

NBDCフォーラム：生命科学のデータベース活用法2020

Protein Data Bank Japan (PDBj)

大阪大学蛋白質研究所

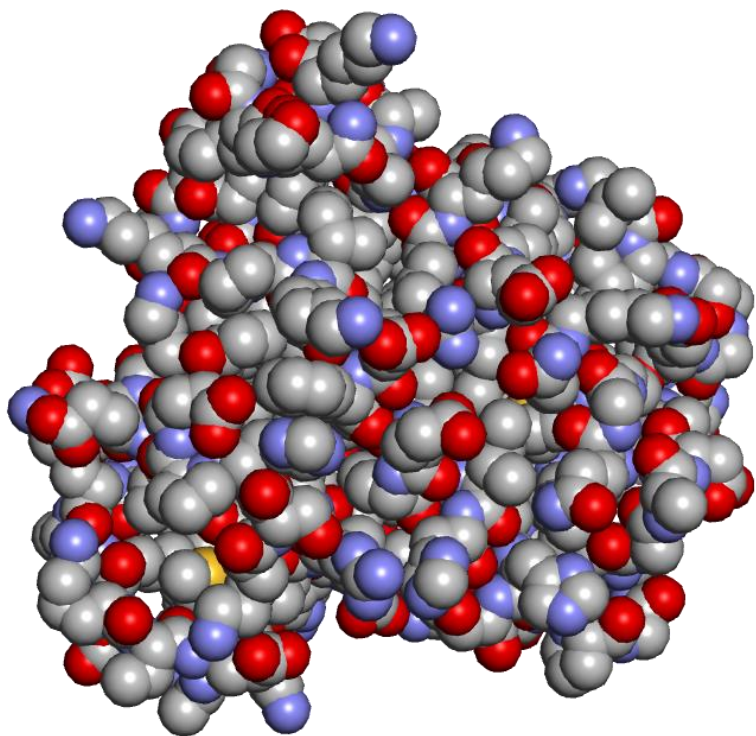
工藤 高裕 (KUDOU Takahiro)

2020年12月2日 (水)

Protein Data Bank Japan (PDBj)

1. PDB (データベース) とは
2. PDBj (組織) のサービス

1. PDBとは



ミオグロビン(1mbn)

タンパク質などの分子の
「かたち」に関する情報を集めた
データベース
蛋白質構造データバンク
(Protein Data Bank)

1. PDBとは

データはテキスト形式

PDBx/mmCIF (現行型式、可変長)

PDBフォーマット (旧型式、固定長)

```
data_1MBN
#
_entry.id      1MBN
#
_audit_conform.dict_name      mmcif_pdbx.dic
_audit_conform.dict_version    5.287
_audit_conform.dict_location
http://mmcif.pdb.org/dictionaries/ascii/mmcif_pdbx.dic
#
loop_
_database_2.database_id
_database_2.database_code
PDB      1MBN
WWPDB    D_1000174928
...
```

```
HEADER      OXYGEN STORAGE                                05-APR-73      1MBN
TITLE       THE STEREOCHEMISTRY OF THE PROTEIN MYOGLOBIN
COMPND      MOL_ID: 1;
COMPND      2 MOLECULE: MYOGLOBIN;
COMPND      3 CHAIN: A;
COMPND      4 ENGINEERED: YES
SOURCE      MOL_ID: 1;
...
ATOM        1  N   VAL A   1      -2.900  17.600  15.500  1.00  0.00      N
ATOM        2  CA  VAL A   1      -3.600  16.400  15.300  1.00  0.00      C
ATOM        3  C   VAL A   1      -3.000  15.300  16.200  1.00  0.00      C
ATOM        4  O   VAL A   1      -3.700  14.700  17.000  1.00  0.00      O
```

書いてあることは、各原子の三次元座標 (XYZ) のほか
登録者、文献、実験手法、由来生物種など

1. PDBとは

- 1971年発足→2021年で50周年



- 世界で唯一



- 現在は世界で4つのグループが協力して運営



1. PDBとは

- PDBjは2000年に登録処理を開始

→今年で20周年



I. PDBとは

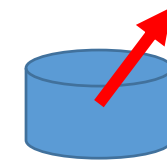
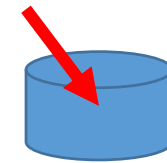
毎年構造解析・バイオインフォマティクスの専門家を招いて諮問委員会を開催し運営方針を検討



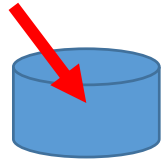
1. PDBとは

主な業務は

- 構造データの登録受付 (Data-in)
- データの公開 (Data-out)

