 - ❖ 原子分解能 (~1Å)に到達するエントリーも始めた

Released Entries By Method/Year (log scale)



Median EM Resolution (Å)



CoreTrustSeal certification renewed through April 2024 ([CoreTrustSeal.org](https://www.CoreTrustSeal.org))

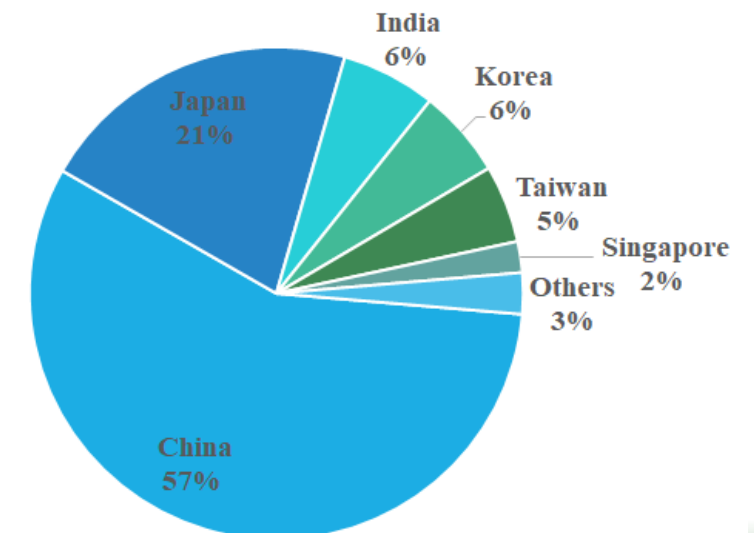
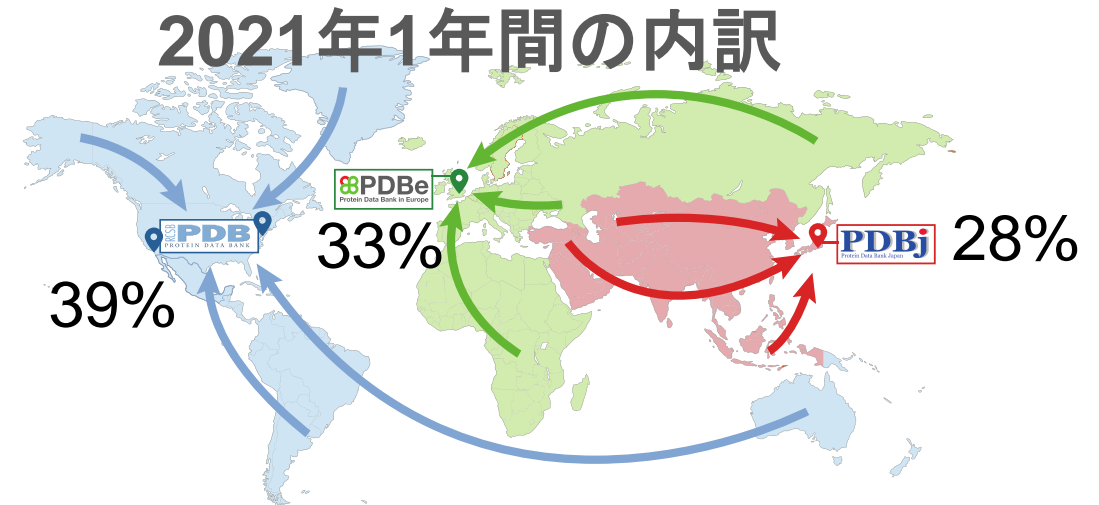
PDB Archiveの推移 (2)

- 2021年1年間に限ると、アジア地区のデータ量増加率が他地域よりも多いので、PDBjの登録・編集の割合は約28%に増加
($4,160/14,570 = 0.286$)

