

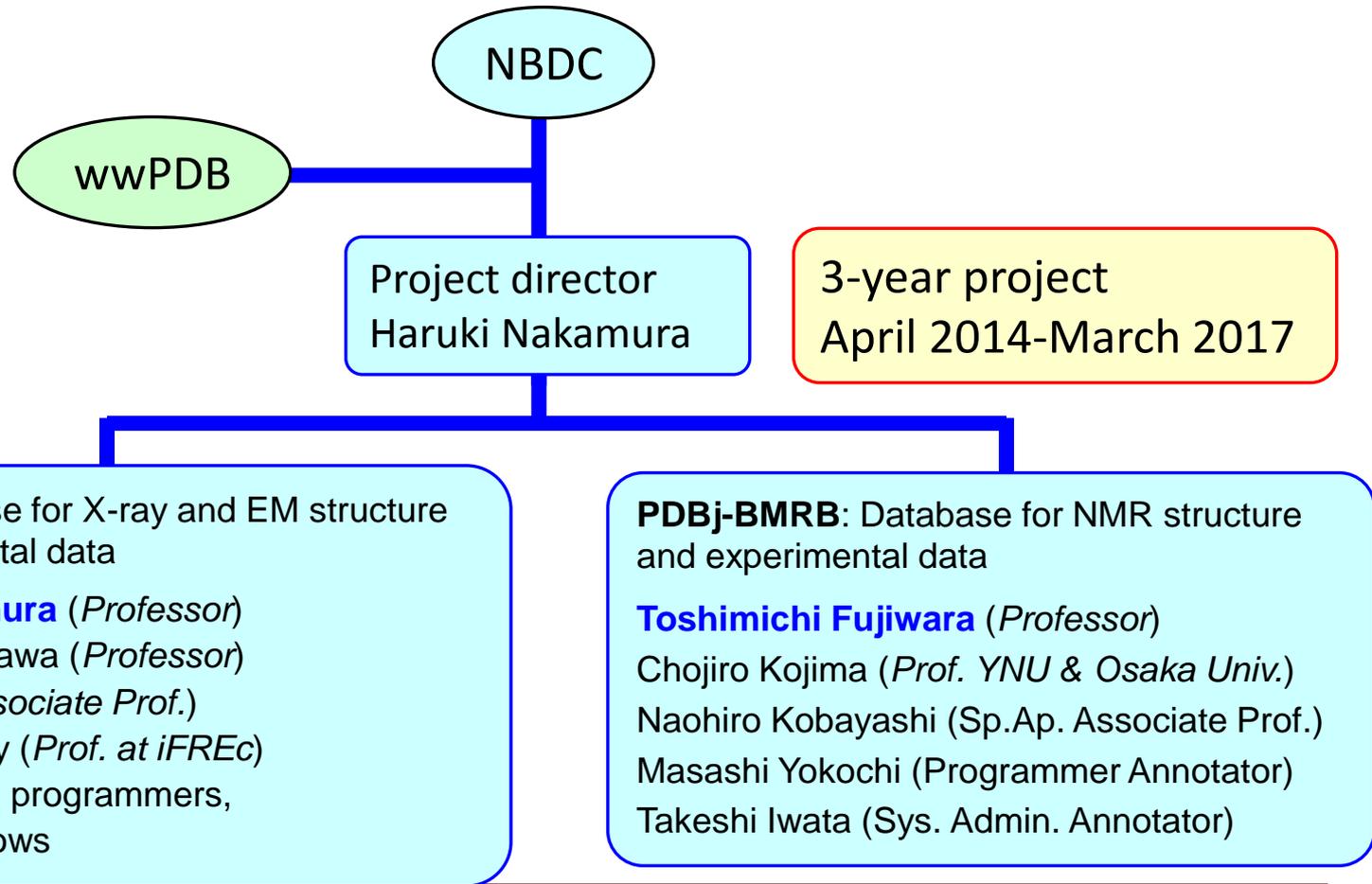
# Application of the integrated NMR database in protein science

Naohiro Kobayashi

PDBj-BMRB group  
Institute for Protein Research  
Osaka university, Japan

# Organization of the project of PDBj at Osaka university supported by JST-NBDC

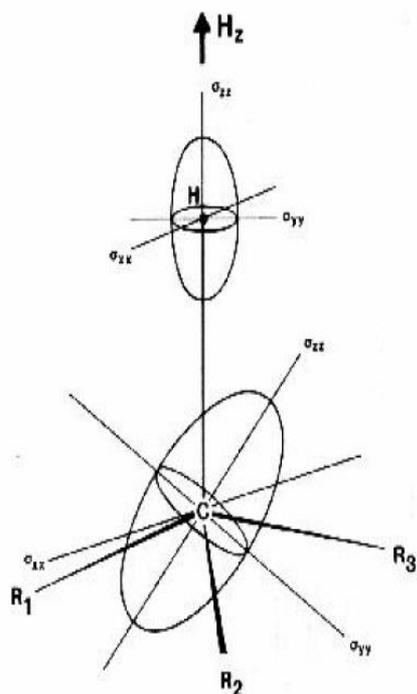
2



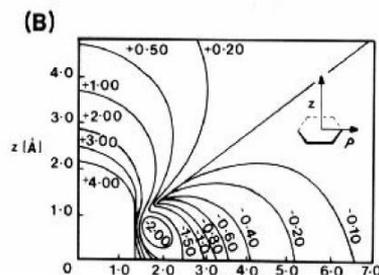
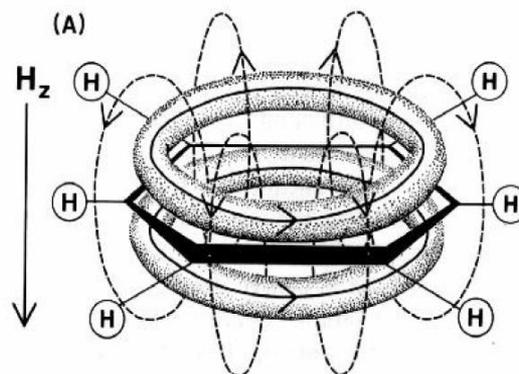
# What is chemical shift?

the most important parameter of NMR...