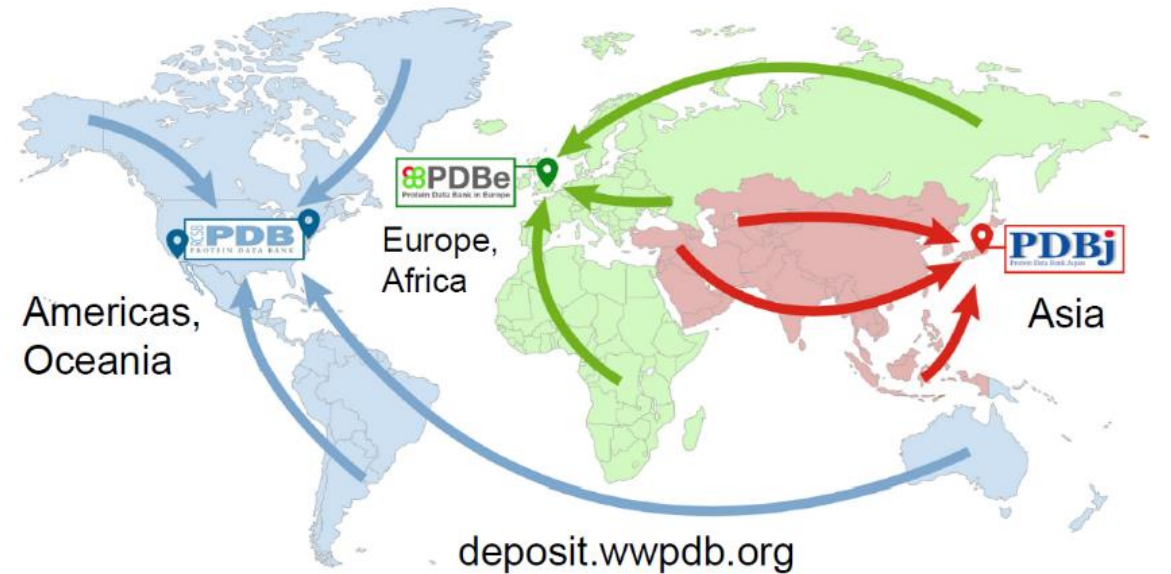


1. PDBとは



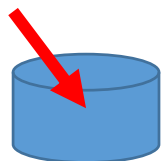
Data-in

データ登録は各拠点で分担
(PDBjはアジア地区担当)

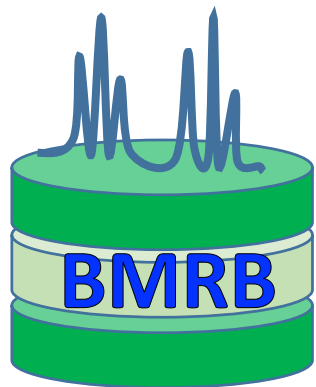


1. PDBとは

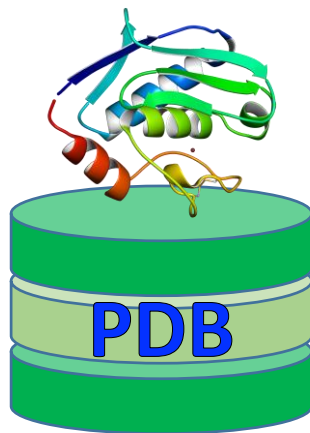
PDBへの登録には実験の根拠が必須
多くの論文雑誌はPDB登録が必須



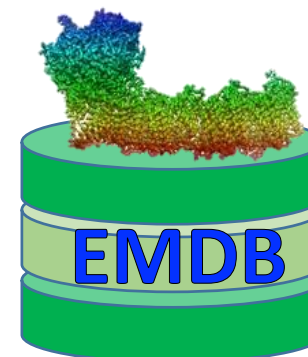
NMR
実験データ



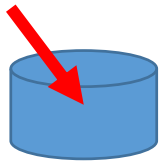
原子
座標



電子顕微鏡
電子密度マップ



1. PDBとは



Data-in

- 新規データは非公開、公開日まで厳重に管理
(入室管理された部屋で専門スタッフが処理)

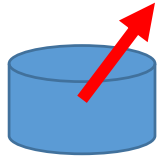


電子錠つきの部屋



専門スタッフが登録・検証作業を実施

1. PDBとは



Data-out

- 全拠点で同時公開（毎週水曜 日本時間9:00）
- 利用は無償・登録不要 人類の財産
- 拠点ごとに独自のツール・サービスを提供

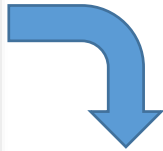
2. PDBjのサービス紹介

簡易検索 (キーワード検索)

日本語OK!

