

# BMRBに登録されたNMRデータの検索 – SPARQLクエリを中心にして

How to retrieve NMR data archived at BMRB with SPARQL

IIBMP2016 (生命医薬情報学連合大会2016) Oct  
1, 2016

PDBj&創薬等ランチョンセミナー

PDBj-BMRB Masashi Yokochi (横地政志)

**PDBj-BMRB**, satellite site of **BMRB**, serves deposition, annotation and distribution sites of **experimental NMR data** written in **NMR-STAR** format



**Members:**



# PDBj-BMRB web services to enhance interoperability of NMR data



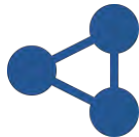
## **BMRB/XML, BMRB/RDF**

representations of BMRB NMR-STAR data in XML and RDF formats



## **PDBj-BMRB integrated search service**

search biological/biochemical DBs at once



## **PDBj-BMRB SPARQL server**

a programmable API for federated search



## **BMRB SQL on your PC**

periodically updated BMRB relational DB for intensive search

# PDBj-BMRB web services to enhance interoperability of NMR data



## **BMRB/XML, BMRB/RDF**

representations of BMRB NMR-STAR data in XML and RDF formats



## **PDBj-BMRB integrated search service**

search biological/biochemical DBs at once



## **PDBj-BMRB SPARQL server**

a programmable API for federated search



## **BMRB SQL on your PC**

periodically updated BMRB relational DB for intensive search



## BMRB/XML: The most comprehensive NMR-STAR data repository as a single format

We have converted canonical NMR-STAR data and diverse repositories such as atomic coordinate of NMR structures, LACS chemical shift validation, PACSY structural annotations, etc. into a single BMRB/XML format.

### BMRB/XML

#### canonical NMR-STAR

1. Conventional NMR-STAR
  2. BMRB+PDB NMR-STAR
  3. PACSY structural annotation
4. LACS chemical shift validation report
  5. Protein Blocks structural annotation
  6. Completeness of assigned chemical shifts

### URLs of each data repository

1. [http://bmrbl.pdbj.org/ftp/pub/bmrbl/entry\\_lists/nmr-star3.1/](http://bmrbl.pdbj.org/ftp/pub/bmrbl/entry_lists/nmr-star3.1/)
2. [http://bmrbl.pdbj.org/ftp/pub/bmrbl/nmr\\_pdb\\_integrated\\_data/coordinates\\_restraints\\_chemshifts/bmrbl\\_plus\\_pdb/](http://bmrbl.pdbj.org/ftp/pub/bmrbl/nmr_pdb_integrated_data/coordinates_restraints_chemshifts/bmrbl_plus_pdb/)
3. <http://pacsy.nmrham.wisc.edu/>
4. [http://bmrbl.pdbj.org/ftp/pub/bmrbl/validation\\_reports/LACS/](http://bmrbl.pdbj.org/ftp/pub/bmrbl/validation_reports/LACS/)
5. <http://bmrblpub.protein.osaka-u.ac.jp/archive/pb/>
6. [http://bmrblpub.protein.osaka-u.ac.jp/archive/cs\\_complete/](http://bmrblpub.protein.osaka-u.ac.jp/archive/cs_complete/)



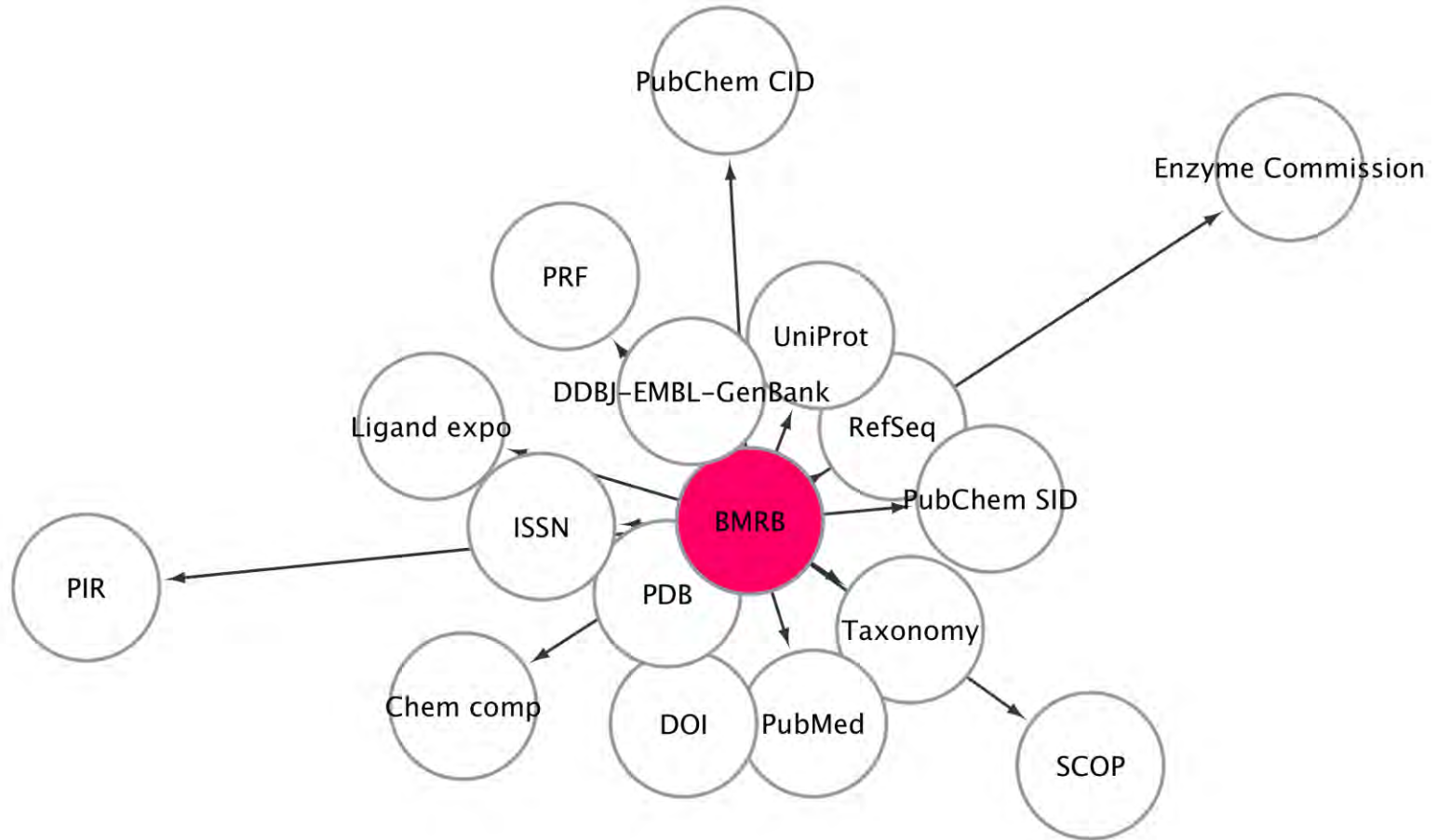
## BMRB/RDF: In compliance with standards of ‘Linked Data’ and biological DB community

```
<BMRBo:entity_db_link rdf:about="http://bmrpub.protein.osaka-u.ac.jp/rdf/bmr11300/entity_db_link/NP_001008202,REF,1,11300">
  <BMRBo:of_datablock rdf:resource="http://bmrpub.protein.osaka-u.ac.jp/rdf/bmr11300"/>
  <BMRBo:entity_db_link.accession_code>NP_001008202</BMRBo:entity_db_link.accession_code>
  <BMRBo:entity_db_link.database_code>REF</BMRBo:entity_db_link.database_code>
  <BMRBo:entity_db_link.entity_id>1</BMRBo:entity_db_link.entity_id>
  <BMRBo:entity_db_link.entry_id>11300</BMRBo:entity_db_link.entry_id>
  <rdfs:seeAlso rdf:resource="http://www.ncbi.nlm.nih.gov/protein/NP_001008202"
    rdfs:label="info:refseq/NP_001008202"/>
  <rdfs:seeAlso rdf:resource="http://identifiers.org/refseq/NP_001008202"
    rdfs:label="urn:miriam:refseq:NP_001008202"/>
  <BMRBo:entity_db_link.author_supplied>no</BMRBo:entity_db_link.author_supplied>
  <BMRBo:entity_db_link.entry_mol_name>cell division cycle 5-like protein [Xenopus (Silurana) tropicalis]</BMRBo:entity_db_link.entry_mol_name>
  <BMRBo:entity_db_link.ordinal>16</BMRBo:entity_db_link.ordinal>
  <BMRBo:entity_db_link.seq_homology_expectation_val>8.17E-32</BMRBo:entity_db_link.seq_homology_expectation_val>
  <BMRBo:entity_db_link.seq_identity>100.00</BMRBo:entity_db_link.seq_identity>
  <BMRBo:entity_db_link.seq_positive>100.00</BMRBo:entity_db_link.seq_positive>
  <BMRBo:entity_db_link.seq_query_to_submitted_percent>82.86</BMRBo:entity_db_link.seq_query_to_submitted_percent>
  <BMRBo:entity_db_link.seq_subject_length>804</BMRBo:entity_db_link.seq_subject_length>
</BMRBo:entity_db_link>
```

