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

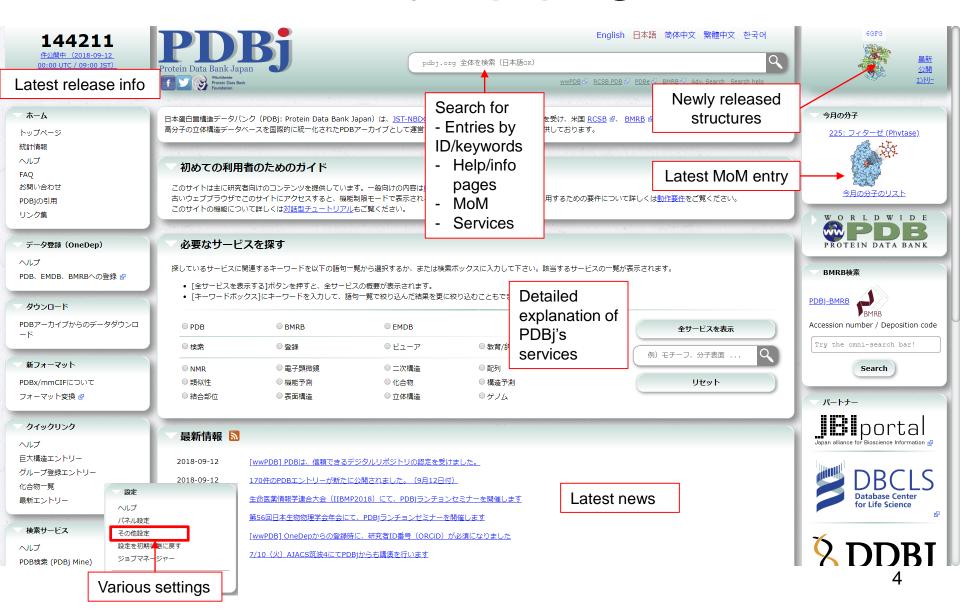
Services offered by PDBj (I)

- 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



Molecule of the Month (I)

ホーム 公開の新しい順 アルファベット順 カテゴリー別 五十音順 トップページ 今月の分子 🔝 統計情報 ヘルプ このサイトはRCSB 🗗の David S. Goodsell博士による「Molecule of the Month 🗗 を日本語に訳したものです。社会で話題となっている内容に関わる分子を蛋白質構造データバンク(PDB)から選び、機能と構造に関して解説してい ます。転載・引用については利用規約でをご覧下さい。 FAQ お問い合わせ 2018 2017 2016 2015 2014 2013 2012 2011 2010 2009 2008 2007 2006 2005 2004 2003 2002 2001 2000 PDBjの引用 リンク集 2018 225: フィターゼ (Phytase) 224: レグマイン (Legumain) データ登録 (OneDep) 223: ピエゾ1機械受容チャネル (Piezo1 Mechanosensitive Channel) 222: タンパク質とナノ粒子 (Proteins and Nanoparticles) ヘルプ PDB、EMDB、BMRBへの登録 🐶 221: ヒトパピローマウイルスとワクチン (Human Papillomavirus and Vaccines) 220: 脱ハロゲン酵素 (Dehalogenases) 219: 液胞型ATPアーゼ (Vacuolar ATPase) ダウンロード 218: EPSP合成酵素と除草剤 (EPSP Synthase and Weedkillers) PDBアーカイブからのデータダウンロ 217: オピオイド受容体 (Opioid Receptors) ページのトップへ 新フォーマット 2017 216: 生分解性プラスチック (Biodegradable Plastic) 215: アスパラギン酸カルバモイル転移酵素 (Aspartate Transcarbamoylase, ATCアーゼ) PDBx/mmCIFについて 214: キメラ抗原受容体 (Chimeric Antigen Receptors) フォーマット変換 🗗 213: サーチュイン (Sirtuins) 212: グルタチオン転移酵素 (Glutathione Transferases) クイックリンク 211: 線手機械 (Pilus Machine) ヘルプ 210: アデニンリボスイッチの動き (Adenine Riboswitch in Action) 巨大構造エントリー 209: 組織グルタミン転移酵素とセリアック病 (Tissue Transglutaminase and Celiac Disease) グループ登録エントリー 208: グルコース輸送体 (Glucose Transporters) 化合物一覧 207: 光活動性黄色タンパク質 (Photoactive Yellow Protein) 最新エントリー 206: グロビンの進化 (Globin Evolution) 205: 核膜孔複合体 (Nuclear Pore Complex) 検索サービス ページのトップへ ヘルプ 204: PD-1 (PD-1, プログラム細胞死タンパク質1) PDB検索 (PDBj Mine) 203: アミノペプチダーゼ1とオートファジー (Aminopeptidase 1 and Autophagy) PDB詳細検索 202: ジペプチジルペプチダーゼ4 (Dipeptidyl Peptidase 4) 化合物検索 (Chemie) 201: イソプレン合成酵素 (Isoprene Synthase) BMRB検索 🚱 200: 正二十面体型ウイルスの準対称性 (Quasisymmetry in Icosahedral Viruses) Sequence-Navigator 199: モネリン (Monellin) Structure-Navigator 198: β-ガラクトシダーゼ (Beta-galactosidase) EM Navigator 🐶

Molecule of the Month (II)

ホーム

トップページ

統計計書報

ヘルプ

FAQ

お問い合わせ

PDBjの引用

リンク集

データ登録 (OneDep)

ヘルプ

PDB、EMDB、BMRBへの登録 🔮

ダウンロード

PDBアーカイブからのデータダウンロード

新フォーマット

PDBx/mmCIFについて フォーマット変換 🖟

クイックリンク

ヘルプ

巨大構造エントリー

グループ登録エントリー

化合物一覧

最新エントリー

検索サードス

ヘルプ

PDB検索 (PDBj Mine)

PDB詳細検索

化合物検索 (Chemie)

BMRB検索 🚱

Sequence-Navigator

Structure-Navigator

EM Navigator @

225: フィターゼ (Phytase)

このページはRCSB 🖤の David S. Goodsell博士による「<u>Molecule of the Month</u> 🗗 2018年9月の記事を日本語に訳したものです。転載・引用については<u>利用規約</u>をご覧下さい。 「今月の分子」一覧に戻る 🖊 この記事のRCSBオリジナルサイト(英語)を見る 🗗

: 翻訳 工藤高裕 (PDBi)

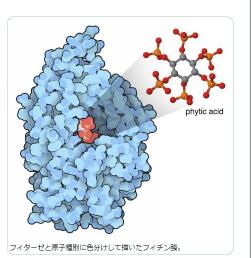
競争がある生態系では、使えるリンの量によって成長に制約が生じるのは珍しいことではなく、生き物はそれを分け合う方法を見つけてきた。我々は食物からリンを摂取しているが、それはヌクレオチドやリン酸イオンなどの役に立つ分子の中にうまく取り込まれている。植物は手元にめぐってくる資源だけでやっていく必要があり、 フィチン酸 (phytic acid、専門的にはイノシトール6リン酸 inositol hexakisphosphate)と呼ばれる珍しい分子にリンを貯蔵していることがよくある。この分子は安定で、有用な分子へと分解する適当な辞素がない限りほとんど消化されない。このことは農業において特別な問題を引き起こす。