自動英訳して検索

The screenshot shows the PDBj homepage. At the top, there are navigation links for English, 日本語, 簡体中文, 繁體中文, and 한국어. A search bar contains the text 'c型肝炎ウイルス'. Below the search bar, there are several sections: 'ホーム', 'データ登録 (OneDep)', 'ダウンロード', and '標準フォーマット'. On the right side, there are sections for '7KFX 最新公開', '今月の分子', and '新型コロナウイルス最新構造情報'.



The screenshot shows the search results page for 'c型肝炎ウイルス'. The search bar contains '(hcv | "hepatitis c virus")'. The results are displayed in a list format. The first result is '2MTS' with the title 'THREE-DIMENSIONAL STRUCTURE AND INTERACTION STUDIES OF HEPATITIS C VIRUS P7 IN 1,2-DIHEXANOYL-SN-GLYCERO-3-PHOSPHOCHOLINE BY SOLUTION NUCLEAR MAGNETIC RESONANCE'. The second result is '1CSJ' with the title 'CRYSTAL STRUCTURE OF THE RNA-DEPENDENT RNA POLYMERASE OF HEPATITIS C VIRUS'. On the right side, there is a '検索結果' section showing the total number of hits (743) and a '表示順' section with options for relevance, date, and resolution.

<https://pdbj.org/>

2. PDBjのサービス紹介

簡易検索の検索結果例

The screenshot shows the search results for 'c型肝炎ウイルス' (Hepatitis C Virus). The search filters are: PDB: 743件, ウェブページ: 6件, ステータス検索: 25件, 化合物検索: 1件, BIRD: 0件. The search term is 'c型肝炎ウイルス' and the search criteria are '(hcv | "hepatitis c virus")'.

Search filters: PDB: 743件, ウェブページ: 6件, ステータス検索: 25件, 化合物検索: 1件, BIRD: 0件

Search term: c型肝炎ウイルス

変換クエリ: (hcv | "hepatitis c virus")

2. PDBjのサービス紹介

詳細検索で絞り込み

Mine: 詳細条件検索

PDBID:

キーワード:

タイトル:

公開日: 以降: 2020 - 1 - 1

以前: 2020 - 12 - 2

登録日: 以降: - -

以前: - -

文献著者:

実験手法:

分解能: -

検索結果

全ヒット件数: 743
表示件数: 25
表示順: 関連性が高い順
Auto-pager:

表示順

関連性が高い順

PDBID昇順 (0→9, a→z)
PDBID降順 (z→a, 9→0)
登録日の古い順
登録日の新しい順
公開日の古い順
公開日の新しい順
分解能の高い順

Mine: 詳細検索結果ページ

詳細検索:
キーワード: C型肝炎ウイルス → hcv | "hepatitis c virus"
公開日 以降: 2020-01-01
公開日 以前: 2020-12-02
実験手法: ELECTRON MICROSCOPY
分解能: 0.0 - 3

キーワードの追加:

6YGI

分子名称: Capsid protein, Capsid protein
著者: Makbul, C., Bottcher, B.
登録日: 2020-03-27
公開日: 2020-09-02
実験手法: ELECTRON MICROSCOPY (3 Å)
主引用文献: Slowly folding surface extension in the prototypic avian hepatitis B virus capsid governs stability. Elife, 9, 2020

例: 2020年以降公開、実験手法は電子顕微鏡、分解能は3 Å以下

2. PDBjのサービス紹介

糖鎖表現の改善

- 命名を標準化 (IUPAC-IUBMB)
- 一貫性のある表記法
- 糖鎖科学で用いられる線形記述法を採用



PDBにおける糖鎖の検索性向上

PDB ID 1b5f

WORLDWIDE PDB PROTEIN DATA BANK
wwPDB.org

Improving Carbohydrates in the PDB for 2020

~10% of PDB structures contain carbohydrates.

To improve the ability to search for these structures and their complex chemistry (e.g., stereo-isomers, anomeric configurations, branched chains), wwPDB is embarking on a new remediation effort that will:

- Standardize Chemical Component Dictionary definitions
- Adopt carbohydrate nomenclature following IUPAC-IUBMB recommendations
- Provide a uniform representation for oligosaccharides
- Adopt glycoscience community naming conventions
- Incorporate glycoscience community tools to properly identify, validate, and biocurate glycosylation sites in PDB structures

Implementation

New sequence descriptions will be included in PDB structure data files:

Condensed IUPAC:
LFucpa1-6[DManpa1-3DManpb1-4DGlcNacb1-4][LFucpa1-3]DGlcNacb1-ASN

LINUCS:
[[ASN][[(4+1)][b-D-GlcpNAc][[(3+1)][a-L-Fucp]{}][(4+1)][b-D-GlcpNAc][[(4+1)][b-D-Manp][[(3+1)][a-D-Manp]{}]][(6+1)][a-L-Fucp]{}]]

These descriptions can be translated into Symbolic representations used by glycoscientists:

Symbolic:

Example based on PDB Entry 6cmg