

タンパク質構造データベース・PDBの 使い方

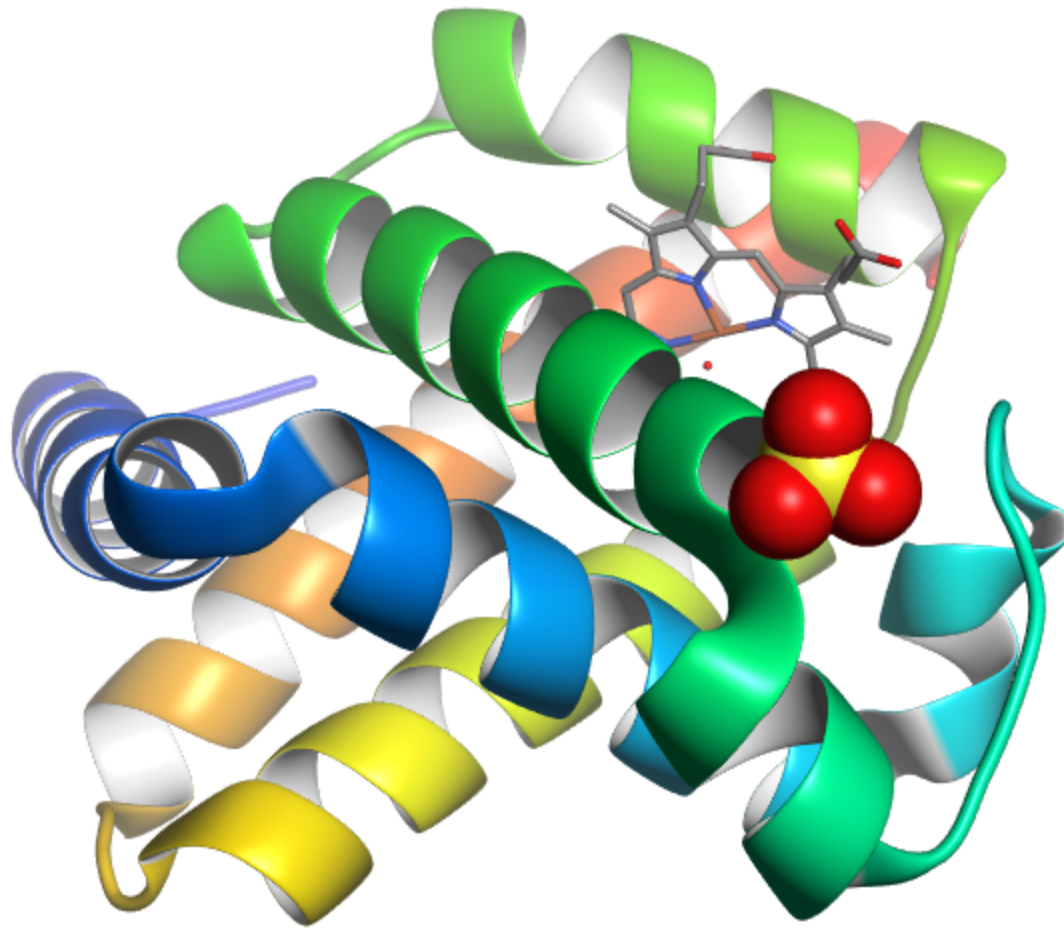
All-in-one 合同講習会 2018／統合データベース講習会：AJACS 番町 2

2019-01-28

JST東京本部別館

鈴木 博文 (早稲田大学 先進理工)

どんなデータの話？



⇒ タンパク質などの生体分子の立体構造のデータ

構成原子の座標(X,Y,Z)と付随情報

(図: PDB-3vau ミオグロビン)

自己紹介

- 氏名: 鈴木 博文
- 研究のバックグラウンド
 - 電子顕微鏡を使った立体構造解析
- 2018年9月まで
 - 大阪大学 蛋白質研究所に所属
 - PDBjの開発スタッフ
- 現在
 - 早稲田大学 先進理工
 - バイオインフォマティクス・データベース
- 今回は...
 - 過去、AJACS講習会の講師を数回
 - 今回は会場が東京なので...

想定する聴衆

構造生物学・バイオインフォマティクス

- 専門外 ～ 初学者

コンピューターを利用した研究

- ある程度経験あり

この講習の方針

PDBデータは...

- 難しい面
 - マニアック (独特の用語、理論体系など)
 - 非専門家には、研究利用は難しい
- 簡単な面
 - データで「遊んでみる」のは簡単、割と楽しい
 - 教育利用、画像をアートとして利用など

なので...

- 「基礎からの知識の積み上げ」より「断片的」「よもやま話的」
- 他のデータベースを使うときにも役立つような情報

PDBjのウェブサイトの紹介

1. PDBjのウェブサイトの紹介 
2. 構造ビューアの使い方
3. PDBの紹介
4. データフォーマット
5. 補足

PDBj トップページ

- <https://pd bj.org/>
- 「pd bj」を検索するとヒット
- スマートフォンにも対応
 - 画面の広さ次第の表示

PDBj 141415
エントリ (2018-06-20)

English 日本語 简体中文 繁體中文 한국어

pd bj.org 全体を検索 (日本語OK)

Protein Data Bank Japan
Worldwide Protein Data Bank Foundation

日本蛋白質構造データバンク (PDBj: Protein Data Bank Japan) は、[JST-NBDC](#) と [大阪大学蛋白質研究所](#) の支援を受け、米国 [RCSB](#)、[BMRB](#)、および欧州 [PDBe](#) と協力して、生体高分子の立体構造データベースを国際的に統一化されたPDBアーカイブとして運営するとともに、様々な解析ツールを提供しております。

▼ **初めての利用者のためのガイド**

このサイトは主に研究者向けのコンテンツを提供しています。一般向けの内容は[PDBj入門](#)をご覧ください。
古いウェブブラウザでこのサイトにアクセスすると、機能制限モードで表示される場合があります。すべての機能を利用するための要件について詳しくは[動作要件](#)をご覧ください。
このサイトの機能について詳しくは[対話型チュートリアル](#)もご覧ください。

▼ **必要なサービスを探す**

探しているサービスに関連するキーワードを以下の語句一覧から選択するか、または検索ボックスに入力して下さい。該当するサービスの一覧が表示されます。

- [全サービスを表示する]ボタンを押すと、全サービスの概要が表示されます。
- [キーワードボックス]にキーワードを入力して、語句一覧で絞り込んだ結果を更に絞り込むこともできます。

6EMX
最新公開エントリ

▼ **今月の分子**

[222: タンパク質とナノ粒子 \(Proteins and Nanoparticles\)](#)

今月の分子のリスト

WORLDWIDE PDB PROTEIN DATA BANK

▶ **BMRB検索**

▶ **パートナー**

IBIportal

ページの構成

上部に検索ボックス

- データエントリ、ドキュメントなども検索可能

左にメインメニュー

- 画面の広さによっては非表示、メニューボタンで表示切り替え

「初めての利用者のためのガイド」

- 文字通り初心者向けへのガイド
- 「対話型チュートリアル」
- 「PDBj入門」非専門家向けサービス

「必要なサービスを探す」

- たくさんのサービスがあるので、サービスを探すためのサービスがある
- キーワードか、カテゴリで検索

「PDBj入門」

- <https://numon.pdbj.org/>
- 教材などへのリンク集
- PDBjがアウトリーチ活動で利用しているサービスなど
 - スマートフォンでVRが見られるサービス
 - 赤青メガネで立体視
- 各種ドキュメント
- 「今月の分子」

PDBj の生体高分子学習ポータルサイト

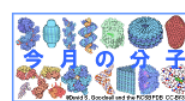
PDBj入門

日本語 English

PDBj入門はPDBj(日本蛋白質構造データバンク)の一般向けサイトです。タンパク質などの生体分子について理解するためのサービスを提供しています。研究者向けサイトは[こちら](#)。



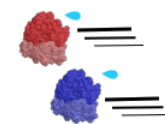
PDBに登録されている生体高分子を取り上げ解説している蛋白質構造百科事典



月替わりの分子構造紹介記事



過去の講習会などの資料



神経衰弱とスネークゲーム

[利用規約](#)・[プライバシーポリシー](#) [お問い合わせ](#)