飼料に含まれるリン

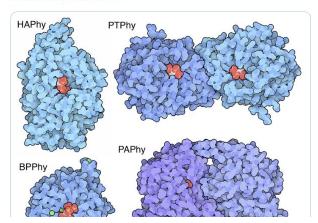
家畜には大豆など植物中心の飼料が与えられることがよくあるが、これにはフィチン酸が豊富に含まれている。多くの胃を持つウシは、フィチン酸を分解し栄養として使えるリン酸分子にしてくれる細菌を持っている。一方、ブタやエワトリは胃を一つしか持っていなくて、ほとんどのフィチン酸は素適りしてしまう。この制約によって2つの問題が生じる。一つは動物の生育に他のリン酸源が必要になることである。そしてもう一つは、肥料に含まれるフィチン酸が環境中に放出されると有毒な藻類が繁茂してしまうなどのの問題を引き起こしてしまうことである。この二つの問題を解決するため、家畜の飼料には酵素のフィターゼ(phytase)がよく添加されている。この酵素は動物の胃でフィチン酸を分解してくれる。

フィターゼ

フィターゼは小さく酸に対し安定な酵素で、フィチン酸から徐々にリン酸基を取り出す。飼料への添加物としてよく使われるのは主に次の2種類である。互いに似ているが一方は細菌由来でここに示すのはPDBエントリー 1dkg の構造、もう一つはカビ由来でここに示すのはPDBエントリー 1hp の構造である。バイオテクノロジーの業界ではより熱に安定で活性の高いフィターゼを作り出し、より大規模に農業で利用できるようにする懸命な努力が続けられている。



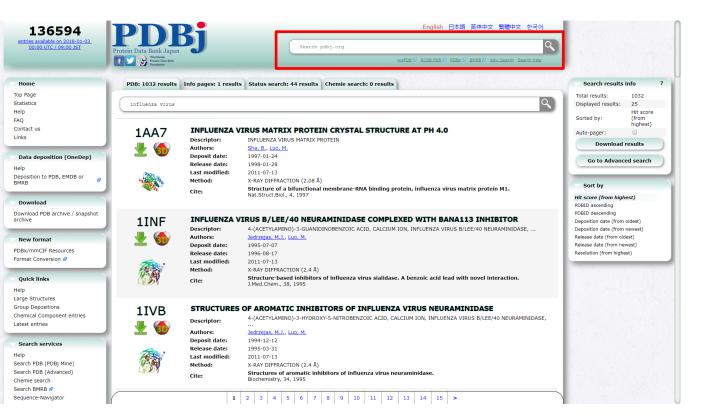
さまざまなタイプのフィターゼ



非常に多くの生き物が植物を食べるので、さまざまな種類のフィターゼが進化してきたことは別に驚くようなことではない。研究者は 横造に基づいてこれらを4つのグループに分類した。ヒスチジン酸フィターゼ(Histidine Acid Phytase、HAPhy)は飼料に添加される フィターゼ(に似ていて、反応を行うヒスチジンを持っている。 β -プロペラフィターゼ(Beta-Propeller Phytase、BPPhy、PDBエント リー $3 \underline{\text{amr}}$)はその珍しい折りたたみ様式から名づけられた。PDBエントリー $3 \underline{\text{mmj}}$ のようなPTPhyフィターゼはタンパク質のチロシン脱リン酸化酵素(Protein Tyrosine phosphatase)と似ている。紫色酸性フィターゼ(Purple Acid Phytase、PAPhy、PDBエントリー $4 \underline{\text{kbp}}$)は金属イオンを含んでおり、紫色をしている。

Searching by PDBj Mine (I)

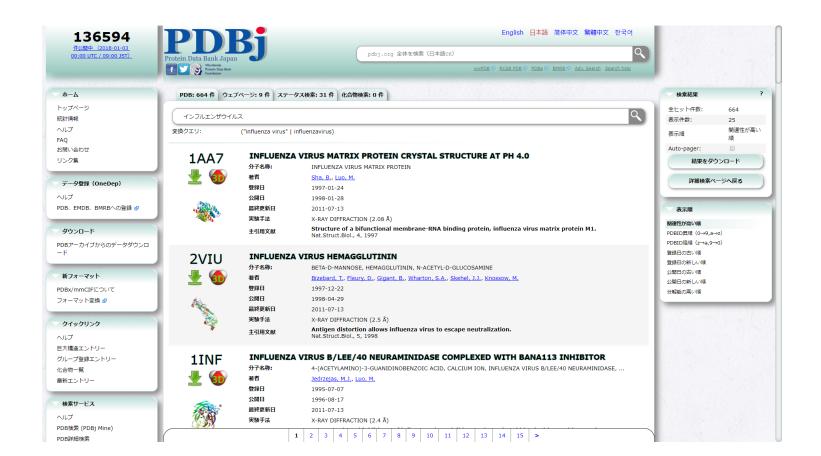
 For simple searches, simply type your query in the search bar at the top of PDBj's home page



Searches various components of the PDBj website, including IDs of various services

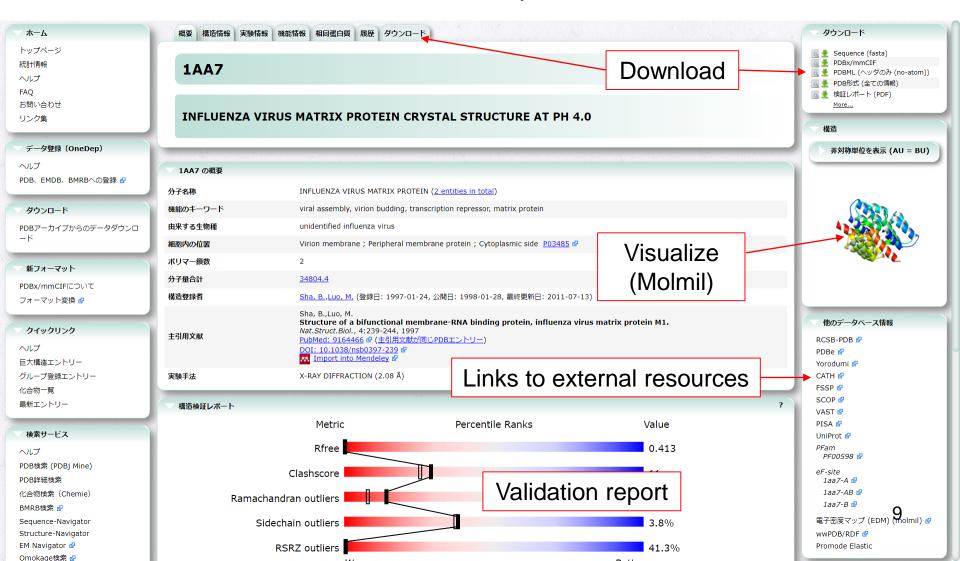
Searching by PDBj Mine (I)

