

PDBj サテライトワークショップ

中級編

PDBj Mine のSQL検索とREST API

金城玲

PDBj/大阪大学蛋白質研究所

2011-06-06

この講習の目的

- PDBのデータは「PDBファイル」だけではないことを理解する。
- PDBのアーカイブ形式の概略を理解する。
- PDBj Mine の SQL インターフェイスを活用できるようにする。

PDBのデータ形式

- PDBでは3種類の異なるデータ形式を提供している。
 - PDB format ... 昔から使われている形式
 - mmCIF ... 正式なアーカイブ形式
 - PDBML ... mmCIFをXML形式に直訳したもの

PDBフォーマットファイル

- そろそろ使うのはやめにしましょう！

ATOM	1	N	ALA	A	1	38.840	0.236	1.012	1.00	34.65	N
ATOM	2	CA	ALA	A	1	38.356	-0.999	0.357	1.00	42.26	C
ATOM	3	C	ALA	A	1	37.098	-1.547	1.056	1.00	41.25	C
ATOM	4	O	ALA	A	1	36.619	-0.946	2.028	1.00	29.44	O
ATOM	5	CB	ALA	A	1	39.398	-2.114	0.379	1.00	40.70	C
ATOM	6	N	SER	A	2	36.610	-2.666	0.495	1.00	32.67	N
ATOM	7	CA	SER	A	2	35.411	-3.244	1.202	1.00	34.90	C
ATOM	8	C	SER	A	2	35.683	-4.740	1.081	1.00	38.30	C
ATOM	9	O	SER	A	2	36.827	-5.147	0.747	1.00	28.59	O
ATOM	10	CB	SER	A	2	34.063	-2.660	0.823	1.00	24.49	C
ATOM	11	OG	SER	A	2	33.031	-3.308	1.686	1.00	20.37	O
ATOM	12	N	ALA	A	3	34.660	-5.537	1.334	1.00	35.91	N
ATOM	13	CA	ALA	A	3	34.815	-6.995	1.246	1.00	33.38	C
ATOM	14	C	ALA	A	3	33.416	-7.594	1.259	1.00	21.71	C
ATOM	15	O	ALA	A	3	32.529	-6.833	1.679	1.00	30.82	O
ATOM	16	CB	ALA	A	3	35.687	-7.414	2.433	1.00	25.44	C

mmCIFフォーマット

- これが基本データです。

```
data_1GOF
#
_entry.id      1GOF
#
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 1993-09-30 ? 1GOF 0
2 2003-04-01 ? ? 1GOF 1
3 2009-02-24 ? ? 1GOF 1
#
loop_
_audit_author.name
_audit_author.pdbx_ordinal
'Ito, N.' 1
'Phillips, S.E.V.' 2
'Knowles, P.F.' 3
```

```
ATOM      1      N   N      . ALA A 1 1      ? 38.840  0.236   1.012   1.00 34.65 ? ? ? ? ? ? 1      ALA A N      1
ATOM      2      C   CA     . ALA A 1 1      ? 38.356 -0.999   0.357   1.00 42.26 ? ? ? ? ? ? 1      ALA A CA     1
ATOM      3      C   C      . ALA A 1 1      ? 37.098 -1.547   1.056   1.00 41.25 ? ? ? ? ? ? 1      ALA A C      1
ATOM      4      O   O      . ALA A 1 1      ? 36.619 -0.946   2.028   1.00 29.44 ? ? ? ? ? ? 1      ALA A O      1
ATOM      5      C   CB     . ALA A 1 1      ? 39.398 -2.114   0.379   1.00 40.70 ? ? ? ? ? ? 1      ALA A CB     1
ATOM      6      N   N      . SER A 1 2      ? 36.610 -2.666   0.495   1.00 32.67 ? ? ? ? ? ? 2      SER A N      1
ATOM      7      C   CA     . SER A 1 2      ? 35.411 -3.244   1.202   1.00 34.90 ? ? ? ? ? ? 2      SER A CA     1
```

mmCIFの概略

Welcome to PDBj - トップページ

http://www.pdbj.org/index_j.html

English Japanese simplified Chinese traditional Chinese Korean 統計情報 ヘルプ

トップページ

データ登録 >>

ADIT: PDB Deposition
ADIT-NMR

検索 >>

Search PDB (Mine/xPSSS)
Latest Released Search
Sequence-Navigator
Structure-Navigator
SeSAW
Ligand Binding Sites (GIRAF)
EM Navigator
Search NMR Data (BMRB)
Status Search

