大阪大学蛋白研セミナー
PDBj & BINDS 合同講習会
2018/02/10 @大阪大学銀杏会館

PDBjのWEBサービスの紹介

大阪大学蛋白質研究所

くどう　たかひろ
工藤　高裕

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ

1. 演習（1、2）
2. より詳細な条件で検索 - SQL検索
3. URLで直接情報取得 - RESTサービス
4. PDBデータに書かれていること
5. データの書式（PDBフォーマット、PDBx/mmCIF、PDBML）
6. 立体構造データの計測法

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

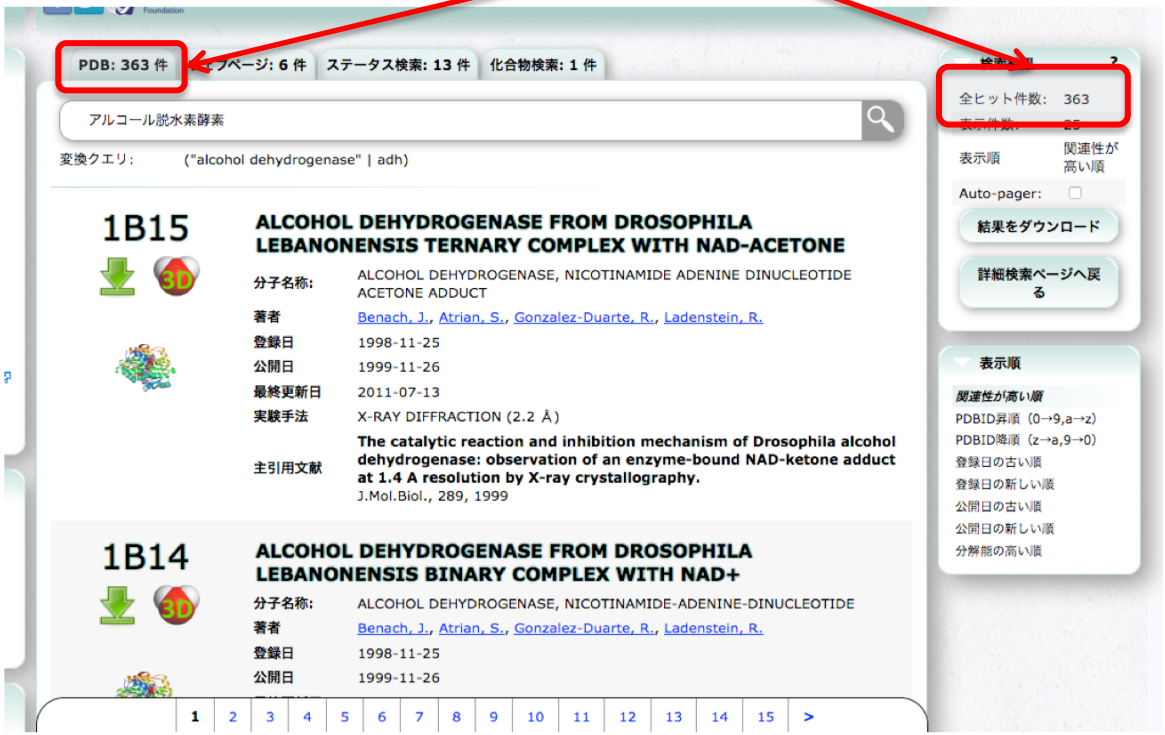
おまけ1 - 演習 -1

- キーワード「アルコール脱水素酵素」で検索
  - 何件ヒットするか？
  - 公開日が最も古いエントリーは？
  - その公開日は？
  - そのエントリーの論文は何年に発表された？
  - その分子の由来生物種は？
  - そのリガンド結合部位となる残基は？