 PDBj Mine can also search in Japanese (via translation of the query):



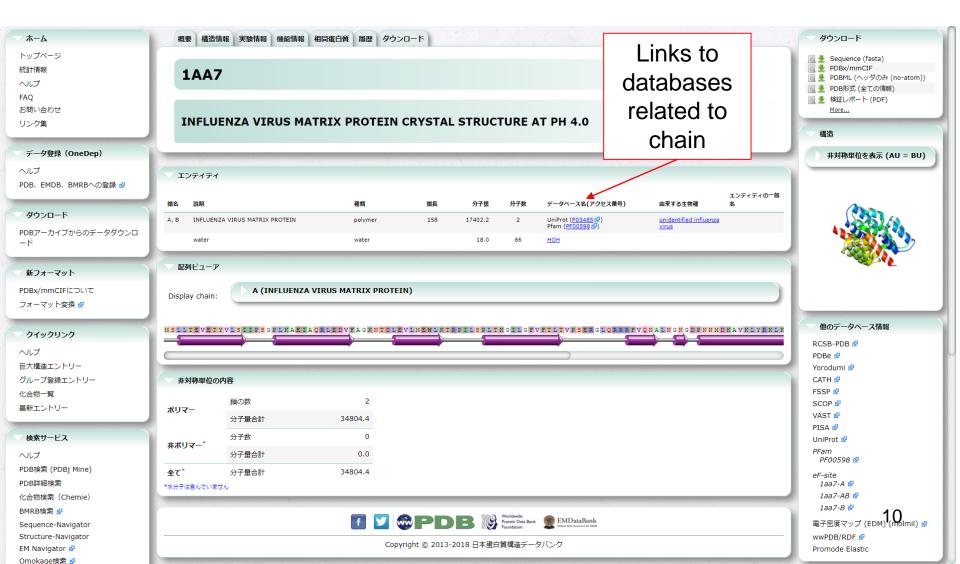
Explore PDB entries (I)

Summary



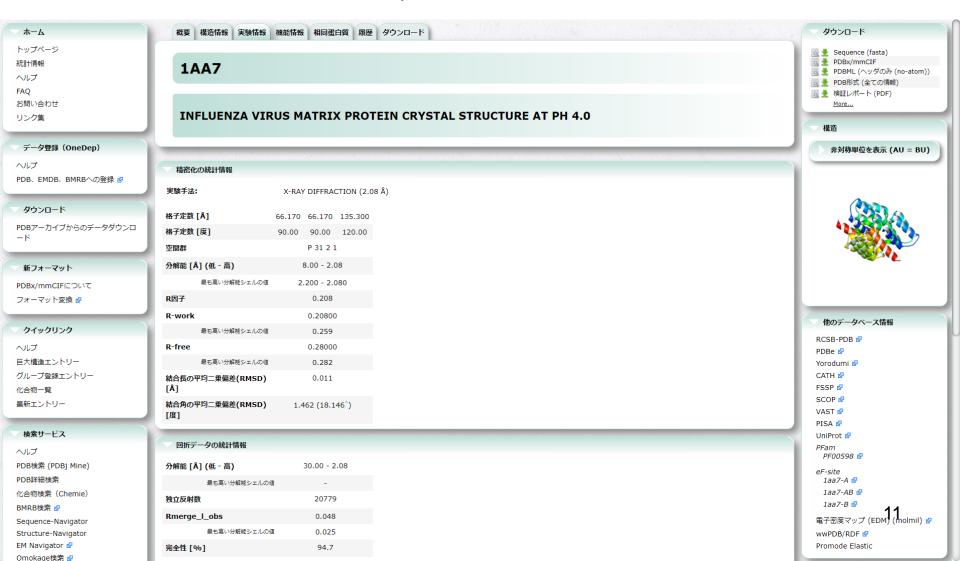
Explore PDB entries (II)

Structural details



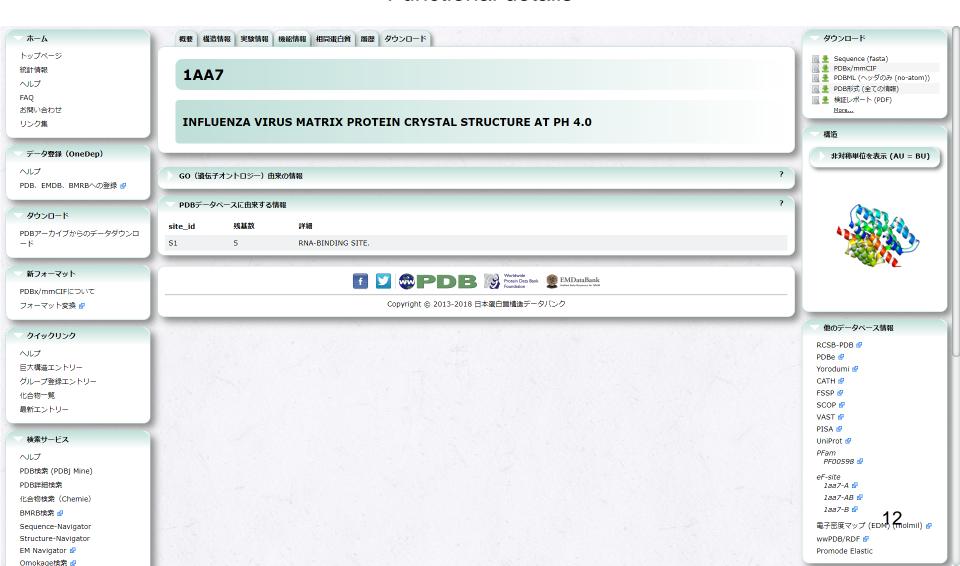
Explore PDB entries (III)