サービス&ソフトウェア >>

JV: Graphic Viewer
万見 (Yorodumi)
Protein Globe
ASH
MAFFTash
SEALA

Structure Prediction >>

CRNPRED
Spanner
SFAS

二次データベース >>

eF-site/eF-seek/eF-surf
eProtS
ProMode / ProMode
Elastic / ProMode
Oligomer
Molecule of the Month

ダウンロード >>

PDB Archive/Snapshot
Archive

リンク集

日本蛋白質構造データバンク (PDBj: Protein Data Bank Japan)は、JST-NBDCと大阪大学の支援を受け、米国RCSB、BMRB、および欧州PDB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。

データ登録

データ登録のご案内 >>

PDB登録   NMRデータ登録 

検索

PDB検索   NMRデータ検索 

Mine日本語ページについて

PDB ID or Keyword

詳細条件検索 >>

☐ Accession number
☐ Deposition code

最新情報

2011/5/27
2011年6月6日(月)14時～16時に、第11回日本蛋白質科学会年會にて、PDBjサテライトワークショップを開催します。(詳細...)

2011/4/20
新しいPDBMLplusスキーマを公開しました。詳細はヘルプページをご覧ください。

2011/4/13
PDBアーカイブのバージョン4.0が、2011年の夏に公開されます。(wwPDBニュース...)

2011/1/18
2011年1月17日(月)、18日(火)に、「DDBJing 講習会 & PDBj 講習会 in 長浜」を長浜バイオ大学にて開催いたしました。(セミナー資料)

2011/1/12
2011年1月15日(JST)より、次週に公開が予定されているPDB ID一覧を、事前にお知らせします (wwPDBニュース...)

2010/12/22
PDBで公開済みのエントリー数が、7万件を超えました。(wwPDBニュース...)

2010/12/6
本日から、NMRによる構造の登録には、同時に化学シフトの登録が必須事項となりました。(詳細...)

Welcome to PDBj - PDBj Mine トップページ

http://service.pdbj.org/mine/index_j.html

English Japanese 統計情報 ヘルプ FAQ お問い合わせ

トップページ

データ登録 >>

ADIT: PDB Deposition
ADIT-NMR

検索 >>

Search PDB (Mine/xPSSS)
Latest Released Search
Sequence-Navigator
Structure-Navigator
SeSAW
Ligand Binding Sites (GIRAF)
EM Navigator
Search NMR Data (BMRB)
Status Search

サービス&ソフトウェア >>

JV: Graphic Viewer
万見 (Yorodumi)
Protein Globe
ASH
MAFFTash
SEALA

Structure Prediction >>

CRNPRED
Spanner
SFAS

二次データベース >>

eF-site/eF-seek/eF-surf
eProtS
ProMode / ProMode
Elastic / ProMode
Oligomer
Molecule of the Month

ダウンロード >>

PDB Archive/Snapshot
Archive

リンク集

日本蛋白質構造データバンク (PDBj: Protein Data Bank Japan)は、JST-NBDCと大阪大学の支援を受け、米国RCSB、BMRB、および欧州PDB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。

Mine PDBj Mine (PDBエントリーの検索)

日本語ページについて
PDBj Mineについて
更新情報

簡易検索

PDB ID or Keyword 1ページの表示件数

最新情報の検索

1ページの表示件数

詳細検索

SQL 検索

PDBMLの構造について → こちらの説明をご覧ください
カテゴリとアイテムについて → mmCIFディクショナリをご参照下さい
PDBMLplusの構造について → PDBMLplusスキーマファイル
PDBMLデータを処理するプログラム例 → ダウンロードページ
SQLクエリを入力して下さい → クエリの例

1ページの表示件数

PDB exchange dictionary



The screenshot shows a web browser window with the URL http://mmcif.pdb.org/dictionaries/mmcif_pdbx.dic/Index/index.html. The page features the RCSB PDB logo and a navigation bar with links: Dictionary Resources Home, Dictionary, Category Groups, Categories, Items, and Data. The 'Categories' link is circled in pink. Below the navigation bar, the page displays the title 'Dictionary Title', the version 'mmcif_pdbx.dic', the version number '1.0694', and the revision history 'history_file'. The footer includes the copyright notice '© RCSB PDB'.