Chemical shift  
anisotropy



Ring current effect



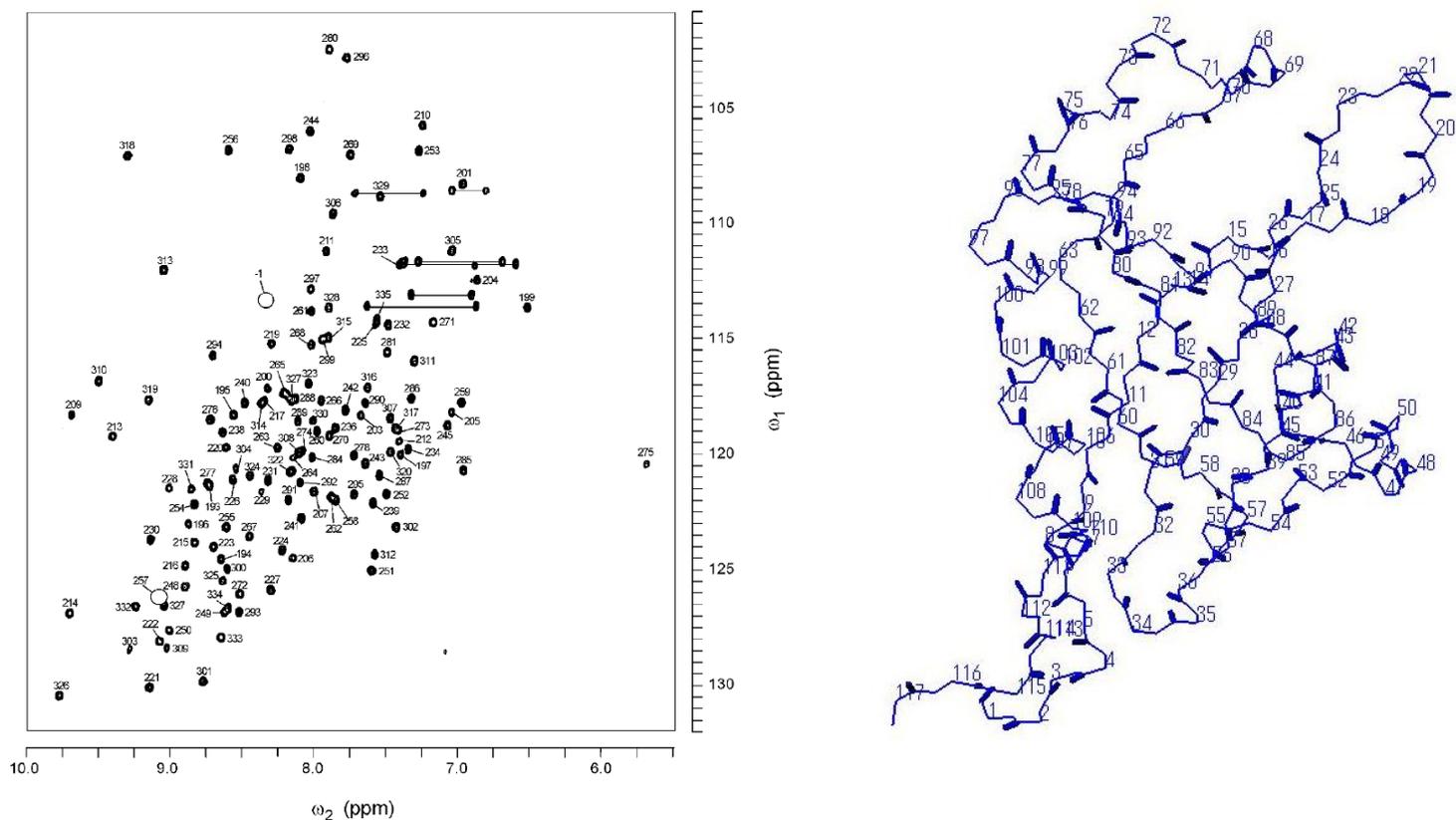
Johnson & Bovey, 1962

**Very sensitive to the  
structure of protein!**

But very difficult to predict  
from structure...

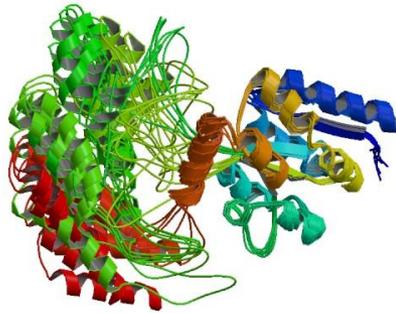
that's why we are collecting  
chemical shift data!

# What you can do with chemical shifts?

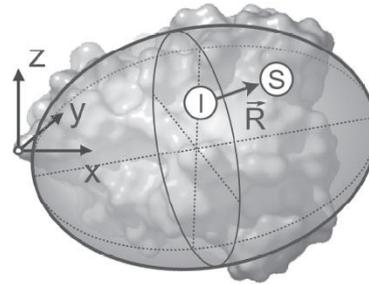


You can use the NMR signal as 100 of probes at atomic resolution

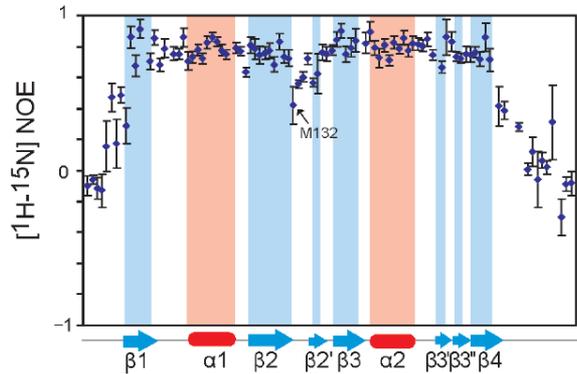
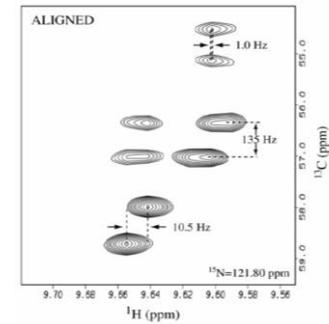
# You can get a variety of information using chemical shifts 6