Experimental details



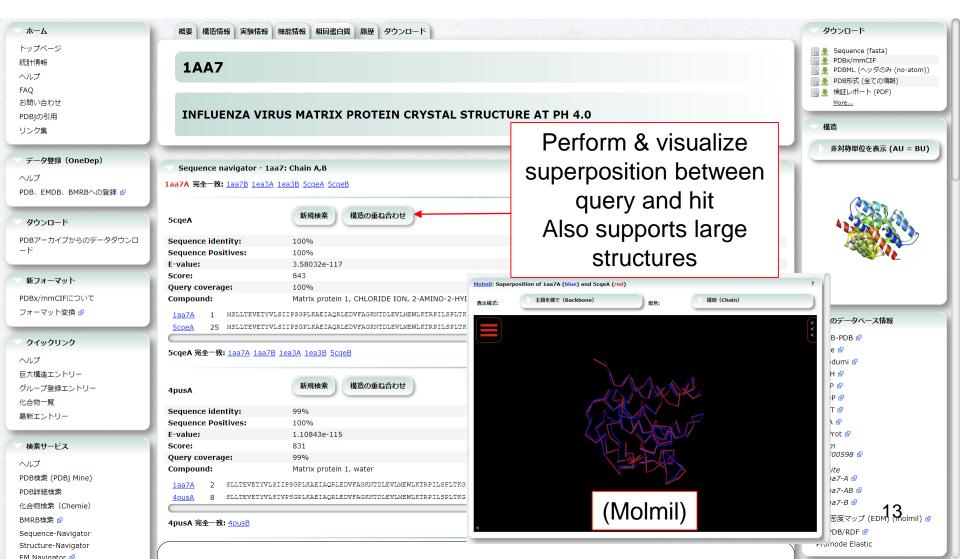
Explore PDB entries (IV)

Functional details



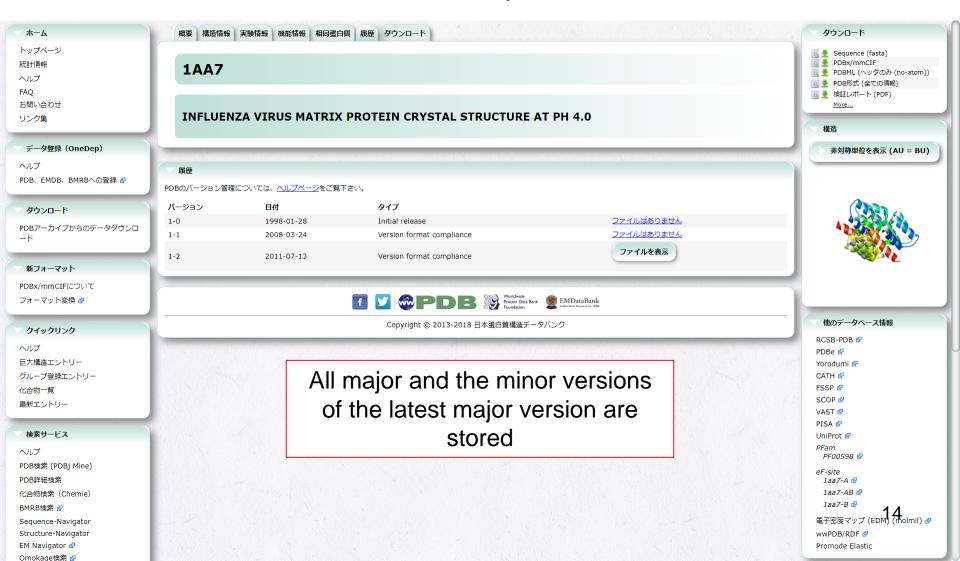
Explore PDB entries (V)

Sequence navigator (PDB BLAST)



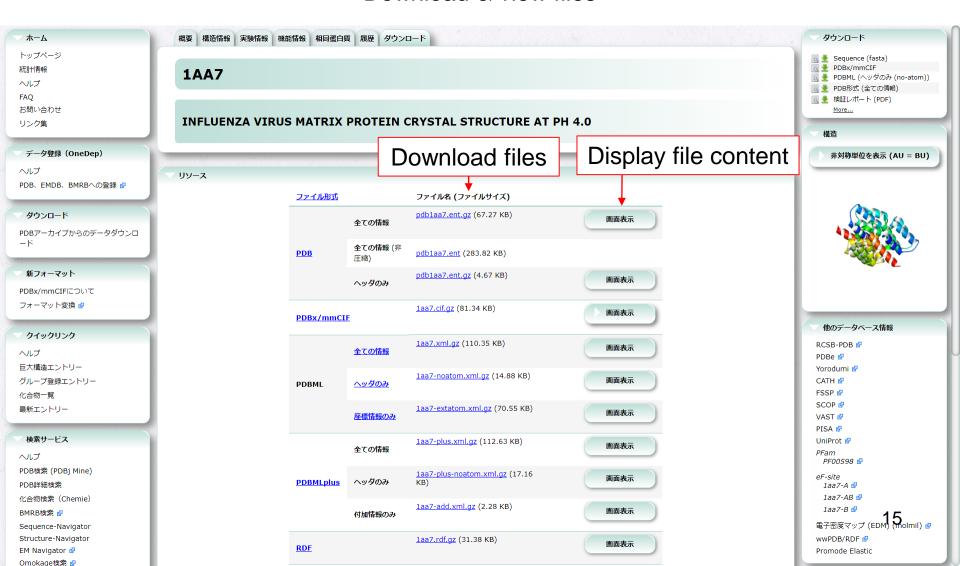
Explore PDB entries (VI)

History

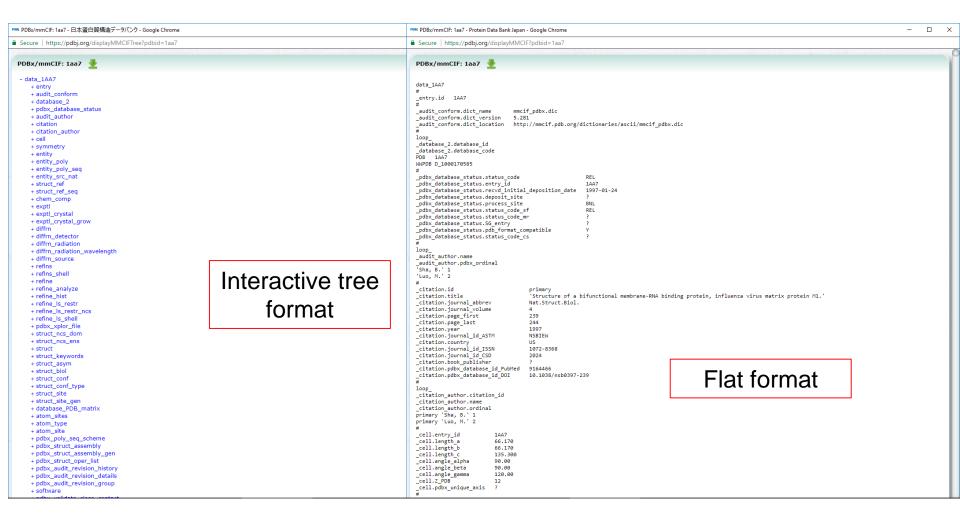


Explore PDB entries (VII)