Categories をクリックして
カテゴリのリストを見る

詳細は

<http://mmcif.pdb.org/>
に文書があります。

「カテゴリー」の見方

atom_site	カテゴリー名
group_PDB	ATOM/HETATM
type_symbol	“C” とか “N” とか(元素記号)
label_atom_id	原子の通し番号
label_comp_id	残基名など(化合物名)
label_asym_id	鎖のラベル
label_seq_id	残基の通し番号
Cartn_x	原子のX座標
Cartn_y	同Y座標
Cartn_z	同Z座標
.....	その他諸々

各カテゴリーは一つのテーブルと見なせる。例えば、“atom_site”カテゴリーは←の様な感じになる。

カテゴリーの見方 (2)

entity (1GOFの場合)

id	type	src_method	pdtx_description	formula_weight	pdtx_number_of_molecules	pdtx_ec
1	polymer	man	GALACTOSE OXIDASE	68579.250	1	1.1.3.9
2	non-polymer	syn	COPPER(II) ION	63.546	1	?
3	non-polymer	syn	SODIUM ION	22.990	1	?
4	non-polymer	syn	ACETIC ACID	60.052	1	?
5	non-polymer	nat	water	18.015	316	?

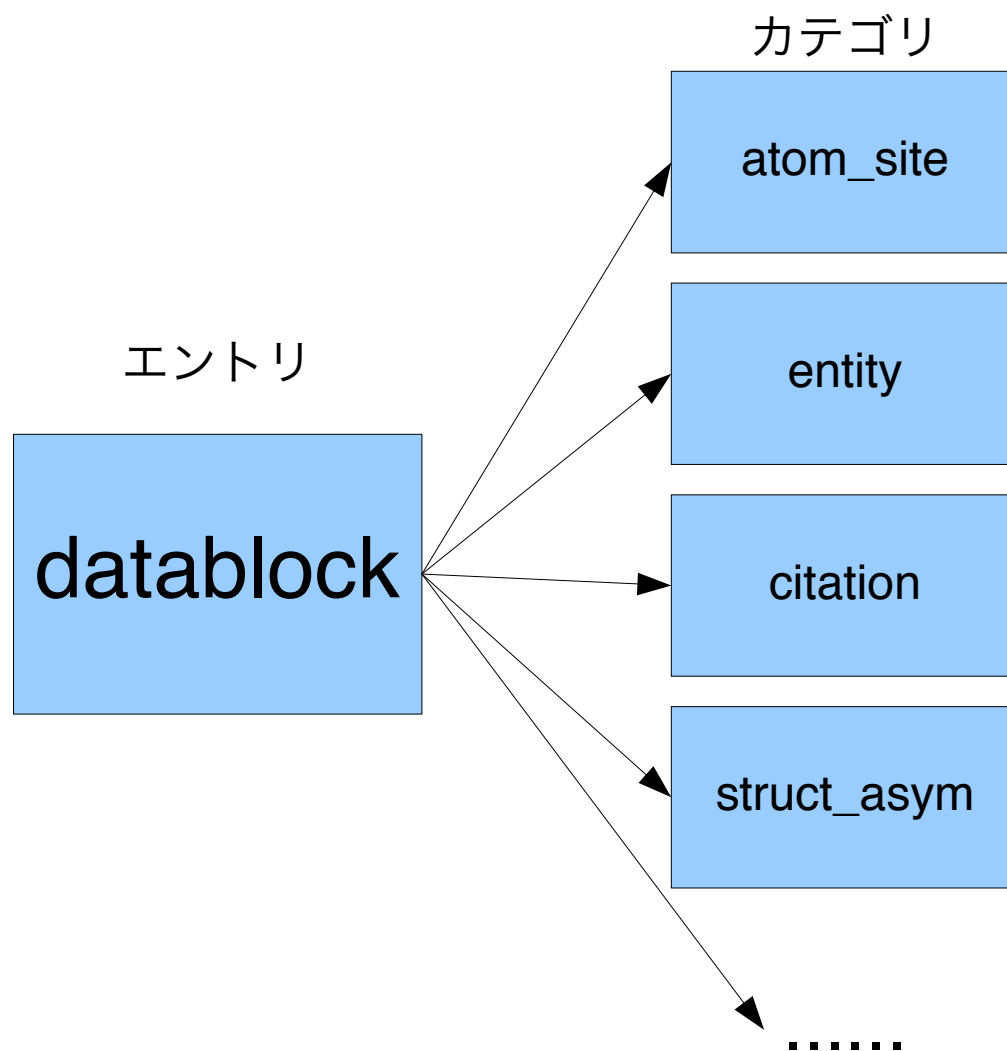
カテゴリー名=テーブル名

↓
xxx.yyy