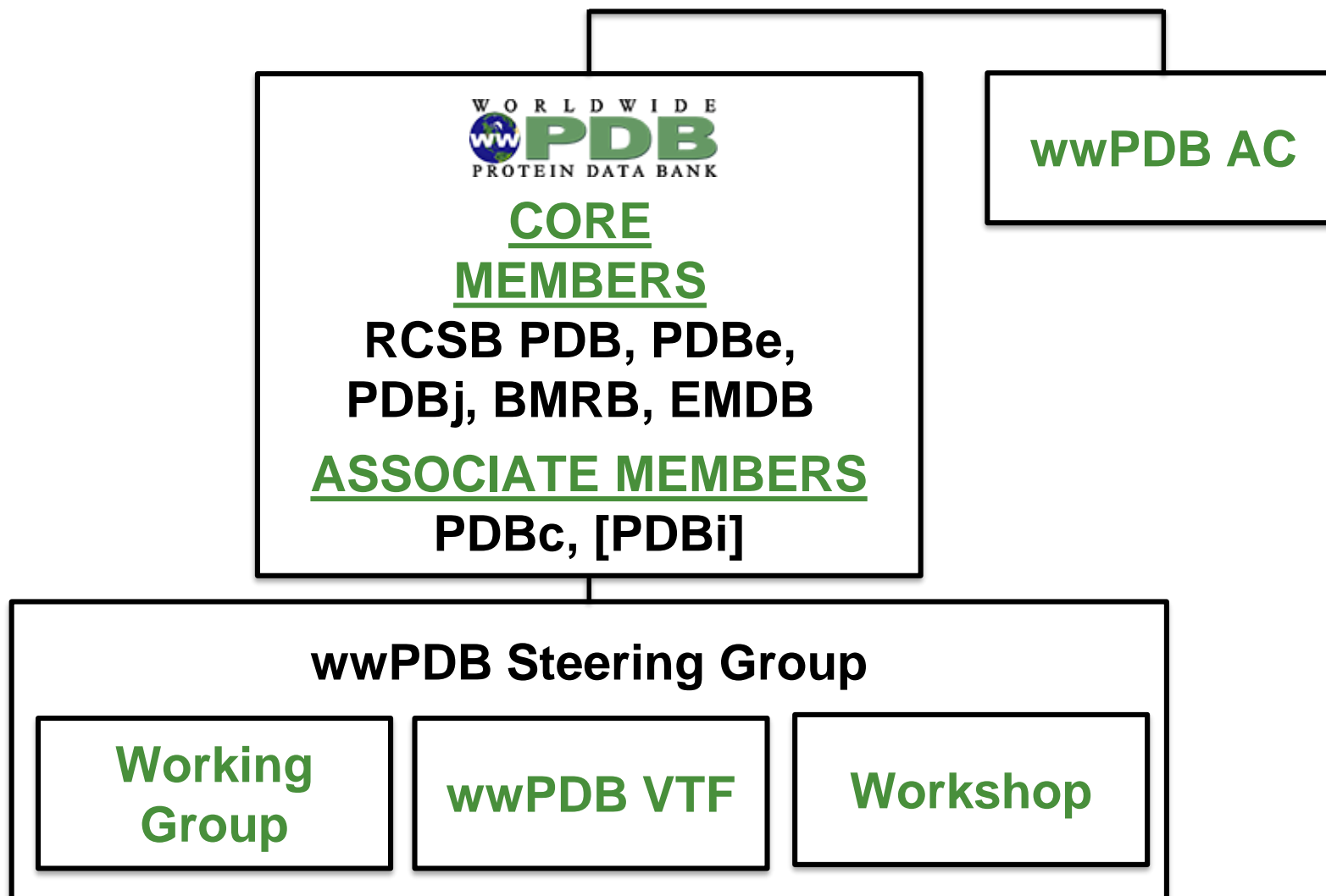
➡ 中国発のエントリー数の急増

- 機械学習による予測構造を併用した構造解析の増加。結晶解析の位相問題解決にも貢献。

➡ 構造予測手法（AlphaFold2等）を併用したIntegrated/Hybrid構造解析に対応した検証レポートの構築が課題

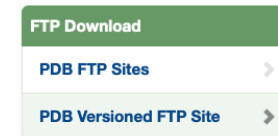


2022に国際組織wwPDBに中国が参加



アナウンス 1 : PDB座標が更新されているかも？

- Improve the quality of PDB structures
- Preserve original PDB IDs and maintain connections to the scientific literature
- Reasoning captured
- The latest versioned file can be accessed at FTP
- The latest minor version of each major version is available at versioned FTP



PDB Versioned FTP Site

Since October 2017, the wwPDB versions PDB entries and distributes the latest and the prior versions of each entry via a versioned FTP archive accessible at ftp-versioned.wwpdb.org and its mirrors in the USA, UK and Japan.