Download & view files



File content visualization



mmCIF format

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

```
data_1AA7
entry.id 1AA7
audit conform.dict name
                              mmcif pdbx.dic
audit conform.dict version
audit conform.dict location
                              http://mmcif.pdb.org/dictionaries/ascii/mmcif_pdbx.dic
loop_
_database_2.database_id
database 2.database code
PDB 1AA7
WWPDB D 1000170585
_pdbx_database_status.status_code
_pdbx_database_status.entry_id
                                                      1997-01-24
_pdbx_database_status.recvd_initial_deposition_date
pdbx_database_status.deposit_site
_pdbx_database_status.process_site
_pdbx_database_status.status_code_sf
pdbx database status.status code mr
pdbx database status.SG entry
pdbx database status.pdb format compatible
_pdbx_database_status.status_code_cs
audit author.name
audit author.pdbx ordinal
'Sha, B.' 1
'Luo, M.' 2
citation.id
citation.title
                                    'Structure of a bifunctional membrane-RNA binding p
citation.journal abbrev
                                    Nat.Struct.Biol.
_citation.journal_volume
citation.page first
                                    239
__citation.page_last
                                    244
                                    1997
_citation.year
citation.journal id ASTM
                                    NSBIEW
citation.country
citation.journal id ISSN
                                    1072-8368
citation.journal id CSD
                                    2024
citation.book_publisher
citation.pdbx database id PubMed
                                   9164466
_citation.pdbx_database_id_DOI
                                    10.1038/nsb0397-239
_citation_author.citation_id
citation author.name
citation author.ordinal
primary 'Sha, B.' 1
primary 'Luo, M.' 2
_cell.entry_id
                         66.170
cell.length a
cell.length b
                         66.170
_cell.length_c
                        135.300
                         90.00
_cell.angle_alpha
_cell.angle_beta
                         90.00
_cell.angle_gamma
                         120.00
cell.Z PDB
```

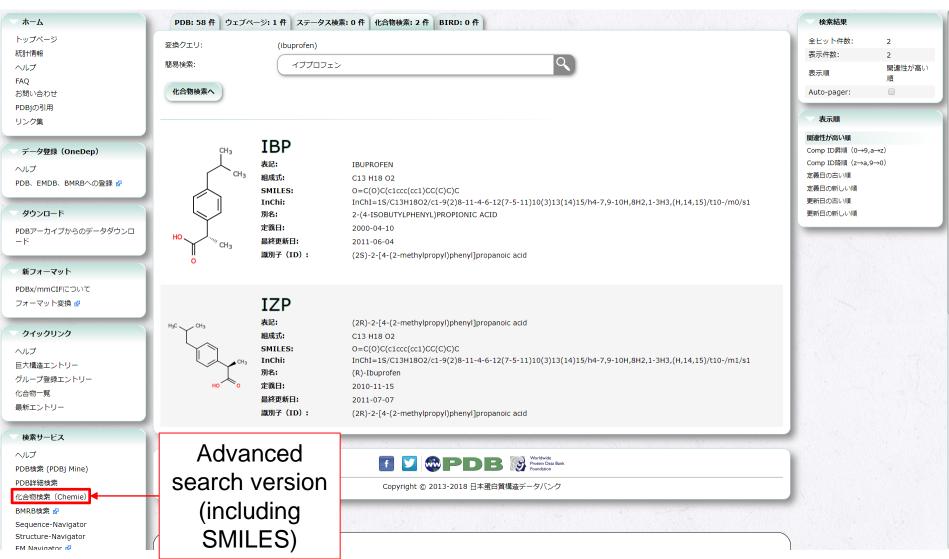
_cell.pdbx_unique_axis

mmJSON format

- mmCIF is difficult to parse
- mmJSON is easy to parse in modern languages
- Typing of data is included in the mmJSON format

```
"data 1AA7": {
         "entry": {
           "id": ["1AA7"]
         "audit conform": {
           "dict name": ["mmcif pdbx.dic"],
           "dict version": ["5.281"],
           "dict location": ["http://mmcif.pdb.org/dictionaries/ascii/mmcif_pdbx.dic"]
         "database 2": {
           "database id": ["PDB", "WWPDB"],
           "database code": ["1AA7", "D 1000170585"]
14
15
         "pdbx database status": {
16
           "status code": ["REL"],
           "entry id": ["1AA7"],
17
           "recvd initial deposition date": ["1997-01-24"],
19
           "deposit site": [null],
           "process site": ["BNL"],
21
           "status code sf": ["REL"],
           "status code mr": [null],
           "SG entry": [null],
24
           "pdb format_compatible": ["Y"],
           "status code cs": [null]
26
27
         "audit author": {
           "name": ["Sha, B.", "Luo, M."],
29
           "pdbx ordinal": [1, 2]
         "citation": {
           "id": ["primary"],
           "title": ["Structure of a bifunctional membrane-RNA binding protein, influenz
34
           "journal abbrev": ["Nat.Struct.Biol."],
           "journal volume": ["4"],
36
           "page first": ["239"],
           "page last": ["244"],
           "year": [1997],
           "journal id_ASTM": ["NSBIEW"],
39
40
           "country": ["US"],
41
           "journal id ISSN": ["1072-8368"],
           "journal id CSD": ["2024"],
43
           "book publisher": [null],
44
           "pdbx database id PubMed": [9164466],
45
           "pdbx database id DOI": ["10.1038/nsb0397-239"]
46
                                                                                 18
47
         "citation author": {
48
           "citation id": ["primary", "primary"],
           "name": ["Sha, B.", "Luo, M."],
```

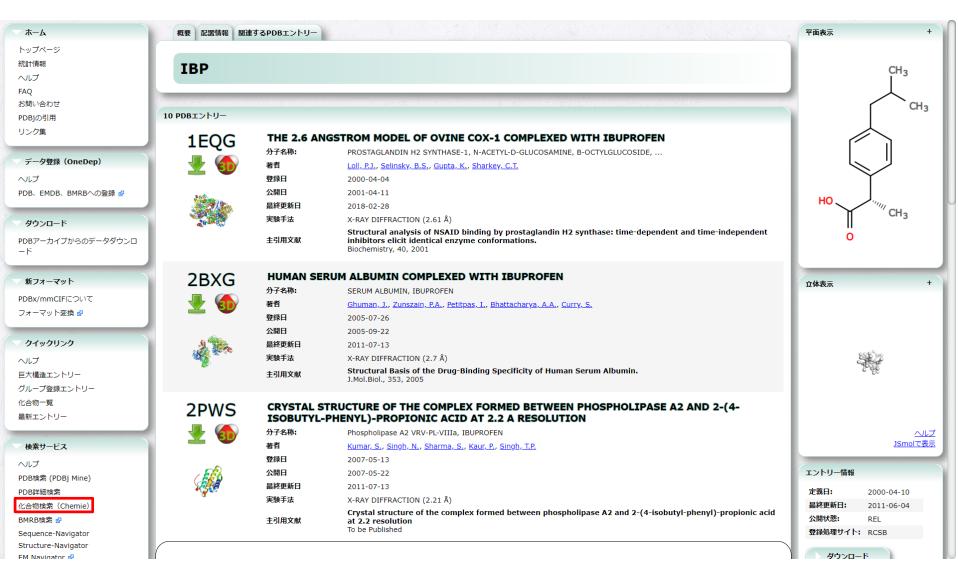
Chemie search (I)