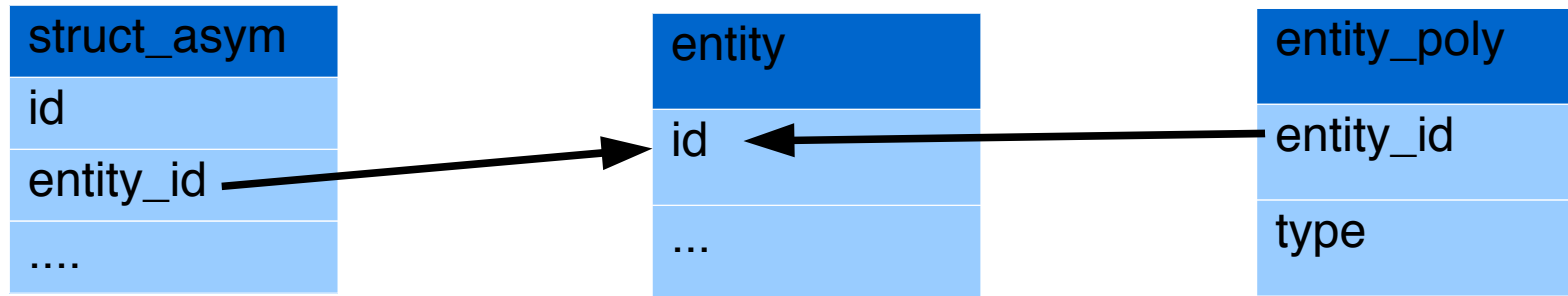
entity.id
entity.type
entity.src_method
...

↑
アイテム名=コラム名

PDBエントリはテーブルの集まり



カテゴリー間の関係



1	polymer	man	'GALACTOSE OXIDASE'	68579.250	1	?	?	?	1.1.3.9
2	non-polymer	syn	'COPPER (II) ION'	63.546	1	?	?	?	?
3	non-polymer	syn	'SODIUM ION'	22.990	1	?	?	?	?
4	non-polymer	syn	'ACETIC ACID'	60.052	2	?	?	?	?
5	water	nat	water	18.015	316	?	?	?	?

```

loop_
_struct_asym.id
_struct_asym.pdbx_blank_PDB_chainid_flag
_struct_asym.pdbx_modified
_struct_asym.entity_id
_struct_asym.details
  
```

A	N	N	1	?
B	N	N	2	?
C	N	N	3	?
D	N	N	4	?
E	N	N	4	?
F	N	N	5	?

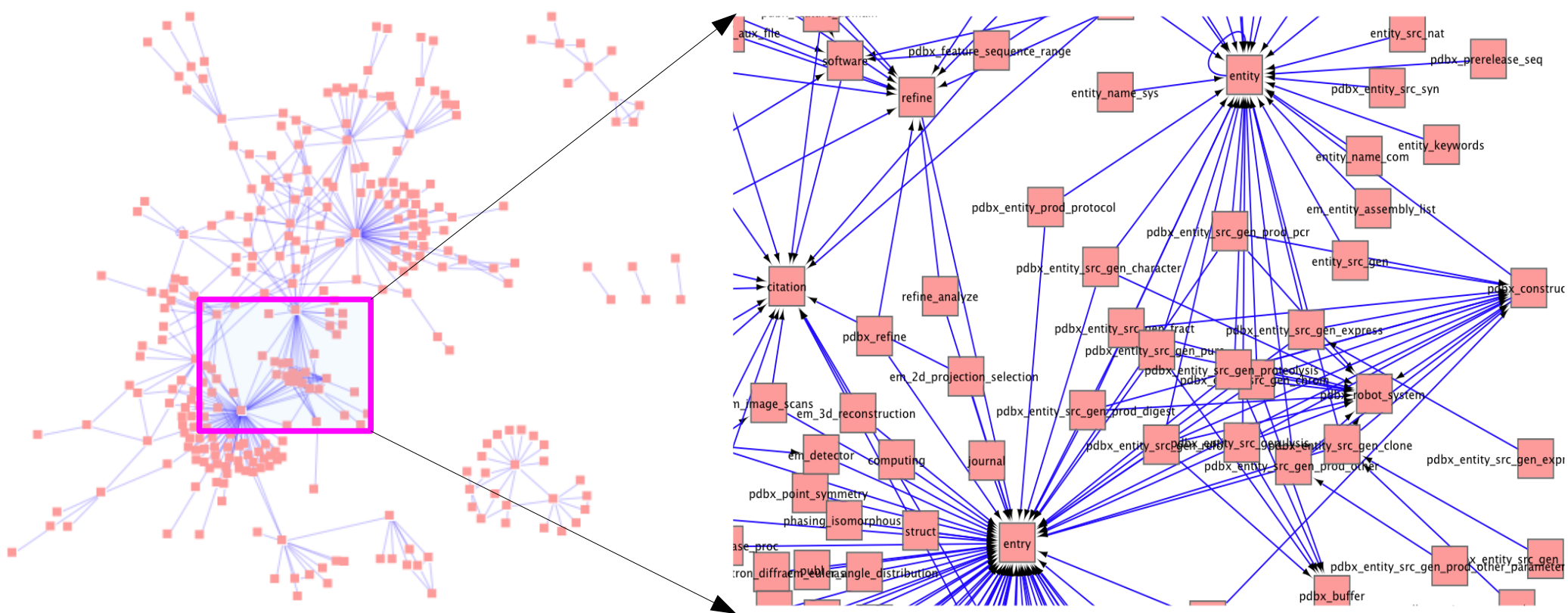
```

_entity_poly.entity_id 1
_entity_poly.type      'polypeptide(L)'
_entity_poly.nstd_linkage no
_entity_poly.nstd_monomer no
  
