

トーゴーの日シンポジウム2019 ワークショップ  
つないで使って役に立てよう Protein Data Bank

# PDBj Mine (PDB検索サービス) の紹介

大阪大学蛋白質研究所

工藤 高裕 (KUDOU Takahiro) ・ 横地 政志 (YOKOCHI Masashi)

2019年10月5日 (土)

# PDBj Mine (PDB検索サービス) の紹介

1. 簡易検索 (キーワード検索)
2. 詳細検索
3. SQL検索・SPARQL検索

# 1. 簡易検索 (キーワード検索)

The screenshot displays the PDBj website interface. At the top, there is a search bar with the text 'pdbj.org 全体を検索 (日本語ca)'. Below the search bar, a red arrow points to the search input field containing '(dengue | "dengue fever")', with a red text overlay that reads '日本語も英語に変換して検索' (Search in Japanese and English). The search results are displayed in a grid format, showing three entries:

- 4CCT**: DENGUE 1 CRYO-EM RECONSTRUCTION. 分子名称: DENGUE VIRUS 1 E PROTEIN, DENGUE VIRUS 1 M PROTEIN. 著者: Kostyuchenko, V.A., Zhang, Q., Tan, J.L., Ng, T.S., Lok, S.M. 登録日: 2013-10-28. 公開日: 2013-11-06. 最終更新日: 2017-08-30. 実験手法: ELECTRON MICROSCOPY (4.5 Å). 主要用文献: Immature and Mature Dengue Serotype 1 Virus Structures Provide Insight Into the Maturation Process. J.Virol., 87, 2013.
- 4C11**: DENGUE VIRUS RNA DEPENDENT RNA POLYMERASE WITH RESIDUES FROM THE NSS LINKER REGION. 分子名称: DENGUE VIRUS TYPE 3 RNA DEPENDENT RNA POLYMERASE, ZINC ION. 著者: Lim, S.P., Lescar, J. 登録日: 2013-08-09. 公開日: 2013-08-21. 最終更新日: 2018-04-04. 実験手法: X-RAY DIFFRACTION (2.6 Å). 主要用文献: A Crystal Structure of the Dengue Virus Ns5 Polymerase Delineates Inter-Domain Amino Acids Residues that Enhance Its Thermostability and De Novo Initiation Activities. J.Biol.Chem., 288, 2013.
- 4UIF**: CRYO-EM STRUCTURE OF DENGUE VIRUS SEROTYPE 2 IN COMPLEX WITH ANTIGEN-BINDING...

On the right side of the search results, there is a '検索結果' (Search Results) panel showing 211 total hits, 25 displayed, and a '結果をダウンロード' (Download Results) button. Below the search results, there is a '表示順' (Sort by) panel with options for '関連性が高い順' (Sort by relevance), '登録日の新しい順' (Sort by registration date), '公開日の新しい順' (Sort by publication date), and '分野別の新しい順' (Sort by field).

# 1. 簡易検索 (キーワード検索)

PDBエントリー検索

PDBjサイト検索

未公開エントリー状態検索

化合物検索

BIRD検索

The screenshot shows the PDBj search results page for the keyword 'デング熱'. The search bar contains 'デング熱' and the search criteria are '(dengue | "dengue fever")'. The results are displayed in a table with columns for PDB ID, title, author, date, and method. The first result is 4CCT, titled 'DENGUE 1 CRYO-EM RECONSTRUCTION'. The second result is 4C11, titled 'DENGUE VIRUS RNA DEPENDENT RNA POLYMERASE WITH RESIDUES FROM THE NS5 LINKER REGION'. The third result is 4UIF, titled 'CRYO-EM STRUCTURE OF DENGUE VIRUS SEROTYPE 2 IN COMPLEX WITH ANTIGEN-BINDING'. The page also includes a sidebar with navigation links and a search results panel on the right with filters like '全ヒット件数: 211' and '表示件数: 25'.