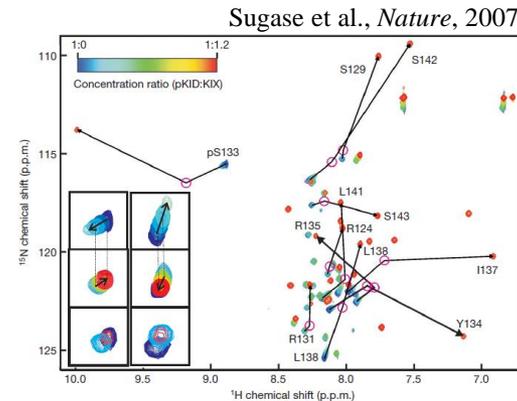
Structure



Domain orientation



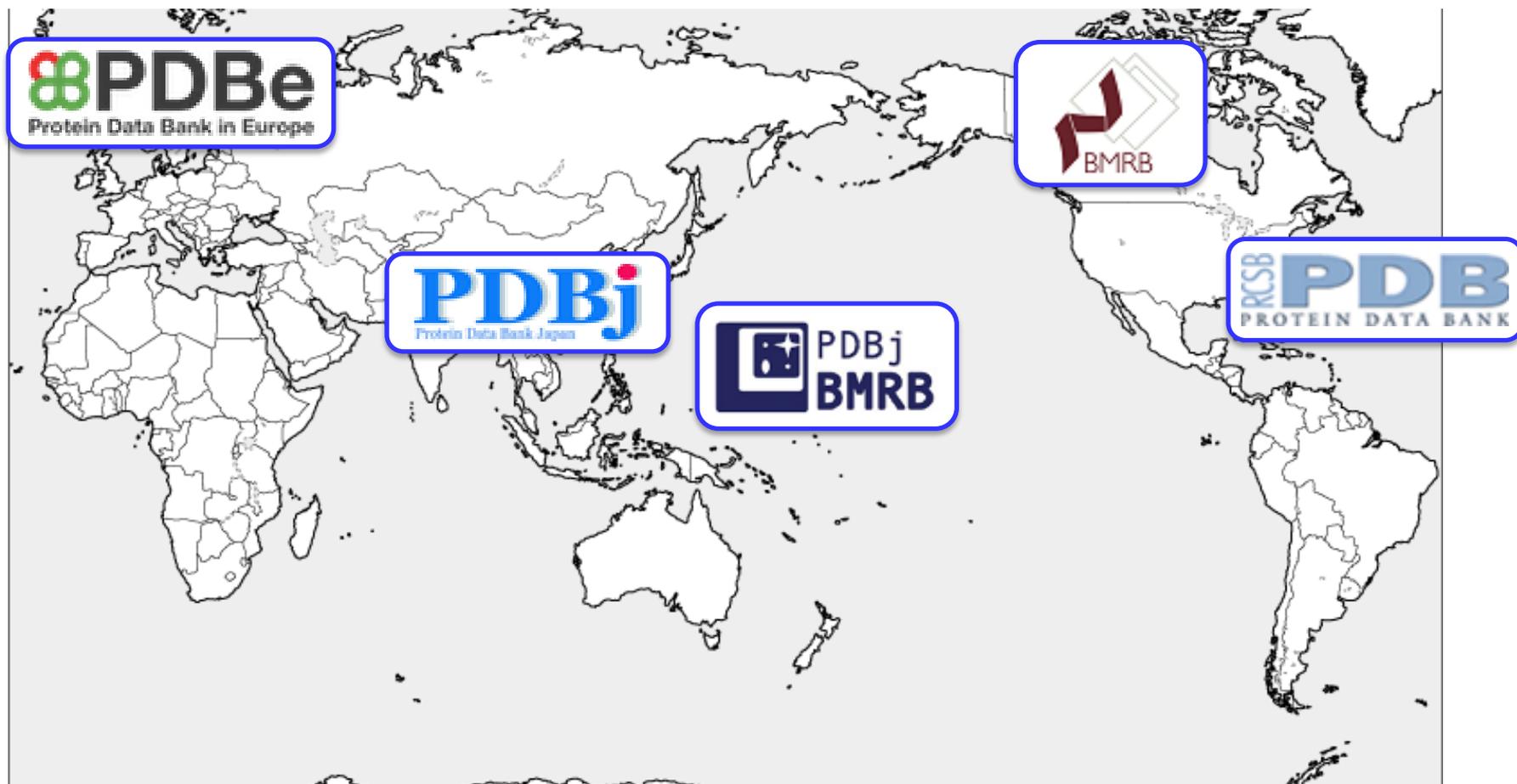
Dynamics



Interaction with ligand

We are collecting chemical shift data in collaboration with Univ. Wisconsin and wwPDB

7

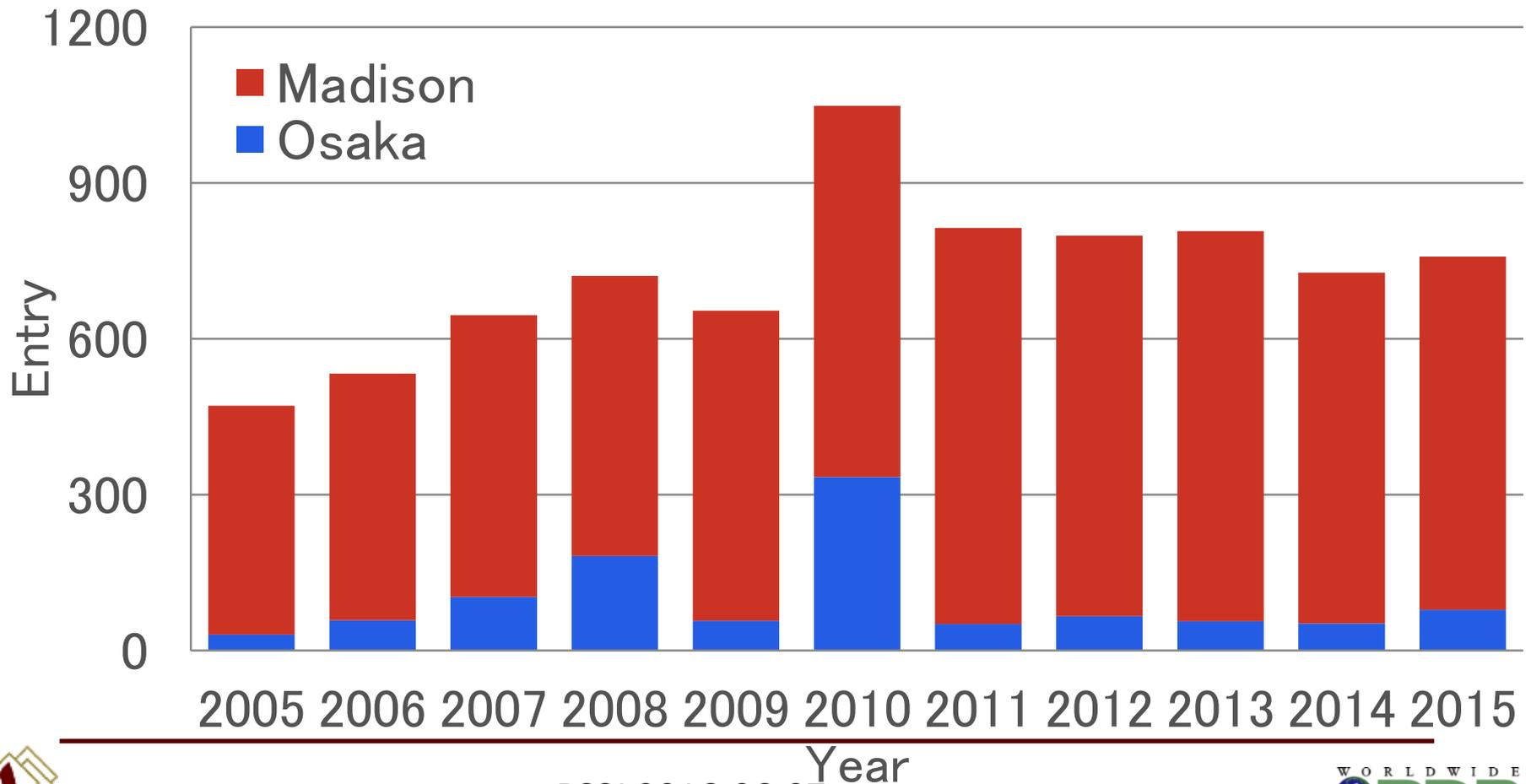


PSSJ 2016-06-07



# Our activities on BMRB processing (2015)

- 78 entries processed at Osaka (~10%)
- BMRB total 759 entries



# Conversion of BMRB data into machine readable format

Yokochi et al., *J. Biomed. Sem.* (2016)