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

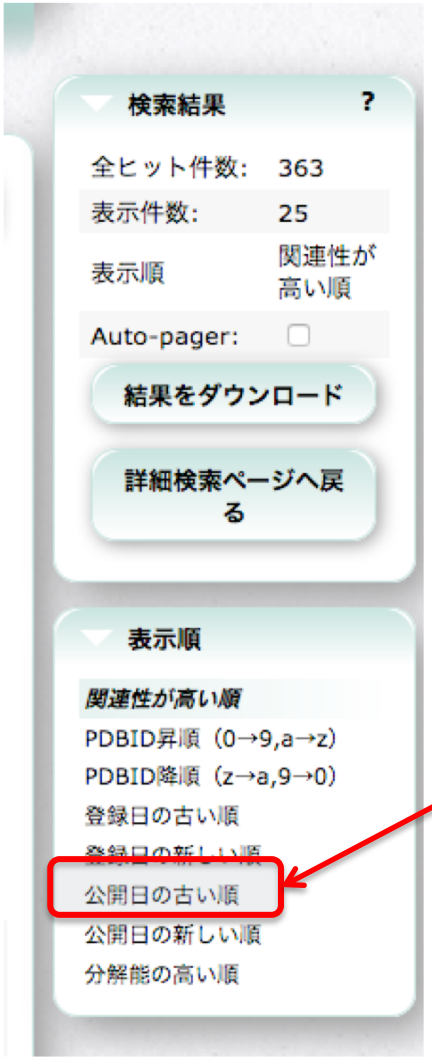
おまけ1 - 演習 -1 解答例

363件ヒット
（2018年2月14日現在）



©2018 PDBj講習会 licensed under CC 表示 4.0 国際

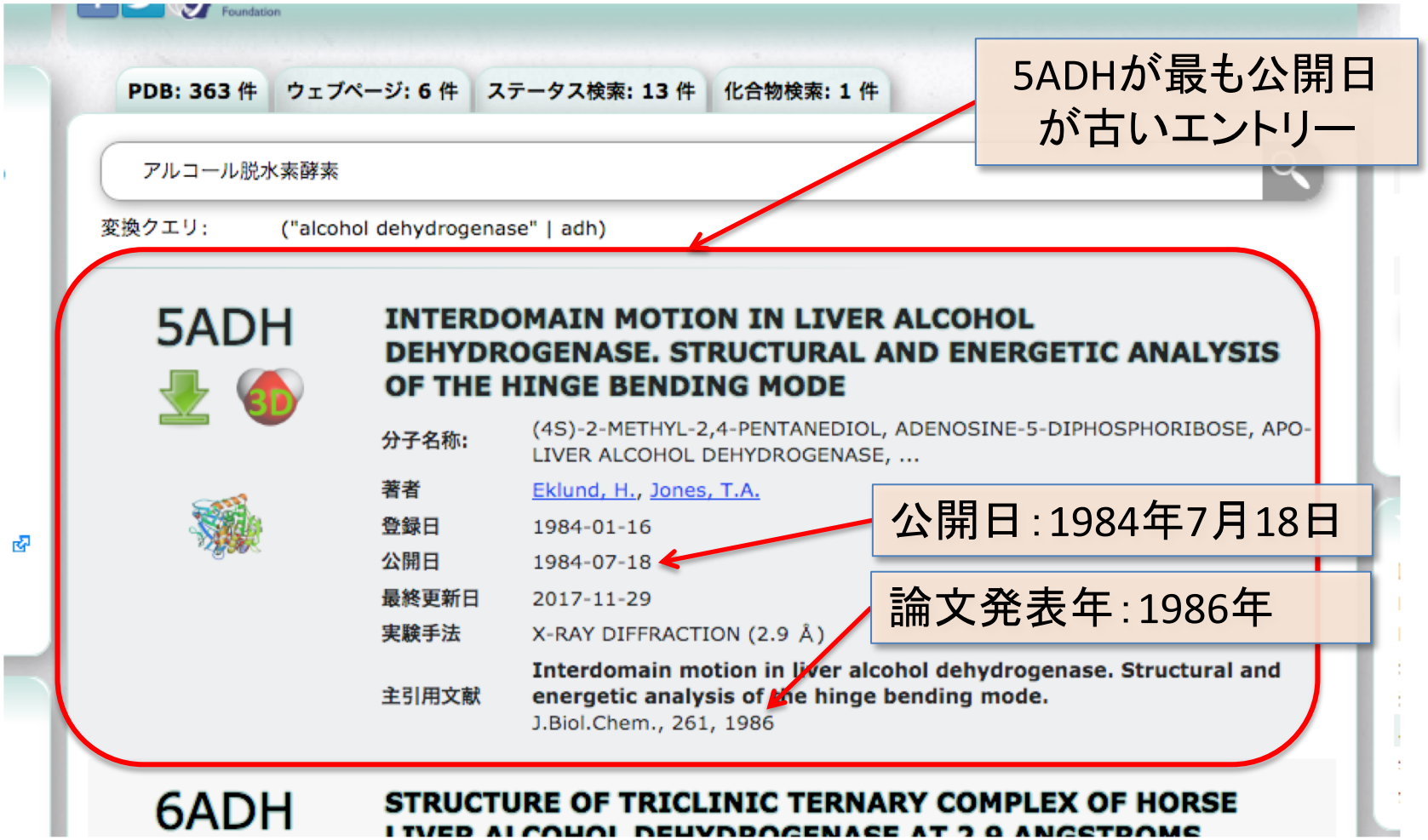
おまけ1 - 演習 -1 解答例



公開日の古い順
に並べ替え

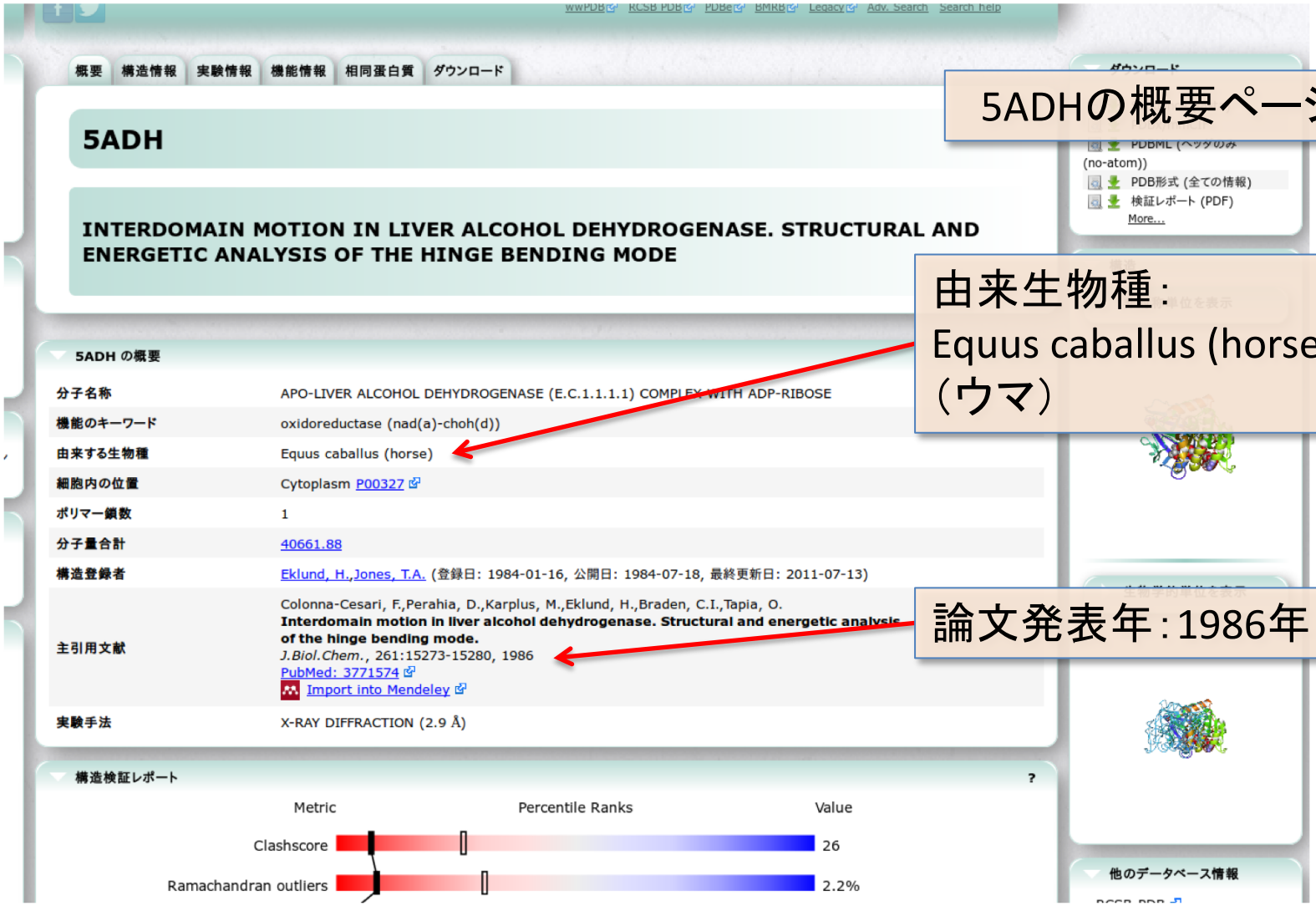
©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ1 - 演習 -1 解答例