# 2. 詳細検索

The screenshot displays the PDBj website interface. At the top, there is a search bar with the text "pdbj.org 全体を検索 (日本語ca)". Below the search bar, the main content area shows search results for "DENGUE 1 CRYO-EM RECONSTRUCTION" (PDBID: 4CCT) and "DENGUE VIRUS RNA DEPENDENT RNA POLYMERASE WITH RESIDUES FROM REGION" (PDBID: 4C11). The 4CCT entry is highlighted with a red box around the "詳細検索ページへ戻る" button. The 4C11 entry is also visible below it. The interface includes a sidebar with navigation links like "ホーム", "データ登録 (OneDep)", "ダウンロード", "新フォーマット", and "クイックリンク". At the bottom, there is a "Mine: 詳細条件検索" section with various filters and a "検索" button.

# 2. 詳細検索

**156587**  
件公開中 (2019-10-02 00:00 UTC / 09:00 JST)

▼ ホーム

- トップページ
- 統計情報
- ヘルプ
- FAQ
- お問い合わせ
- PDBjの引用・利用規約
- 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

**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:

含まれるポリマー鎖の種類:

- ポリペプチド(D体)  含む  含まない  どちらでもよい
- ポリペプチド(L体)  含む  含まない  どちらでもよい
- ポリデオキシリボヌクレオチド(DNA)  含む  含まない  どちらでもよい
- ポリリボヌクレオチド(RNA)  含む  含まない  どちらでもよい
- 多糖(D体)  含む  含まない  どちらでもよい
- 多糖(L体)  含む  含まない  どちらでもよい
- DNA/RNA 複合体  含む  含まない  どちらでもよい
- 環状疑似ペプチド  含む  含まない  どちらでもよい
- その他  含む  含まない  どちらでもよい

実験手法:

分解能:  -

▼ 詳細条件を選択して下さい

PDBID

キーワード

タイトル

公開日

登録日

最終更新日

文献著者

論文題名

雑誌名

発行年

巻番号

主引用文献のみ

含まれるポリマー鎖の種類

分子名称

外部データベース

リガンドと補欠分子族

ポリマー鎖の数

ポリマー鎖の長さ

実験手法

分解能

由来する生物種

宿主生物種

▼ 詳細条件を選択して下さい

PDBID

キーワード

タイトル

公開日

登録日

最終更新日

文献著者

論文題名

雑誌名

発行年

巻番号

主引用文献のみ

含まれるポリマー鎖の種類

分子名称

外部データベース

リガンドと補欠分子族

ポリマー鎖の数

ポリマー鎖の長さ

実験手法

分解能

由来する生物種

宿主生物種

# 2. 詳細検索

### Mine: 詳細検索結果ページ

**詳細検索:**

キーワード:

含まれるポリマー鎖の種類:

含まれるポリマー鎖の種類:

実験手法:

公開日以降:

公開日以前:

キーワードの追加:

[詳細検索ページへ戻る](#)

---

**5DTO**




**DENGUE VIRUS FULL LENGTH NS5 COMPLEXED WITH VIRAL CAP 0-RNA AND SAH**

分子名称: NSS, RNA (5'-R(P\*AP\*GP\*UP\*U)-3'), ZINC ION, ...

著者: [Zhao, Y., Soh, T.S., Lim, S.P., Chung, K.Y., Swaminathan, K., Vasudevan, S.G., Shi, P.-Y., Lescar, J., Luo, D.](#)

登録日: 2015-09-18

公開日: 2015-11-25

最終更新日: 2015-12-16

実験手法: X-RAY DIFFRACTION (2.603 Å)

主引用文献: **Molecular basis for specific viral RNA recognition and 2'-O-ribose methylation by the dengue virus nonstructural protein 5 (NS5)**  
Proc.Natl.Acad.Sci.USA, 112, 2015

---

**5XC6**




**DENGUE VIRUS 4 NS3 HELICASE IN COMPLEX WITH SSRNA SLA12**

分子名称: NS3 Helicase, RNA (5'-R(\*AP\*GP\*UP\*GP\*UP\*UP\*AP\*GP\*UP\*CP\*U)-3'), PHOSPHATE ION

著者: [Swarbrick, C.M.D., Basavannacharya, C., Chan, K.W.K., Chan, S.A., Singh, D., Wei, N., Phoo, W.W., Luo, D., Lescar, J., Vasudevan, S.G.](#)

登録日: 2017-03-22

公開日: 2017-11-08

最終更新日: 2017-12-27

実験手法: X-RAY DIFFRACTION (2.9 Å)