PDB Versioned FTP Sites

The PDB Versioned FTP sites are updated every Wednesday at 00:00 UTC.

wwPDB: ftp-versioned.wwpdb.org
 RCSB PDB (USA): ftp-versioned.rcsb.org
 PDBe (UK): ftp.ebi.ac.uk/pub/databases/pdb_versioned/
 PDBj (Japan): ftp-versioned.pdbj.org

What is PDB Entry Versioning

Changes made to a PDB entry after its initial release are considered to be either “major” or “minor”. Updates to atomic coordinates, polymer sequence, or chemical description in the coordinate file trigger a major version increment, retaining the originally issued PDB accession code. Other changes to the metadata in the coordinate file are considered minor. **Currently, no changes are permitted to the experimental data from which the coordinates are derived.** To keep track of the changes between versions, a set of new revision categories were defined in the PDBx/mmCIF dictionary (http://mmcif.wwpdb.org/dictionaries/mmcif_pdbx_v50.dic/Groups/audit_group.html). The revision trail is included in the PDBx/mmCIF formatted coordinate files.

ftp-versioned.pdbj.org

アナウンス 1 : 座標更新履歴をご活用下さい

- >200 entries versioned
- Enabled rapid correction of >30 SARS-CoV-2 structures
- Top reasons
 - Coordinate completeness
 - Ligand geometry
 - Ligand identity
 - Atom clashes
- Depositors are encouraged to update their structures

履歴

PDBのバージョン管理については、[ヘルプページ](#)をご覧ください。

バージョン	日付	タイプ	
1-0	2020-02-05	Initial release	ファイルを表示
2-0	2020-02-12	Coordinate replacement, Advisory, Atomic model, Data collection, Database references, Derived calculations, Refinement description, Structure summary	ファイルはありません
2-1	2020-02-19	Database references, Structure summary	ファイルはありません
2-2	2020-02-26	Data collection	ファイルはありません
2-3	2020-03-11	Source and taxonomy, Structure summary	ファイルはありません
2-4	2020-03-18	Database references	ファイルはありません
2-5	2020-04-22	Database references	ファイルはありません
2-6	2020-05-06	Database references, Source and taxonomy, Structure summary	ファイルはありません
2-7	2020-05-27	Database references	ファイルはありません
2-8	2020-06-10	Structure summary	ファイルはありません
2-9	2020-06-24	Database references	ファイルを表示
3-0	2020-07-29	Coordinate replacement, Advisory, Atomic model, Author supporting evidence, Data collection, Database references, Derived calculations, Refinement description, Structure summary	ファイルはありません
3-1	2021-03-10	Structure summary	ファイルを表示

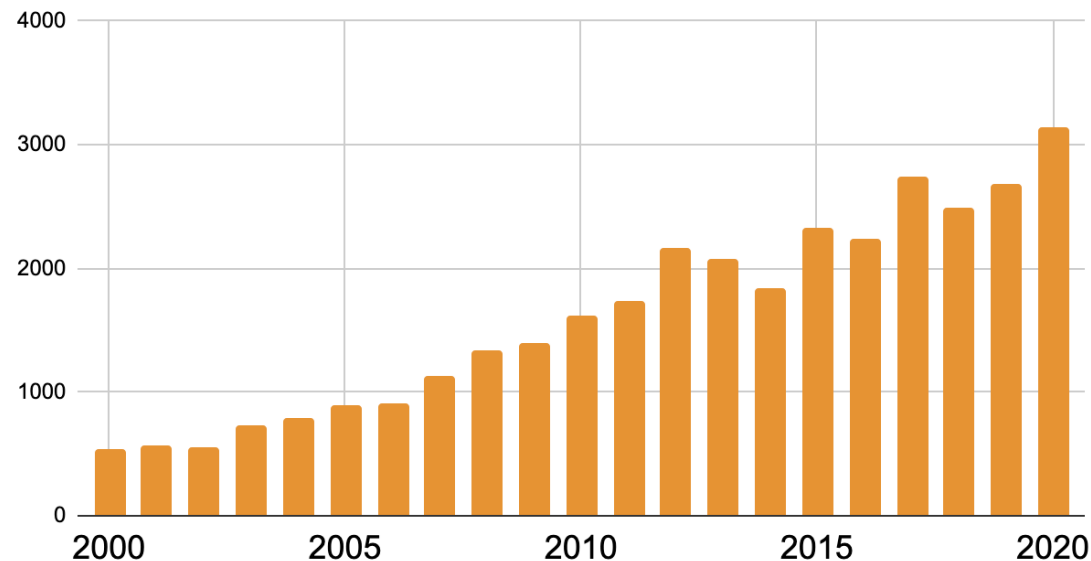
First SARS-CoV-2 structure in the PDB (6LU7)