©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ1 - 演習 -1 解答例



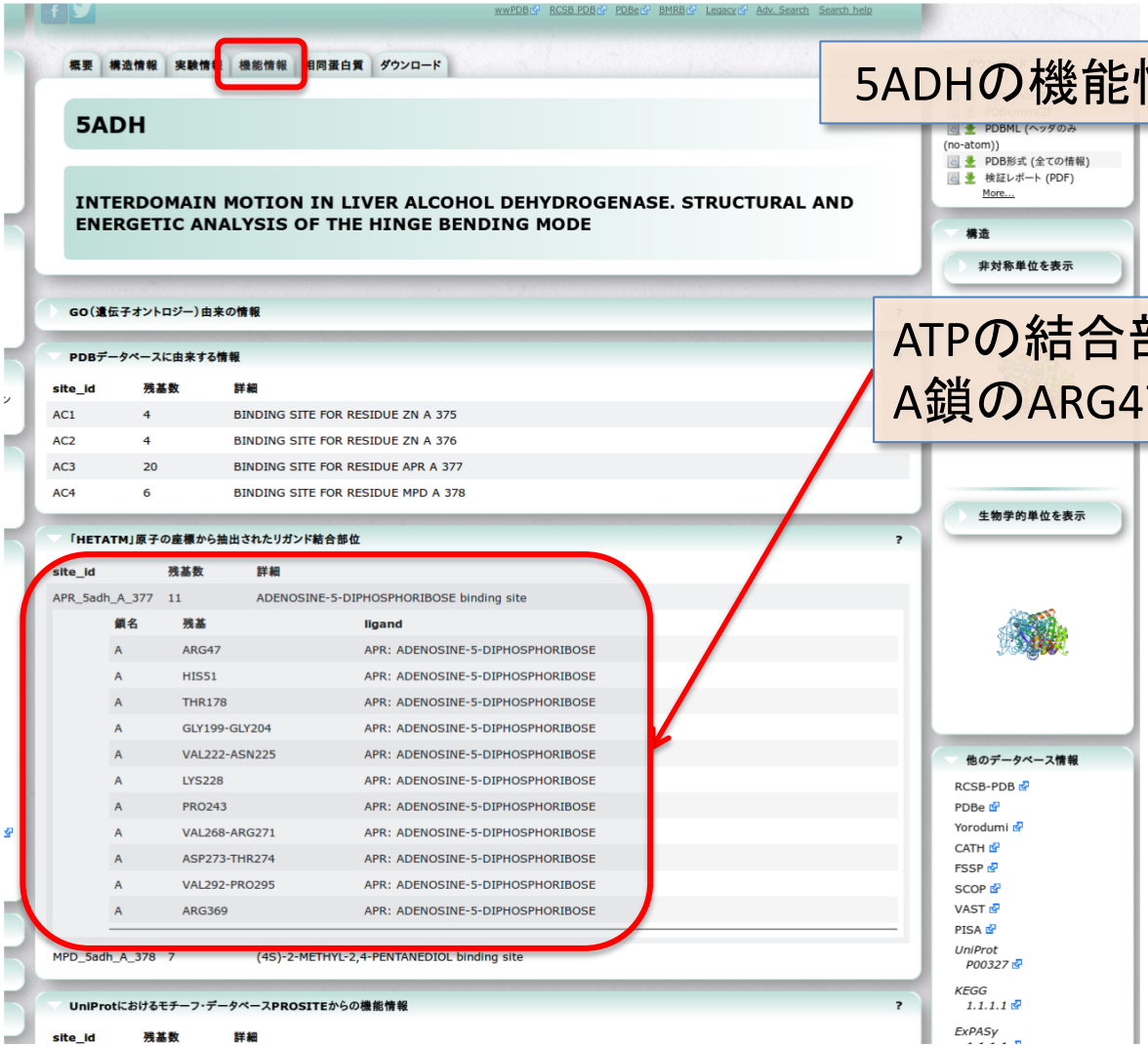
5ADHの概要ページ

由来生物種：
Equus caballus (horse)
（ウマ）

論文発表年：1986年

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ1 - 演習 -1 解答例



5ADHの機能情報ページ

ATPの結合部位は
A鎖のARG47、HIS51など

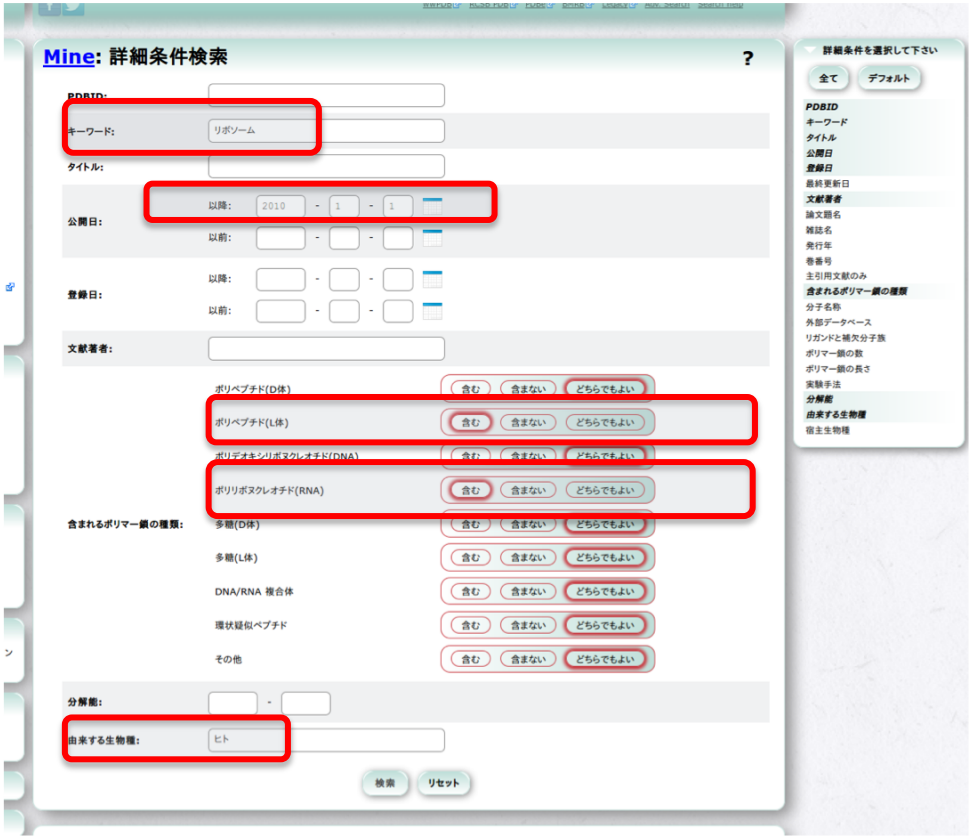
©2018 PDBj講習会 licensed under CC 表示 4.0 国際

