Introduction to PDBj services for searching and exploring the PDB
IIBMP2018, 2018-09-20

Gert-Jan Bekker
Protein Data Bank Japan
Institute for Protein Research
Osaka University
Services offered by PDBj (I)

- PDBj Mine (search & explore entries)
  - Search by PDB/EMDB/comp_id
  - Search PDB/chem_comp/BIRD by keywords
  - Advanced search for filtering
  - SQL search
- Sequence Navigator (PDB BLAST)
- Structure Navigator (structure comparison)
- Downloading structures (FTP/HTTP)
Services offered by PDBj (II)

• Deposition of structures via OneDep
  – Uniform across wwPDB members

• EM Navigator / Omokage search
  – For searching EMDB (keyword & shape)

• Promode Elastic
  – Protein dynamics database (NMA)

• Molecule of the Month
  – Japanese translation of RCSB’s MotM by David S. Goodsell
PDBj top page

- Latest release info
- Search for:
  - Entries by ID/keywords
  - Help/info pages
  - MoM
  - Services
- Newly released structures
- Latest MoM entry
- Detailed explanation of PDBj's services
- Latest news
- Various settings
Molecule of the Month (I)
Molecule of the Month (II)

225: フィターゼ (Phytase)

フィターゼは、植物性細胞内のフィチン酸を取り除く酵素で、主要な役割は植物性の食物を動物の食物の一部として利用する際にフィチン酸を分解することにある。フィチン酸は、植物の細胞壁構成成分であるフィチンとリン酸から形成される。フィターゼは、この化合物を分解することで、動物にフィチン酸が吸収されずに溶け落ちることを防ぎ、動物の栄養価を向上させる。特に、家禽や家畜の飼料にフィターゼを添加すると、フィチン酸の影響を最小限に抑え、栄養価が大幅に向上する。

フィターゼの構造は、二つの主要な部分から成り立っている。一つは、フィチン酸の分解に直接関与する酵素活性部位である。二つ目の部分は、フィチン酸の構造を認識し、酵素がフィチン酸を認識可能にする役割を果たしている。

フィターゼの役割は、農業や飼料産業において重要である。フィチン酸を分解するフィターゼの存在は、動物の飼料の利用性を向上させ、飼料の効率的な利用を可能にしている。これにより、飼料のコストを削減し、環境負荷を低減することができる。

フィターゼは、植物性飼料を効率的に利用するたために不可欠な酵素である。今後も、この酵素の応用が進むことが期待されている。
Searching by PDBj Mine (I)

• For simple searches, simply type your query in the search bar at the top of PDBj’s home page.

Searches various components of the PDBj website, including IDs of various services.
Searching by PDBj Mine (I)

- PDBj Mine can also search in Japanese (via translation of the query):
Explore PDB entries (I)

Summary

Download

Visualize (Molmil)

Links to external resources

Validation report
Explore PDB entries (II)

Structural details

Links to databases related to chain
Explore PDB entries (III)

Experimental details

1AA7

INFLUENZA VIRUS MATRIX PROTEIN CRYSTAL STRUCTURE AT pH 4.0
Explore PDB entries (IV)

Functional details

**1AA7**

**INFLUENZA VIRUS MATRIX PROTEIN CRYSTAL STRUCTURE AT PH 4.0**

<table>
<thead>
<tr>
<th>site_id</th>
<th>residue</th>
<th>detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>5</td>
<td>RNA-BINDING SITE.</td>
</tr>
</tbody>
</table>
Explore PDB entries (V)

Sequence navigator (PDB BLAST)

Perform & visualize superposition between query and hit
Also supports large structures
All major and the minor versions of the latest major version are stored
Explore PDB entries (VII)

Download & view files

<table>
<thead>
<tr>
<th>ファイル形式</th>
<th>ファイル名 (ファイルサイズ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB</td>
<td>pdb1aa7.ent.gz (67.27 KB)</td>
</tr>
<tr>
<td></td>
<td>pdb1aa7.ent.gz (283.82 KB)</td>
</tr>
<tr>
<td></td>
<td>pdb1aa7.ent.gz (4.67 KB)</td>
</tr>
<tr>
<td>PDBx/mmCIF</td>
<td>1aa7.cif.gz (81.34 KB)</td>
</tr>
<tr>
<td>PDBML</td>
<td>1aa7-noatom.xml.gz (110.35 KB)</td>
</tr>
<tr>
<td></td>
<td>1aa7-noatom.xml.gz (14.98 KB)</td>
</tr>
<tr>
<td></td>
<td>1aa7-extatom.xml.gz (70.55 KB)</td>
</tr>
<tr>
<td>PDBMLplus</td>
<td>1aa7-plus.xml.gz (112.63 KB)</td>
</tr>
<tr>
<td></td>
<td>1aa7-plus-noatom.xml.gz (17.16 KB)</td>
</tr>
<tr>
<td></td>
<td>1aa7-add.xml.gz (2.28 KB)</td>
</tr>
<tr>
<td>RDF</td>
<td>1aa7.rdf.gz (31.38 KB)</td>
</tr>
</tbody>
</table>