**NMR-STAR v3**

BMRBxTool

**BMRB/XML**

```
#####  
# Entry information #  
#####  
  
save_entry_information  
  _Entry.Sf_category      entry_information  
  _Entry.Sf_framecode    entry_information  
  _Entry.ID              15400  
  _Entry.Title  
;  
Backbone and side chain chemical shift assignments of the F153-to-5-flurotryptophan mutant of human cardiac troponin C  
;  
  _Entry.Version_type    new  
  _Entry.Submission_date 2007-07-20  
  _Entry.Accession_date  2007-07-20  
  _Entry.Last_release_date .  
  _Entry.Original_release_date .  
  _Entry.Origination     author  
  _Entry.NMR_STAR_version 3.0.8.100  
  _Entry.Original_NMR_STAR_version 3.0.8.100  
  _Entry.Experimental_method NMR  
  _Entry.Experimental_method_subtype solution  
  _Entry.Details         .  
  _Entry.BMRB_internal_directory_name .
```

```
<BMRBx:entryCategory>  
  <BMRBx:entry id="15400">  
    <BMRBx:accession_date>2006-12-07+09:00</BMRBx:accession_date>  
    <BMRBx:bmr Internal_directory_name xsi:nil="true"/>  
    <BMRBx:details xsi:nil="true"/>  
    <BMRBx:experimental_method>NMR</BMRBx:experimental_method>  
    <BMRBx:experimental_method_subtype>SOLUTION</BMRBx:experimental_method_subtype>  
    <BMRBx:last_release_date xsi:nil="true"/>  
    <BMRBx:nmr_star_version>3.0.8.100</BMRBx:nmr_star_version>  
    <BMRBx:original_nmr_star_version>3.0.8.100</BMRBx:original_nmr_star_version>  
    <BMRBx:original_release_date xsi:nil="true"/>  
    <BMRBx:origination>author</BMRBx:origination>  
    <BMRBx:sf_category>entry_information</BMRBx:sf_category>  
    <BMRBx:sf_framecode>entry_information</BMRBx:sf_framecode>  
    <BMRBx:submission_date>2006-12-07+09:00</BMRBx:submission_date>  
    <BMRBx:title>Backbone and side chain chemical shift assignments of the F153-to-5-flurotryptophan mutant of human cardiac troponin C</BMRBx:title>  
    <BMRBx:version_type>original</BMRBx:version_type>  
  </BMRBx:entry>  
</BMRBx:entryCategory>
```

bmr15400.str

### BMRB/XML

```
<BMRBx:entryCategory>
  <BMRBx:entry id="15400">
    <BMRBx:accession_date>2006-12-07+09:00</BMRBx:accession_date>
    <BMRBx:bmr_internal_directory_name xsi:nil="true"/>
    <BMRBx:details xsi:nil="true"/>
    <BMRBx:experimental_method>NMR</BMRBx:experimental_method>
    <BMRBx:experimental_method_subtype>SOLUTION</BMRBx:experimental_method_subtype>
    <BMRBx:last_release_date xsi:nil="true"/>
    <BMRBx:nmr_star_version>3.0.8.100</BMRBx:nmr_star_version>
    <BMRBx:original_nmr_star_version>3.0.8.100</BMRBx:original_nmr_star_version>
    <BMRBx:original_release_date xsi:nil="true"/>
    <BMRBx:origination>author</BMRBx:origination>
    <BMRBx:sf_category>entry_information</BMRBx:sf_cate
    <BMRBx:sf_framecode>entry_information</BMRBx:sf_fro
    <BMRBx:submission_date>2006-12-07+09:00</BMRBx:subr
    <BMRBx:title>Backbone and side chain chemical shift
5-flurotryptophan mutant of human cardiac troponin C</BMR
    <BMRBx:version_type>original</BMRBx:version_type>
  </BMRBx:entry>
</BMRBx:entryCategory>
```

bmr15400.xml

BMRBoTool

### BMRB/RDF

```
<BMRBo:has_entryCategory>
  <BMRBo:entryCategory rdf:about="http://bmrpub.protein.osaka-u.ac.jp/rdf/bmr15
    <BMRBo:has_entry>
      <BMRBo:entry rdf:about="http://bmrpub.protein.osaka-u.ac.jp/rdf/bmr15400
        <BMRBo:of_datablock rdf:resource="http://bmrpub.protein.osaka-u.ac.jp
        <BMRBo:entry.id>15400</BMRBo:entry.id>
        <BMRBo:entry.accession_date>2006-12-07+09:00</BMRBo:entry.accession_d
        <BMRBo:entry.experimental_method>NMR</BMRBo:entry.experimental_method
        <BMRBo:entry.experimental_method_subtype>SOLUTION</BMRBo:entry.experi
        <BMRBo:entry.nmr_star_version>3.0.8.100</BMRBo:entry.nmr_star_version
        <BMRBo:entry.original_nmr_star_version>3.0.8.100</BMRBo:entry.origina
        <BMRBo:entry.origination>author</BMRBo:entry.origination>
        <BMRBo:entry.sf_category>entry_information</BMRBo:entry.sf_category>
        <BMRBo:entry.sf_framecode>entry_information</BMRBo:entry.sf_framecode
        <BMRBo:entry.submission_date>2006-12-07+09:00</BMRBo:entry.submission
        <BMRBo:entry.title>Backbone and side chain chemical shift assignments
tryptophan mutant of human cardiac troponin C</BMRBo:entry.title>
        <BMRBo:entry.version_type>original</BMRBo:entry.version_type>
      </BMRBo:entry>
    </BMRBo:has_entry>
  </BMRBo:entryCategory>
</BMRBo:has_entryCategory>
```