Two statements using **rdfs:seeAlso** appear, the former one represents **polite URL** pointing original resource of NCBI RefSeq database and the resource has a label written in the **formal URN**, the latter one is a statement semantically equivalent to the former one, but utilizes a **persistent URI** resolving system of Identifiers.org with the **MIRIAM URN** widely accepted by biological database community.



## BMRB/RDF: Linked external information resources via RDF links



Shorter distances from BMRB indicate closer relationships with BMRB

# PDBj-BMRB web services to enhance interoperability of NMR data



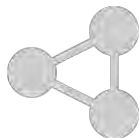
## BMRB/XML, BMRB/RDF

representations of BMRB NMR-STAR data in XML and RDF formats



## PDBj-BMRB integrated search service

search biological/biochemical DBs at once



## PDBj-BMRB SPARQL server

a programmable API for federated search



## BMRB SQL on your PC

periodically updated BMRB relational DB for intensive search





# PDBj-BMRB integrated search service: Search biological/biochemical DBs at once

<http://bmrdep.pdbj.org>



Deposition ▾ Search ▾ Archives ▾ Mirrors ▾ Language ▾ About ▾

BMRB ID or Dep. Code

11046 BMRB entries and 1324 Metabolomics entries are available on Sep 05, 2016

## Search

Everything

Sequence



e.g., Entry ID, Macromolecule, Gene Ontology or Author



[Search help](#)

PDBj-BMRB group in Osaka University supported by [JST-NBDC \(Japan Science and Technology Agency - National Bioscience Database Center\)](#) contributes to the community for Biomolecular NMR by collection, annotation, validating and archiving experimental NMR data in collaboration with [wwPDB \(WorldWide Protein Data Bank\)](#). Our mission is to empower scientists in their analysis of the structure, dynamics, and chemistry of biological systems and to support further development of biomolecular NMR spectroscopy. Please deposit your NMR structure and related NMR experimental data through our deposition site, [ADIT-NMR](#). The released data are available from our [BMRB mirror server](#). Please try our new search service above which allows you to quickly find rich information of biomolecule such as function, sequence, structure, NMR data as well as interaction in multiple databases including [BMRB](#), [PDB](#), [EMDB](#), [UniProt](#), [IntAct](#), [BMRB-Metabolomics](#) and [Ligand Expo](#).

## Contents

Overview of data deposition process  
Flowchart about NMR data deposition

How to submit your data via ADIT-NMR  
Step by step guide for data deposition

Should I choose ADIT-NMR or SMSDep?  
Differences in two deposition systems

FAQ  
Frequently Asked Questions

NMR Tool Box  
Tools for NMR analysis

Related tools  
Tools for validation and deposition

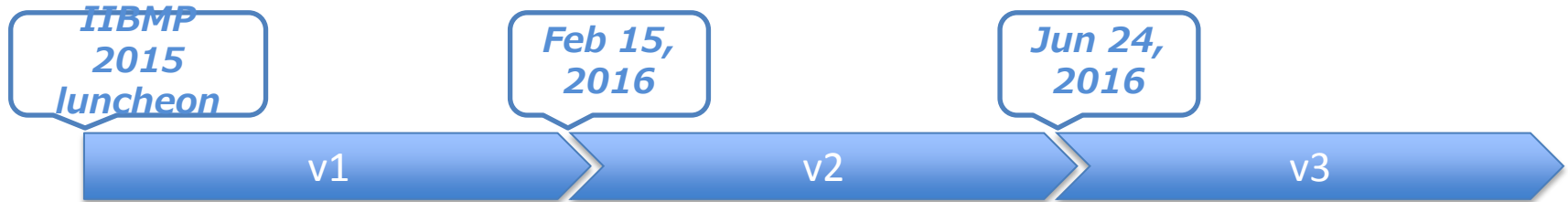
BMRBxTool and BMRBoTool  
Tools for interoperability of BMRB

BMRB/XML & BMRB/RDF Data Server  
Web services utilizing XML and RDF

## Information



# Timeline of PDBj-BMRB search service



- Keyword search
  - Sequence search
  - Auto suggest
  - Search settings
  - Search filter
  - BMRB, PDB, Swiss-Prot, Ligand expo and BMRB-Metabolomics
  - Mutual reference using PubMed, BLAST, IntAct and PubChem
  - Entry list view
  - Sort search results
- Lower latency (4~100 times faster)
  - System stability
  - EMDB integration
- Entry page with static URL
  - Entry tracking by Dep. ID
  - Query history
  - Search help page
  - Unified file exportation
  - Chart displays
  - Entries sharing articles
  - Details of experiment
  - Hit context with highlight
  - Statistics on search results



## Basic query composition

A query is broken up into single term, phrases (a group of terms surrounded by double quotes) and operators. Multiple terms and phrases separated by a white space are automatically conjugated by "AND" operator.

The query syntax allows you to use:

- **phrases** in the specified attribute (e.g. `data_set.type:order_parameters`)
- **wild card** (either ? or \* character in term)
- **regular expression** surrounded by slashes
- **fuzzy search** based on Demerau-Levenshtein distance using a tilde character in term end, proximity search(e.g. `"ph domain"~3`, which searches for "ph" and "domain" within 3 words.)
- **range search** for decimal or date attributes (e.g. `entry.original_release_date:[20010101 TO 20051231]`)
- a series of Boolean operators ("**OR**", "**AND**" and "**NOT**")
- **grouping** using parentheses to form sub queries that is useful if you want to control Boolean logic for a query.

**What entries can be retrieved using the query below?**

`"ph domain"~3 AND data_set.type:order_parameters`



# Search results


Found 467 Documents (0.735 - 1.365 seconds) Order by Export as

**BMRB** 98 **PDB** 104 **EMDB** 3 **Swiss-Prot** 262 **Metabolomics** **Ligand Expo**

Showing 1 - 10 of 98 Documents, order by Relevance

1 **BMRB: 17103**

The structure of the calcium-sensitizer, dfbp-o, in complex with the N-domain of **troponin** C and the switch region of **troponin** I



2L1R  
score: 1750.000

**Authors:** [Robertson, I.M.](#), [Sun, Y.](#), [Li, M.X.](#), [Sykes, B.D.](#)

**Assembly:** calcium-sensitizer, dfbp-o, in complex with **troponin** C and the switch region of **troponin** I

**Entity:**

1. **cTnI**(144-163) (polymer), 89 monomers, 10062.23 Da [Detail](#)
2. **cTnI**(144-163) (polymer), 20 monomers, 2214.660 Da [Detail](#)
3. **CA** (non-polymer), 40.078 Da
4. **dfbp-o** (non-polymer), 197.239 Da

**Total weight:** 12514.208 Da

**Max. entity weight:** 10062.23 Da

**Source organism:** [Homo sapiens](#)

**Exptl. method:** [SOLUTION NMR](#)

**Refine. method:** [simulated annealing](#)

**Data set:** [assigned\\_chemical\\_shifts](#)

**Chem. Shift Compl.:** Sequence coverage: **97.2** %, Completeness: **72.7** %, Completeness (bb): **73.4** % [Detail](#)