Download files
Display file content
File content visualization

Interactive tree format

Flat format
mmCIF format

- PDB flat-format has been deprecated
  - 99999 atoms
  - 62 chains
- Replaced by PDBx:mmCIF
  - No such limits
  - Clear formatting
mmJSON format

- mmCIF is difficult to parse
- mmJSON is easy to parse in modern languages
- Typing of data is included in the mmJSON format
Chemie search (I)

Advanced search version (including SMILES)
Chemie search (II)

Interactive 3D viewer (Molmil)

Download & display files
Chemie search (III)
BIRD/PRD search (I)
BIRD/PRD search (II)

Interactive 3D viewer (Molmil)

Download & display files
BIRD/PRD search (III)
Searching by PDBj Mine (II)

Advanced search
Mine 2 RDB Search (I)

- Very flexible way to search all metadata (i.e. excluding coordinates) of the PDB
- All PDB data is stored in an RDB (Relational Data Base)
  - Table structure is the same as in the mmCIF format
- SQL (Structured Query Language) is used to query the RDB
Mine 2 RDB Search (II)

Searches for PDB entries with Pfam ID “PF00171” and a resolution better than 2 Å.
Mine 2 RDB Search (III)

Searches for PDB entries that contain compounds with a specific InChIKey.
Mine 2 RDB Search (IV)

Searches for PDB entries that contain compounds with a specific CSD ID.
Mine 2 RDB Search (V)

Searches for PDB entries that contain compounds of a specific class (antibiotic) and a molecular weight less than 1 kD.
RDF & SPARQL

- https://rdf.wwpdb.org/
- Integrated search: https://integbio.jp/rdf/sparql
- https://integbio.jp/rdf/?view=detail&id=pdbj
REST APIs

- PDBj website = REST + JS interface
- Download files (mmCIF, PDBML, PDB, mmJSON, etc)
- Perform search queries (Quick, Advanced, SQL, etc)
- https://pdbj.org/help/rest-interface
- https://pdbj.org/help/rest-interface-examples

```python
import urllib2
try:
    import json
except:
    import simplejson as json

url = "https://pdbj.org/rest/advancedSearch"
params = "r_date_year_after=2014&r_date_month_after=1&r_date_day_after=1&method=1&res_max=1.8&sortBy=b&field=pdbid&field=struct_title&field=resolution"

json_data = urllib2.urlopen(url+"?"+params).read()
data = json.loads(json_data)

total = data["total"]
results = data["results"]
print "Total number of results:", total
print "Printing the first 10 results:" 
for i in xrange(10):
    print results[i][0], results[i][1], results[i][2]
```
Visualization using Molmil (I)

- select (select sc12, resi 12 and sidechain)
- color (color cyan, model #1 and symbol C)
- cartoon_color (cartoon_color cyan, model #1)
- set_color (set_color mycolor 12 12 12)
- show (show sticks, sidechain)
- hide (hide cartoon, model #1)
- turn (turn x, 90)
- move (move x, 90)
- fetch (fetch 1crn)
- load (load https://pdbj.org/rest/displayPromodeEfile?format=anm&id=1u
  bq_1, format=pdb)
- mplay
- mstop
- origin (origin chain A)
- set (stick_radius f, depth_cue 1/0, orthoscopic on/off,
cartoon_smooth_loops 0/2)
- bg_color (bg_color cyan)
- label (label resi 12 and sidechain, Res12)
- save (save filename.pdb, model #1 and name CA, 0,
pdb)
- viewport (viewport 500, 500)
- findseq (findseq ACDEF, model #1, my_seq)
Visualization using Molmil (II)

- Drag-and-drop multiple files
- Load in user-specified files of various formats:
  - PDB (.pdb/.ent)
  - mmCIF (.cif)
  - GROMACS (.gro, .trr, .xtc)
  - myPRESTO (.cod, .cor, .mnt)
  - CCP4 (.ccp4)
  - MOL MDL (.mdl, .mol, .sdl)
  - MOL2 (.mol2)
  - XYZ (.xyz)
  - efvet (.efvet)
  - MPBF (.mpbf)
  - MJS (.mjs)
- Embedding of Molmil
- Scripting of Molmil (.mjs & pymol-like commands)
- High quality image & movies

Bekker et al, 2017
Visualization using Molmil (III)

Numoto et al., 2018