```

カテゴリー関係の例

- chain ID (label_asym_id) が “A” の分子名は？
 - struct_asym.id = “A” を探す。
 - struct_asym.id = “A” の行の entity_id をみる。
 - 例: struct_asym.id = “A” → struct_asym.entity_id = 1
 - entity.id = “1” を探す。
 - entity.pdbx_description をみる。
 - 例 : entity.pdbx_description = “GALACTOSE OXIDASE”

カテゴリー関係の全貌



これらの関係を利用することで、欲しいデータを一網打尽に検索できる！

エントリから関係データベースへ

- 各PDBエントリはテーブルの集まり。
- 関係データベースもテーブルの集まり。
- PDBの全エントリーを関係データベースに納めるのは単純！
 - ただし、各テーブルに PDB ID (に相当するもの) の
コラムを付け足す必要がある。
 - 1GOF の entity.id = “1”と1A00 の entity.id = “1” は異なる！

SQL超入門

- 関係データベース検索のための国際標準言語はSQL(エスキューエル)。
- 各テーブルを集合（集合論で言う意味）と見なして、取得したいデータの条件を論理演算として表現する。

SELECT コラム名, ...

FROM テーブル名, ...

WHERE テーブルに対する条件

PDBj MineでSQLを実行してみる

The screenshot shows the PDBj Mine website. The browser address bar displays 'http://service.pdbj.org/mine/index_j.html'. The page has a header with 'PDBj' logo and navigation links. A left sidebar contains various search and service links. The main content area is titled 'PDBj Mine (PDBエントリーの検索)' and includes sections for '簡易検索' (Simple Search), '最新情報の検索' (Search for latest information), '詳細検索' (Detailed Search), and 'SQL 検索' (SQL Search). The 'SQL 検索' section contains instructions and a text input field where the query 'select pdbid from xmldoc' has been entered. Below the input field are buttons for 'SQLを送信' (Send SQL) and 'リセット' (Reset). A black arrow points from the text 'ここに、SQL文を入力して「SQLを送信」をクリック！' to the 'SQLを送信' button.

Welcome to PDBj - PDBj Mine トップページ

http://service.pdbj.org/mine/index_j.html

English Japanese 統計情報 ヘルプ FAQ お問い合わせ

トップページ

データ登録 >>

- ADIT: PDB Deposition
- ADIT-NMR

検索 >>

- Search PDB (Mine/xPSSS)
- Latest Released Search
- Sequence-Navigator
- Structure-Navigator
- SeSAW
- Ligand Binding Sites (GIRAF)
- EM Navigator
- Search NMR Data (BMRB)
- Status Search

サービス&ソフトウェア >>

- JV: Graphic Viewer
- 万見 (Yorodumi)
- Protein Globe
- ASH
- MAFFTash
- SEALA

Structure Prediction >>

- CRNPRED
- Spanner
- SFAS

二次データベース >>

- eF-site/eF-seek/eF-surf
- eProtS
- ProMode / ProMode Elastic / ProMode Oligomer
- Molecule of the Month

ダウンロード >>

- PDB Archive/Snapshot Archive

リンク集

日本語ページについて PDBj Mineについて 更新情報

PDBj Mine (PDBエントリーの検索)

