



# PDBとBMRB: データベースの高度化と 機械学習への応用

大阪大学蛋白質研究所  
小林直宏

# JST-NBDCにより支援される 大阪大学 PDBj, PDBj-BMRBグループ

NBDC

wwPDB

栗栖源嗣:教授

2017年4月より  
5年間のプロジェクト

PDBj: X線解析、EM解析、NMR解析による生体高分子構造データベース

栗栖源嗣 (大阪大学:教授)

中村春木 (大阪大学:教授)

中川敦史 (大阪大学:教授)

金城玲 (大阪大学:准教授)

藤 博幸 (関西学院大学:教授)

PDBj-BMRB: 生体高分子NMRの実験データベース

藤原敏道 (大阪大学:教授)

宮ノ入洋平 (大阪大学:准教授)

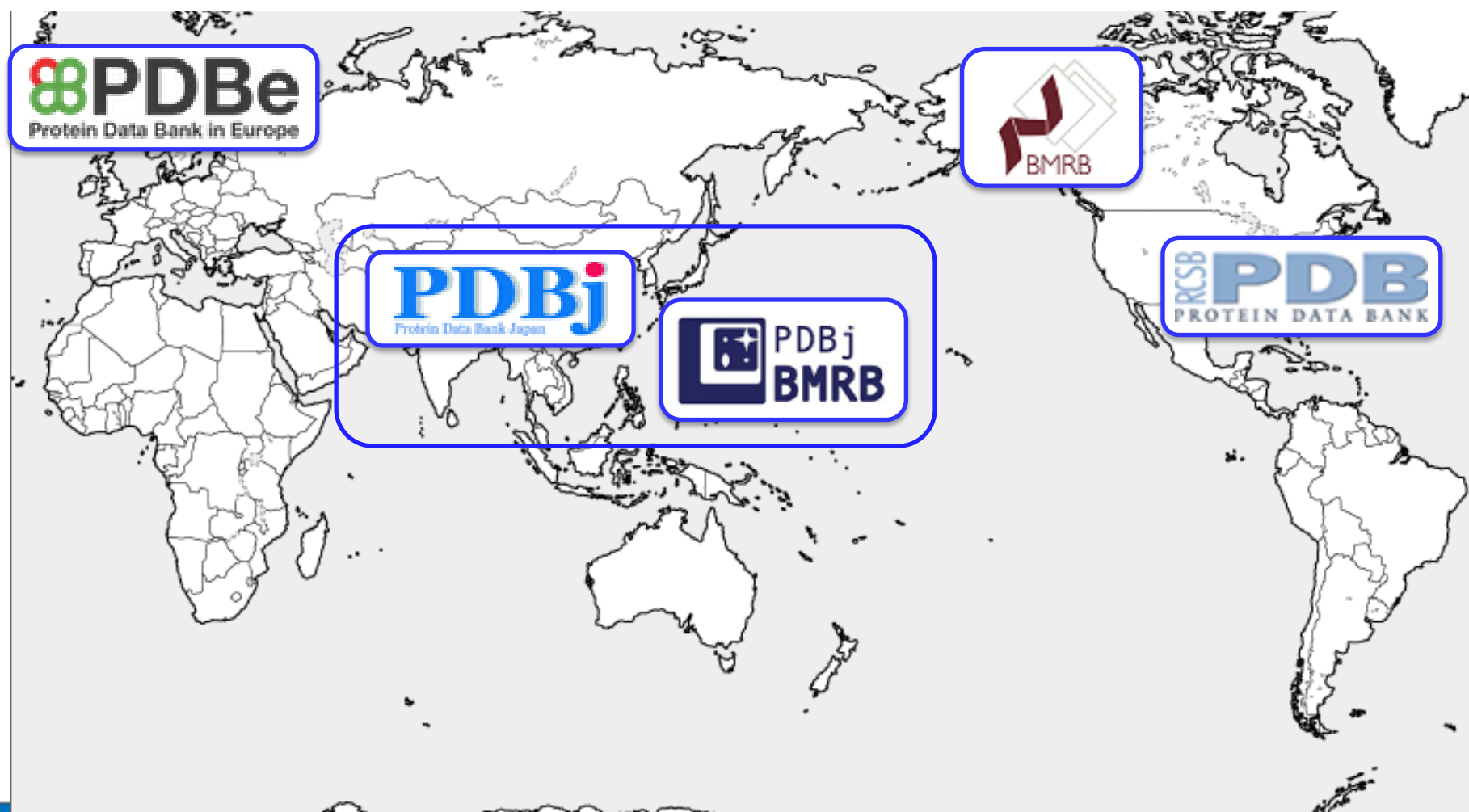
児嶋長次郎 (大阪大学・横浜国大:教授)