**Release date:** 2009-12-08

**Citation:** **A structural and functional perspective into the mechanism of Ca<sup>2+</sup>-sensitizers that target the cardiac **troponin** complex** [↗](#)  
[Robertson, I.M.](#), [Sun, Y.](#), [Li, M.X.](#), [Sykes, B.D.](#)  
*J. Mol. Cell. Cardiol.* (2010), **49**, 1031-1041, PubMed: [20801130](#) [↗](#), DOI: [10.1016/j.yjmcc.2010.08.019](#) [↗](#), [Abstract](#)

**Entries sharing articles:** PDB: **1** entries [Detail](#)

**Related entities:** 2. **cTnI**(144-163), **1** **4** **10** **18** entities [Detail](#)

**Interaction partners:** 2. **cTnI**(144-163), **18** interactors [Detail](#)

**Experiments performed:** **13** experiments [Detail](#)

**Keywords:** Ca<sup>2+</sup>-sensitizer, dfbp-o, N-domain, **troponin** C, **troponin** I

2 **BMRB: 16752**

Mapping **1** 2 3 4 5 6 7 8 9 10



# Unified file exportation of derived data

## BMRB Mirror server:

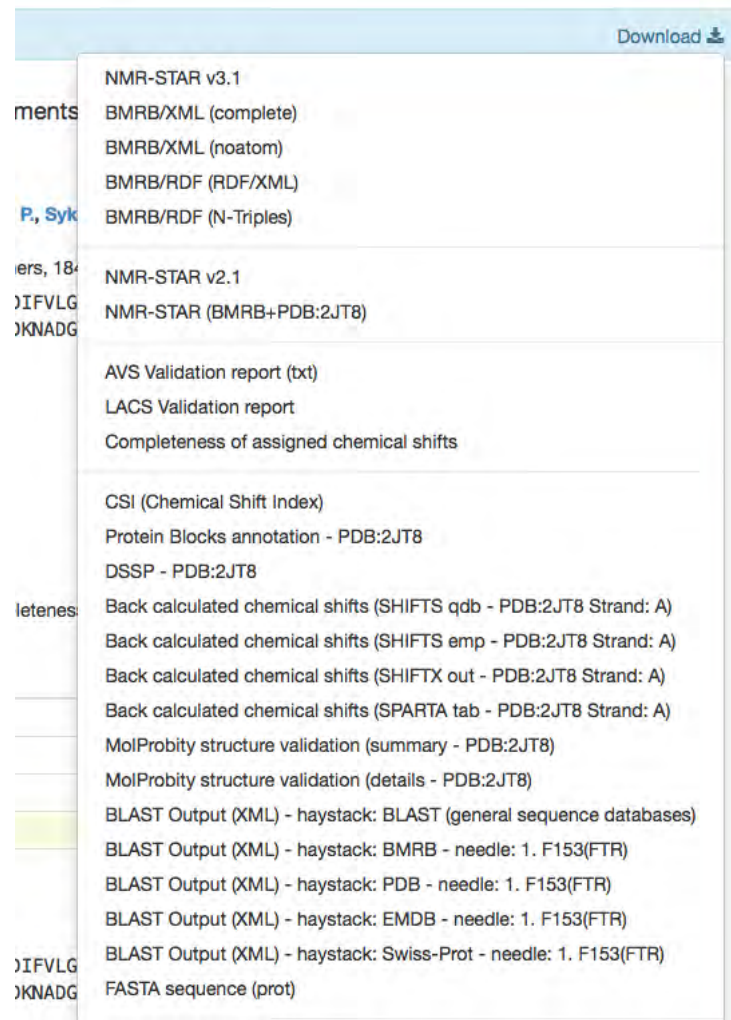
- NMR-STAR
- [PDB+BMRB\\*](#)
- AVS validation report
- LACS validation report
- CSI
- CS-Rosetta structures
- [DSSP\\*](#)
- [Watson Click basepairs \(Wattos\)\\*](#)
- [Backcalculated CS \(SHIFTS, SHIFTX, SPARTA\)\\*](#)
- [MolProbity validation\\*](#)
- BLAST Output (general)
- Timedomain

## BMRB/XML & BMRB/RDF Data server:

- BMRB/XML
- BMRB/RDF
- CS completeness
- Protein Blocks (NMR structure)
- [Protein Blocks \(related X-ray structure\)\\*](#)
- BLAST Output (DB specific)
- FASTA

## PDBj-BMRB entry page:

Download menu button



\* represents derived data accessible by PDB ID

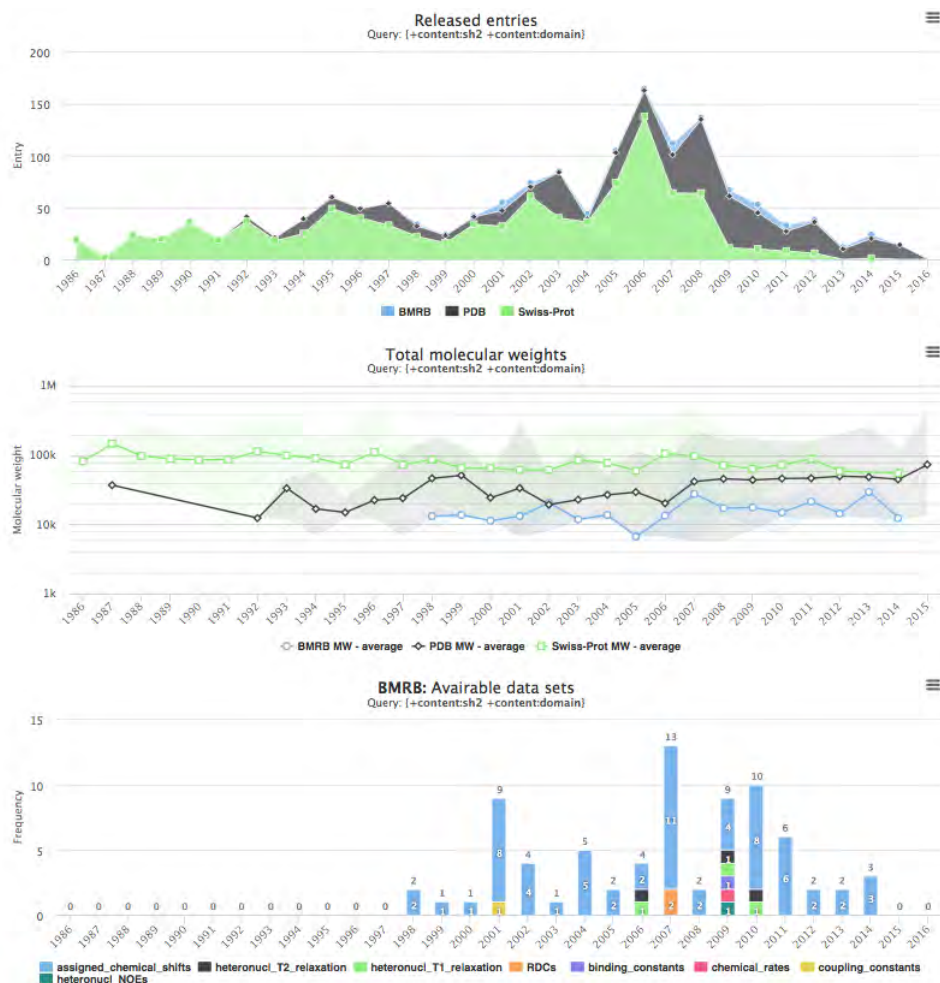




# Statistics on search results

Statistics on search results are shown as charts summarizing the following points:

- Released entries
- Total molecular weights
- Available data sets (BMRB)
- Completeness of assigned chemical shifts
- Highest resolution (PDB, EMDB)
- R-value (PDB)
- Applied experimental methods
- Applied refinement methods
- Cited journal names
- Most frequent keywords
- Most frequent entry author names