簡易検索

PDB ID or Keyword 検索 1ページの表示件数 16

最新情報の検索

新規公開されたエントリー 更新されたエントリー 1ページの表示件数 16

詳細検索

SQL 検索

PDBMLの構造について → [こちらの説明をご覧ください](#)
カテゴリとアイテムについて → [mmCIFディクショナリをご参照下さい](#)
PDBMLplusの構造について → [PDBMLplusスキーマファイル](#)
PDBMLデータを処理するプログラム例 → [ダウンロードページ](#)
SQLクエリを入力して下さい → [クエリの例](#)

1ページの表示件数 16

SQLを送信 リセット

ここに、SQL文を入力して
「SQLを送信」をクリック！

```
select pdbid from xmldoc
```

PDBj Mineのテーブル構成

- “xmldoc”
 - pdbid, PDBMLplus, キーワード(tsv)
- “brief_summary”
 - Advanced search で使う項目をまとめたテーブル
- “カテゴリー名”
 - 各カテゴリーに対応するテーブル
- “xpath”
 - PDBMLの個々のXML要素に対応する（上級者向け）

entityカテゴリー(1)

http://mmcif.pdb.org/dictionaries/mmcif_pdbx.dic/Categories/entity.html

```
SELECT d.pdbid , e.id
, e.pdbx_description
, e.formula_weight, e.pdbx_ec
, e.pdbx_number_of_molecules
, e.details
FROM xmldoc d , entity e
WHERE d.docid = e.docid
```

xmldocテーブルとentityテーブルで同じPDBエントリに相当するものを docid で対応づける

テーブルの結合

TableA

id	val
1	aaa
2	bbb
3	ccc

TableB

entity_id	eval
1	xxx
2	yyy

SELECT * FROM TableA, TableB

id	val	entity_id	eval
1	aaa	1	xxx
1	aaa	2	yyy
2	bbb	1	xxx
2	bbb	2	yyy
3	ccc	1	xxx
3	ccc	2	yyy

SELECT * FROM TableA, TableB
WHERE TableA.id = TableB.entity_id

id	val	entity_id	eval
1	aaa	1	xxx
2	bbb	2	yyy

entityカテゴリー(2)

http://mmcif.pdb.org/dictionaries/mmcif_pdbx.dic/Categories/entity.html

```
SELECT d.pdbid , e.id  
      , e.pdbx_description  
      , e.formula_weight, e.pdbx_ec  
      , e.pdbx_number_of_molecules  
      , e.details  
FROM xmldoc d , entity e  
WHERE d.docid = e.docid  
      AND e.pdbx_ec IS NOT NULL
```

酵素番号がついているものだけをリストする

entityカテゴリー(3)

http://mmcif.pdb.org/dictionaries/mmcif_pdbx.dic/Categories/entity.html

```
SELECT d.pdbid , e.id
, e.pdbx_description
, e.formula_weight, e.pdbx_ec
, e.pdbx_number_of_molecules
, e.details
FROM xmldoc d , entity e
WHERE d.docid = e.docid
      AND e.pdbx_ec IS NOT NULL
ORDER by e.pdbx_ec
```

酵素番号でソートする

二つのカテゴリーをつなぐ(1)

```
SELECT d.pdbid, e.id, a.id  
      , e.pdbx_description  
FROM xmldoc d, entity e  
      , struct_asym a  
WHERE d.docid = e.docid  
      AND e.docid = a.docid  
      AND e.id = a.entity_id
```

二つのカテゴリーをつなぐ(2)

```
SELECT d.pdbid, e.id, a.id  
      , e.pdbx_description  
FROM xmldoc d  
JOIN entity e ON d.docid = e.docid  
JOIN struct_asym a  
  ON e.docid = a.docid  
   AND e.id = a.entity_id
```

WHEREの代わりにJOINも使える。
(前のページのクエリと同じこと)

二つのカテゴリーをつなぐ(3)

```
SELECT d.pdbid, e.id AS entity_id
      , a.id AS asym_id
      , e.pdbx_description
FROM xmldoc d
JOIN entity e ON d.docid = e.docid
JOIN struct_asym a
  ON e.docid = a.docid
  AND e.id = a.entity_id
WHERE d.pdbid = '2v47'
```

特定のエントリーに関する情報だけを抜き出す

UniProtとの関係を探す(1)

```
SELECT d.pdbid, e.id AS entity_id  
      , r.pdbx_db_accession  
FROM xmldoc d  
JOIN entity e ON e.docid = d.docid  
JOIN struct_ref r  
      ON r.docid = e.docid  
      AND r.entity_id = e.id  
WHERE r.db_name = 'UNP'
```