# Example of SPARQL search script (Uniprot -> OMIM)

11

```
PREFIX uniprot_c: <http://purl.uniprot.org/core/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX omim_v: <http://bio2rdf.org/omim_vocabulary:>

SELECT DISTINCT ?label ?omim_id ?dbsnp_id ?mutation ?phenotype
FROM <http://purl.uniprot.org/uniprot>
FROM <http://bio2rdf.org/omim_resource:bio2rdf.dataset.omim.R3>
WHERE {
  BIND (IRI(CONCAT("http://purl.uniprot.org/uniprot/", "P51608"))) AS ?s_uniprot)

  SERVICE <http://uniprot.bio2rdf.org/sparql>
  {
    ?s_uniprot uniprot_c:recommendedName ?s_name .
    ?s_name uniprot_c:fullName ?label .
    ?s_uniprot 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:x-dbsnp ?s_dbsnp ;
      omim_v:mutation ?mutation ;
      rdfs:label ?phenotype .
  }
  BIND (STRAFTER(STR(?s_dbsnp), "http://bio2rdf.org/dbsnp:") AS ?dbsnp_id)
}
```

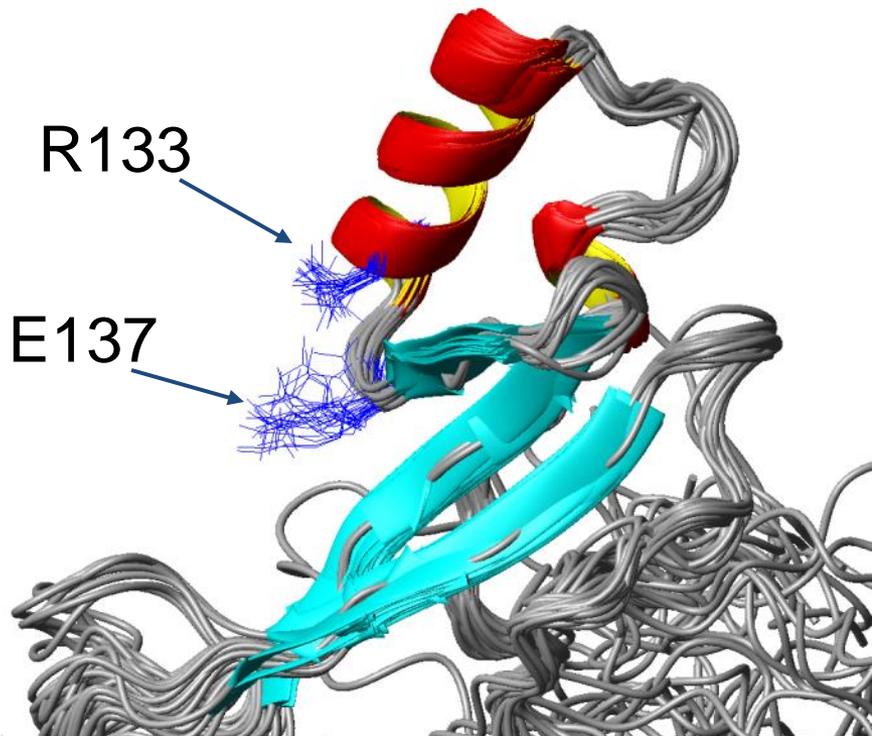
Uniprot ID

dbSNP ID

- Similar to SQL language
- Multiple database search can be applied with a simple script

## Data analysis using BMRB/PDB/OMIM database

BMRB	Mutation	OMIM	dbSNP	2nd	SASA%	$\Delta\Delta G_{hydr}$	
4280	L100V	300005	rs28935168	Coil	22.5	3.9	
4280	R106W	300005	rs28934907	Strand	8.0	11	
4280	R133C	300005	rs28934904	Coil	49.4	7.2	→ High SASA
4280	E137G	300005	rs61748392	Helix	41.7	6.6	→ High $\Delta\Delta G_{hydr}$
4280	A140V	300005	rs28934908	Helix	42.0	1.7	



-- detailed information from OMIM--

MeCP2: methylated CpG binding protein  
NMR structure 1QK9: Wakefield et al., 1999

The mutations, R133C and E137G have been found in patients with X-linked mental retardation.

# [Application of BMRB/XML]

## Multiple database search on the PDBj-BMRB portal

A screenshot of the PDBj search interface. The search bar is highlighted with a blue border. The search filter is set to "Sequence". The search box contains the text "e.g., Entry ID, Macromolecule, Gene Ontology or Author". The search button is a blue square with a magnifying glass icon. Red arrows point from text labels to these elements: "Search tabs & settings" points to the gear icon, "Search filter" points to the "Sequence" tab, "Universal search box" points to the search input field, and "Search button" points to the magnifying glass icon. A green "Feedback" button is visible on the left side of the search bar.

Search

Everything Sequence  ← Search tabs & settings

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

As for query syntax, refer to [this](#)

Search filter Universal search box Search button

BMRB 75

PDB 543

UniProtKB 954

Metabolomics

Ligand Expo 2

Showing 1 - 10 of 75 Documents, order by Relevance

1 **BMRB: 6503**

score: 1761.000

**1H, 13C, and 15N complete chemical shift assignments for the apo v-Src SH2 domain****Authors:** Taylor, J.D., Williams, M.A., Ababou, A., Ladbury, J.E.**Entity:** 1. apo v-Src SH2 (polymer), 106 monomers, 14078.34 Da [Detail](#)**Formula weight:** 14078.34 Da**Organism:** Rous sarcoma virus [↗](#)**Exptl. method:** NMR**Data set:** [assigned\\_chemical\\_shifts](#)**Chem. Shift Compl.:** Assigned residue coverage: **100.0 %**, Completeness: **92.6 %**, Completeness (bb): **98.2 %** [Detail](#)Polymer type: **polypeptide(L)**

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
<b>All</b>	92.6 % (1172 of 1266)	91.6 % (602 of 657)	92.2 % (450 of 488)	92.2 % (302 of 311)
<b>Backbone</b>	98.2 % (612 of 623)	99.5 % (208 of 209)	97.1 % (302 of 311)	98.2 % (302 of 311)
<b>Sidechain</b>	88.8 % (659 of 742)	87.9 % (394 of 448)	89.5 % (247 of 276)	89.5 % (247 of 276)
<b>Aromatic</b>	54.1 % (66 of 122)	54.1 % (33 of 61)	53.3 % (32 of 60)	53.3 % (32 of 60)
<b>Methyl</b>	98.0 % (100 of 102)	98.0 % (50 of 51)	98.0 % (50 of 51)	98.0 % (50 of 51)

**Release date:** 2004-12-01**Citation:** NMR [↑](#) 1 2 3 4 5 6 7 8 **12** domain from the Rous sarcoma virus[Taylor, J.D., Fawaz, P.R., Ababou, A., Williams, M.A., Ladbury, J.E.](#)