主引用文献: **NS3 helicase from dengue virus specifically recognizes viral RNA sequence to ensure optimal replication**  
Nucleic Acids Res., 45, 2017

**検索結果** ?

全ヒット件数: 2

表示件数: 2

表示順:

Auto-pager:

---

**表示順**

関連性が高い順

PDBID昇順 (0→9,a→z)

PDBID降順 (z→a,9→0)

登録日の古い順

登録日の新しい順

公開日の古い順

公開日の新しい順

分解能の高い順

# 3. SQL検索

**156587**  
件公開中 (2019-10-02  
00:00 UTC / 09:00 JST)



Protein Data Bank Japan

Worldwide Protein Data Bank Foundation

English 日本語 简体中文 繁體中文 한국어

pdbj.org 全体を検索 (日本語OK)

wwPDB RCSB PDB PDBe BMRB Adv. Search Search help

**SQL 検索** ?

検索クエリを入力してください:

注: 検索に時間がかかり過ぎると処理に失敗しエラーが表示されることがあります。その時は処理が軽くなるようクエリ内容を見直してください。

```
SELECT e.pdbid, e.entity_id, s.chain, s.sp_primary
FROM entity_poly e
JOIN sifts.pdb_chain_uniprot s
ON s.pdbid = e.pdbid AND s.chain = ANY(STRING_TO_ARRAY(e.pdbx_strand_id, ','))
```

ここに任意のSQLクエリを入力

検索

**Total number of results: 468549**

**101m**  
entity\_id: 1

chain: A

sp\_primary: P02185

ヒットしたPDBエントリー及び結果のリスト

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	>
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	---

**Download**

- Custom XML format
- Comma-separated values
- Tab-separated values
- TSV (unescaped special characters)

検索結果のファイルもダウンロードできます

# 3. SQL検索

SQL検索可能なデータセット

1. PDBコアアーカイブ
2. PDBアーカイブ内の化合物ライブラリ (CCD: Chem Comp Dictionary)
3. PDBアーカイブ内の抗生物質、阻害剤 (BIRD: Biologically Interesting Molecule Reference Dictionary)
4. PDBとUniProtKBの相互参照 (SIFTS: Structure Integration with Function, Taxonomy and Sequences)
5. wwPDB検証レポート (VRPT)



# 3. SQL検索／SQL例文集

Structure Navigator

EM Navigator [↗](#)

Omokage search [↗](#)

wwPDB/RDF [↗](#)

SeSAW [↗](#)

Status search

---

**Molecular Viewers**

---

**Services & Softwares**

---

**Derived databases**

---

**Educational service**

---

**About PDBj**

---

**Customizations**

Help

Configure panels

Configure customizations

Reset all customizations

Job manager

---

A- A A+

**SQL Search** ← ここをクリックすると以下のSQLクエリ例文集が表示されます

Enter search query:

Note: The search will fa

Search Res

**PDBj Mine 2: SQL Search**

This page is also available in: [日本語](#)

This page provides some examples of SQL queries that can be executed from the [PDBj Mine](https://pdbj.org/mine) interface at <https://pdbj.org/mine>. You can use these samples at **SQL Search** of PDBj Mine or [PDBj Mine:REST API](#). Documentation describing the SQL schema can be found on the [Mine 2 RDB Documentation](#) page.

The result are returned in web pages or a test file such as XML, CSV, and TSV, respectively. (See also [PDBj Mine: search result.](#))

- Search for all PDB IDs:
- Search for all count (PDB IDs):
- Search by keyword:
- Search for all PDB IDs (1g\*):
- Search for citation\_author category (1gof):
- Search for citation\_author ('Ito, N.')
- Search for citation\_author ('Ito, N.' and 'Phillips, S.E.V.')
- Search for citation\_author ['Ito, N.'] (primary):
- Search for journal name ['molecular cell']:
- Search for journal year ['2001']:
- Search for journal volume ['555']:
- Search for exprimental method [SOLUTION NMR]:
  - Note on experimental methods
- Search for journal year[1991] volume[266] exp\_method [X-RAY DIFFRACTION]:
- Search for exptl\_crystal/density\_percent\_sol (1iqq):
- Search for citation/pdbx\_database\_id\_DOI (1iqq):
- Search the PubMed ID of 1iqq (citation/pdbx\_database\_id\_PubMed):
- Search the PubMed ID of 1a08 and 1ai9 (citation/pdbx\_database\_id\_PubMed):