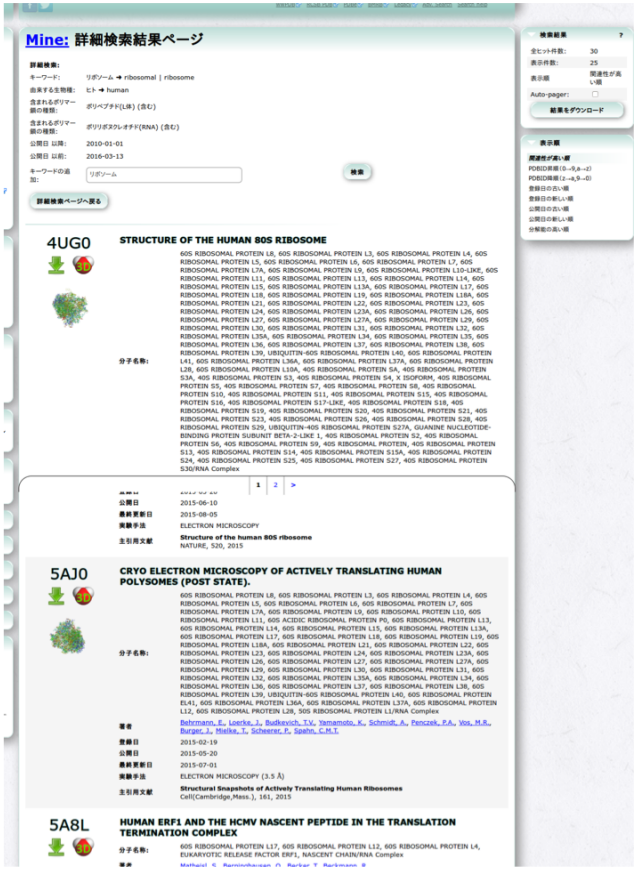
おまけ1 - 演習 -2

- 以下の条件で検索
  - キーワード「リボソーム」
  - 由来生物種に「ヒト」を含む
  - L体ペプチド鎖とRNA鎖の両方を含む
  - 2010年以降に公開
- ヒット数は？
- どんな鎖で構成されているか？

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ1 - 演習 -2





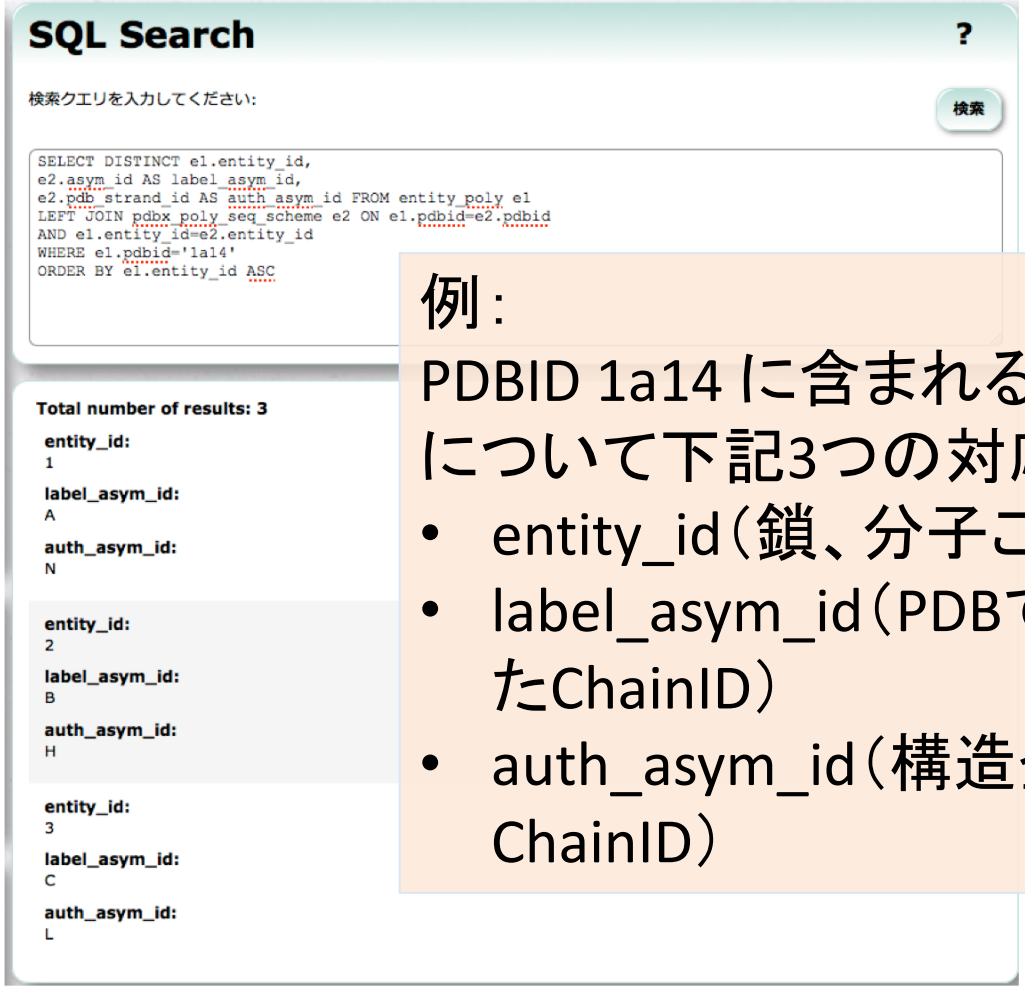
©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ2
より詳細な条件で検索 – SQL検索