UniProtとの関係を探す(2)

```
SELECT d.pdbid, e.id AS entity_id
      , a.id AS asym_id
      , r.pdbx_db_accession
FROM xmldoc d
JOIN entity e ON e.docid = d.docid
JOIN struct_ref r
      ON r.docid = e.docid
      AND r.entity_id = e.id
JOIN struct_asym a
      ON a.docid = e.docid
      AND a.entity_id = e.id
WHERE r.db_name = 'UNP'
```

UniProtとの関係を探す(3)

```
SELECT d.pdbid, e.id AS entity_id
      , a.id AS asym_id
      , p.pdbx_strand_id
      , r.pdbx_db_accession
FROM xml doc d
JOIN entity e ON e.docid = d.docid
JOIN struct_ref r
      ON r.docid = e.docid
      AND r.entity_id = e.id
JOIN struct_asym a
      ON a.docid = e.docid
      AND a.entity_id = e.id
JOIN entity_poly p
      ON p.docid = e.docid
      AND p.entity_id = e.id
WHERE r.db_name = 'UNP'
```

PubMed ID を取得する

```
SELECT d.pdbid  
      , c."pdbx_database_id_PubMed"  
FROM xml doc d  
JOIN citation c ON c.docid = d.docid  
WHERE c.id = 'primary'
```

数を数える(1)

```
SELECT d.pdbid
      , SUM(e.pdbx_number_of_molecules)
AS nmol
FROM xmldoc d
JOIN entity e ON e.docid = d.docid
JOIN entity_poly p
  ON p.docid = e.docid
  AND p.entity_id = e.id
WHERE p.type = 'polypeptide(L)'
GROUP BY d.pdbid
```

エントリに含まれる蛋白質（ポリペプチド鎖）の総数を数える。

数を数える(2)

```
SELECT d.pdbid
      , SUM(e.pdbx_number_of_molecules) AS nmol
FROM xmldoc d
JOIN entity e ON e.docid = d.docid
JOIN entity_poly p
  ON p.docid = e.docid
  AND p.entity_id = e.id
WHERE p.type = 'polypeptide(L)'
GROUP BY d.pdbid
HAVING SUM(e.pdbx_number_of_molecules) > 10
ORDER BY nmol DESC
```

エントリに含まれる蛋白質（ポリペプチド鎖）の総数が10より多いものを選ぶ

同一鎖内SS結合の数を数える

```
SELECT d.pdbid, c.ptnr1_label_asym_id
, COUNT(c.id) AS nss
FROM xml doc d
JOIN struct_conn c
  ON c.docid = d.docid
WHERE c.conn_type_id = 'disulf'
  AND c.ptnr1_label_asym_id =
c.ptnr2_label_asym_id
GROUP BY d.pdbid, c.ptnr1_label_asym_id
ORDER BY nss DESC
```

異なる鎖間SS結合の数を数える

```
SELECT d.pdbid
, c.ptnr1_label_asym_id
, c.ptnr2_label_asym_id
, COUNT(c.id) AS nss
FROM xmldoc d
JOIN struct_conn c
  ON c.docid = d.docid
WHERE c.conn_type_id = 'disulf'
  AND c.ptnr1_label_asym_id <>
c.ptnr2_label_asym_id
GROUP BY d.pdbid, c.ptnr1_label_asym_id
, c.ptnr2_label_asym_id
ORDER BY nss DESC
```

ヘリックスを列挙する

```
SELECT d.pdbid, c.beg_label_asym_id
      , c.beg_label_seq_id
      , c.end_label_seq_id
      , c."pdbx_PDB_helix_length"
FROM xmldoc d
JOIN struct_conf c
  ON c.docid = d.docid
WHERE c.conf_type_id = 'HELX_P'
      AND c."pdbx_PDB_helix_length" > 5
      AND d.pdbid = '1a00'
```

ヘリックスの個数を列挙する

```
SELECT d.pdbid, c.beg_label_asym_id
      , COUNT(c.id) AS nhelix
FROM xmldoc d
JOIN struct_conf c
  ON c.docid = d.docid
WHERE c.conf_type_id = 'HELX_P'
      AND d.pdbid = '1a00'
GROUP BY d.pdbid, c.beg_label_asym_id
ORDER BY nhelix DESC
```

練習：実験手法

- exptl

カテゴリーに実験手法(X-RAY CRYSTALLOGRAPHY, SOLUTION NMR, etc.)を記述しますが、それはどのアイテムでどのような手法が（現在のところ）許されているか調べてください。