小林直宏 (大阪大学:特任准教授)

横地政志 (大阪大学:特任研究員)

岩田武史 (大阪大学:特任研究員)

# 世界4拠点によって組織される wwPDBの構造データベースの国際的連携



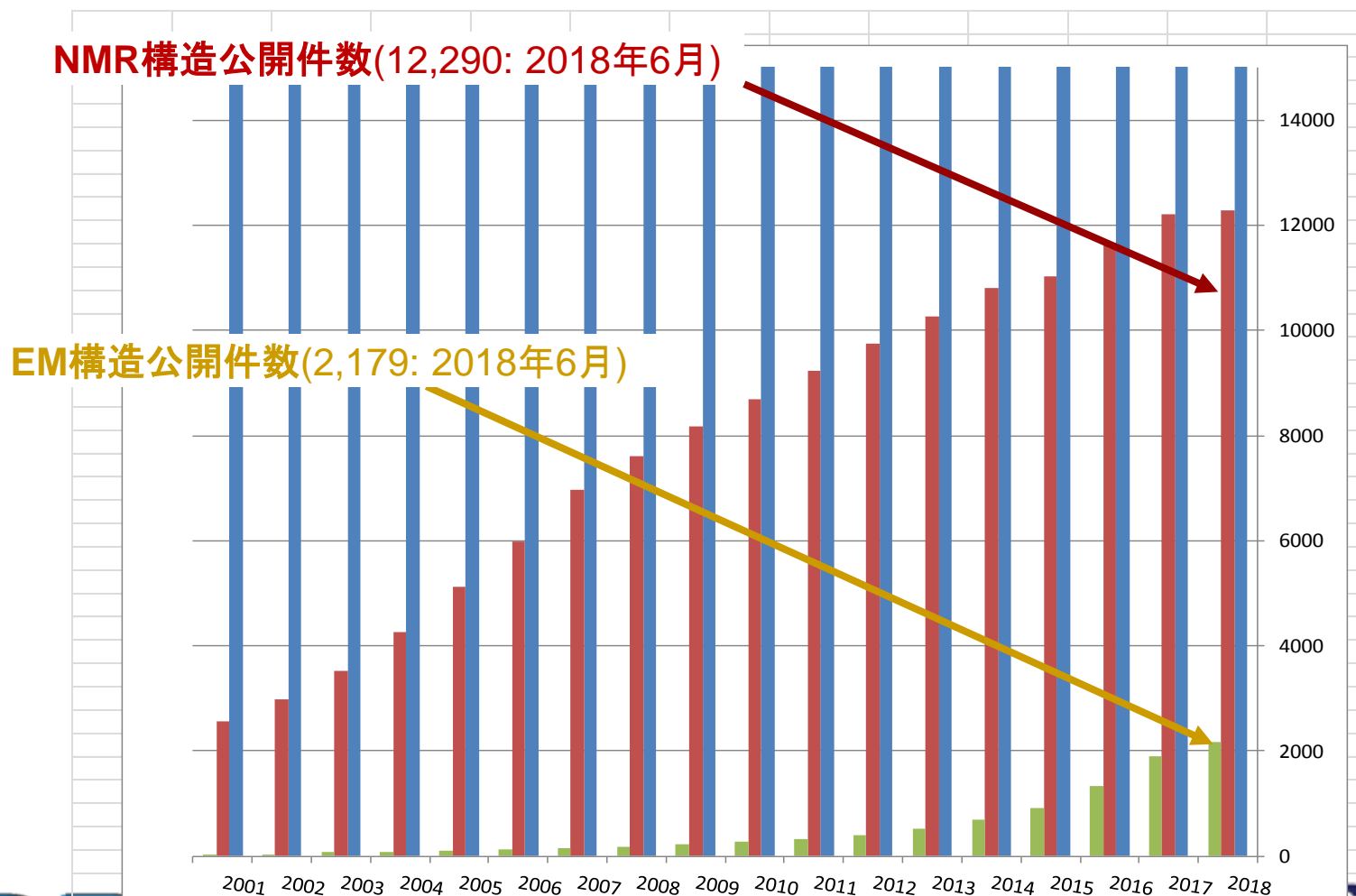