今月の分子

- <https://numon.pdbj.org/mom/>
- 生体分子の構造の美しい画像と丁寧な解説文
- 米国 RCSB-PDBの David S. Goodsellによる「Molecule of the Month」の日本語訳

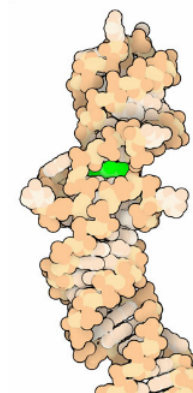
229: 蛍光RNAアプタマー(Fluorescent RNA Aptamers)

著者: David S. Goodsell 翻訳: 工藤 高裕(PDBj)

細胞の中をより詳しく調べる新たな道具を探す試みは常に続けられている。緑色蛍光タンパク質(green fluorescent protein, GFP)は全く新たな分野を切り開いた道具の一例である。これを使えばある特定のタンパク質に印をつけることができ、生きた細胞の中でそのタンパク質が何をしているのかをみることができる。最近、これと似た方法でRNAをみる新たな道具の開発が進められている。RNA自体は蛍光を発しないので、蛍光体(fluorophore、小さな蛍光分子)に結合してその蛍光を強める短いRNAをつくるという手法が用いられた。このRNAはリボソーム(ribosome)のような天然のRNAに組み込むことができる。蛍光体を細胞に加えると、改変型リボソームに結合し、それを我々は観察できるようになる。

アプタマーの獲得

SELEX(Systematic Evolution of Ligands by EXponential enrichment、急速増幅法によるリガンドの計画的進化) はこのような役に立つ蛍光体結合性RNA分子を見つけるのに使われてきた。この手法では、まずある蛍光体にさまざまなRNA配列を混ぜ、次にそれと結合するものを単離する。そしてアミン基に修飾を加え、その中から最もよいものを



トップページからの検索

- トップページの上部の検索ボックス
- キーワード検索
- 日本語もOK
- PDBデータだけでなく、各種サービス、ドキュメントも検索可能



PDBj Mine 詳細検索

- <https://pd bj.org/advanced-search>
- 多様な条件での絞り込み検索可能

Mine: 詳細条件検索

PDBID:

キーワード:

タイトル:

公開日:

以降: - - 

以前: - - 

登録日:

以降: - - 

以前: - - 

文献著者:

分解能:

 -

更に高度な検索

- <https://pd bj.org/mine>
- **PDBj Mine**: PDBj検索エンジンの総称
- **XPath Search**: XML構造に基づく検索
- **CIF Query Service**: 関係モデルに基づくデータ構造を利用した検索
- **SQL検索**: SQL文で検索

PDBj Mine (PDBエントリーの検索)

Mine quick search



Mine entries

[新規公開されたエントリー](#)[更新されたエントリー](#)[最新の廃止/置換されたエントリー](#)[全ての廃止/置換されたエントリー](#)[巨大構造エントリー](#)

Specialized search

[詳細条件検索](#)[ステータス検索](#)[Chemie search \(化合物の検索\)](#)

More information

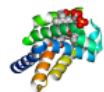
検索結果

- 一般的な「いわゆる検索結果」
- 上部にページ種類のタブ
 - PDBエントリ
 - ウェブページ
 - 化合物
 - その他いろいろ

PDB: 4 件 ウェブページ: 38 件 ステータス検索: 0 件 化合物検索: 0 件 BIRD: 0 件

肉

変換クエリ: (meat | flesh)

3VAU   

MYOGLOBIN NITRITE STRUCTURE: NITRIHEME MODIFIED

分子名称: Myoglobin, [3,3'-(7-ethenyl-3,8,13,17-tetramethyl-12-[(E)-2-nitroethenyl]porphyrin-2,18-diyl-kappa~4~N~21~,N~22~,N~23~,N~24~}dipropionato(2-))iron, NITRITE ION, ...

著者 [Yi, J., Richter-Addo, G.B.](#)

登録日 2011-12-29

公開日 2012-04-11

実験手法 X-RAY DIFFRACTION (1.7 Å)

主引用文献 **Unveiling the three-dimensional structure of the green pigment of nitrite-cured meat**
Chem.Commun.(Camb.), 48, 2012

データエントリのページ

- タブ切り替えで表示情報を切り替え

- 構造情報
- 実験情報
- 機能情報
- など

- 右カラム

- 構造ビューア
- ダウンロード
- 他のデータベースへのリンク

- 図はPDB-3VAU（ミオグロビン）のページ

概要 構造情報 実験情報 機能情報 相同蛋白質 履歴 ダウンロード

3VAU

Myoglobin nitrite structure: nitriheme modified

▼ 3VAU の概要

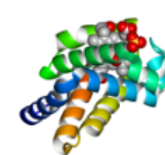
分子名称	Myoglobin, [3,3'-{7-ethenyl-3,8,13,17-tetramethyl-12-[(E)-2-nitroethenyl]porphyrin-2,18-diyl-kappa~4~N~21~,N~22~,N~23~,N~24~}dipropanoato(2-)]iron, NITRITE ION, ... (5 entities in total)
機能のキーワード	myoglobin, nitrite, nitriheme, oxygen transport
由来する生物種	Equus caballus (horse)
ポリマー鎖数	1

▼ ダウンロード

- Sequence (fasta)
- PDBx/mmCIF
- PDBML (ヘッダのみ (no-atom))
- PDB形式 (全ての情報)
- 検証レポート (PDF) [More...](#)

▼ 構造

▶ 非対称単位を表示 (AU = BU)



構造ビューアの使い方

1. PDBjのウェブサイトを紹介
2. 構造ビューアの使い方 
3. PDBの紹介
4. データフォーマット
5. 補足

PDBエントリのページからMolmil

- 構造エントリページの右側の画像をクリックすると、Molmilのウィンドウがポップアップ
- 図はPDB-3cl0の例



Molmil

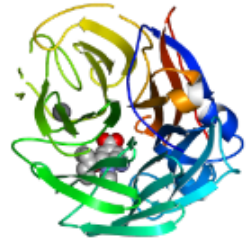
- <https://pd bj.org/help/molmil>
- PDBjで開発している構造ビューア
- ウェブブラウザ上で動作 (Javascript/WebGL)
- マウスのドラッグで分子の回転、ホイール回転(または右ボタンドラッグ)で拡大縮小、など
- スマートフォンでも動作、タッチ操作（フリックで回転、ピンチで拡大縮小、など）
- メニュー操作で各種表示スタイルへの切り替え
- コマンドラインインターフェースも

2種類のビューアリンク

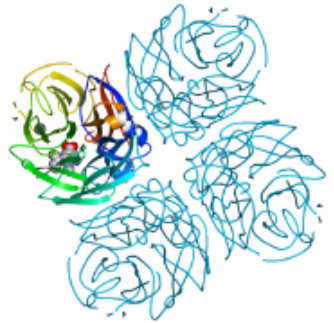
- 構造データによっては「生物学的単位を表示」という項目も表示される
- 図はPDB-3cl0の例
- PDB-3cl2の場合、同様の構造データで同様の生物学的単位だが、非対称単位が異なる