## Chart displays: Assigned chemical shifts

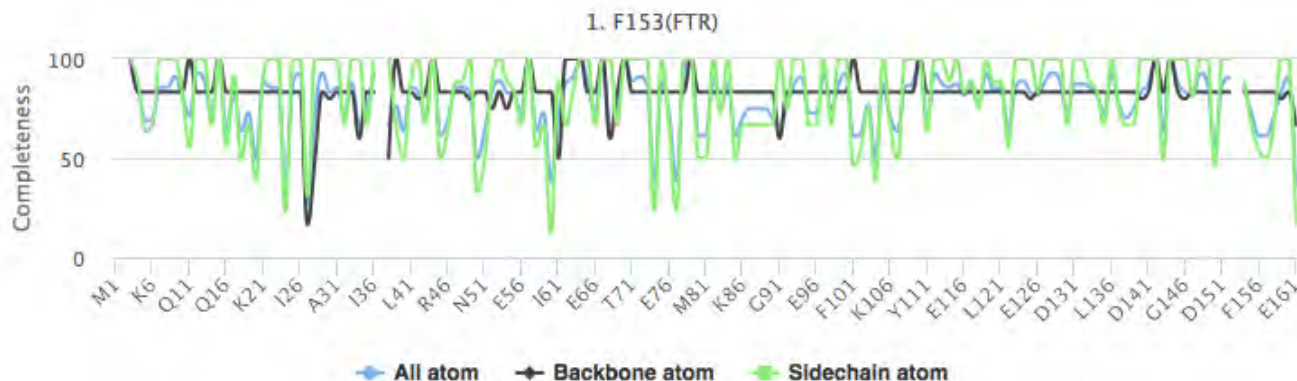
**Chem. Shift Compl.:** Sequence coverage: **97.5 %**, Completeness: **75.9 %**, Completeness (bb): **81.7 %** [Detail](#)   
Polymer type: **polypeptide(L)**

[Download](#)

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
All	75.9 % (1398 of 1842)	76.8 % (737 of 960)	71.7 % (509 of 710)	88.4 % (152 of 172)
Backbone	81.7 % (771 of 944)	95.6 % (304 of 318)	67.5 % (316 of 468)	95.6 % (151 of 158)
Sidechain	73.8 % (772 of 1046)	67.6 % (434 of 642)	86.4 % (337 of 390)	7.1 % (1 of 14)
Aromatic	24.0 % (25 of 104)	25.0 % (13 of 52)	23.1 % (12 of 52)	
Methyl	100.0 % (148 of 148)	100.0 % (74 of 74)	100.0 % (74 of 74)	

### 1. F153(FTR)

MDDIYKAAVE QLTEEQKNEF KAAFDIFVLG AEDGSISTKE LGKVMRMLGQ NPTPEELQEM IDEVDEDGSG TVDFDEFLVM  
MVRSMKDDSK GKSEEELSDL FRMFDKNADG YIDLDELKIM LQATGETITE DDIEELMKDG DKNNDGRIDY DEXLEFMKGV  
E



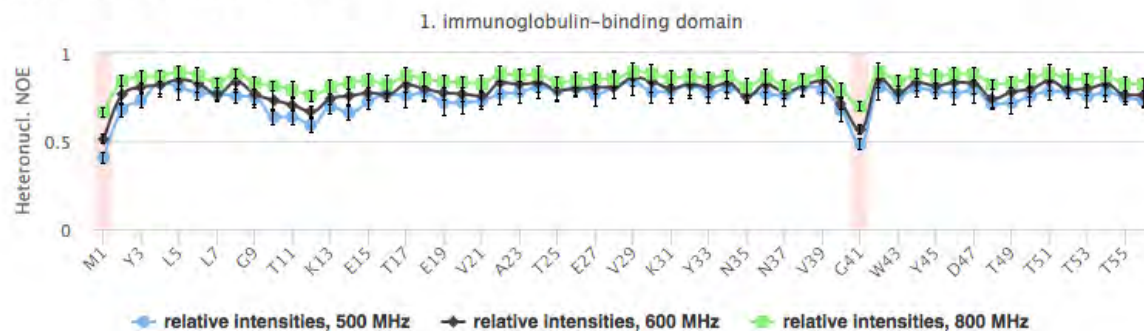
[+ Show sample condition](#)



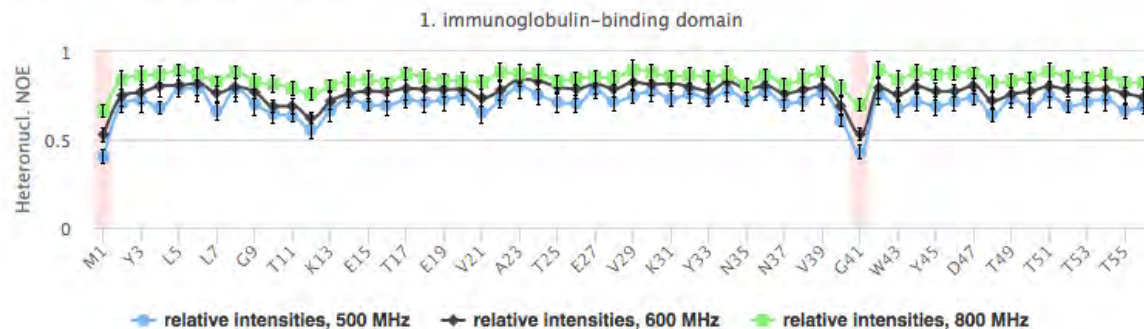
# Chart displays:

## Heteronuclear $T_1$ , $T_2$ , NOE and $S^2$

**Exptl. method:** NMR  
**Data set:** heteronucl\_NOEs, heteronucl\_T1\_relaxation, heteronucl\_T2\_relaxation  
**Heteronucl.  $T_1$ :** 1680  $T_1$  values in 30 lists  
**Coherence:** Sz, **Field strength ( $^1\text{H}$ ):** 500 MHz, 600 MHz, 800 MHz, **Temperature:** 278 ( $\pm 0.1$ ) K, 283 ( $\pm 0.1$ ) K, 288 ( $\pm 0.1$ ) K, 293 ( $\pm 0.1$ ) K, 298 ( $\pm 0.1$ ) K, 303 ( $\pm 0.1$ ) K, 308 ( $\pm 0.1$ ) K, 313 ( $\pm 0.1$ ) K, 318 ( $\pm 0.1$ ) K, 323 ( $\pm 0.1$ ) K, **pH:** 5.25 ( $\pm 0.1$ ) [Detail](#)  
**Heteronucl.  $T_2$ :** 1680  $T_2$  values in 30 lists  
**Coherence:** S(+,-), **Field strength ( $^1\text{H}$ ):** 500 MHz, 600 MHz, 800 MHz, **Temperature:** 278 ( $\pm 0.1$ ) K, 283 ( $\pm 0.1$ ) K, 288 ( $\pm 0.1$ ) K, 293 ( $\pm 0.1$ ) K, 298 ( $\pm 0.1$ ) K, 303 ( $\pm 0.1$ ) K, 308 ( $\pm 0.1$ ) K, 313 ( $\pm 0.1$ ) K, 318 ( $\pm 0.1$ ) K, 323 ( $\pm 0.1$ ) K, **pH:** 5.25 ( $\pm 0.1$ ) [Detail](#)  
**Heteronucl. NOE:** 1680 NOE values in 30 lists  
**Value type:** relative intensities, **Field strength ( $^1\text{H}$ ):** 500 MHz, 600 MHz, 800 MHz, **Temperature:** 278 ( $\pm 0.1$ ) K, 283 ( $\pm 0.1$ ) K, 288 ( $\pm 0.1$ ) K, 293 ( $\pm 0.1$ ) K, 298 ( $\pm 0.1$ ) K, 303 ( $\pm 0.1$ ) K, 308 ( $\pm 0.1$ ) K, 313 ( $\pm 0.1$ ) K, 318 ( $\pm 0.1$ ) K, 323 ( $\pm 0.1$ ) K, **pH:** 5.25 ( $\pm 0.1$ ) [Detail](#)  
**Temperature:** 278 ( $\pm 0.1$ ) K, **pH:** 5.25 ( $\pm 0.1$ )



Temperature: 283 ( $\pm 0.1$ ) K, pH: 5.25 ( $\pm 0.1$ )





# PDBj-BMRB web services to enhance interoperability of NMR data



## **BMRB/XML, BMRB/RDF**

representations of BMRB NMR-STAR data in XML and RDF formats



## **PDBj-BMRB integrated search service**

search biological/biochemical DBs at once



## **PDBj-BMRB SPARQL server**

a programmable API for federated search



## **BMRB SQL on your PC**

periodically updated BMRB relational DB for intensive search



# PDBj-BMRB SPARQL server: Graphical interface for querying RDF data

<http://bmrpub.protein.osaka-u.ac.jp/search/rdf>

## Virtuoso SPARQL Query Editor

[About](#) | [Namespace Prefixes](#) | [Inference rules](#)