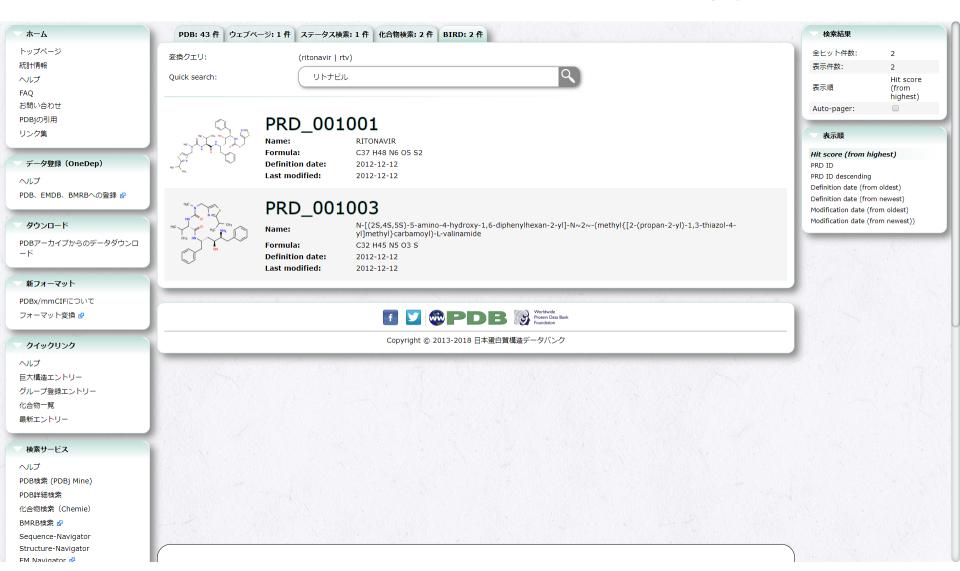
Chemie search (II)



Chemie search (III)



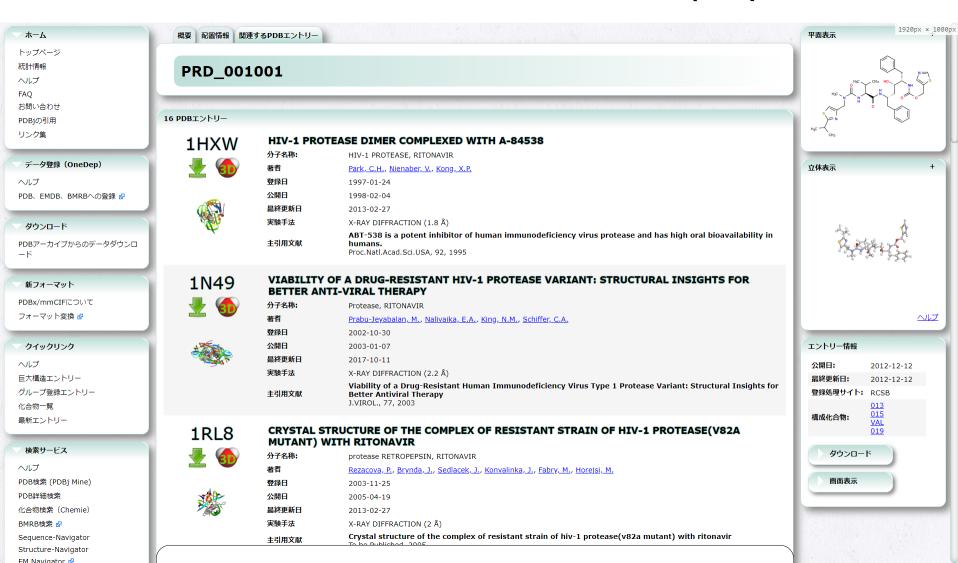
BIRD/PRD search (I)



BIRD/PRD search (II)



BIRD/PRD search (III)



Searching by PDBj Mine (II)

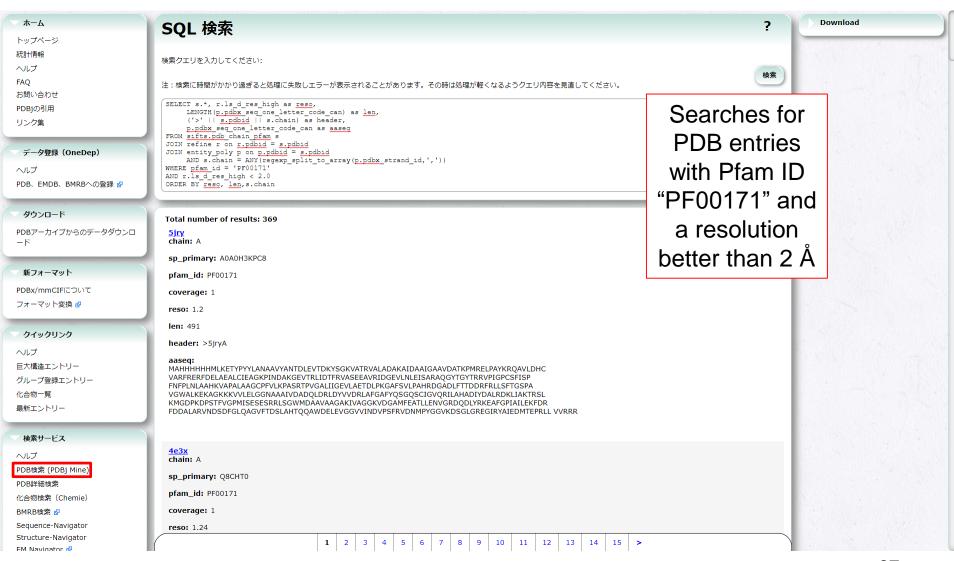
Advanced search



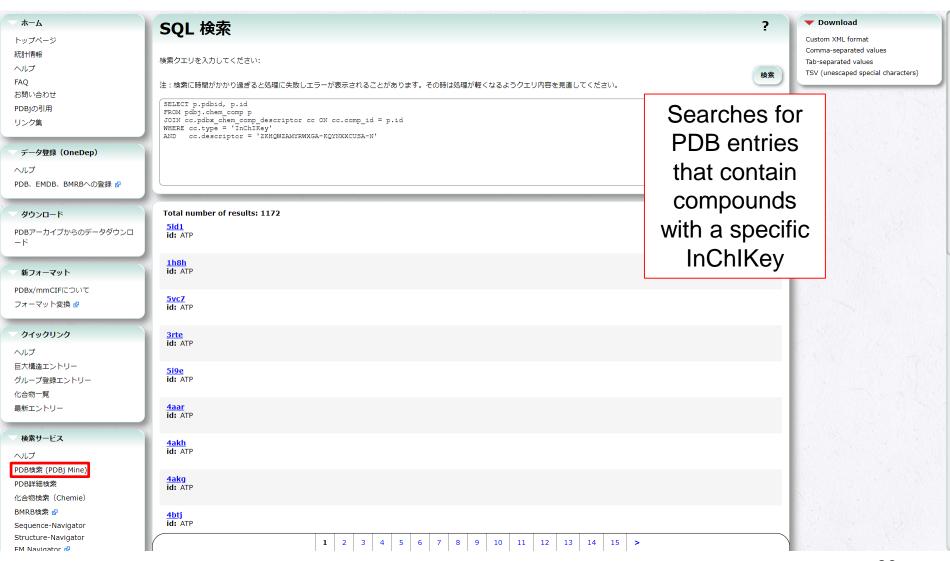
Mine 2 RDB Search (I)