▼ 構造

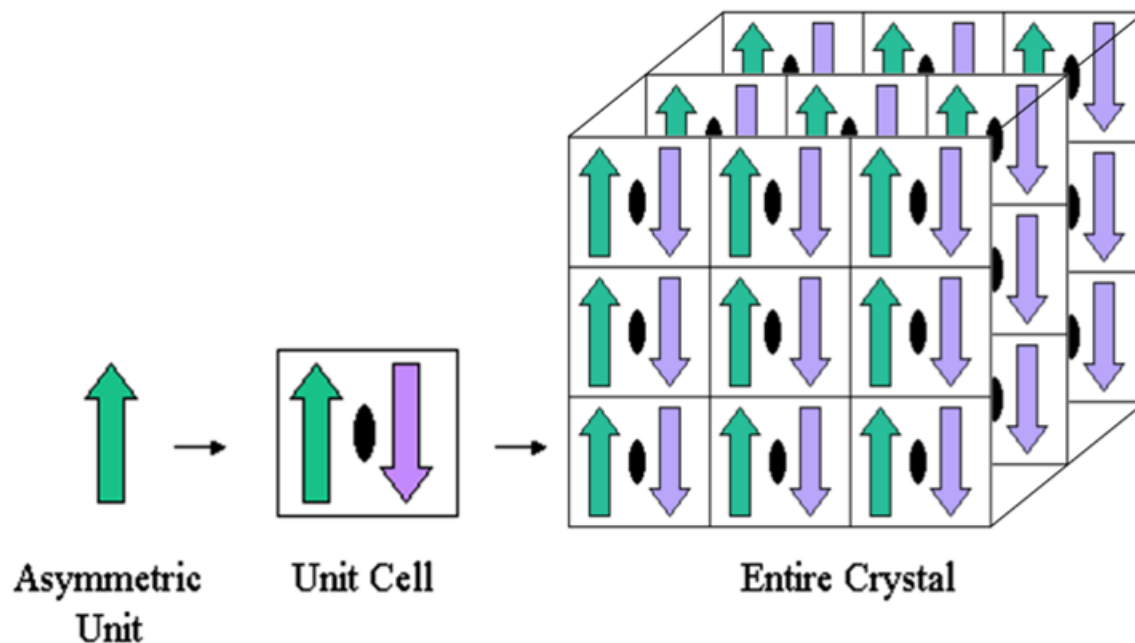
▶ 非対称単位を表示



▶ 生物学的単位を表示



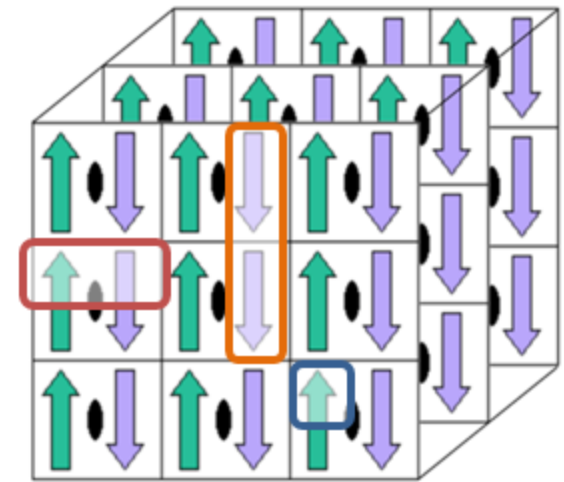
非対称単位



- PDBに登録されているデータのほとんどは結晶構造
- データ登録者が繰り返しの基本単位とみなした構造・**非対称単位 (Asymmetric unit)** の原子配置と、その対称性情報が登録されている
- らせん対称や、正20面体型対称など、結晶学以外の非対称単位もある

生物学的単位（生物学的集合体）

- 「生体内や溶液中での構造に近い」と考えられる構造単位を、結晶構造から切り出したもの
- データ登録者が定義したものと、計算によって見積もられたものが登録されている
- 生物学的単位とみなした根拠（クロマトグラフィーなど）も記述されている
- 非対称単位との関係（大きい・小さい）は、データごとに異なる



その他の構造ビューア

PyMOL

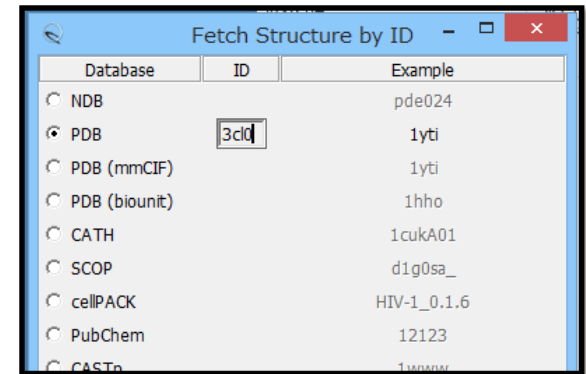
- 高機能なビューアとしては、デファクトスタンダード
- オープンソース
- バイナリ完全版は有償
- 高度な操作にはスクリプトの知識が必要

UCSF-Chimera

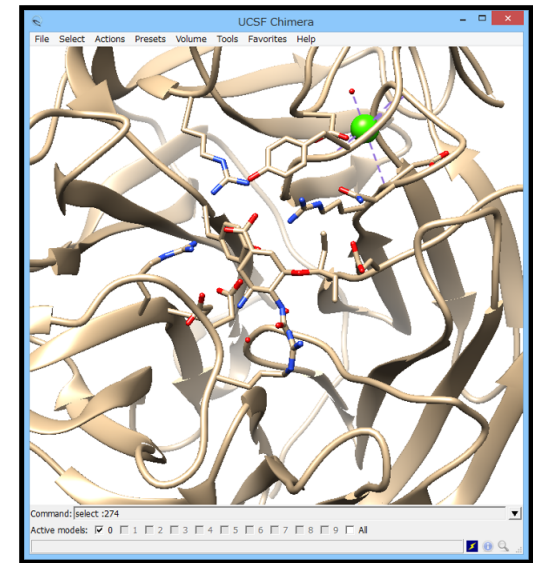
- 非営利利用ではフリー、各種OS向けインストーラが配布
- ソースは非公開
- GUIが良くできている
 - 適当に触っていると割とスゴイ作図や解析ができてしまう
- ヘルプ・ドキュメントも充実

UCSF-Chimera

- メニュー -> open -> Fetch by ID -> PDB-IDを入力するだけでデータが開く

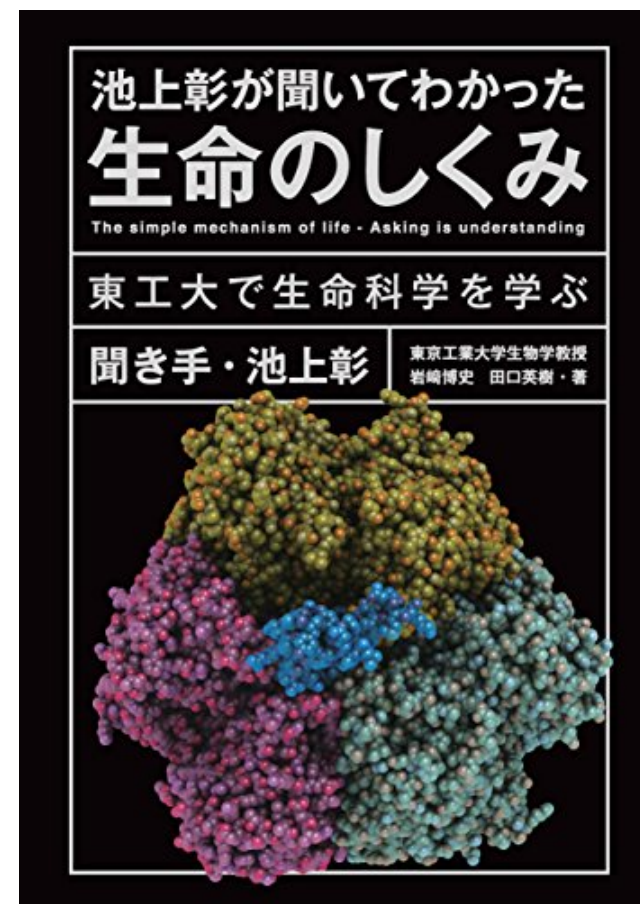


- 注目部位へのフォーカス、複数の構造の重ね合わせ、モーフィングアニメーションの作成、配列ビューアとの連携表示など、簡単なGUIのみで実現可能
- 以前のAJACS講習会で、変異残基の構造変化などの実習を実施
 - <https://events.biosciencedbc.jp/training/ajacs70/>



QuteMol

- きれいな絵を描くことに特化したソフト
- 低機能だが、その分簡単に「凄い絵」が描ける
- 右図の本の表紙画像はQuteMolで作成



データと画像の利用について

権利の話

- PDBデータはフリー（パブリックドメイン）
- 各ウェブサイトの画像などについてはそれぞれのドキュメントを参照
 - PDBjの利用規約: <https://pdbj.org/info/terms-conditions>

科学の文脈でのマナー

- データの出処は明記する、PDBの場合はPDB-ID

ソフトウェアのライセンスにも注意

- 文献引用の義務・要請がある場合も

PDBの紹介

1. PDBjのウェブサイトを紹介
2. 構造ビューアの使い方
3. PDBの紹介 
4. データフォーマット
5. 補足

2つの側面

PDBは...