練習：βシートの情報はどこ？

- `struct_sheet`
- `struct_sheet_range`
- `struct_sheet_order`

の3つのカテゴリーにまたがって定義されています。これらの関係を調べてください。

練習：生物学的単位は？

- `pdbx_struct_assembly`
- `pdbx_struct_assembly_gen`
- `pdbx_struct_oper_list`

の3つにまたがって定義されています。これらの関係を調べてください。

REST API

- Xpath
 - 個々のエントリのPDBMLの内容をxpath で指定して取得するサービス
- SQL
 - SQLクエリをプログラムで投げて結果を返すサービス。

XPath

<http://service.pdbj.org/mine/xpath>

を開くと簡易説明が表示されます。

<http://service.pdbj.org/mine/xpath/1gof/PDBx:datablock/PDBx:entityCategory/PDBx:entity>

[http://service.pdbj.org/mine/xpath/1gof/PDBx:datablock/PDBx:entityCategory/PDBx:entity\[@id='1'\]](http://service.pdbj.org/mine/xpath/1gof/PDBx:datablock/PDBx:entityCategory/PDBx:entity[@id='1'])

↑の末尾にさらに [/PDBx:pdbx_description](#) をつけてみる

さらにその後ろに [/text\(\)](#) をつけてみる

Xpath APIの実例

たとえば、シェルで wget コマンドを使って、

```
% wget "http://service.pdbj.org/mine/xpath/1gof//PDBx:struct/PDBx:title/text()" -O test.out
```

とすると、test.out に結果が保存される。

```
% cat test.out  
NOVEL THIOETHER BOND REVEALED BY A 1.7 ANGSTROMS CRYSTAL  
STRUCTURE OF GALACTOSE OXIDASE
```

REST SQL API

- <http://service.pdbj.org/mine/sql/> にアクセスすると簡単な説明が表示されます。
- パラメータは2つ：
 - q: SQL文そのもの
 - format: 結果の出力フォーマット
 - xml, csv, tsv が選べます。

REST SQL APIプログラム例(perl)

```
#!/usr/bin/env perl
use LWP::UserAgent;
use HTTP::Request::Common;
$sua = new LWP::UserAgent;

my $url = 'http://service.pdbj.org/mine/sql';
my $q = <<EOF
SELECT s.pdbid , p.entity_id , p.pdbx_seq_one_letter_code_can
FROM xmldoc s
JOIN entity_poly p ON p.docid = s.docid
WHERE s.pdbid like 'lm%'
EOF
;

# make request
my $req = POST($url, Content_Type => 'form-data',
               Content => [ 'format' => 'csv', 'q' => "$q" ]);

my $res = $sua->request($req);

# show response.
if ($res->is_success) {
    printf "success!\n";
    print $res->content;
} else {
    print "failed!\n";
}
```

REST SQL API使用例

```
% cat demo_sql.pl
#!/usr/bin/env perl
use LWP::UserAgent;
use HTTP::Request::Common;
$ua = new LWP::UserAgent;

my $url = 'http://service.pdbj.org/mine/sql';
...

% ./demo_sql.pl
success!
pdbid,entity_id,pdbx_seq_one_letter_code_can
1m00,1,"RFLKVKNWETDVVLTDTLHLKSTLETGCTEHICMGSIMLPSQHTRKPEDVRTKDQLFPLAKEFLDQ
YYSSIKRFGSKAHM
DRLEEVNKEIESTSTYQLKDTELIYGAKHAWRNASRCVGRIQWSKLQVFDARDCTTAHGMFNYICNHVKYATNK
GNLRSA
ITIFPQRTDGKHDFRVWNSQLIRYAGYKQPDGSTLGDPANVQFTEICIQQGWKAPRGRFDVLPLLLQANGNDPE
LFQIPP
....
```

まとめ

- mmCIF/PDBMLをつかうと、細かな情報を系統立てて理解することができる。
- PDBj Mine の SQL は mmCIF/PDBMLの構造に即して設計されている。
- SQLを（うまく）使うと、様々なPDBのデータに関する様々な情報をまとめて取り出すことができる。
- REST APIを使うと、プログラムからPDBj Mineを検索できる。

おまけ

- PDBj Mine の関係データベースは丸ごとダウンロードできます。
- PostgreSQL (<http://postgresql.jp>) を多少カスタマイズしてインストールする必要があります。
- 全体のダンプファイルと、毎週の更新の差分が提供されています。
- セットアップの仕方等は

http://www.pdbj.org/doc/help_jp.cgi?PDBj%20Mine%3aローカルインストール

を参照してください。