- 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
- Dump files: http://ftp.pdbj.org/mine2/

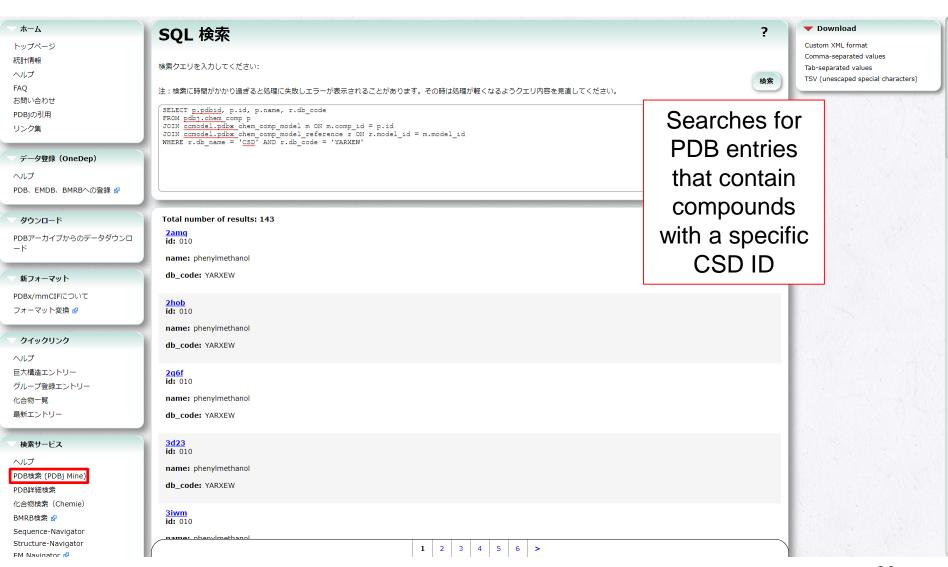
Mine 2 RDB Search (II)



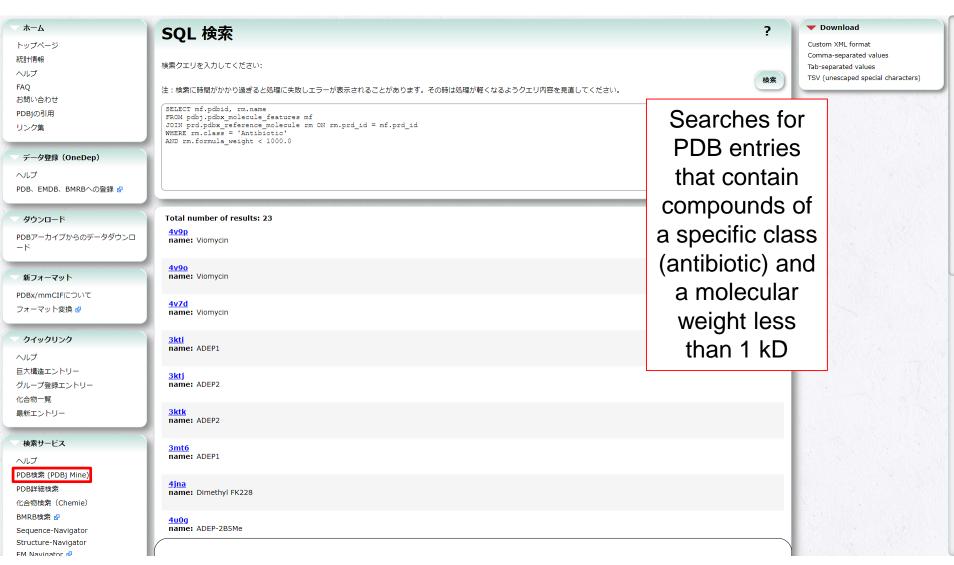
Mine 2 RDB Search (III)



Mine 2 RDB Search (IV)



Mine 2 RDB Search (V)



RDF & SPARQL

- https://rdf.wwpdb.org/
- http://ftp.pdbj.org/RDF/
- Integrated search: https://integbio.jp/rdf/sparql





Results Format:

JSON

Run Query

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-interfaceexamples

```
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=8&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 "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]
```

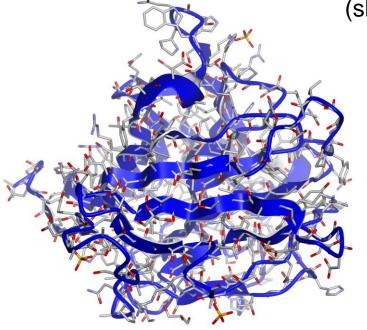
Visualization using Molmil (I)



https://pdbj.org/molmil_beta/

Active for services via https://pdbj.org/appCustomizations

(slide #4)

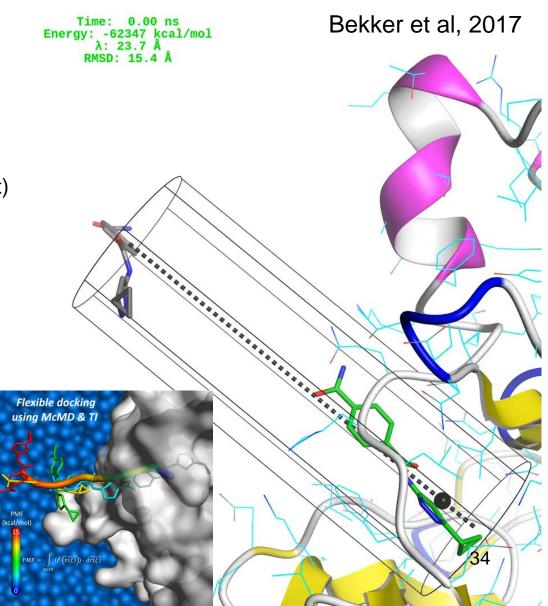


Pymol commands

- **Select** (select sc12, resi 12 and sidechain)
- COIOr (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 1cm)
 - 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