「SQL Search」をクリック
→クエリ事例

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ2
より詳細な条件で検索 – SQL検索



例：

PDBID 1a14 に含まれる各ポリマー鎖のIDについて下記3つの対応情報を得る

- entity_id（鎖、分子ごとの識別ID）
- label_asym_id（PDBで系統的に定義したChainID）
- auth_asym_id（構造登録者が定義したChainID）

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ2
より詳細な条件で検索 – SQL検索

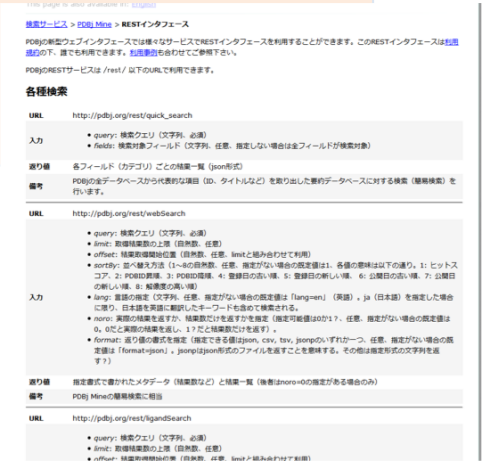
- 以下のページにさまざまなSQL検索の事例を掲載しています。

  https://pdbj.org/help/mine2-sql

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ3 - URLで直接情報取得 - RESTサービス

- URL中で**欲しい内容**や**結果取得フォーマット**を指定することができる
- **ブラウザ**のアドレス欄に入力して利用できる他、**curlコマンド**などを使ってプログラム/スクリプトで情報を取得できる
- 詳しくは下記ページを参照下さい
  http://pdbj.org/help/rest-interface



This page is also available in: English

検索サービス > PDBj Mine > **RESTインタフェース**

PDBjの新型ウェブインタフェースでは様々なサービスでRESTインタフェースを利用することができます。このRESTインタフェースは利用規約の下、誰でも利用できます。利用事例も合わせてご参照下さい。

PDBjのRESTサービスは /rest/ 以下のURLで利用できます。

各種検索

URL	http://pdbj.org/rest/quick_search
入力	• *query*: 検索クエリ（文字列、必須） • *fields*: 検索対象フィールド（文字列、任意、指定しない場合は全フィールドが検索対象）
返り値	各フィールド（カテゴリ）ごとの結果一覧（json形式）
備考	PDBjの全データベースから代表的な項目（ID、タイトルなど）を取り出した要約データベースに対する検索（簡易検索）を行います。
URL	http://pdbj.org/rest/webSearch
入力	• *query*: 検索クエリ（文字列、必須） • *limit*: 取得結果数の上限（自然数、任意） • *offset*: 結果取得開始位置（自然数、任意、limitと組み合わせて利用） • *sortby*: 並べ替え方法（1～8の自然数、任意、指定がない場合の既定値は1。各値の意味は以下の通り。1: ヒットスコア、2: PDBID昇順、3: PDBID降順、4: 登録日の古い順、5: 登録日の新しい順、6: 公開日の古い順、7: 公開日の新しい順、8: 解像度の高い順） • *lang*: 言語の指定（文字列、任意、指定がない場合の既定値は「lang=en」（英語）。ja（日本語）を指定した場合に限り、日本語を英語に翻訳したキーワードを含めて検索される。 • *noro*: 実際の結果を返すか、結果数だけを返すかを指定（指定可能値は0かと1？、任意、指定がない場合の既定値は0。0だと実際の結果を返し、1？だと結果数だけを返す）。 • *format*: 返り値の書式を指定（指定できる値はjson、csv、tsv、jsonpのいずれか一つ、任意、指定がない場合の既定値は「format=json」。jsonp(jsonp形式のファイルを返すことを意味する。その他は指定形式の文字列を返す）
返り値	指定書式で書かれたメタデータ（結果数など）と結果一覧（後者はnoro=0の指定がある場合のみ）
備考	PDBj Mineの簡易検索に相当
URL	http://pdbj.org/rest/ligandSearch
	• *query*: 検索クエリ（文字列、必須） • *limit*: 取得結果数の上限（自然数、任意） • *offset*: 結果取得開始位置（自然数、任意、limitと組み合わせて利用）

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ3 - URLで直接情報取得 - RESTサービス

例：PDBID 3bvkのデータをmmCIF形式でダウンロード

https://pdbj.org/rest/downloadPDBfile?id=3bvk&format=mmcif

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ4
PDBデータに書かれていること

PDBの各エントリーに格納されている情報の中でも基本的な項目を以下に挙げます。

PDBID	PDBに登録されている各エントリーに割り振られている一意な値で、書式は「1～9の1文字＋英数字3文字」です（例：1mbn）。
鎖識別子（Chain ID）	PDBの各エントリーには複数の分子（鎖）で構成されていることがあります。その各分子を識別するための一意な値です。 原則として「1文字の英数字」です（例：A）。 Aからアルファベット順に命名される場合が多いですが、抗体などで重鎖をH、軽鎖をLとするなど例外もあります。