DBスキーマの説明文書が表示されます

様々なSQLクエリ例

# 3. SQL検索／PDB検証レポートの検索

## Search wwPDB validation report (VRPT)

### [Search for rotamer outliers]

[Run SQL query](#)

```
select pdbid, rotamer_outliers_percent from vrpt.pdbx_dcc_geometry where rotamer_outliers_percent is not null
```

### [Search for RSRZ outliers]

[Run SQL query](#)

RSRZは、以下で定義されたRSR (Real-space R)のZスコア：

$$RSR = \frac{\sum_{RS} |\rho_{obs} - \rho_{calc}|}{\sum_{RS} |\rho_{obs} + \rho_{calc}|}$$

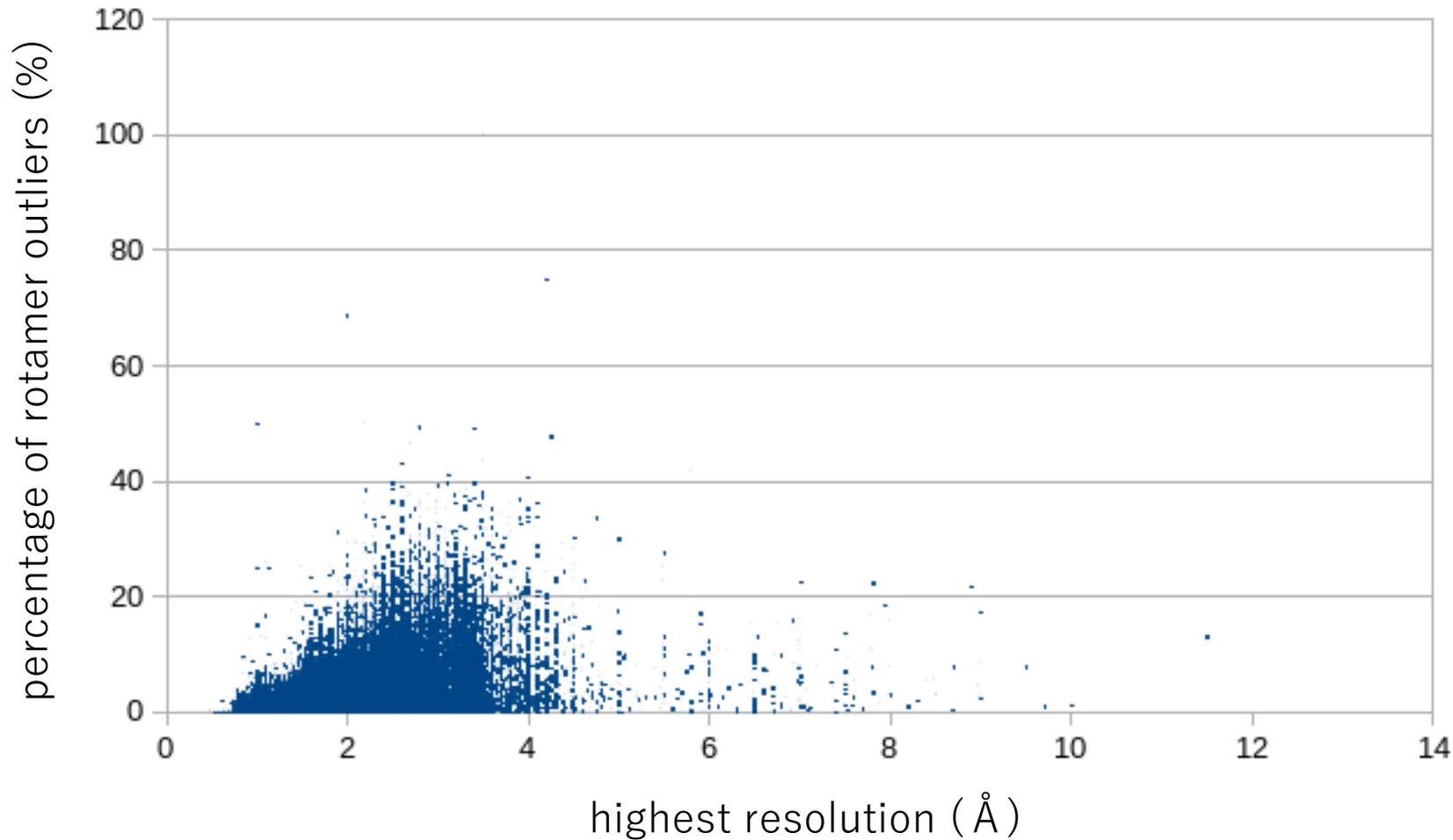
```
select pdbid, "RSRZ_outliers_percent" from vrpt.pdbx_dcc_map_overall where "RSRZ_outliers_percent" is not null
```