世界でたったひとつのデータベース

- 厳重に秘匿
- 厳格に管理

世界中にたくさんあるデータベース

- 容易に入手
- 自由に利用
- 勝手に公開

「用語」についてよもやま話 - Gmailの例

- 「Gmail」という名称が重複
 - サービス名
 - アプリ(メールクライアント)名
- それも良し悪し
 - 日常の利用では便利、覚えやすい
 - ちょっとマニアックな話をするときは混乱を招く
 - 例「GmailでOutlookが使えるし、OutlookでGmailが使えるよ」

いろいろな「PDB」

実態	名称
データベース名 (アーカイブ名)	PDB (Protein Data Bank)
組織名	wwPDB, PDBj, RCSB-PDB, PDBe
サービス名	PDBjなど
ファイル形式 (旧・公式フォーマット)	PDB形式 (拡張子:ent/pdbなど)

- データのファイルフォーマットについては後述

「用語」についてよもやま話 - 「データベース」

文脈で意味が変わる

- 辞書的には「整理されたデータの集まり」
- 検索などができるインターネット上のサービス
- 狭義ではデータベース管理システムを指すことも

データベースの種類

- 外部からのデータ登録で成り立っているデータベース運営を「**データバンク**」と呼ぶことも
- 検索できない、単なるファイルの集まりは「**データアーカイブ**」と呼ぶことも
- 一次データベース、派生データベース、統合データベースなどいろいろ
 - GenBank、PubMedで有名なNCBIのサイト、たくさんのデータベースがある
 - structureというデータベースがある => PDB派生

PDBとは



メインページ
コミュニティ・ポータル
最近の出来事
新しいページ
最近の更新
おまかせ表示
練習用ページ
アップロード (ウィキメディア・コモンズ)

ヘルプ

ヘルプ
井戸端
お知らせ
バグの報告
寄付
ウィキペディアに関するお問い合わせ

ログインしていません トーク 投稿記録 アカウント作成 ログイン

ページ ノート

閲覧

編集

履歴表示

Wikipedia内を検索



蛋白質構造データバンク

蛋白質構造データバンク（たんぱくしつこうぞうデータバンク、**PDB**; Protein Data Bank）は、**蛋白質**（タンパク質）と**核酸**の3次元構造の構造座標（**立体配座**）を蓄積している国際的な公共の**データベース**である。PDBに蓄積されている構造データは、X線結晶解析法、NMR法（核磁気共鳴法）などによって実験的に決定されたデータである。なお、理論的な予測（**蛋白質構造予測**）で推定されたデータは蓄積していない。

世界中の**生物学者**や**生化学者**たちが、PDBに構造データを登録する。PDBに登録されたデータは**パブリックドメイン**のもとで公開され、誰もが無償でアクセスすることができる。**日本**では**大阪大学蛋白質研究所**にその支所がある。

PDBは、生物学的構造データの中心的なデータベースである。**構造生物学**の研究で欠かせない情報源であり、また近年では**構造ゲノミクス**の研究でも重要なデータベースである。**バイオインフォマティクス**の研究でも、PDBに代表される3次元分子構造データベースは重要な研究対象である。PDBから派生したデータベースとプロジェクトは非常に多く、蛋白質の構造、機能、進化のそれぞれの側面から、PDBの構造データの統合や分類を行っている。

目次 [非表示]

- 歴史
 - 1.1 構造データの増加
- 構造データの内容
 - 2.1 統計情報

PDBとは

PDBとは

- (英)Protein Data Bank
- (日)蛋白質構造データバンク
- 「生物学的構造データの中心的なデータベース」

「Protein Data Bank」という名前だけど...

- タンパク質の情報を網羅的に集めたものではない
- タンパク質だけではない、核酸、糖鎖など
- 生体分子の立体構造(原子座標)のデータベース
 - **実験により観測された**立体構造
 - 同じタンパク質でも別の実験のデータは**別エントリ**
 - 理論のみの構造モデルは受け入れていない

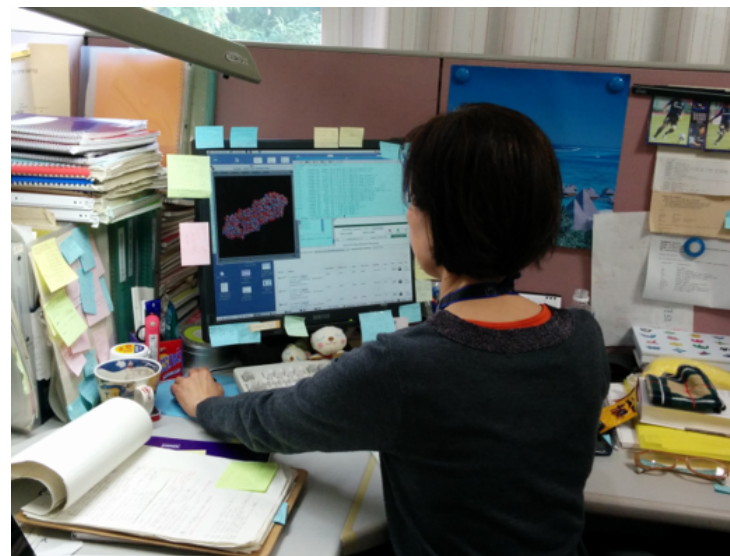
PDBの歴史

年	出来事
1971	設立 ブルックヘブン国立研究所(米)
1998	RCSB(米)へ移管
2003	wwPDB設立 (RCSB-PDB/BMRB(米), PDBe(欧), PDBj(日))

- 非常に長い歴史
 - インターネット・データベースの「黎明期」から
- 運営組織の移管を経ている
- 国際的な運営組織、日米欧にメンバー

データの受け入れ (登録・検証・品質管理)

- 世界中の研究者から集める
 - PDBへの登録は事実上、義務
- 登録受付・検証・品質管理
 - PDBの3拠点で**分担**
 - 専任のアノテーターによる、共通ルールでの品質管理と注釈付
 - 全拠点で**共通**のデータアーカイブを作成
- 新規データは**非公開**、公開日まで厳重に管理
 - 競争が激しい研究テーマが多い



PDBへの登録が義務

nature誌が論文を受理する条件

	PDBへの登録
Macromolecular structure	Worldwide Protein Data Bank (wwPDB)
	Biological Magnetic Resonance Data Bank (BMRB)
	Electron Microscopy Data Bank (EMDB)