BMRB 75

PDB 543

UniProtKB 954

Metabolomics

Ligand Expo 2

Showing 1 - 10 of 543 Documents, order by Relevance

## 1 PDB: 1YTB



score: 1752.000

## CRYSTAL STRUCTURE OF A YEAST TBP/TATA-BOX COMPLEX

**Authors:** [Kim, Y., Geiger, J.H., Hahn, S., Sigler, P.B.](#)**Entity:** 1 (A, B). [DNA \(29MER\)](#) (polymer), 29 monomers, 8987.892 × 2 Da [Detail](#) ▶  
2 (C, D). [PROTEIN \(TATA BINDING PROTEIN \(TBP\)\)](#) (polymer), 180 monomers,  
3 (E, F, G, H). [water](#) (water), 18.015 × 4 Da**Total weight:** 58217.6 Da**Max. entity weight:** 20120.908 Da**Organism:** [Saccharomyces cerevisiae](#) [↗](#)**Exptl. method:** [X-RAY DIFFRACTION](#)**Refine. method:** Rfree **26.1** %, Rwork **20.1** %, Resolution **1.800** - 6.000 Å**Release date:** 1994-09-28**Citation:** [Crystal structure of a yeast TBP/TATA-box complex](#) [↗](#)[Kim, Y., Geiger, J.H., Hahn, S., Sigler, P.B.](#)[Nature](#) (1993), **365**, 512-520, PubMed: [8413604](#) [↗](#), DOI: [10.1038/365512a0](#) [↗](#), [Abstract](#) ▶**Keywords:** [TRANSCRIPTION/DNA](#), [PROTEIN-DNA COMPLEX](#), [TRANSCRIPTION-DNA COMPLEX](#)

**Release date:** 2004-12-01

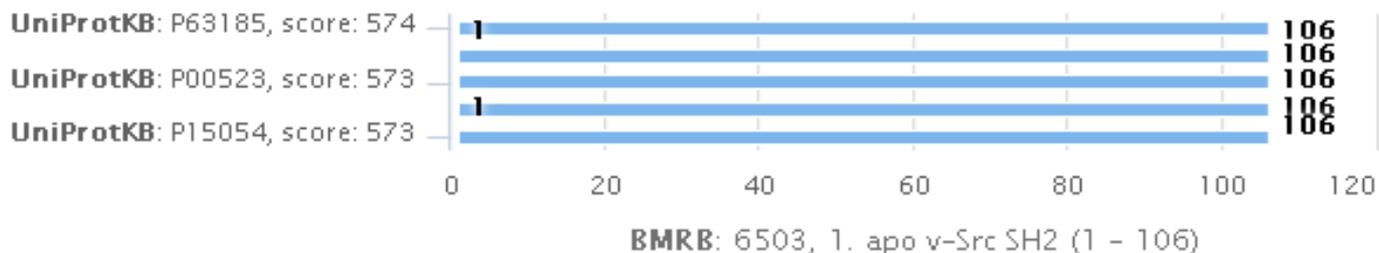
**Citation:** **NMR assignment of the apo and peptide-bound SH2 domain from the Rous sa**

Taylor, J.D., Fawaz, R.R., Ababou, A., Williams, M.A., Ladbury, J.E.

*J. Biomol. NMR* (2005), **32**, 339-339, PubMed: 16211495 [↗](#), DOI: 10.1007/s10858-005-0471-7 [↗](#)

**Related entries:** 1. apo v-Src SH2, ★: 8 ☆: 1 ☆: 14 🚩: 174 entries [Detail](#) ▾

Aligned regions of top 5 related entries



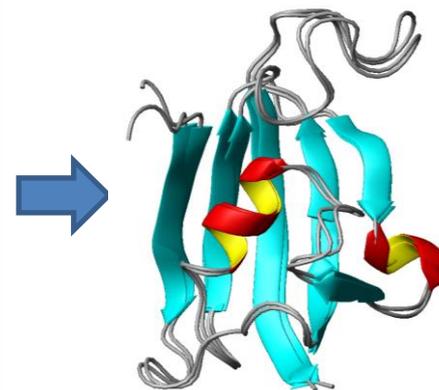
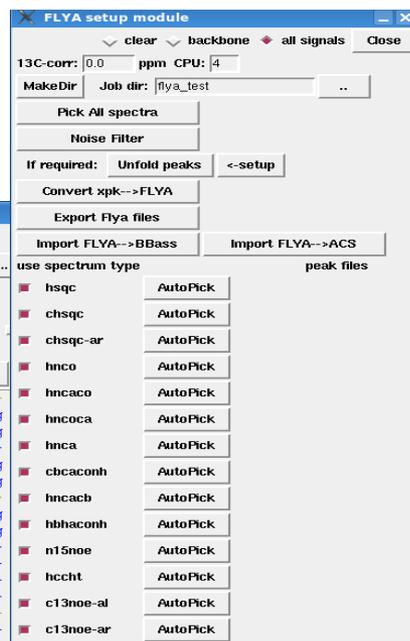
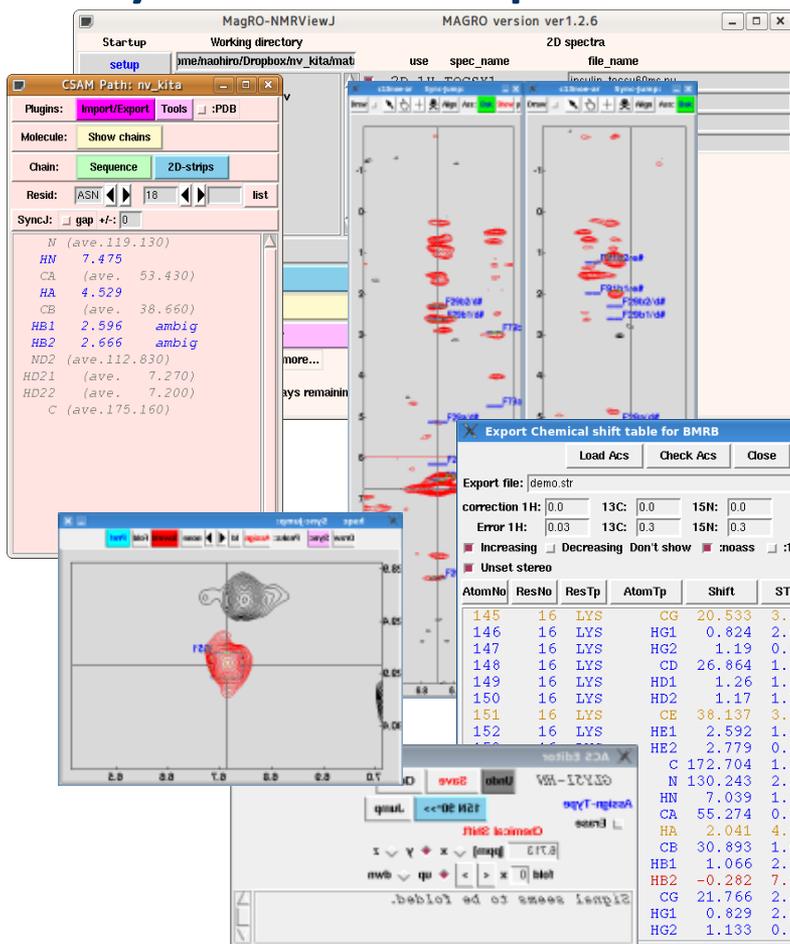
[+ More details for 197 related entries](#)