### [Find correlation between rotamer outliers and resolution]

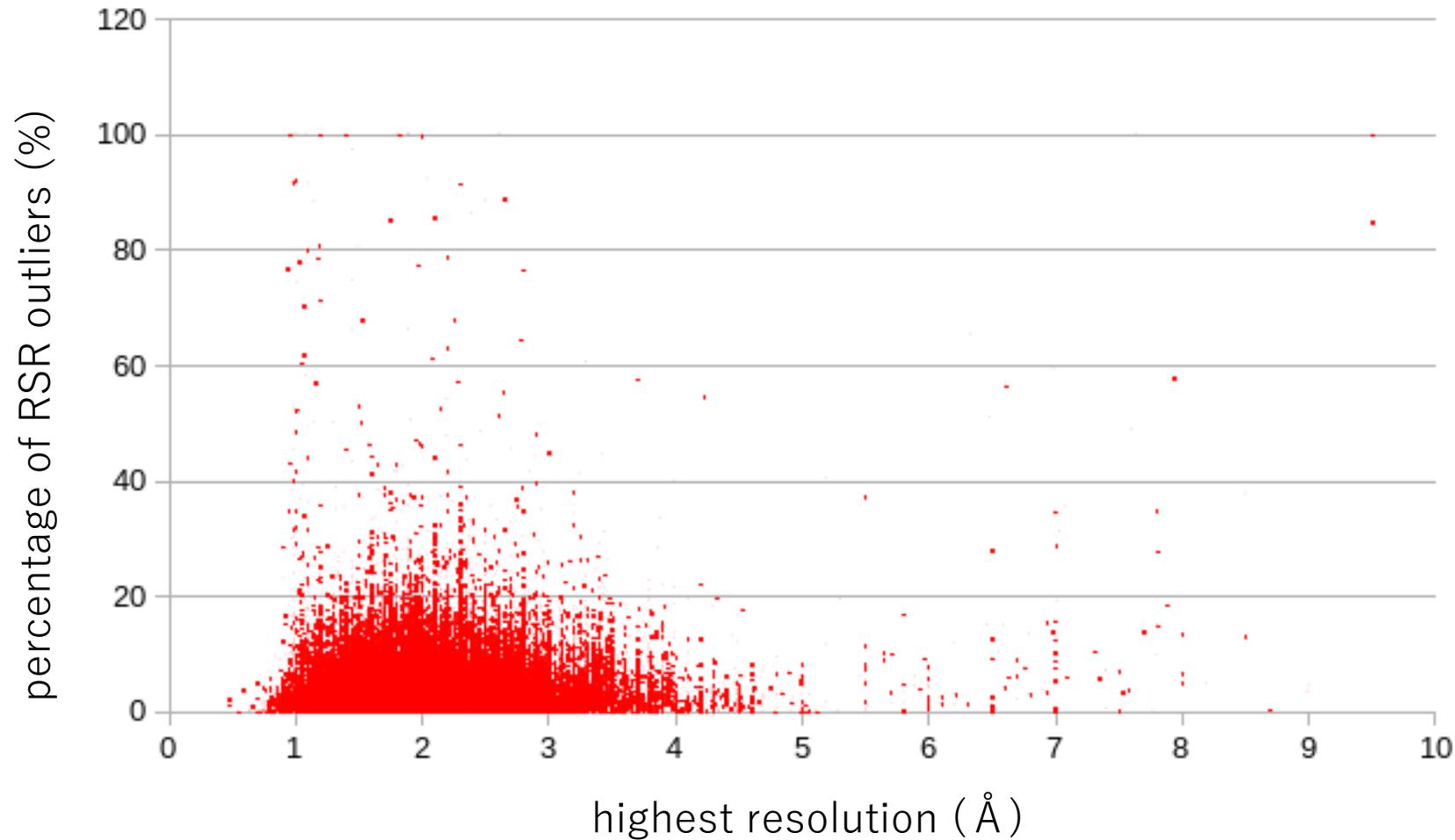
[Run SQL query](#)

```
select g.pdbid, g.rotamer_outliers_percent, d.ls_d_res_high
from vrpt.pdbx_dcc_geometry as g, vrpt.pdbx_dcc_density as d
where g.pdbid = d.pdbid
```