<http://www.nature.com/authors/policies/availability.html> より

他の多くの論文ジャーナルも同様の条件

→ PDBにデータを登録しないと、論文発表できない

wwPDBの「諮問委員会」



wwPDB Advisory Committee Meetings 2014

メンバーは世界中の構造解析・バイオインフォマティクスの専門家

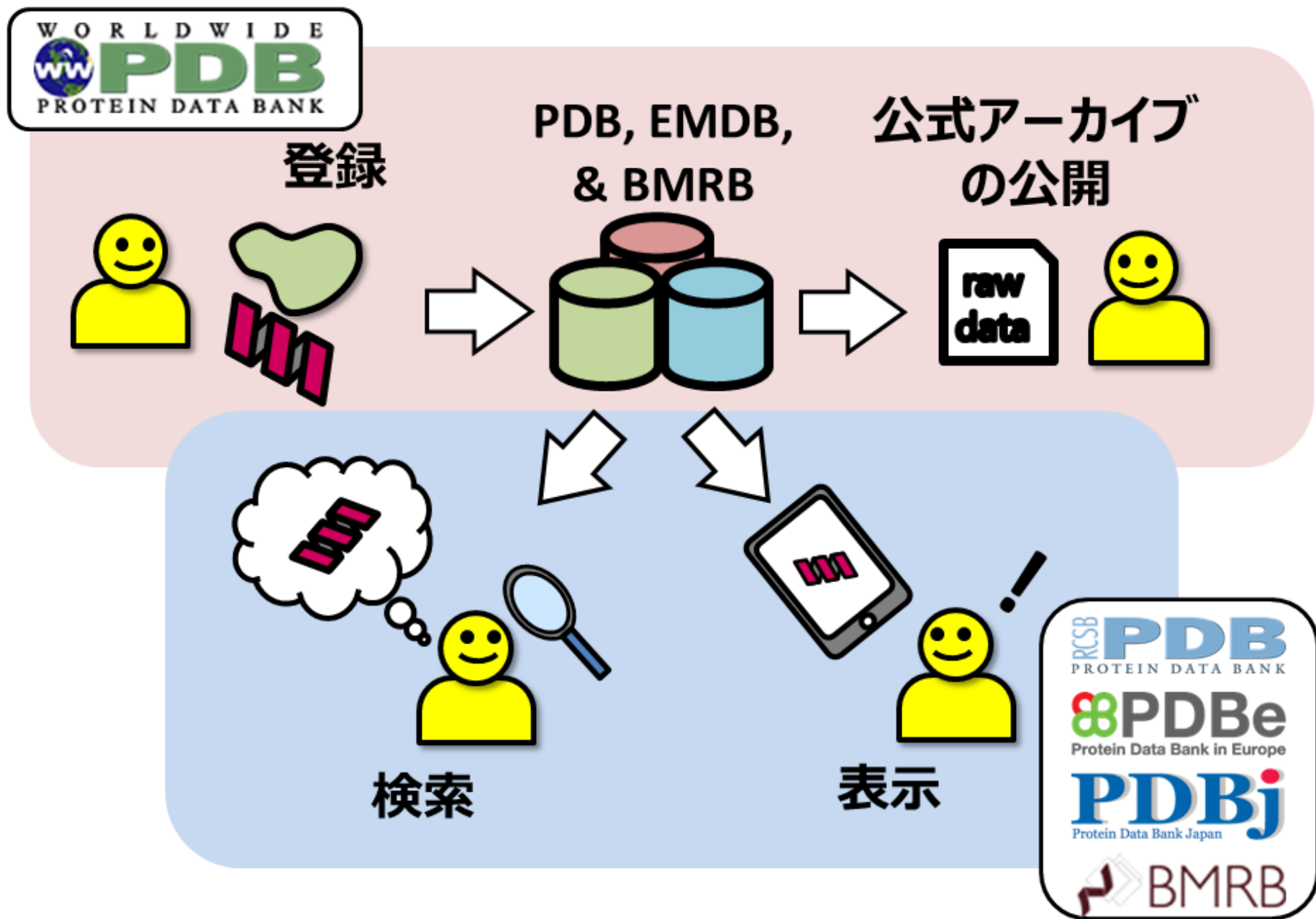
<http://www.wwpdb.org/about/advisory>

データの公開

- 世界同時公開
 - 毎週水曜0:00 (日本時間9:00)
- 公開後は**パブリックドメイン**
 - 用途は問わない、人類の財産
- 公開データ
 - 世界共通のデータアーカイブ
 - 各拠点独自のツール・独自のサービスを通じて公開



PDBの運営



2つの側面

PDBは...

世界でたったひとつのデータベース

- 厳重に秘匿
- 厳格に管理

世界中にたくさんあるデータベース

- 容易に入手
- 自由に利用
- 勝手に公開

データフォーマット

1. PDBjのウェブサイトを紹介
2. 構造ビューアの使い方
3. PDBの紹介
4. データフォーマット 
5. 補足

ダウンロードできるファイル

▼ リソース		
ファイル形式	ファイル名 (ファイルサイズ)	
PDB	全ての情報	pdb3cl0.ent.gz (69.57 KB) 画面表示
	全ての情報 (非圧縮)	pdb3cl0.ent (291.81 KB)
	ヘッダのみ	pdb3cl0.ent.gz (7.05 KB) 画面表示
	PDBx/mmCIF	3cl0.cif.gz (87.16 KB) ▶ 画面表示

個別ページのダウンロードタブ

- 多くの種類のファイルがある、PDBとPDBx/mmCIFに注目
- ファイル形式名がその定義のページへのリンクになっている
- ファイル名をクリックでダウンロード
- [画面表示]ボタンでブラウザ内での表示

PDB形式

特徴

- テキスト形式、行ごとのデータ区切り、文字数で情報内容を定義
- 以前の公式フォーマット、段階的に廃止していく予定

利点

- コンパクト
- 読み込みやすい

欠点

- 複雑な情報は表記できない
- 巨大な構造が表記できない (分子数、原子数に制限がある)
- 不完全な情報

PDB形式

- 行の先頭が、その行の内容を定義
- ファイルの前の方は"REMARK"などのメタ情報(構造データの付随情報)
- その後、殆どは"ATOM"という単語で始まる原子座標の情報

```

ATOM 747 CE2 TRP A 178 -28.573 -65.728 10.661 1.00 16.97 C
ATOM 748 CE3 TRP A 178 -28.822 -63.356 11.119 1.00 16.48 C
ATOM 749 CZ2 TRP A 178 -29.867 -66.006 11.136 1.00 17.08 C
ATOM 750 CZ3 TRP A 178 -30.103 -63.625 11.591 1.00 15.26 C
ATOM 751 CH2 TRP A 178 -30.612 -64.946 11.602 1.00 17.35 C
ATOM 752 N SER A 179 -26.140 -61.689 13.002 1.00 17.78 N
ATOM 753 CA SER A 179 -26.798 -60.641 13.776 1.00 17.83 C
ATOM 754 C SER A 179 -26.204 -60.739 15.172 1.00 18.02 C
ATOM 755 O SER A 179 -25.941 -61.842 15.652 1.00 18.22 O
ATOM 756 CB SER A 179 -28.317 -60.833 13.796 1.00 17.95 C
ATOM 757 OG SER A 179 -28.969 -59.763 14.473 1.00 17.70 O
ATOM 758 N ALA A 180 -25.961 -59.595 15.809 1.00 18.06 N
ATOM 759 CA ALA A 180 -25.153 -59.563 17.025 1.00 18.49 C
ATOM 760 C ALA A 180 -25.458 -58.430 18.002 1.00 18.93 C
ATOM 761 O ALA A 180 -26.024 -57.401 17.637 1.00 18.74 O
ATOM 762 CB ALA A 180 -23.651 -59.561 16.664 1.00 18.19 C
ATOM 763 N SER A 181 -25.071 -58.653 19.250 1.00 19.53 N
ATOM 764 CA SER A 181 -25.019 -57.626 20.275 1.00 20.54 C
ATOM 765 C SER A 181 -23.834 -57.940 21.191 1.00 20.88 C
ATOM 766 O SER A 181 -23.350 -59.076 21.215 1.00 20.41 O
ATOM 767 CB SER A 181 -26.319 -57.596 21.080 1.00 20.99 C
ATOM 768 OG SER A 181 -26.343 -56.466 21.937 1.00 22.51 O
ATOM 769 N ALA A 182 -23.356 -56.933 21.926 1.00 21.38 N
ATOM 770 CA ALA A 182 -22.237 -57.115 22.853 1.00 22.01 C
ATOM 771 C ALA A 182 -22.223 -56.016 23.899 1.00 22.80 C
ATOM 772 O ALA A 182 -22.582 -54.872 23.607 1.00 22.97 O
ATOM 773 CB ALA A 182 -20.914 -57.139 22.106 1.00 21.87 C
ATOM 774 N CYS A 183 -21.827 -56.367 25.123 1.00 23.40 N
ATOM 775 CA CYS A 183 -21.589 -55.378 26.174 1.00 23.83 C
ATOM 776 C CYS A 183 -20.773 -55.950 27.331 1.00 23.42 C
ATOM 777 O CYS A 183 -20.618 -57.166 27.458 1.00 23.61 O
ATOM 778 CB CYS A 183 -22.904 -54.767 26.673 1.00 23.92 C
ATOM 779 SG CYS A 183 -24.149 -55.962 27.209 1.00 27.49 S
ATOM 780 N HIS A 184 -20.264 -55.056 28.172 1.00 22.82 N
ATOM 781 CA HIS A 184 -19.398 -55.411 29.288 1.00 21.91 C
ATOM 782 C HIS A 184 -20.060 -54.993 30.605 1.00 21.88 C
ATOM 783 O HIS A 184 -20.484 -53.834 30.758 1.00 21.63 O
ATOM 784 CB HIS A 184 -18.056 -54.704 29.106 1.00 21.42 C