**Default Data Set Name (Graph IRI)**

**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:**

**Execution timeout:**  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](#))*



# PDBj-BMRB SPARQL server:

## A programmable API for federated search

### 📄 Graph URIs

- <http://bmrpub.protein.osaka-u.ac.jp/rdf/bmr> (for BMRB entries)
- <http://bmrpub.protein.osaka-u.ac.jp/rdf/bms> (for Metabolomics entries)
- <http://rdf.wwpdb.org/pdb> (for PDB entries)

### 📄 Prefixes

- `rdf:` <<http://www.w3.org/1999/02/22-rdf-syntax-ns#>>
- `rdfs:` <<http://www.w3.org/2000/01/rdf-schema#>>
- `owl:` <<http://www.w3.org/2002/07/owl#>>
- `BMRBo:` <[http://bmrpub.protein.osaka-u.ac.jp/schema/mmcif\\_nmr-star.owl#](http://bmrpub.protein.osaka-u.ac.jp/schema/mmcif_nmr-star.owl#)>
- `PDBo:` <<http://rdf.wwpdb.org/schema/pdbx-v40.owl#>>

### 📄 Reference Ontologies

- [BMRB/OWL](#): BMRB/RDF ontology
- [PDBx OWL](#): PDB/RDF ontology

### 📄 How to run SPARQL query via **curl** command. Please include **FROM** clause in a query to select Graph URI.

```
% curl -F "query=QUERY" -F "format=FORMAT" http://bmrpub.protein.osaka-u.ac.jp/search/rdf
```

```
% curl -F "query=@QFILE" -F "format=FORMAT" http://bmrpub.protein.osaka-u.ac.jp/search/rdf
```

, where **QUERY** and **QFILE** stand for actual query strings and query file name, respectively. **FORMAT** represents MIME type selected from either **text/html**, **application/vnd.ms-excel**, **application/results-xml**, **application/results-json**, **text/plain**, **application/rdf+xml**, **text/csv** or **text/tab-separated-values**.



# Basic SPARQL query composition

# Count the number of primary publications cited by BMRB entries per year:

**PREFIX** BMRBo: <http://bmrbbpub.protein.osaka-u.ac.jp/schema/mmcif\_nmr-star.owl#>

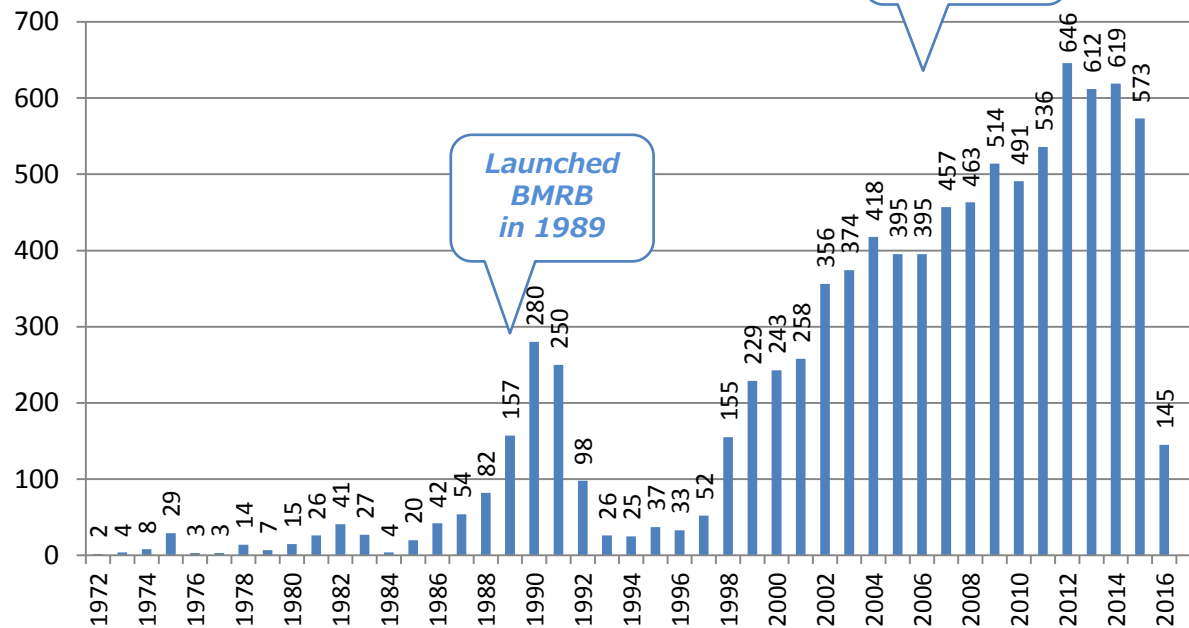
**SELECT** ?year (COUNT(?year) AS ?count)

**FROM** <http://bmrbbpub.protein.osaka-u.ac.jp/rdf/bmr>

**WHERE** {

    ?s\_cit BMRBo:citation.id "1" ;  
    BMRBo:citation.year ?year .

} **GROUP BY** ?year **ORDER BY** ?year



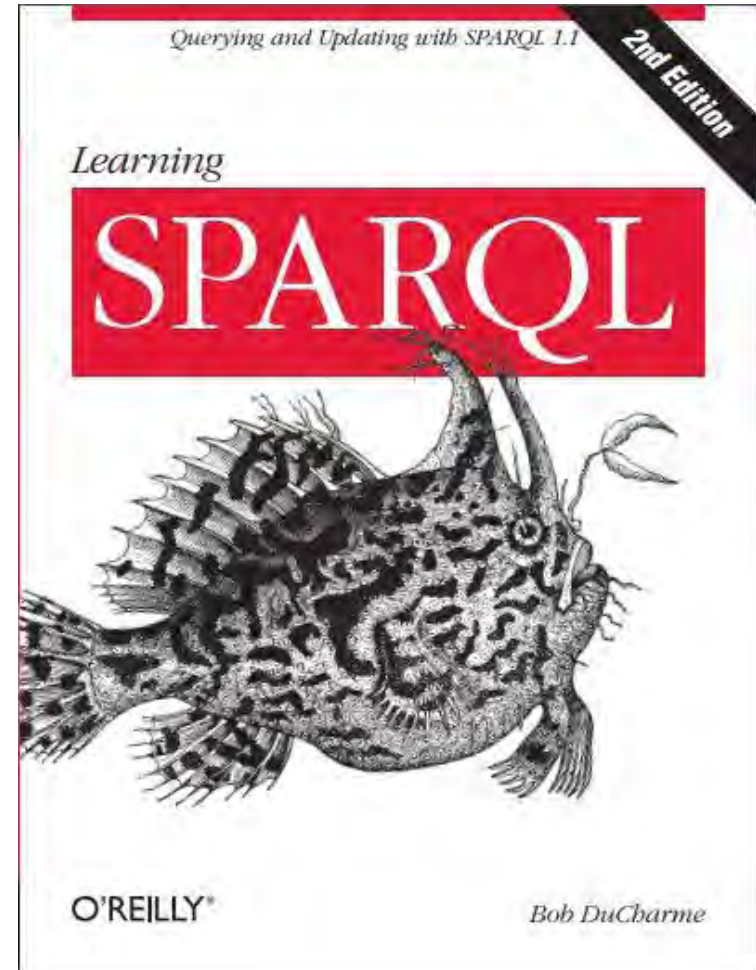


# Federated SPARQL query using 'SERVICE' clause

... While typical SPARQL query or subquery retrieves data from somewhere local or remote and applies a query to it, the **SERVICE** keyword lets you say “send this query off to the specified SPARQL endpoint service, which will run the query, and then retrieve the result.” It’s a great keyword to know, ...

*Learning SPARQL by Bob DuCharme*

## Reference





# Retrieve a list of MeSH (Medical Subject Headings) words in publications of a period of time:

**PREFIX** BMRBo: <[http://bmrpub.protein.osaka-u.ac.jp/schema/mmcif\\_nmr-star.owl#](http://bmrpub.protein.osaka-u.ac.jp/schema/mmcif_nmr-star.owl#)>

**PREFIX** pubmed\_v: <[http://bio2rdf.org/pubmed\\_vocabulary:](http://bio2rdf.org/pubmed_vocabulary:)>

**SELECT** ?word (COUNT(?word) **AS** ?count)

**FROM** <<http://bmrpub.protein.osaka-u.ac.jp/rdf/bmr>> **FROM** <<http://bio2rdf.org/pubmed>>

**WHERE** {

**SELECT DISTINCT** ?pubmed\_id ?word

**WHERE** {

?s\_citation BMRBo:citation.pubmed\_id ?pubmed\_id ; BMRBo:citation.year ?year .