アナウンス 2 : 化合物IDが5桁になる

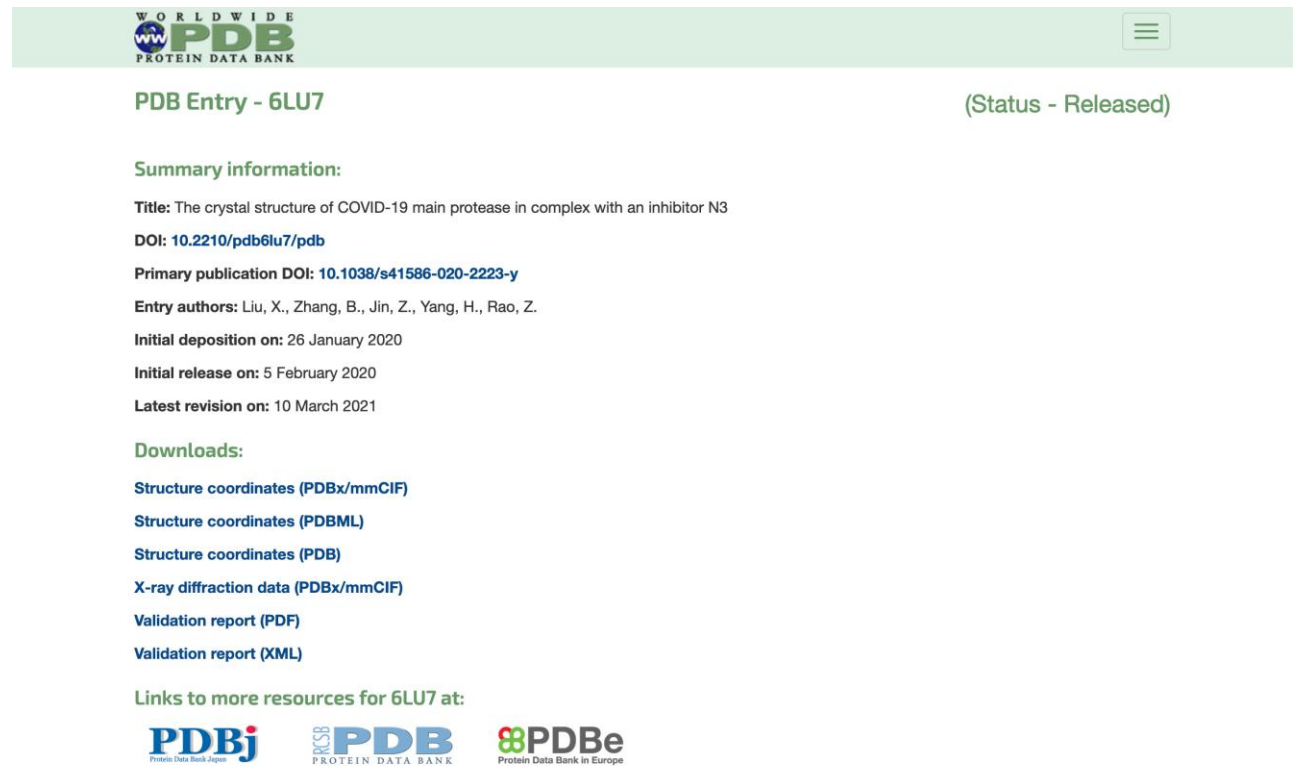
- PDBフォーマットでつかう3文字の化合物IDが2023年中に足りなくなります
- 以降の新規データはPDBx/mmCIFでのみ提供されます
- PDBフォーマットでは提供できなくなります。

Number of New Chemical Component Entries Created Each Year



アナウンス 3 : PDB IDは8桁になる

- 4文字のPDB IDも近い将来枯渇します
- 既に8文字のPDB IDでデータが提供されています
wwpdb.org/pdb?id=pdb_00006lu7
- 既に提供を開始しています！



The screenshot shows the PDB entry page for 6LU7. At the top, there is a navigation bar with the PDB logo and a menu icon. The title of the entry is "PDB Entry - 6LU7" with a status of "(Status - Released)". Below the title, there is a "Summary information:" section containing the following details:

- Title:** The crystal structure of COVID-19 main protease in complex with an inhibitor N3
- DOI:** [10.2210/pdb6lu7/pdb](https://doi.org/10.2210/pdb6lu7/pdb)
- Primary publication DOI:** [10.1038/s41586-020-2223-y](https://doi.org/10.1038/s41586-020-2223-y)
- Entry authors:** Liu, X., Zhang, B., Jin, Z., Yang, H., Rao, Z.
- Initial deposition on:** 26 January 2020
- Initial release on:** 5 February 2020
- Latest revision on:** 10 March 2021

Below the summary information, there is a "Downloads:" section with the following links:

- [Structure coordinates \(PDBx/mmCIF\)](#)
- [Structure coordinates \(PDBML\)](#)
- [Structure coordinates \(PDB\)](#)
- [X-ray diffraction data \(PDBx/mmCIF\)](#)
- [Validation report \(PDF\)](#)
- [Validation report \(XML\)](#)

At the bottom, there is a section "Links to more resources for 6LU7 at:" with logos for PDBj, RCSB PDB, and PDBe.