```

PDBx/mmCIF形式

特徴

- 以前は「mmCIF」と呼ばれていた形式
- **テキスト形式**、**STAR形式**(空白区切り、表ベース)をベースにした**CIF形式**をさらに拡張したもの
- **STAR (Self-defining Text Archive and Retrieval)**: ひとつのテキストファイルにたくさんの表を記述
 - PDBx/mmCIFでは各々のエントリに約60個程度のテーブルが入っている
- **CIF**は"Crystallographic Information File"、**mm**は "macro molecule"(巨大分子)
- PDBの現在の公式フォーマット、多くのソフトウェアが対応済み

PDBx/mmCIF形式

利点

- **関係モデル**に従い情報を格納、複雑な情報を厳密に記述可能
- PDBのデータを完全に記述可能

欠点

- STAR形式は普及していない、プログラムを自作する場合は面倒
- PDB形式のみに対応するソフトウェアも
 - コンバーターはある（完全な変換は原理的に不可能）

PDBx/mmCIF形式

- "_"で始まる文字列がデータの種類を定義 (表のヘッダのようなもの)
- 一文字以上の空白がデータの区切り、改行は空白と同じとみなす
- 空白を含む文字列はクォーテーション、改行を含む文字列はセミコロンで囲む
- 表形式になっている部分は、定義された情報の繰り返し
- PDB同様、ほとんどは"ATOM"という単語で始まる原子座標の情報

```

_atom_site.label_alt_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_entity_id
_atom_site.label_seq_id
_atom_site.pdbx_ins_code
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.occupancy
_atom_site.B_iso_or_equiv
_atom_site.pdbx_formal_charge
_atom_site.auth_seq_id
_atom_site.auth_comp_id
_atom_site.auth_asym_id
_atom_site.auth_atom_id
_atom_site.pdbx_PDB_model_num
ATOM 1 N N . VAL A 1 1 ? -15.199 -55.103 39.829 1.00 39.20 ? 83 VAL A N 1
ATOM 2 C CA . VAL A 1 1 ? -16.505 -54.611 40.373 1.00 39.21 ? 83 VAL A CA 1
ATOM 3 C C . VAL A 1 1 ? -17.424 -54.194 39.214 1.00 38.72 ? 83 VAL A C 1
ATOM 4 O O . VAL A 1 1 ? -16.955 -53.671 38.199 1.00 38.68 ? 83 VAL A O 1
ATOM 5 C CB . VAL A 1 1 ? -16.309 -53.459 41.435 1.00 39.66 ? 83 VAL A CB 1
ATOM 6 C CG1 . VAL A 1 1 ? -17.629 -52.738 41.767 1.00 39.48 ? 83 VAL A CG1 1
ATOM 7 C CG2 . VAL A 1 1 ? -15.665 -54.008 42.725 1.00 39.71 ? 83 VAL A CG2 1
ATOM 8 N N . LYS A 1 2 ? -18.726 -54.430 39.395 1.00 37.86 ? 84 LYS A N 1
ATOM 9 C CA . LYS A 1 2 ? -19.740 -54.244 38.361 1.00 36.89 ? 84 LYS A CA 1
ATOM 10 C C . LYS A 1 2 ? -20.042 -52.785 38.025 1.00 35.58 ? 84 LYS A C 1
ATOM 11 O O . LYS A 1 2 ? -19.997 -51.912 38.888 1.00 35.42 ? 84 LYS A O 1
ATOM 12 C CB . LYS A 1 2 ? -21.040 -54.964 38.759 1.00 37.26 ? 84 LYS A CB 1
ATOM 13 C CG . LYS A 1 2 ? -20.899 -56.476 39.033 1.00 38.99 ? 84 LYS A CG 1
ATOM 14 C CD . LYS A 1 2 ? -20.489 -57.276 37.782 1.00 41.28 ? 84 LYS A CD 1
ATOM 15 C CE . LYS A 1 2 ? -20.485 -58.799 38.053 1.00 42.37 ? 84 LYS A CE 1
ATOM 16 N NZ . LYS A 1 2 ? -21.853 -59.410 38.036 1.00 42.26 ? 84 LYS A NZ 1
ATOM 17 N N . LEU A 1 3 ? -20.348 -52.543 36.753 1.00 34.20 ? 85 LEU A N 1
ATOM 18 C CA . LEU A 1 3 ? -20.848 -51.254 36.288 1.00 32.93 ? 85 LEU A CA 1
ATOM 19 C C . LEU A 1 3 ? -22.211 -50.980 36.925 1.00 31.89 ? 85 LEU A C 1
ATOM 20 O O . LEU A 1 3 ? -23.153 -51.754 36.756 1.00 31.63 ? 85 LEU A O 1
ATOM 21 C CB . LEU A 1 3 ? -20.948 -51.240 34.758 1.00 32.86 ? 85 LEU A CB 1
ATOM 22 C CC . LEU A 1 3 ? -19.822 -50.611 33.002 1.00 32.02 ? 85 LEU A CC 1

```

入っている情報

- 主に原子座標、原子の種類、XYZ
- 階層的な構成要素の情報
 - 原子 < 分子(残基) < ポリマー < 複合体・結晶
 - 分子はポリマー（タンパク質・核酸）と小分子単体で異なる情報種
 - 分子数、種類数は無制限
 - 各タンパク質には由来生物情報なども
- タイトル、キーワードなど、エントリの情報
- データの登録者
- 文献情報（ジャーナル、号、著者）

関係モデル

- 複数の表で複雑なデータを表現
- 例、人物のデータベースを考える
 - 一人につき、氏名は一つ、電話番号は複数
 - 表形式で電話番号一つのセルに入れない、無駄なカラムは作りたくない
- 解決策
 - 電話番号の表を別で作る、氏名の表とIDで結びつける
 - 関係データベースで用いられる様式

PDBx 辞書

- <https://mmcif.pdbj.org/>
- 辞書自体もSTAR形式で表記されている
- CIF形式ではデータの「キー名」は、「カテゴリ名.(ピリオド)アイテム名」と表記
 - 例: atom_site.Cartn_x
- 関係データベースで言うところの、テーブル名、カラム名に相当
- 500種類以上のカテゴリが定義されている



The screenshot shows the PDBx/mmCIF website interface. At the top, there is a navigation bar with links: PDBx/mmCIF, トップ, 辞書, 文書, ダウンロード, フォーマット変換, お問い合わせ. Below this is a search bar with a green '検索' button and the text '現行の辞書で検索'. The PDB logo is also visible. Below the search bar, there are tabs: 辞書, カテゴリグループ, データカテゴリ, データアイテム, Supporting Data, and a 'Browse:' link. The 'カテゴリグループ' tab is selected, showing a list of categories. The list is titled '- A -' and 'カテゴリ数 13'. The categories listed are: atom_site, atom_site_anisotrop, atom_sites, atom_sites_alt, atom_sites_alt_ens, atom_sites_alt_gen, atom_sites_footnote, and atom_type. Each category has a small icon to its right.