**FILTER** (bound(?pubmed\_id) && xsd:integer(?year) >= 2001 && xsd:integer(?year) <= 2010)

**BIND** (IRI(CONCAT("http://bio2rdf.org/pubmed:", ?pubmed\_id)) **AS** ?s\_pubmed)

**SERVICE** <<http://cu.pubmed.bio2rdf.org/sparql>> {

?s\_pubmed pubmed\_v:mesh\_heading ?s\_meshhd .

?s\_meshhd pubmed\_v:mesh\_descriptor\_name ?mesh\_descriptor .

}

**FILTER NOT EXISTS** { ?s\_meshhd pubmed\_v:mesh\_qualifier\_name ?mesh\_qualifier .

**BIND** ((IF (CONTAINS(?mesh\_descriptor, ","), STRBEFORE(?mesh\_descriptor, ","), ?mesh\_descriptor)) **AS** ?word)

**FILTER** (?word **NOT IN** ("Magnetic Resonance Spectroscopy", "Nuclear Magnetic Resonance"))

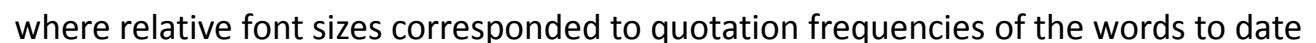
**FILTER** (?word **NOT IN** ("X-Ray Diffraction", "X-rays", "Crystallography", "Crystallization"))

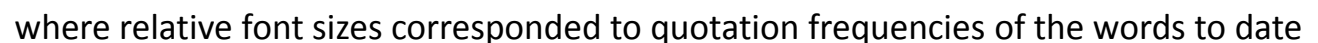
}

**ORDER BY DESC**(?count)

Total query execution time: about 10 minutes





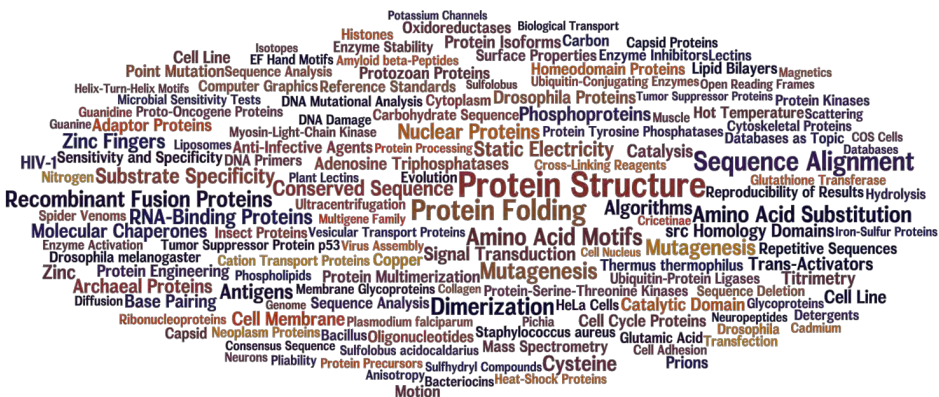




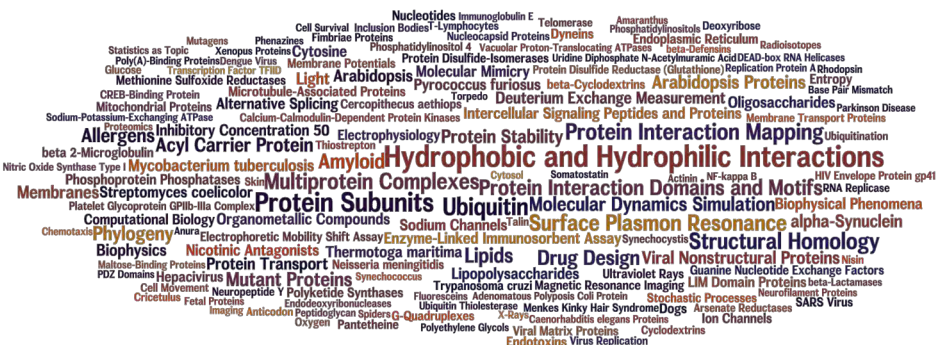


# Compering BMRB and PDB by the MeSH words in the same period

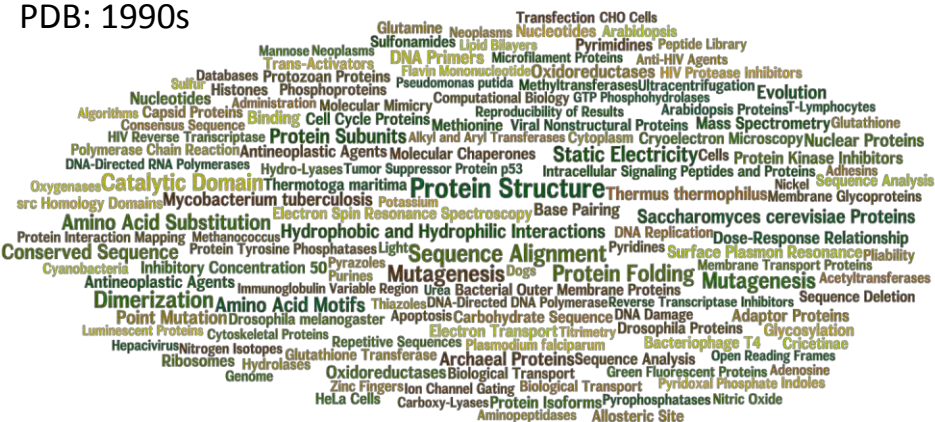
BMRB: 1990s



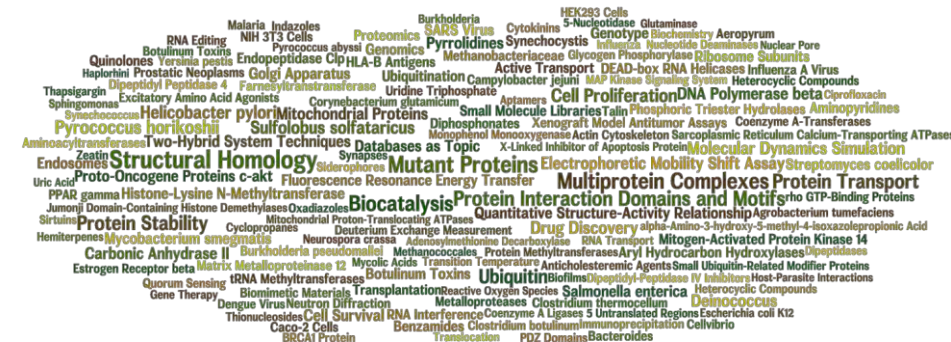
BMRB: 2000s



PDB: 1990s



PDB: 2000s



where relative font sizes corresponded to quotation frequencies of the words to date



# Retrieve phenotypes annotated with the information for SNPs from the human genome in BMRB entities:

**PREFIX** BMRBo: <http://bmrpub.protein.osaka-u.ac.jp/schema/mmcif\_nmr-star.owl#>

**PREFIX** rdfs: <http://www.w3.org/2000/01/rdf-schema#>

**PREFIX** omim\_v: <http://bio2rdf.org/omim\_vocabulary:>

```
SELECT DISTINCT ?entity_id ?uniprot_id ?label ?omim_id ?dbsnp_id ?mutation ?phenotype
FROM <http://bmrpub.protein.osaka-u.ac.jp/rdf/bmr> FROM <http://purl.uniprot.org/uniprot>
FROM <http://bio2rdf.org/omim>
WHERE {
    ?s_up BMRBo:entity_db_link.entry_id "4280" ; BMRBo:entity_db_link.entity_id ?entity_id ;
        BMRBo:entity_db_link.database_code "SP" ; BMRBo:entity_db_link.accession_code ?uniprot_id ;
        rdfs:seeAlso ?s_uniprot .
    ?s_uniprot rdfs:label ?info .
```