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ4
PDBデータに書かれていること

残基番号 （residue number）	各高分子の各構成要素に割り振られる識別子。必ずしも1から始まるとは限らず、負の数や文字を含んだ値である場合もあります。
化合物ID （chem_comp ID）	PDBに登録されている各エントリーに登場する各種化合物（アミノ酸、ヌクレオチドなどの高分子鎖構成要素、その他リガンドなど）は全て3文字以内の識別子が割り振られています。ペプチドの残基名もこの中に含まれます。 例：GLY（グリシン）、A（アデノシン-5’-1リン酸）、ATP、G39（オセルタミビル）

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ4
PDBデータに書かれていること

原子ID （atom number）	各原子を識別するための一意な数値。
原子名 （atom name）	4文字以内で記述される原子の名称。分子（残基）内で重複がないようPDBが定めた辞書に従い命名されている。 アミノ酸の例： CA→α炭素、N→主鎖の窒素原子 CB→β位の炭素

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ5
データの書式 – 各フォーマットの特徴

フォーマット	形式	データの完全性	人が読みやすい	プログラムで処理しやすい	標準化
PDB	flat	×	◎	×	wwPDB
mmCIF	STAR	○	○	○	IUCr wwPDB
PDBML	XML	○	×	◎	W3C wwPDB

- PDBフォーマットは一部データが割愛されているが、主要データを人が読む用途には適している。自由記述されている部分がありプログラムで一貫して扱うにはあまり適さない。
- mmCIFはデータが完全で、人が読む用途にも比較的適している。パーザライブラリ(C++)はRCSBから用意されている。
- PDBMLは適当なブラウザ（PDBjの「Display」機能、Firefoxなど）がないと人が直接読む用途には向かない。一般的なXML処理ライブラリを用いることができる。

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ5 – データの書式 PDBフォーマット

80列

項目名

内容

行番号：値が複数行にまたがる時に表示

```
HEADER    OXYGEN TRANSPORT                                    06-MAY-92   1BAB
TITLE     HEMOGLOBIN THIONVILLE: AN ALPHA-CHAIN VARIANT WITH A SUBSTITUTION OF A
TITLE    2 GLUTAMATE FOR VALINE AT NA-1 AND HAVING AN ACETYLATED METHIONINE NH2
TITLE    3 TERMINUS
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: HEMOGLOBIN THIONVILLE (DEOXY) (ALPHA CHAIN);
COMPND   3 CHAIN: A, C;
COMPND   4 ENGINEERED: YES;
COMPND   5 MOL_ID: 2;
COMPND   6 MOLECULE: HEMOGLOBIN THIONVILLE (DEOXY) (BETA CHAIN);
COMPND   7 CHAIN: B, D;
COMPND   8 ENGINEERED: YES
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE   3 ORGANISM_COMMON: HUMAN;
SOURCE   4 ORGANISM_TAXID: 9606;
SOURCE   5 MOL_ID: 2;
SOURCE   6 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE   7 ORGANISM_COMMON: HUMAN;
SOURCE   8 ORGANISM_TAXID: 9606
KEYWDS    OXYGEN TRANSPORT
EXPDTA    X-RAY DIFFRACTION
AUTHOR    J.S.KAVANAUGH,A.ARNONE
```

構成分子（鎖）
(COMPND)

由来生物種
(SOURCE)

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ5 – データの書式 PDBフォーマット

原子座標部分（ATOM/HETATM）

```
...
HETATM    1  C    ACE A   0       2.779  16.023   3.370  1.00 31.31           C
HETATM    2  O    ACE A   0       2.503  15.263   2.399  1.00 44.36           O
HETATM    3  CH3  ACE A   0       1.802  17.070   3.818  1.00 36.07           C
ATOM      4  N    MET A   1       3.731  15.587   4.153  1.00 37.41           N
ATOM      5  CA   MET A   1       5.138  15.646   4.141  1.00 35.32           C
ATOM      6  C    MET A   1       5.633  16.755   3.199  1.00 36.65           C
ATOM      7  O    MET A   1       5.247  17.091   2.090  1.00 38.17           O
ATOM      8  CB  AMET A   1       5.948  14.392   4.107  0.50 41.69           C
ATOM      9  CB  BMET A   1       5.845  14.342   3.844  0.50 50.59           C
ATOM     10  CG  AMET A   1       5.190  13.146   4.444  0.50 48.68           C
ATOM     11  CG  BMET A   1       7.255  14.502   3.338  0.50 55.46           C
ATOM     12  SD  AMET A   1       5.671  12.457   6.047  0.50 38.49           S
ATOM     13  SD  BMET A   1       7.674  13.001   2.363  0.50 64.47           S
ATOM     14  CE  AMET A   1       5.513  10.705   5.729  0.50 15.99           C
ATOM     15  CE  BMET A   1       6.791  13.369   0.844  0.50 50.65           C
ATOM     16  N    GLU A   2       6.572  17.395   3.874  1.00 29.52           N
ATOM     17  CA   GLU A   2       7.303  18.521   3.312  1.00 27.19           C
ATOM     18  C    GLU A   2       8.797  18.347   3.570  1.00 21.42           C
ATOM     19  O    GLU A   2       9.243  17.888   4.644  1.00 25.36           O
ATOM     20  CB   GLU A   2       6.844  19.851   3.878  1.00 30.27           C
ATOM     21  CG   GLU A   2       6.452  20.870   2.781  1.00 42.56           C
ATOM     22  CD   GLU A   2       6.771  22.291   3.186  1.00 54.97           C
ATOM     23  OE1  GLU A   2       7.058  22.597   4.382  1.00 45.83           O
...
```