バイオフィンフォーマティクス研究者にお願い

・ PDBの豊富な情報を活用して、複数のデータベースを横断的に利用するデータ駆動型研究に資する統合利用ポータルを構築します。

・ PDBjが管理・公開しているwwPDBのRDFサイトも拡充します。

・ 学会で意向調査を行いますので ご協力をお願いします。




PDB/RDF PDB entry : 1E6E <https://rdf.wwpdb.org/pdb/1E6E> About PDB/RDF
PDB/RDF, chem_comp/RDF, PRD/RDF

PREFIX

- rdf: <<http://www.w3.org/1999/02/22-rdf-syntax-ns#>>
- rdfs: <<http://www.w3.org/2000/01/rdf-schema#>>
- owl: <<http://www.w3.org/2002/07/owl#>>
- PDBo: <<https://rdf.wwpdb.org/schema/pdbx-v50.owl#>>
- PDBr: <<https://rdf.wwpdb.org/pdb/>>
- dc: <<http://purl.org/dc/elements/1.1/>>
- dcterms: <<http://purl.org/dc/terms/>>
- skos: <<http://www.w3.org/2004/02/skos/core#>>

Search PDB/RDF, chem-comp/RDF, PRD/RDF

ID: (e.g., '7RSA') PDB ID

All wwPDB/RDF files (RDF/XML format) can be downloaded at <https://data.pdbj.org/pdbjplus/data/pdb/rdf/>.

The SPARQL endpoint for wwPDB/RDF (along with many other databases) is available at the NBDC RDF portal: <https://integbio.jp/rdf/dataset/pdbj>



PDBj Protein Data Bank Japan

English 日本語 簡体中文 繁體中文 한국어

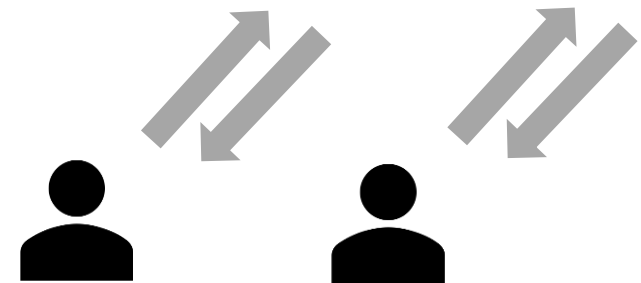
検索:

最新情報

- ✓ 機械学習用データセット
- ✓ 疾患、生物学的現象、化合物などの切り口からのデータ提供

PDBj-PDB-JBIportal@wwpdb.org

Copyright © 2013-2022 日本蛋白質科学データバンク

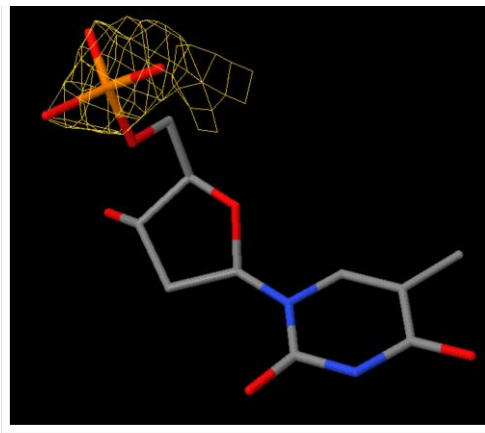


情報科学研究者
データ駆動型研究

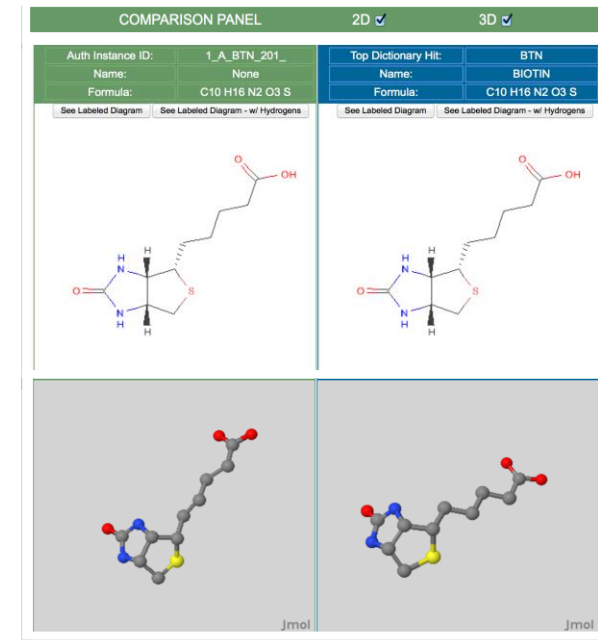
生物学・化学・医薬研究者
現象からのデータ検索

まだ分解能だけで選別していませんか？

- 原子座標だけだと厳密な化合物を区別できないので、編集作業中に厳密に化合物を特定します
- 特に結合次数やキラリティー，原子の価数を厳密にチェック
- しかし，下のような実験データで判断した構造情報でよいのでしょうか？



Local ligand density display (1.5 sigma omit map)
 TMP in entry 3HW4 with LLDF=6.77 (RSR=0.41, CC=0.70)



Deposited instance from
 coordinates (left) and the closest
 match in the dictionary (right)

実は、全てのエントリーに検証レポートがついています

EDITORIAL

nature
structural &
molecular biology

Nature Struct. Mol. Biology, 23 (10), 871, 2016

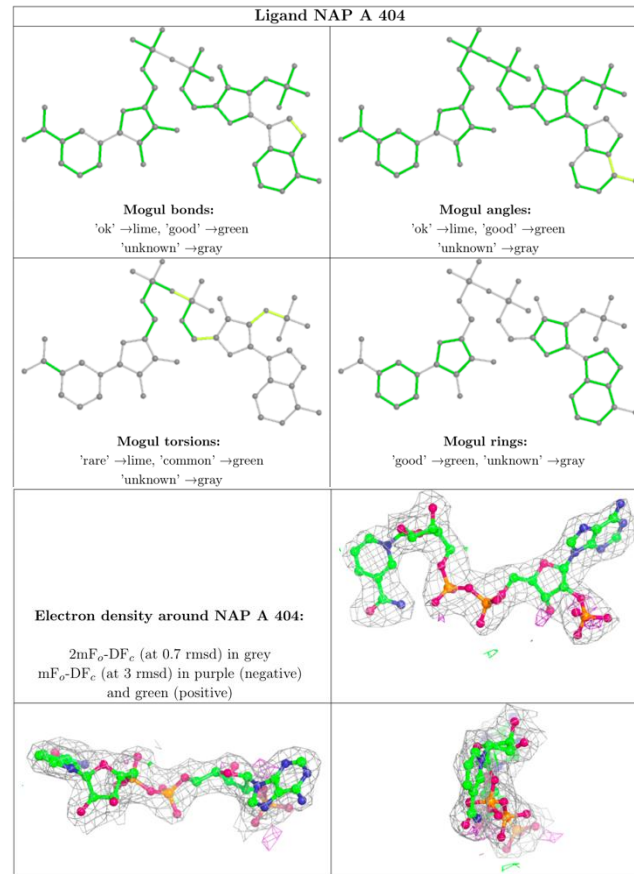
Where are the data?

Here, we announce two policy changes across Nature journals: data-availability statements in all published papers and official Worldwide Protein Data Bank (wwPDB) validation reports for peer review.

We are now taking a further step and are **requesting official wwPDB validation reports for peer review.** These reports are made available by the wwPDB after data deposition (<http://www.wwpdb.org/validation/validation-reports>). Other Nature journals will soon follow suit.