**FILTER** (STRSTARTS(?info, "info:uniprot"))

```
SERVICE <http://sparql.uniprot.org/sparql> {
    ?s_uniprot rdfs:label ?label ; rdfs:seeAlso ?o_purl .
}
```

```
FILTER (STRSTARTS(STR(?o_purl), "http://purl.uniprot.org/mim/"))
BIND (STRAFTER(STR(?o_purl), "http://purl.uniprot.org/mim/") AS ?omim_id)
BIND (IRI(CONCAT("http://bio2rdf.org/omim:", ?omim_id)) AS ?s_omim)
```

```
SERVICE <http://omim.bio2rdf.org/sparql> {
    ?s_omim omim_v:variant ?s_allele .
    ?s_allele omim_v:dbsnp ?s_dbSNP ; omim_v:mutation ?mutation ; rdfs:label ?phenotype .
```

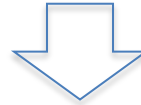
```
    BIND (STRAFTER(STR(?s_dbSNP), "http://bio2rdf.org/dbSNP:") AS ?dbSNP_id)
}
```

Total query execution time: about 125 minutes

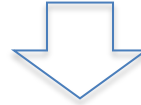


## Federated search beyond DB boundaries: BMRB, UniProt and OMIM

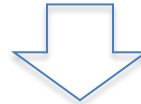
**4597** phenotype annotated SNPs in all BMRB entities



**574** residues having backbone chemical shifts



**74** residues having structural information



Classification by structural parameters archived at BMRB

- Degeneration of protein fold or its function by stop codon: **12**
- Mutation of inherently hydrophilic amino acids (Arg, Lys and Pro) in a hydrophobic environment: **11**
- Mutation of Arg, Lys and Pro in mostly buried residues (rSASA<10%): **15**
- Substitution of residues on the protein surface (rSASA>50%) with different charge: **14**
- Substitution of residues on the protein surface to bulky aromatic residue: **4**
- The other cases having little relation with the structural parameters above: **20**

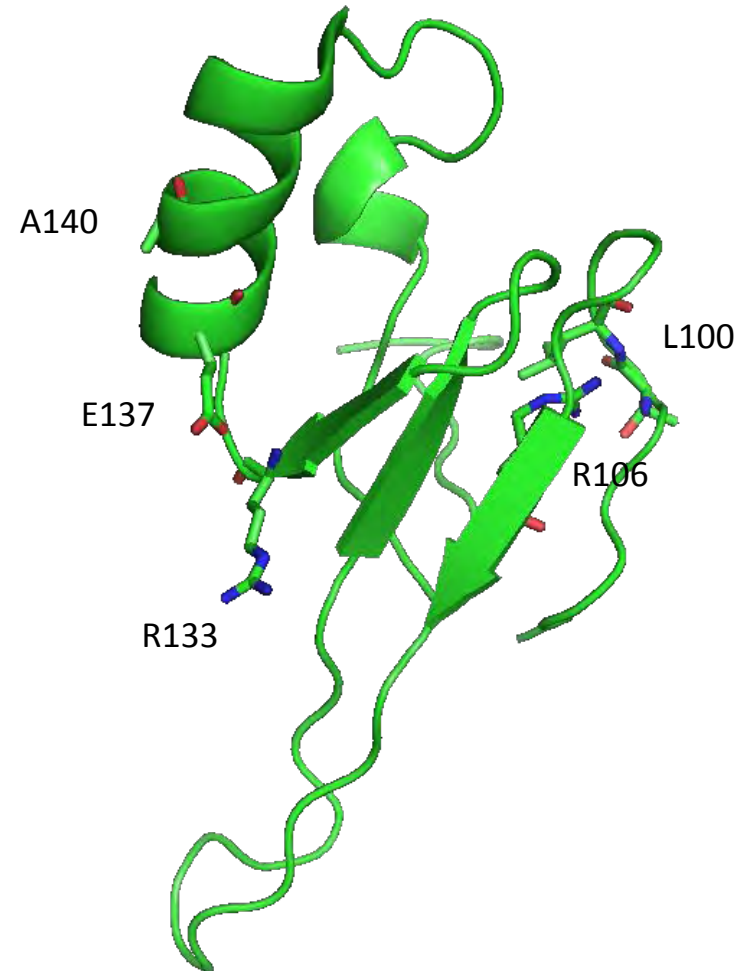


## Federated search beyond DB boundaries: BMRB, UniProt and OMIM

Results performed by the SPARQL query for BMRB entry 4280, showing BMRB ID, mutation, OMIM ID, dbSNP ID, secondary structure, and SASA.

BMRB	Mutation	OMIM	dbSNP	2nd	SASA%
4280	L100V	300005	rs28935168	Coil	22.5
4280	R106W	300005	rs28934907	Strand	8.0
4280	R133C	300005	rs28934904	Coil	49.4
4280	E137G	300005	rs61748392	Helix	41.7
4280	A140V	300005	rs28934908	Helix	42.0

Ribbon models of NMR structure (PDB: 1QK9) of MECP2. The side-chains for the mutation (E137 and A140) exposed to solvent. R133 is critical residue for CpG DNA recognition. L100 and R106 are located in hydrophobic environment. All residues are responsible for X-linked mental retardation (Rett syndrome).