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ5 – データの書式 PDBフォーマット

書式の定義（wwPDB、英語）

http://www.wwpdb.org/documentation/file-format-content/format33/v3.3.html

書式の定義（PDBj、日本語）

http://legacy.pdbj.org/doc/pdbformats/pdb_ja.html

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ5 – データの書式 PDBフォーマットのまとめ

- １行80文字固定
- 1行に収まらない時は次の行に継続（7～10コラム目に継続を示す数字がある）
- 代表的なヘッダは：HEADER, TITLE, COMPND, SOURCE, EXPDTA, AUTHOR, JRNL, REMARK (番号つき), ATOM, HETATM
- 情報は一部だけ・将来的には廃止
- 詳しくは http://www.wwpdb.org/documentation/annotation

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ5 – データの書式 PDBx/mmCIF

可変

項目名（カテゴリ名.要素名）

値

```
data_1BAB
#
_entry.id    1BAB
#
_audit_conform.dict_name        mmcif_pdbx.dic
_audit_conform.dict_version     4.007
_audit_conform.dict_location    http://mmcif.pdb.org/dictionaries/ascii/mmcif_pdbx.dic
#
_database_2.database_id      PDB
_database_2.database_code   1BAB
#
loop_
_database_PDB_rev.num
_database_PDB_rev.date
_database_PDB_rev.date_original
_database_PDB_rev.status
_database_PDB_rev.replaces
_database_PDB_rev.mod_type
1 1994-01-31 1992-05-06 ? 1BAB 0
2 2003-04-01 ?          ? 1BAB 1
3 2009-02-24 ?          ? 1BAB 1
4 2011-07-13 ?          ? 1BAB 1
…
```

書式定義辞書
(audit_confirm)

更新情報
(database_PDB_rev)

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ5 – データの書式 PDBx/mmCIF

項目名（_カテゴリ名.要素名）　値

```
data_1BAB
#
_entry.id    1BAB
#
_audit_conform.dict_name        mmcif_pdbx.dic
_audit_conform.dict_version     4.007
_audit_conform.dict_location    http://mmcif.pdb.org/dictionaries/ascii/mmcif_pdbx.dic
#
_database_2.database_id         PDB
_database_2.database_code    1BAB
#
loop_
_database_PDB_rev.num
_database_PDB_rev.date
_database_PDB_rev.date_original
_database_PDB_rev.status
_database_PDB_rev.replaces
_database_PDB_rev.mod_type
1 1994-01-31 1992-05-06 ? 1BAB 0
2 2003-04-01 ?          ? 1BAB 1
3 2009-02-24 ?          ? 1BAB 1
4 2011-07-13 ?          ? 1BAB 1
…
```

データが1組
項目名　値

データが2組以上
loop
先に項目名を列挙
各データセットをを空白区切りで記述

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ5 – データの書式 PDBx/mmCIF

項目名（カテゴリ名.要素名）

値

各原子の情報
(atom_site)

```
…
loop
_atom_site.group_PDB
_atom_site.id
_atom_site.type_symbol
_atom_site.label_atom_id
_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_PDB_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.Cartn_x_esd
_atom_site.Cartn_y_esd
_atom_site.Cartn_z_esd
_atom_site.occupancy_esd
_atom_site.B_iso_or_equiv_esd
_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    C   C    . ACE A 1 1   ? 2.779   16.023  3.370   1.00 31.31  ? ? ? ? ? ? 0   ACE A C   1
ATOM   2    O   O    . ACE A 1 1   ? 2.503   15.263  2.399   1.00 44.36  ? ? ? ? ? ? 0   ACE A O   1
ATOM   3    C   CH3  . ACE A 1 1   ? 1.802   17.070  3.818   1.00 36.07  ? ? ? ? ? ? 0   ACE A CH3 1
…
```

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ5 – データの書式 PDBx/mmCIFの中身

各情報は「カテゴリ」に分類され、更にカテゴリはグループ化されています。以下に代表的なグループ名とその概要、所属するカテゴリの例を挙げます。

atom_group	原子の属性に関する情報 例：atom_site
chem_comp_group	低分子リガンドなどに関する情報 例：chem_comp
citation_group	文献に関する情報 例：citation

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ5 – データの書式 PDBx/mmCIFの中身

entity_group	化合物・分子に関する情報 例：entity、entity_poly（高分子）、entity_poly_seq（配列情報）、entity_src_gen（遺伝子源生物）、entity_src_nat（分子源生物）、pdbx_entity_src_syn（人工合成分子）
struct_group	構造情報 例：struct_asym（非対称単位）、struct_biol（生物学的単位）、struct_ref（他DB情報）、struct_conf（αらせん・ターン）、struct_sheet（βシート）、struct_site（活性部位・基質結合部位など）

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ5 – データの書式 PDBx/mmCIF関係リンク

書式の概要

http://mmcif.pdbj.org/docs/tutorials/mechanics/pdbx-mmcif-syntax.html

各項目の定義

http://mmcif.pdbj.org/dictionaries/mmcif_pdbx_v40.dic/Index/

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ5 – データの書式 PDBx/mmCIFのまとめ

- 文脈自由文法(STAR[Self-defining Text Archive and Retrieval]形式)
- タグ（カテゴリ名＋要素名）とアノテーションの組。
- タグは「_」で始まる。
- タグとアノテーションの組が複数ある場合には「loop_」構文で繰り返しを指定する。
- 各タグの意味はmmCIF dictionaryで定義されている。
- 色々な種類のデータは「カテゴリー」によってグループ化されている。
- 書式定義など詳しくは http://mmcif.pdbj.org/