# 3. SQL検索／回転異性体の異常率と分解能

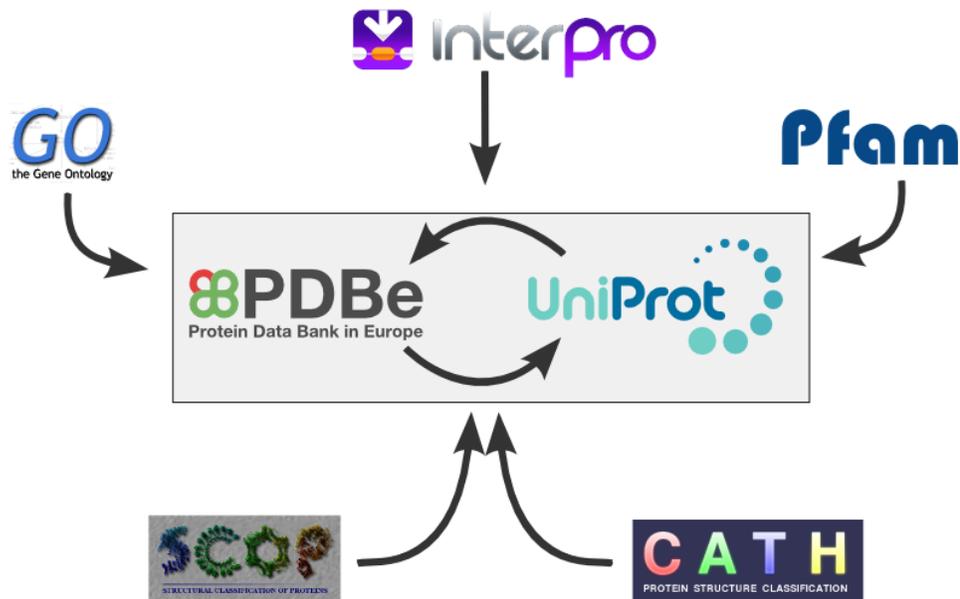


### 3. SQL検索／RSRの異常率(RSRZ>2)と分解能



# 3. SQL検索／機能と品質を同時に検索

SIFTS: Structure Integration with Function, Taxonomy and Sequence



VRPT: wwPDB validation report



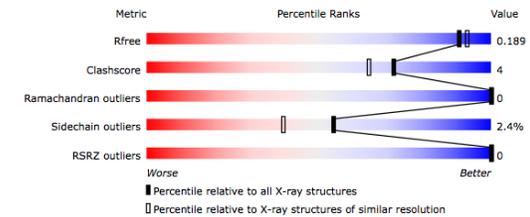
Full wwPDB X-ray Structure Validation Report ⓘ

Feb 12, 2017 – 07:52 pm GMT

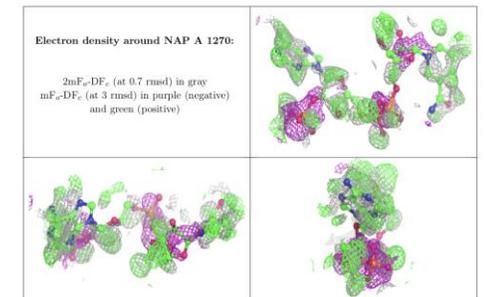
PDB ID : 1CBS  
 Title : CRYSTAL STRUCTURE OF CELLULAR RETINOIC-ACID-BINDING PROTEINS I AND II IN COMPLEX WITH ALL-TRANS-RETINOIC ACID AND A SYNTHETIC RETINOID  
 Authors : Kleywegt, G.J.; Bergfors, T.; Jones, T.A.  
 Deposited on : 1994-09-28  
 Resolution : 1.80 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
 A user guide is available at <http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
 with specific help available everywhere you see the ⓘ symbol.