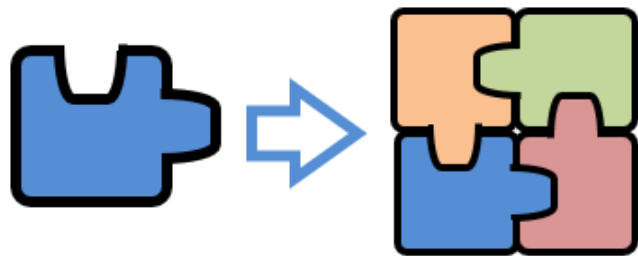
カテゴリ名	アイテム数
atom_site	1
atom_site_anisotrop	1
atom_sites	1
atom_sites_alt	1
atom_sites_alt_ens	1
atom_sites_alt_gen	1
atom_sites_footnote	1
atom_type	1

補足

1. PDBjのウェブサイトを紹介
2. 構造ビューアの使い方
3. PDBの紹介
4. データフォーマット
5. 補足 

Multi-things going multi

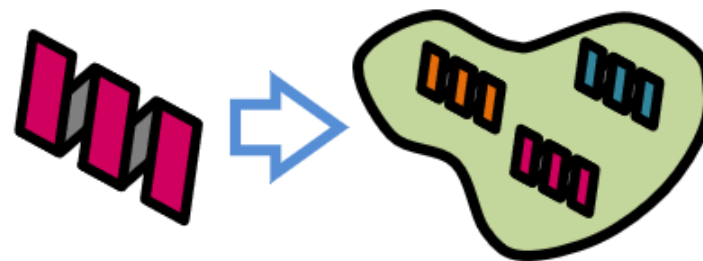
Methodology



X-ray/NMR,
alone

+ 3DEM, etc.,
hybrid

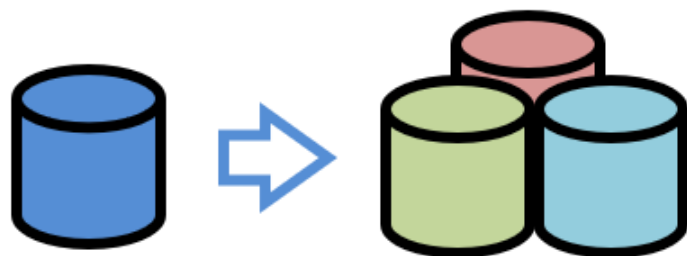
Scale



molecule

+ complex,
organelle, etc.

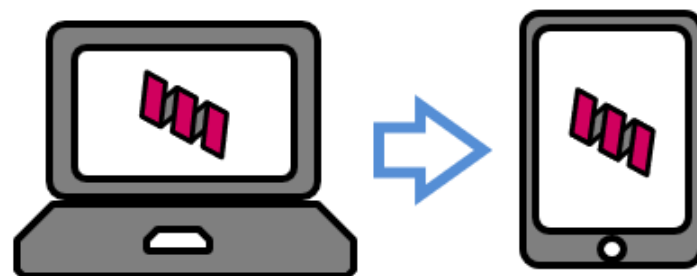
Databank



PDB

+ EMDB, BMRB,
SASBDB, etc.