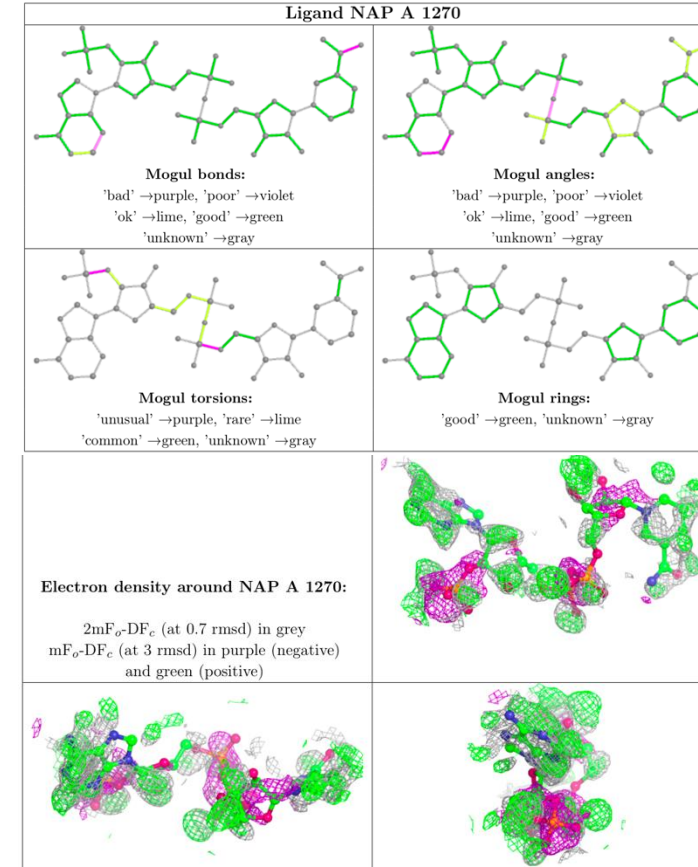
X線：化合物情報を構造化学的・結晶学的に評価

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
3	NAP	A	404	48/48	0.96	0.14	31,43,66,70	0



PDB entry 5zix (Better data quality)

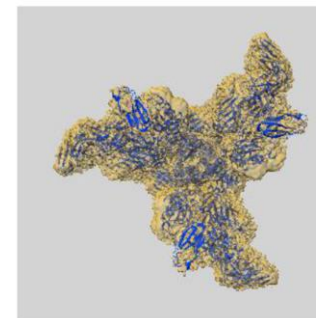
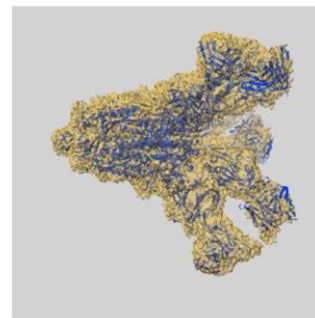
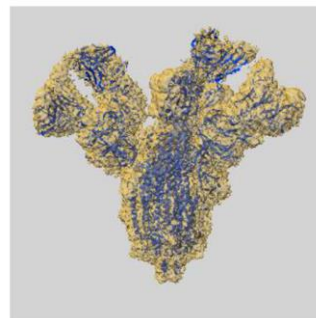
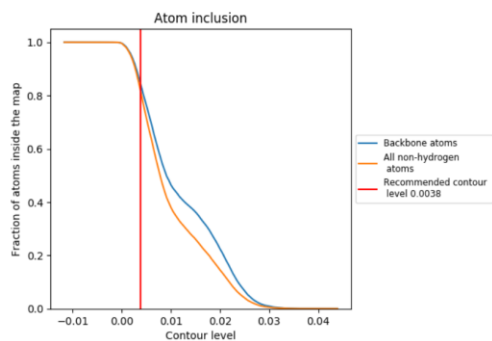
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
3	NAP	A	1270	48/48	-0.06	0.67	87,96,100,100	0



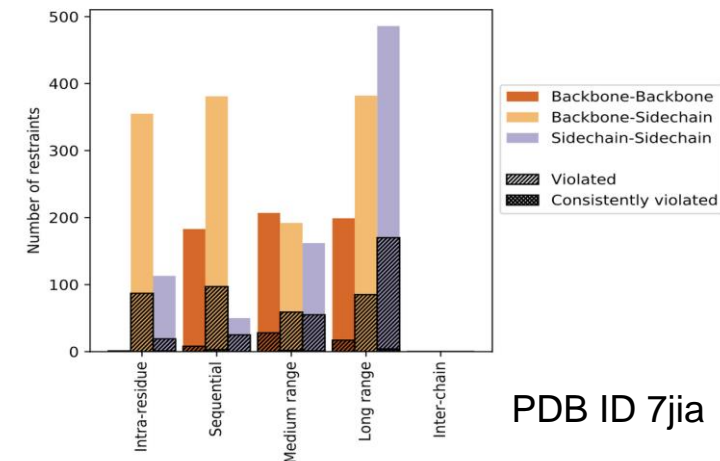
PDB entry 1zk4 (Worse data quality)