In addition to geometric validation for ligands, for X-ray diffraction PDB entries the wwPDB validation report also presents images displaying the ligand and the surrounding electron density map.



# 3. SQL検索／機能と品質を同時に検索

**[Select serine protease inhibitor complexes having good side chain rotamer combining with SIFTS data]**

[Run SQL query](#)

```
select pdbid, rotamer_outliers_percent
from vrpt.pdbx_dcc_geometry
where rotamer_outliers_percent is not null and
pdbid in (
select pdbid from sifts.pdb_chain_go where go_id='GO:0097180' group by pdbid
) order by rotamer_outliers_percent
```

Here, the following subquery is used to extract the SIFTS data:

```
select distinct pdbid from sifts.pdb_chain_go where go_id='GO:0097180'
```

**[Search serine protease inhibitor complexes combining with SIFTS data ordered by 3 validation metrics, side chain rotamer, RSR, and resolution]**

[Run SQL query](#)

```
select g.pdbid,g.rotamer_outliers_percent, m."RSRZ_outliers_percent", d.ls_d_res_high
from vrpt.pdbx_dcc_geometry as g, vrpt.pdbx_dcc_map_overall as m, vrpt.pdbx_dcc_density as d
where g.rotamer_outliers_percent < 2.0 and m."RSRZ_outliers_percent" < 2.0 and
g.pdbid = m.pdbid and g.pdbid = d.pdbid and g.pdbid in (
select pdbid from sifts.pdb_chain_go where go_id='GO:0097180' group by pdbid)
order by g.rotamer_outliers_percent, m."RSRZ_outliers_percent", d.ls_d_res_high
```

# 3. SQL検索／構造品質でソートかけたセリンプロテアーゼ阻害剤複合体の検索結果

## SQL Search

Enter search query:

Note: The search will fail if it takes too long to finish, in which case, please reconsider the query.

```
select g.pdbid,g.rotamer_outliers_percent, m."RSRZ_outliers_percent", d.ls_d_res_high
from vrpt.pdbx_dcc_geometry as g, vrpt.pdbx_dcc_map_overall as m, vrpt.pdbx_dcc_density as d
where g.rotamer_outliers_percent < 2.0 and m."RSRZ_outliers_percent" < 2.0 and
g.pdbid = m.pdbid and g.pdbid = d.pdbid and g.pdbid in (
select pdbid from sifts.pdb_chain_go where go_id='GO:0097180' group by pdbid)
order by g.rotamer_outliers_percent, m."RSRZ_outliers_percent", d.ls_d_res_high
```

**Search**

**Download**

- Custom XML format
- Comma-separated values
- Tab-separated values
- TSV (unescaped special characters)