Browsing device



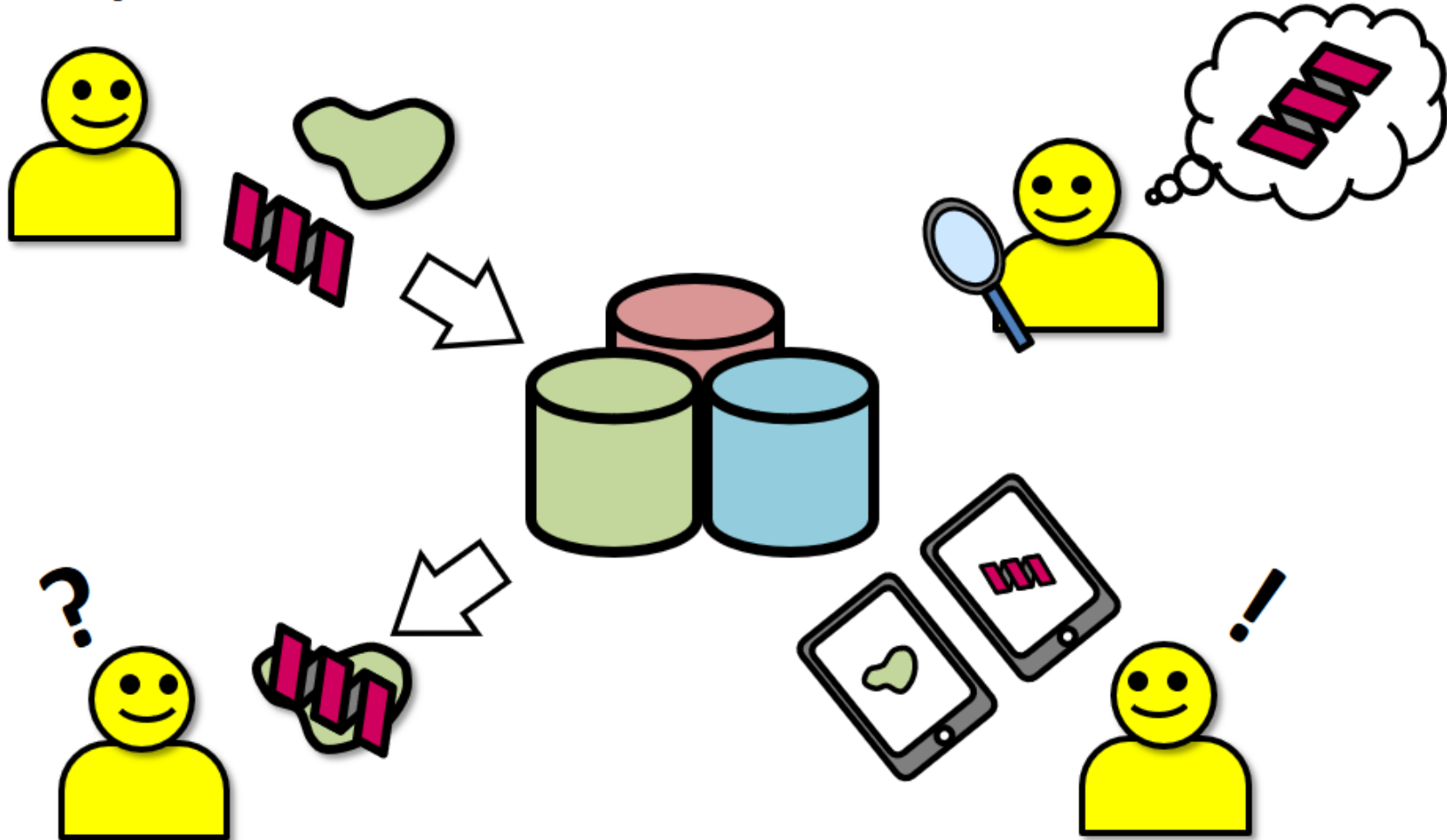
PC

+ phone,
tablet

One for the multi

Deposition

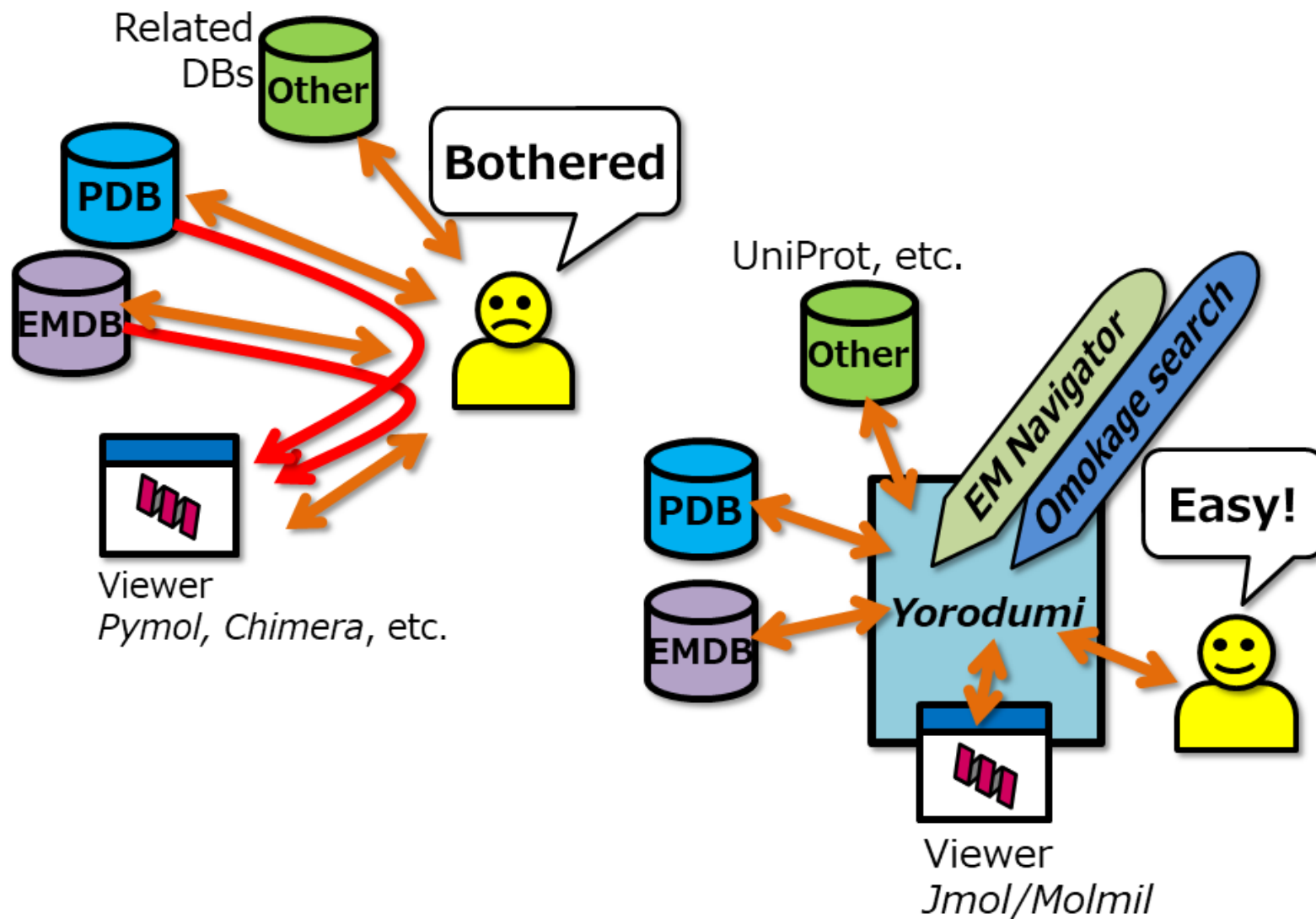
Search & browse



Comparison

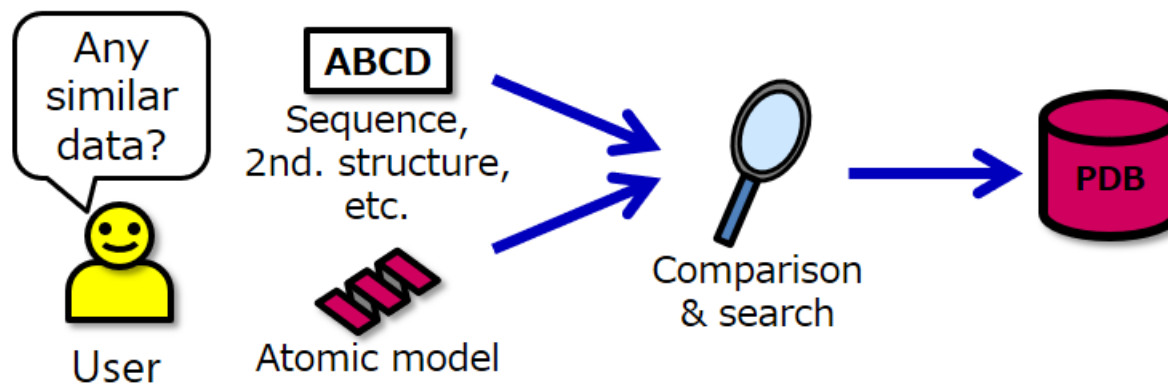
Structure viewing

「万見」とは

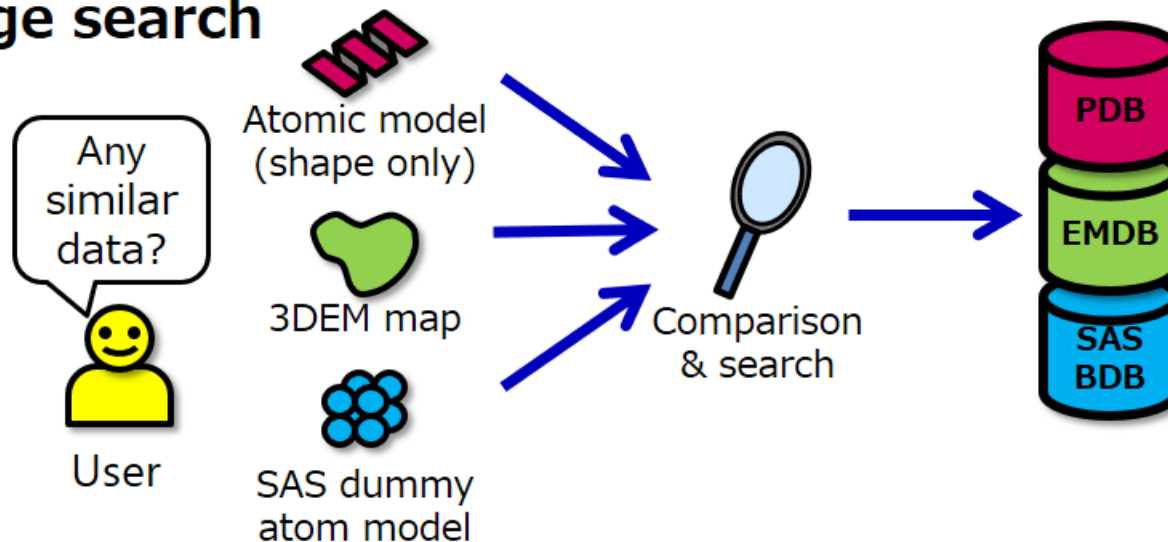


「Omokage検索」とは

Typical structure search (e.g. Structure Navigator of PDBj)



Omokage search



HOMCOS : 相同複合体の検索・モデリングサーバ



HOMCOS : 相同複合体の検索・モデリングサーバ

[\[Go to English page\]](#)  [ヘルプページ](#)

HOMCOS (HOMology modeling of COMplex Structure) は、PDBに収納されている複合体の立体構造データを利用して、分子の類似性・相同性から、構造未知の分子ペアの構造を予測するためのサーバです。アミノ酸配列や化学構造から、PDB内から複合体の立体構造が解けている他の分子を探索したり、PDB内の構造を鋳型にして複合体立体構造を予測することができます。タンパク質の配列類似性検索にはBLASTを、化合物の構造類似性検索にはKCOMBUを用いています。

サービス		クエリ(問い合わせ)	説明
結合分子の検索	 タンパク質に対する結合分子の検索	アミノ酸配列	1本のアミノ酸配列をクエリとして、それと類似したタンパク質と結合している分子を検索します
	 化合物に対する結合タンパク質の検索	化学構造	1つの化合物構造をクエリとして、それと類似した化合物と結合しているタンパク質を検索します
複合体立体構造のモデリング	 ホモ多量体のモデル	アミノ酸配列	1本のアミノ酸配列をクエリとして、そのホモ多量体の立体構造をホモロジーモデリングによって予測します
	 ヘテロ多量体のモデル	2本のアミノ酸配列	2本のアミノ酸配列をクエリとして、そのヘテロ多量体の立体構造をホモロジーモデリングによって予測します
	 化合物-タンパク質複合体のモデル	アミノ酸配列と化学構造	1本のアミノ酸配列と1つの化学構造をクエリとして、その複合体の立体構造をホモロジーモデリングによって予測します

LastUpdate of PDB:20190116, Release of UniProt:2019_01 [[Download data\(PDB:20190116\).\]](#)

<http://homcos.pdbj.org/>

詳しくは . . .

- 過去の講習会の資料
 - <https://pd bj.org/info/previous-workshop>
- HOMCOSを利用した講習
 - HOMCOSでWeb検索で講習会の資料がヒット
- UCSF-Chimeraの基本的な利用法の講習など
 - (電子顕微鏡データを利用した講習会)
 - <https://pd bj.org/workshop/20160315/suzuki.pdf>