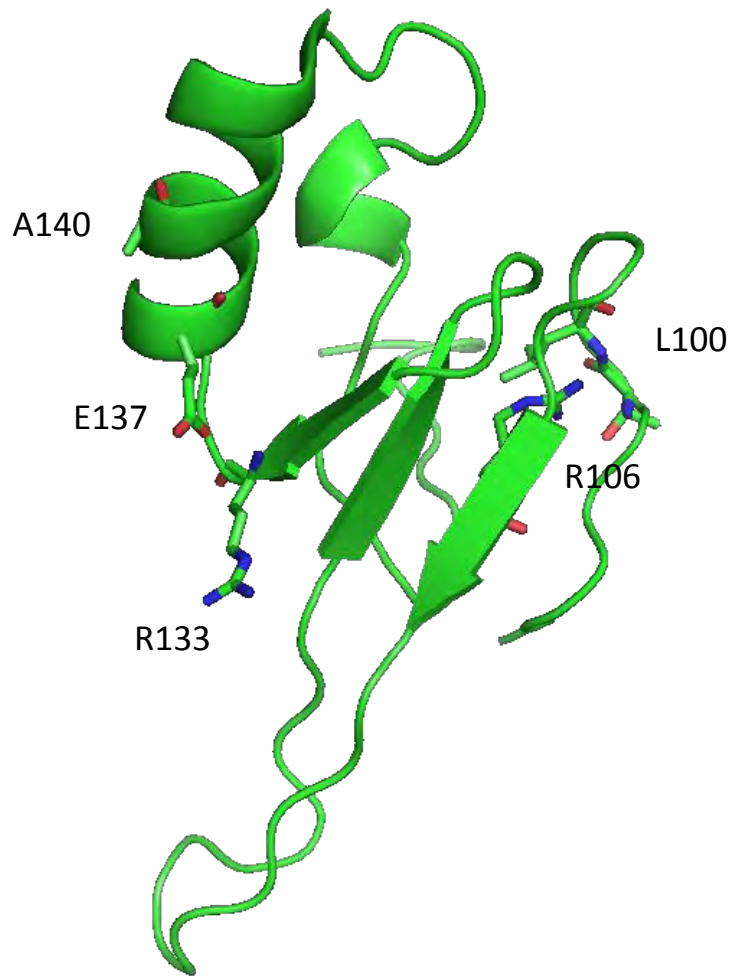


Solution structure of MeCP2 methyl CpG binding domain



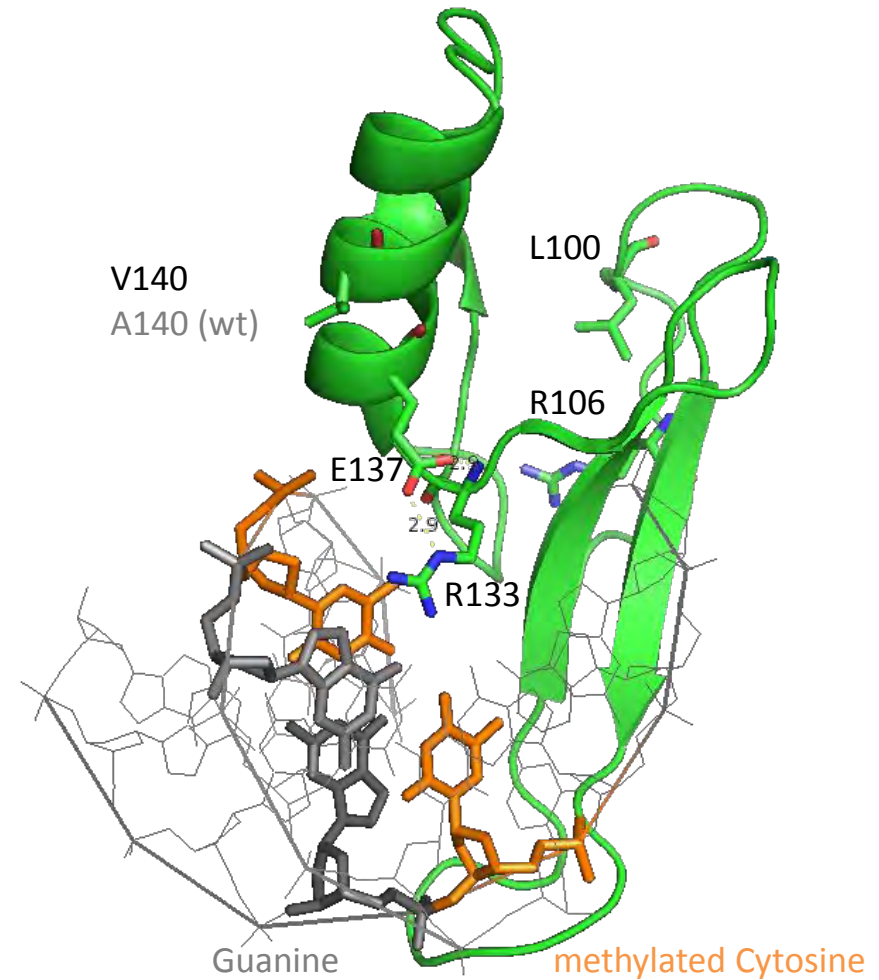


## Rett syndrome related residues of MeCP2 in complex with methylated DNA (CpG motif)



Solution structure of MeCP2 MBD, PDB: 1QK9

*J. Mol. Biol.* (1999) **291**:1055



MeCP2 MBD (A140V) in complex with CpG DNA, PDB: 5BT2

*Scientific report* (2016) **6**:31210



# PDBj-BMRB SPARQL server: Other 30 SPARQL query examples

<http://bmrpub.protein.osaka-u.ac.jp>

## BMRB/XML & BMRB/RDF Data Server:

common open representations of BMRB NMR-STAR data in XML and RDF formats

[Home](#) [Search](#) [Examples](#) [Download](#) [Resources](#) [NEWS](#)

### Virtuoso SPARQL Query Editor

[About](#) | [Namespace Prefixes](#) | [Inference rules](#)

Default Data Set Name (Graph IRI)

Query Text

# 1. Select all category holders of datablock class of BMRB entry 15400:

```
SELECT *  
FROM <http://bmrpub.protein.osaka-u.ac.jp/rdf/bmr>  
WHERE {  
  <http://bmrpub.protein.osaka-u.ac.jp/rdf/bmr15400> ?p ?o .  
}
```

### 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](#)

# PDBj-BMRB web services to enhance interoperability of NMR data



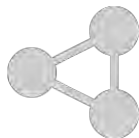
## **BMRB/XML, BMRB/RDF**

representations of BMRB NMR-STAR data in XML and RDF formats



## **PDBj-BMRB integrated search service**

search biological/biochemical DBs at once



## **PDBj-BMRB SPARQL server**

a programmable API for federated search



## **BMRB SQL on your PC**

periodically updated BMRB relational DB for intensive search



BMRB SQL on your PC: periodically updated  
BMRB relational DB for intensive search

### **BMRB PostgreSQL 9.5 dump file (noatom):**

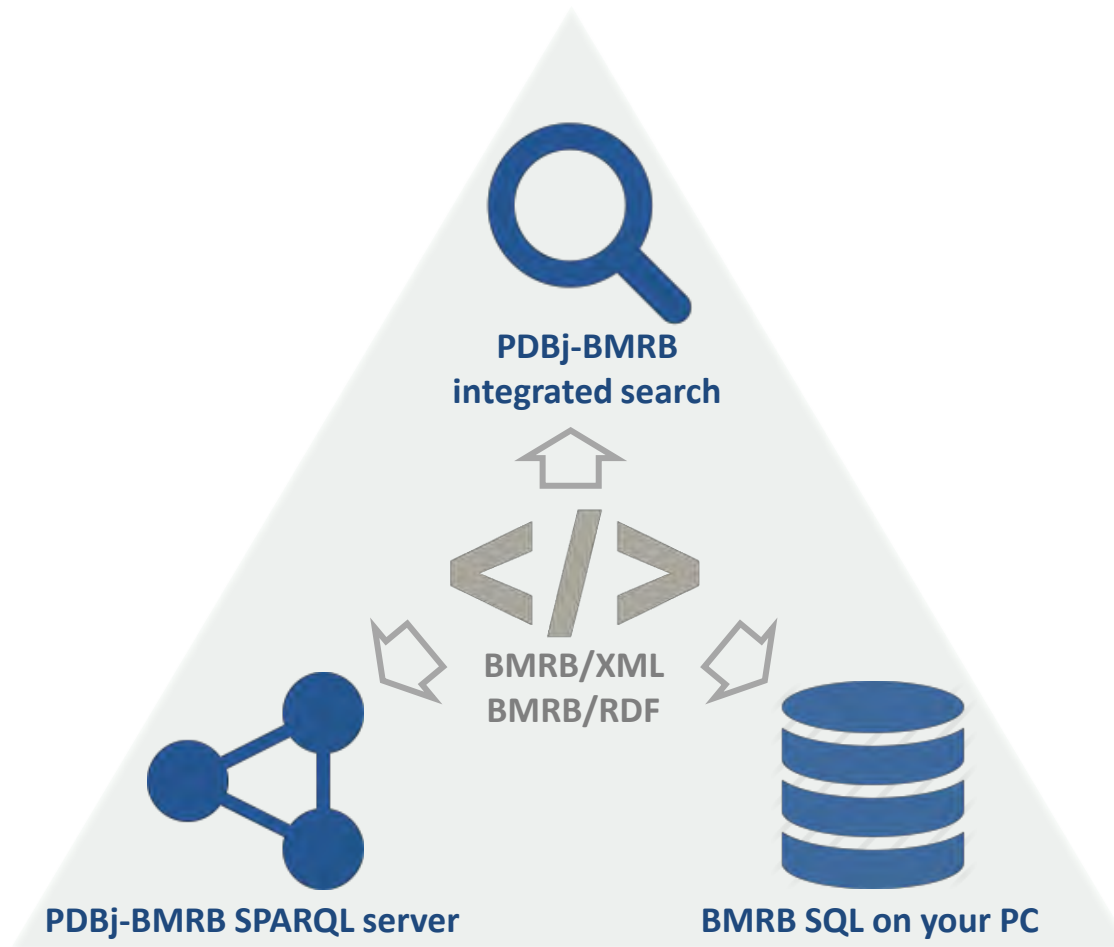
```
% rsync -av --delete rsync://bmrbsub.protein.osaka-u.ac.jp/bmrbsub-pg-dump .  
% pg_restore -c -d $DB_NAME -U $USER bmrbsub_clone.sql
```

- Easy to install
- Updated every week
- Data remediation equivalent to BMRB/XML
- Reducing latency as much as you want



# A selection of PDBj-BMRB web services for researchers

Easy-to-use



Linked data

Low latency

# Collaborators



Takeshi Iwata (岩田武史)

Naohiro Kobayashi, Ph. D. (小林直宏)

Chojiro Kojima, Ph. D. (児島長次郎)

Toshimichi Fujiwara, Ph. D., PI (藤原敏道)



Gert-Jan Bekker

Akira R. Kinjo, Ph. D. (金城玲)

Haruki Nakamura, Ph. D., PI (中村春木)



Ulrich L. Eldon, Ph. D.

John L. Markley, Ph. D., PI