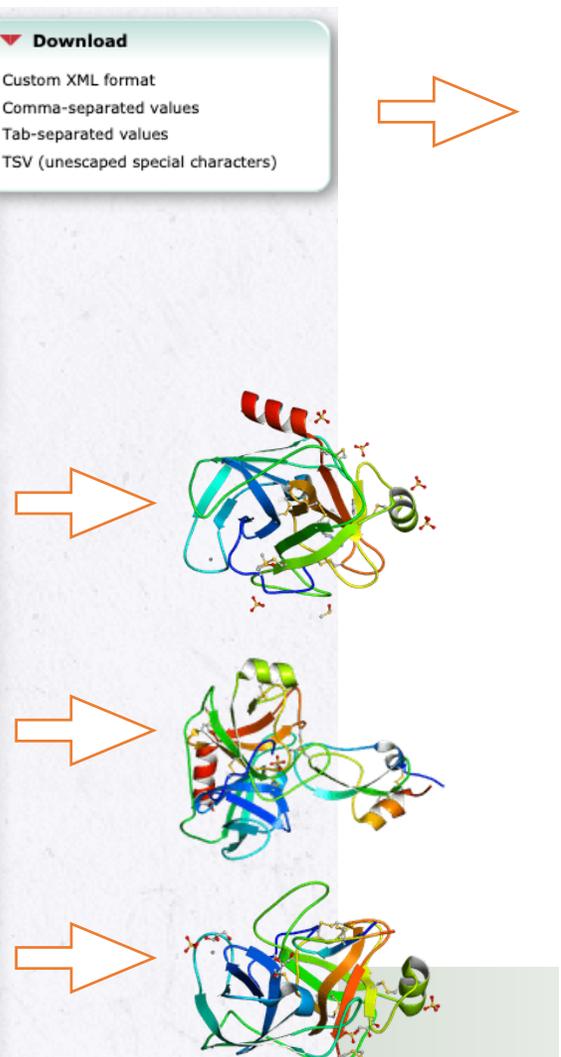
**Total number of results: 198**

**5mng**  
rotamer\_outliers\_percent: 0  
RSRZ\_outliers\_percent: 0  
ls\_d\_res\_high: 1.34

**3m7q**  
rotamer\_outliers\_percent: 0  
RSRZ\_outliers\_percent: 0  
ls\_d\_res\_high: 1.7

**6bvh**  
rotamer\_outliers\_percent: 0  
RSRZ\_outliers\_percent: 0.42  
ls\_d\_res\_high: 1.93

1 2 3 4 5 6 7 8 >



sql\_query

pdbid	rotamer_outliers_percent	RSRZ_outliers_percent	ls_d_res_high
5mng	0.0	0.0	1.34
3m7q	0.0	0.0	1.7
6bvh	0.0	0.42	1.93
6e5m	0.0	0.43	1.61
5mng	0.0	0.45	0.86
5mno	0.0	0.45	0.96
5mnm	0.0	0.45	0.98
5mnf	0.0	0.45	0.99
5mnp	0.0	0.45	1.01
5mnl	0.0	0.45	1.04
3a8a	0.0	0.45	1.4
3plp	0.0	0.45	1.63
1mts	0.0	0.45	1.9
1mtu	0.0	0.45	1.9
6b6q	0.0	0.45	2.0
6b6n	0.0	0.45	2.0
6b6r	0.0	0.45	2.0
6b6t	0.0	0.45	2.0
6eaw	0.0	0.88	1.29
5mnk	0.0	0.9	0.8
5mnn	0.0	0.9	0.86
5mnc	0.0	0.9	0.92

# 3. SPARQL検索／SPARQLエンドポイント

**PDBj-BMRB Data Server:**  
common open representations of BMRB NMR-STAR data in XML, RDF, and JSON formats

Home Search **Examples** Download Resources NEWS

## Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

<https://bmrpub.pdbj.org/rdf/bmr>

Query Text

```
select distinct ?Concept where {[] a ?Concept} LIMIT 100
```

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format:

HTML

Execution timeout:

0

milliseconds (values less than 1000 are ignored)

Options:

Strict checking of void variables

(The result can only be sent back to browser, not saved on the server, see [details](#))

Run Query

Reset

## Query examples

### Category holders

1. Select all category holders of datablock class of BMRB entry 15400: [Show](#)
2. Select all category holders of datablock class of Metabolomics entry bmse000400: [Show](#)

### Entry statistics

3. Count entries per submission year and experimental method (subtype): [Show](#)

### Assembly descriptions

4. Select all assembly names, asym IDs, entity IDs, polymer types, formula weights and functions in a assembly: [Show](#)

### Entity descriptions

5. Select all entity names and sequences of polymer entities expressed using one-letter code: [Show](#)
6. Select all original source information of molecular entities and external links to NCBI

SPARQL検索可能なRDFデータ

1. PDBコアアーカイブ (PDB)\*
2. BMRBコアアーカイブ (BMRB)\*
3. PDBとUniProtKBの相互参照 (SIFTS)
4. wwPDB検証レポート (VRPT)

\*NBDC RDF portalでも利用可能  
毎月1日にデータ更新

wwPDB検証レポートのSPARQL検索例  
については、PDBjのポスターNo. 32-33  
を参照してください。