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ5 – データの書式 PDBML

カテゴリ

要素（データセット）

要素名 ＋ *値*

```
<?xml version="1.0" encoding="UTF-8" ?>

<PDBx:datablock datablockName="1BAB"
   xmlns:PDBx="http://pdbml.pdb.org/schema/pdbx-v40.xsd"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://pdbml.pdb.org/schema/pdbx-v40.xsd pdbx-v40.xsd">

   <PDBx:atom_siteCategory>
      <PDBx:atom_site id="1">
         <PDBx:B_iso_or_equiv>31.31</PDBx:B_iso_or_equiv>
         <PDBx:Cartn_x>2.779</PDBx:Cartn_x>
         <PDBx:Cartn_y>16.023</PDBx:Cartn_y>
         <PDBx:Cartn_z>3.370</PDBx:Cartn_z>
         <PDBx:auth_asym_id>A</PDBx:auth_asym_id>
         <PDBx:auth_atom_id>C</PDBx:auth_atom_id>
         <PDBx:auth_comp_id>ACE</PDBx:auth_comp_id>
         <PDBx:auth_seq_id>0</PDBx:auth_seq_id>
         <PDBx:group_PDB>ATOM</PDBx:group_PDB>
         <PDBx:label_alt_id xsi:nil="true" />
         <PDBx:label_asym_id>A</PDBx:label_asym_id>
         <PDBx:label_atom_id>C</PDBx:label_atom_id>
         <PDBx:label_comp_id>ACE</PDBx:label_comp_id>
         <PDBx:label_entity_id>1</PDBx:label_entity_id>
         <PDBx:occupancy>1.00</PDBx:occupancy>
         <PDBx:pdbx_PDB_model_num>1</PDBx:pdbx_PDB_model_num>
         <PDBx:type_symbol>C</PDBx:type_symbol>
      </PDBx:atom_site>
…
 </PDBx:atom_siteCategory>
```

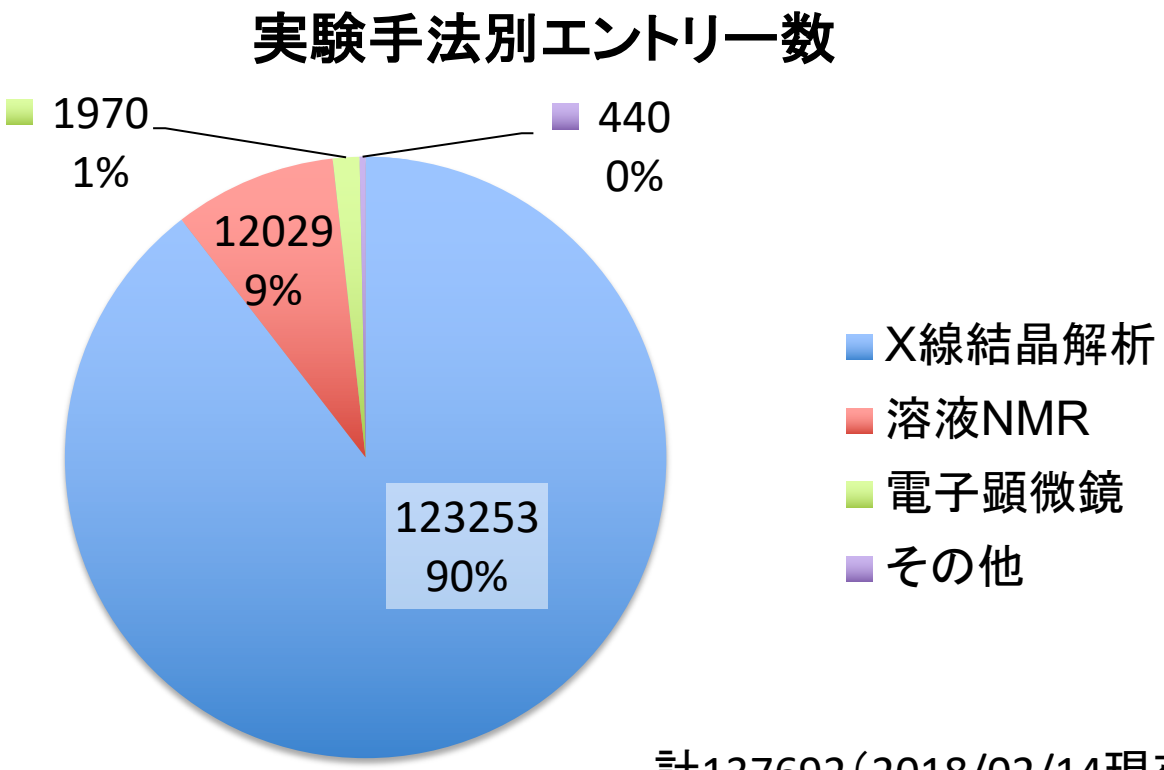
©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ5 – データの書式 PDBMLの概要

- 文脈自由文法(XML)
- PDBx/mmCIFをXMLに翻訳したもの
- XMLはより一般的なフォーマットだが、PDBx/mmCIFよりファイルサイズが大きくなる
- XML要素の型はPDBMLのXML Schemaで定義されている（タグの意味はPDBx/mmCIFから継承されている。）

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

おまけ6 - 立体構造データの計測法



計137692（2018/02/14現在）

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

X線結晶解析

分子の結晶をつくり、それにX線を当てて得られる回折像から構造情報を得る※1。
【長所】
分子量の制約がない
【短所】
結晶をつくる必要がある
結晶状態の分子が生体内で機能している状態と同じとは限らない

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

溶液NMR（核磁気共鳴）

- 分子の溶けた溶液に磁場をかける
- 質量数が奇数の原子（^{1}H、^{13}Cなど）は原子種ごとに特有な周波数で原子核が歳差運動をする
- 同じ周波数で磁場を回転させると共鳴が起きて電磁波の吸収や放出が起こる
- 各原子の結合状態などによってその周波数は少しずれる（化学シフト）
- 医療診断に用いられるMRIも同じ原理。

※1　※2

【長所】
結晶をつくる必要がない
【短所】
分子量に制約あり（～40kDa）
^{13}Cなどを含む材料が高くつく

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

電子顕微鏡（単粒子解析）

1分子の電子顕微鏡写真をたくさんとって、その平均から分子の立体構造を計算する※1 。

【長所】

結晶化はいらない

大きな分子が得意

ウイルス粒子など対称性の高い分子が得意

【短所】

一般に解像度が低い

©2018 PDBj講習会 licensed under CC 表示 4.0 国際

参照情報

※1 神田 大輔「いきなりはじめる構造生物学」学研メディカル秀潤社、2011年、
ISBN 978-4-7809-0842-8

※2 核磁気共鳴 - Wikipedia
https://ja.wikipedia.org/wiki/核磁気共鳴

©2018 PDBj講習会 licensed under CC 表示 4.0 国際