Visualization using Molmil (III)

```
set orthoscopic, off;
set depth cue, 1;
bg color white;
set label size, 30;
load first.gro;
load traj.xtc;
show tube;
cartoon color white, all;
color cyan, symbol C;
show sticks, resi 73 and sidechain;
show sticks, resi 85 and sidechain;
show sticks, resi 110 and sidechain;
 show sticks, resi 131 and sidechain;
show sticks, resi 133 and sidechain;
show sticks, resi 135 and sidechain;
origin resi 73 or resi 85 or resi 110 or resi 131 or resi 133 or resi 135;
show sticks, resi 107 and sidechain;
show sticks, resi 109 and sidechain:
 show sticks, resi 176 and sidechain;
turn x, 60;
move z, 100;
move y, -3;
 set label color, black;
label resi 73, Arg73;
 label resi 85, Thr85;
 label resi 107, Thr107;
label resi 109, Ser109
 label resi 110, Gln110;
 label resi 131, Asp131;
label resi 133, Asn133
label resi 135, Arg135;
label resi 176, Ala176;
set label color, red;
label resi 105-113, β3-α2;
label resi 132-138, β4-α3;
label resi 76-82, β1-β2;
set label color, magenta;
cli soup.renderer.customFogRange = [71, 100];
var pc1 = [-0.9596410989761353, -0.6479037404060364, -0.2896493375301361, 0.9235408902168274, 1.2893149852752686, 1.6075838804244995];
var pc2 = [0.10558268427848816, -0.4552445411682129, -0.0695020779967308, -0.4226403534412384, 0.42642560601234436, -0.22939902544021606];
var data = molmil.calcCenter(molmil.quickSelect("resi 73 or resi 85 or resi 110 or resi 131 or resi 133 or resi 135"));
var xyz = data[0];
var settings = {xyz, fontSize: 38, dx: -20, dy: -8};
var infoLabel = molmil.addLabel("", settings, cli soup), labelText = "PC1: %PC1\nPC2: %PC2";
□cli soup.animation.frameAction = function() {
  var pclval = (pcl[this.frameNo]*10).toFixed(2);
                                                                                                                                  Numoto et al., 2018
  var pc2val = (pc2[this.frameNo]*10).toFixed(2);
  if (pclval > 0 && pclval < 10) pclval = " " + pclval;</pre>
  if (pc2val > 0 && pc2val < 10) pc2val = " " + pc2val;
```

molmil.addLabel(labelText.replace(/%PC1/q, pc1val).replace(/%PC2/q, pc2val), {}, infoLabel);

set movie mode, swing;

35

https://pdbj.org/contact

PDBj-master PDBj annotation staff Home Top Page Please refer to the FAQ page. If you cannot find the answers to your questions, please contact us using the question form below. Please Statistics include the service name in your Message or Subject if you have any Help problems or questions about PDBj services. FAQ Contact Us Cite Us Name: Links Email: Subject: Data deposition (OneDep) Help Deposition to PDB, EMDB Message: or BMRB Download Send Download PDB archive / snapshot archive Please make sure your email address is valid. After submitting inquiry by clicking "Send Mail" button above, you will receive a New format receipt via email. If you don't get a receipt, you may mistype your email address or your receipt may be filed in the PDBx/mmCIF Resources spam/junk folder. Format Conversion 🚱 Moridavide Protein Data Bank Foundation Quick links Help Copyright © 2013-2018 Protein Data Bank Japan Large Structures **Group Depositions** Chemical Component entries Latest entries Search services Heln

ct https://pdbj.org/info/cite-us

| Please cite PDBJ or PDBJ services with the Please also refer to Terms and Conditions Also see "Cite a PDB structure entry" to ci • Cite PDBJ • Cite a PDBJ service • Cite a PDB structure entry | on using contents in PDBj site. | | |
|--|---|-------------------------|------|
| Cite PDBi | | | |
| • | | | |
| Please cite "PDBj" with references (29) (3) | 0) in the list of publications. | | |
| Cite a PDBi service | | | |
| • | | | |
| Please cite "a PDB] service" with its corres | ponding references. When no reference is available, | please cite the service | with |
| Service | URL | citations | |
| Search PDB (PDBj Mine) | pdbj.org/mine ₽ | (29) (30) | |
| Chemie search | pdbj.org/chemie-search @ | (29) (30) | |
| Search BMRB | bmrbdep.pdbj.org ₽ | (30) | |
| Sequence-Navigator | pdbj.org/seq-navi ₽ | - | |
| Structure-Navigator | pdbj.org/struc-navi ₽ | (10) | |
| EM Navigator | <u>pdbj.org/emnavi</u> ⊌ | (29) | |
| Omokage search | pdbj.org/omokage ₽ | (26) (30) | |
| wwPDB/RDF | rdf.wwpdb.org © | (19) | |
| SeSAW | pdbj.org/sesaw @ | (15) | |
| Ligand Binding Sites (GIRAF) | pdbj.org/giraf @ | (8) (13) (14) (18) | |
| jV : Graphic Viewer | pdbj.org/jv ₽ | (3) | |
| Molmil: WebGL Molecular Viewer | pdbj.org/molmil ₽ | (25) | |
| Yorodumi | pdbj.org/yorodumi ₽ | (29) | |
| ASH | pdbj.org/ash ₽ | (5) (10) (11) | |
| MAFFTash | pdbj.org/mafftash ₽ | (11) | |
| NMRToolBox | bmrbdep.pdbj.org/en/nmr_tool_box.html @ | (7) (17) (23) | |
| gmfit | pdbj.org/gmfit ⊌ | (12) | |
| CRNPRED | pdbj.org/crnpred ⊌ | (4) (6) | |
| Spanner | <u>pdbj.org/spanner</u> | (16) | |
| | pdbj.org/sfas ₽ | - | |
| SFAS | homcos.pdbj.org ₽ | (24) | |
| SFAS HOMCOS | | | |
| | pdbj.org/efsite ₽ | (1) (3) | |
| HOMCOS | pdbj.org/efsite 때 pdbj.org/efseek 때 pdbj.org/efsurf 때 | (1) (3) | |
| HOMCOS eF-site eF-seek | pdbj.org/efseek ₪ | | |

Cite a PDB structure entry

Please cite a PDB structure entry using the following information.

An entry with a published primary citation (PDB ID and primary citation)

PDB ID: 102

Cite Us

D.W. Heinz, W.A. Baase, F.W. Dahlquist, B.W. Matthews (1993) How Amino-Acid Insertions are Allowed in an Alpha-Helix of T4 Lysozyme Nature 361:561.

· An entry without a published primary citation (PDB ID, author names, and title)

PDB ID: 1CI

W. Shi, D.A. Ostrov, S.E. Gerchman, V. Graziano, H. Kycia, B. Studier, S.C. Almo, S.K. Burley, New York Structural GenomiX Research Consortium (NYSGXRC). The Structure of PNP Oxidase from S. cerevisiae

• An entry may also be referenced using its Digital Object Identifier (DOI).

36