CryoEM構造もNMR構造も検証レポートを作成

- EM map volume analysis
 - The fit of model to the map at residue level and global visual overlay
 - Map analysis and visualization
- NMR restraints assessments
 - Distance and dihedral angle restraints with graphical and tabular statistics
 - Available for restraints deposited in single NEF/NMR-STAR formats



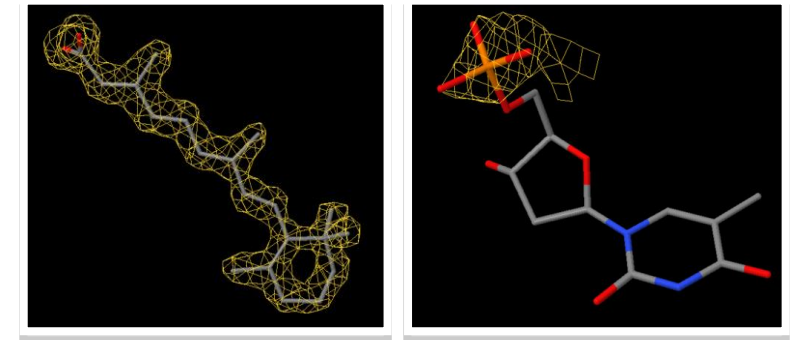
PDB ID 7cww



PDB ID 7jia

化合物に特化した機械学習用データセット公開

- ・ 化合物情報に特化した機械学習用データセットを公開します。
- ・ 蛋白研の実験研究者と協力し、複数の判断基準を導入して機械的にフィルタリングしたデータセットを公開します。
- ・ 各種学会で意向調査を行いますので、ご協力をお願いします。



同分解能でも異なる実験データとの一致度の例

左 : RSR=0.10, CC=0.95

右 : RSR=0.41, CC=0.70

Acknowledgements

スタッフ

このページの他言語版もあります: [English](#)

- **統括責任者**
 - 栗栖 源嗣 (大阪大学蛋白質研究所・教授)
- **PDB/EMDB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)
 - 于 健 (大阪大学蛋白質研究所・特任講師)
 - 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)
 - 金 宙妍 (大阪大学蛋白質研究所・特任研究員)
 - 丹羽 智美 (大阪大学蛋白質研究所・特任研究員)
- **PDB/EMDBデータベース高度化グループ**
 - 水口 賢司 (大阪大学蛋白質研究所・教授)
 - Bekker, Gert-Jan (大阪大学蛋白質研究所・特任助教)
 - 長尾 知生子 (大阪大学蛋白質研究所・助教)
 - 山下 鈴子 (大阪大学蛋白質研究所・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)
- **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)
 - 児嶋 長次郎 (横浜国立大学工学部・教授)
 - 宮ノ入 洋平 (大阪大学蛋白質研究所・准教授)
 - 岩田 武史 (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)
- **EMPIARグループ**
 - 中根 崇智 (大阪大学蛋白質研究所・特任講師)
 - 常住 規代 (大阪大学蛋白質研究所・特任研究員)
- **PRF分室**
 - 栗栖 源嗣 ((財) 蛋白質研究奨励会・招聘研究員)
 - 磯山 正治 ((財) 蛋白質研究奨励会・情報室長)
- **研究協力者**
 - 中村 春木 (大阪大学蛋白質研究所・名誉教授)
 - 由良 敬 (お茶の水女子大学 教授 / 文理融合AI・データサイエンスセンター 副センター長) for [EMPIAR-PDBj](#)
 - 藤 博幸 (関西学院大学理工学部・教授)
 - 川端 猛 (東北大学 情報科学研究科・特任准教授) for [HOMCOS](#), [gmfit](#) and [EMPIAR-PDBj](#)
 - 鈴木 博文 (東北大学 情報科学研究科・研究員) for [EM Navigator](#)
 - 小林 直宏 (理化学研究所横浜放射光科学研究センター・上級研究員)
 - 輪湖 博 (早稲田大学社会科学総合学術院・教授) for [ProMode](#)
 - 猿渡 茂 (北里大学理学部・准教授) for [ProMode](#)
 - 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 [eF-site](#)
 - [Standley, M.Daron](#) (大阪大学微生物病研究所・教授) for [SeqNavi](#), [StructNavi](#), [SeSAW](#), [ASH](#), [MAFFTash](#), [Spanner](#) and [SFAS](#)
 - 加藤 和貴 (大阪大学微生物病研究所・准教授) for [MAFFTash](#)
- **事務職員**
 - 佐久間 量子 (大阪大学蛋白質研究所・特任事務職員)