**Keywords:** Rous sarcoma virus, RSV, SH2, Src homology 2

# [Application of BMRB/RDF]

## MagRO: Tool for highly automated NMR data analysis and deposition to BMRB

## FLYA tool for fully automated Structure analysis



Simple editor for SPARQL Query

Simple editor for SPARQL Query    Load example    Run Query    Close

Compound ID:

```

PREFIX BMRBo:
SELECT ?entry
FROM <http://
WHERE {
?s_ca BMRBo:a
BMRBo:a
BMRBo:a
BMRBo:a
BMRBo:a
BMRBo:a
BMRBo:a
}

```

**Query results**

Target Comp\_ID: PTR    Target Atom\_ID: HD2

Query results Page: 1    <<    <    >    >>    Show    Close

Average: 7.039 (ppm)    STD: 0.313 (ppm)    Total data: 15

ID	BMRB_ID	Entity	Res	Chem
1	19423	2	9	7.060
2	17632	2	5	7.056
3	17632	2	9	7.033
4	17932	1		
5	17680	1		
6	17930	1		
7	17931	1	52	7.089
8	17931	1	64	7.064
9	5212	2	7	7.060
10	17080	2	12	7.106
11	17080	2	24	7.327
12	18182	2	5	7.126
13	4263	2	6	7.380
14	7061	2	4	6.090
15	7150	2	26	6.610

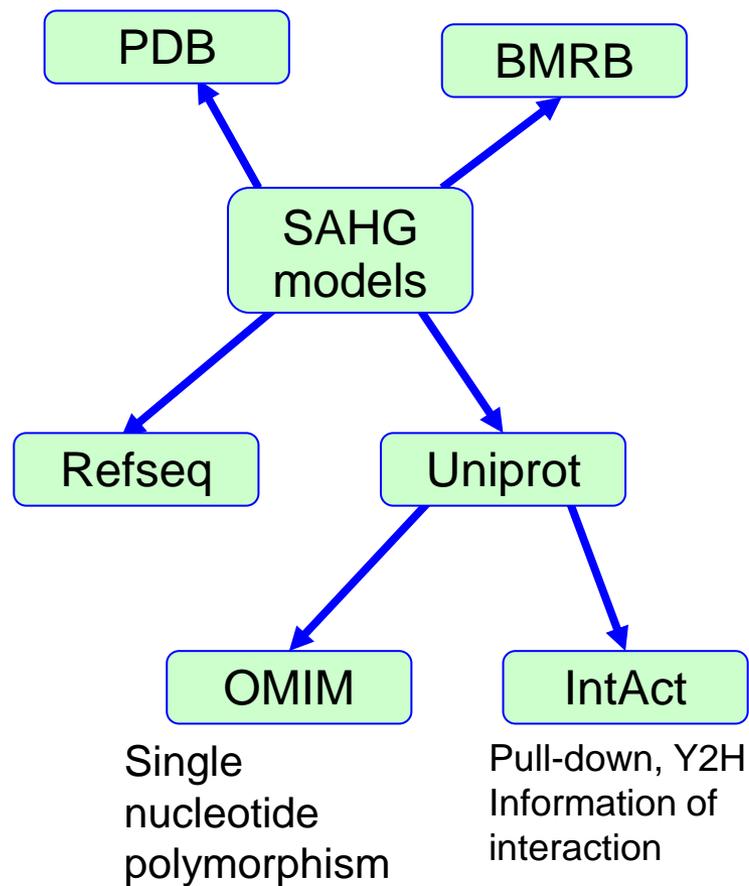
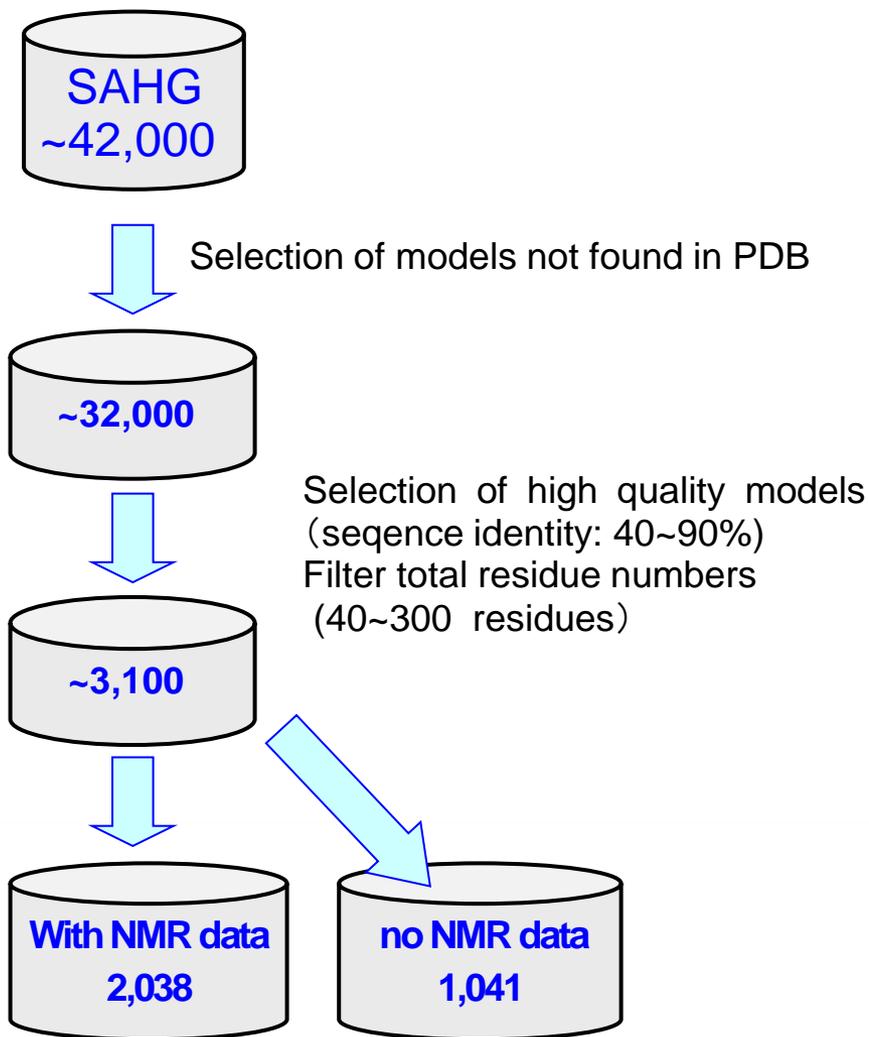
Easy to get averaged chemical shifts for target Comp\_ID and Atom\_ID

BMRB

PROTEIN DATA BANK

# [Application of BMRB/RDF]

Secondary database of modeled structures linked with the other life science databases



**Multiple Protein Model Search**  
(Multiple Protein Model Search for NMR Drug Discovery)

This search engine has been developed for assisting NMR based drug discovery using modeled structures. The modeled structures are based on the entries archived in SAHG server which has been developed by Chie Motono Et al. (AIST). The SAHG project has comprehensively modeled structures for the sequences coded on human genome which are available from RefSeq database and using a lot of highly trustful homology search and modelling methods such as Modeller. Our server can search protein modeled structures by key words and the searched results will be linked with not only the original PDB entry but also the other life science databases such as BMRB, UniProt, OMIM and IntAct. Using the chemical shift data derived from BMRB will be useful for NMR based drug screening.

The search method is easy, so just type key words of protein in the entry and press search button. The search results will show ribbon model and the link of modeled structure archived in SAHG database and PDB like which has been used for homology modeling and sequence identity. If available, the links for BMRB, UniProt, OMIM and IntAct entries will appear.

Motono et. al. Nucleic Acids Res. 2011 Jan;39(Database Issue):D487-93. doi: 10.1093/nar/gkq1057. Epub 2010 Nov 3.

Key word search:

Show only entries have OMIM info  
 Show only entries have BMRB\_ID  
 Show only entries have interaction >

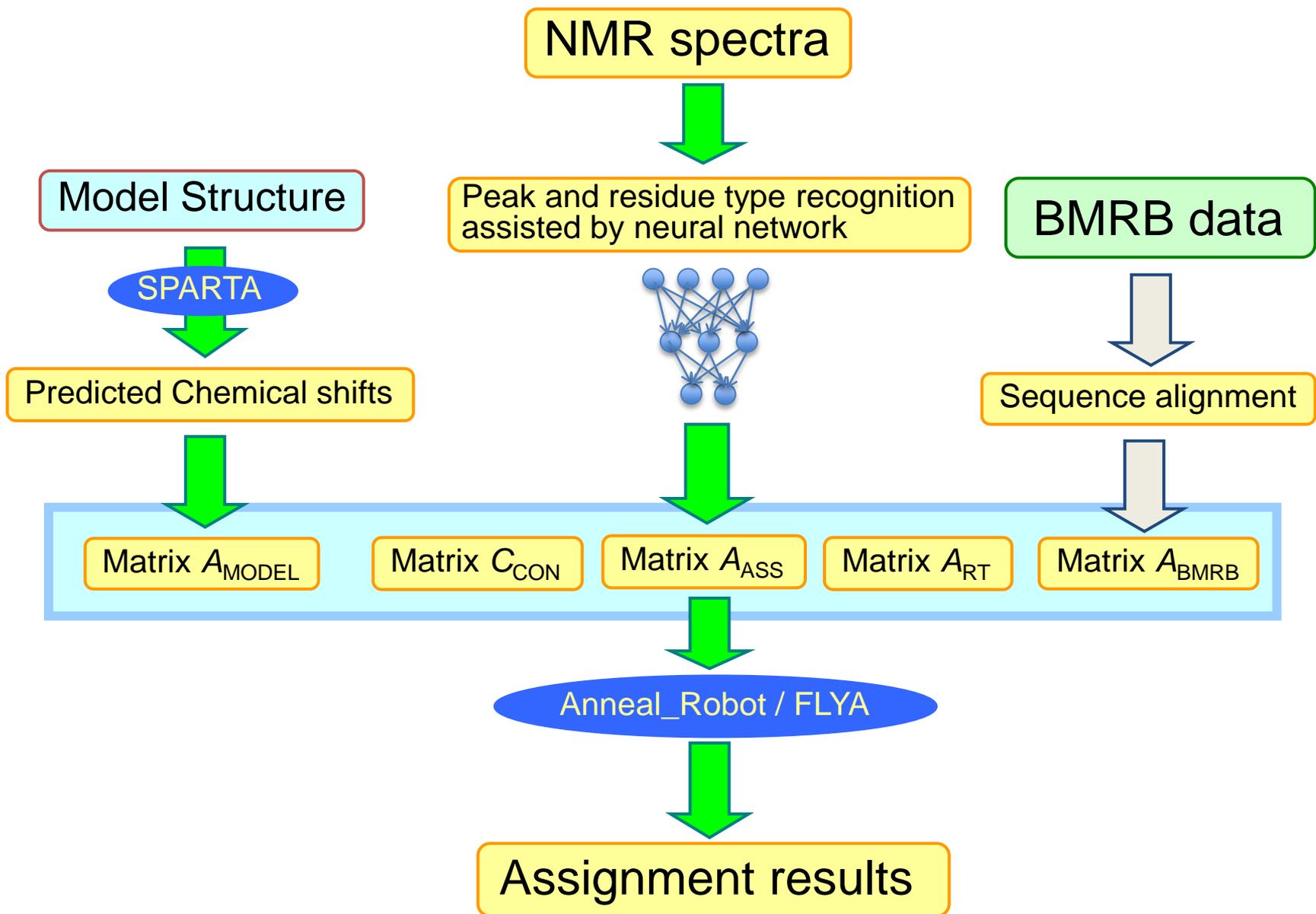
ID	Mutation	Phenotype
	MET 175 → ARG	CLEIDOCRANIAL DYSPLASIA
	SER 191 → ASN	CLEIDOCRANIAL DYSPLASIA
	ARG 225 → GLN	CLEIDOCRANIAL DYSPLASIA
	ARG 225 → TRP	CLEIDOCRANIAL DYSPLASIA
Q13950	170 – 293	<a href="#">60Q211</a>
	THR 200 → ALA	CLEIDOCRANIAL DYSPLASIA

.html



1000-2010-05-07

# Automated assign system of **MagRO** with modeled structure and BMRB data



## Acknowledgement

### **Osaka University:**

Toshimichi Fujiwara, Chojiro Kojima (YNU, Osaka Univ.)  
Masashi Yokochi, Takeshi Iwata

**Wisconsin University:** Eldon L. Ulrich, John L. Markley

**AIST (Tokyo):** Chie Motono

**Nagoya University:** Hidekazu Hiroaki

**Kyoto University:** Masato Katahira, Takashi Nagata