# 各年における構造総公開件数とその内訳

PDB構造公開件数 (138,994: 2018年6月)

NMR構造公開件数(12,290: 2018年6月)



# 各年における構造総公開件数とその内訳

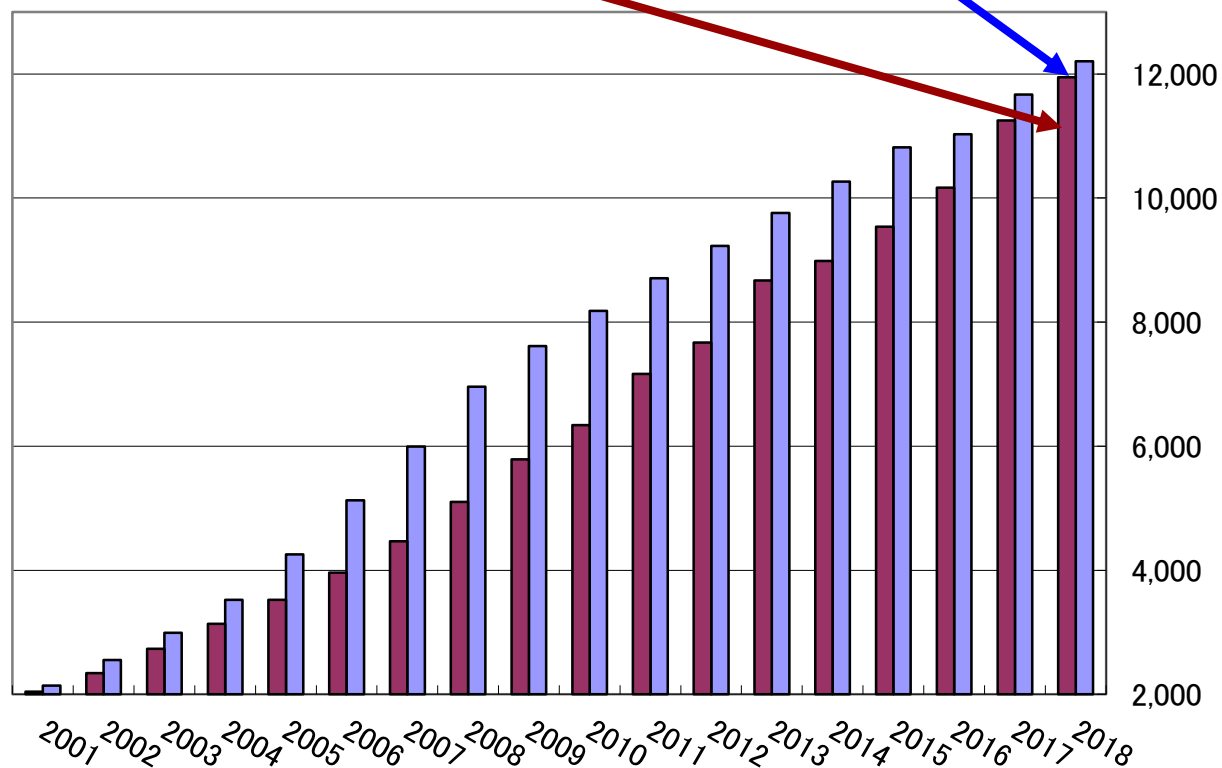


# 各年におけるNMR構造公開件数とBMRB公開件数の推移

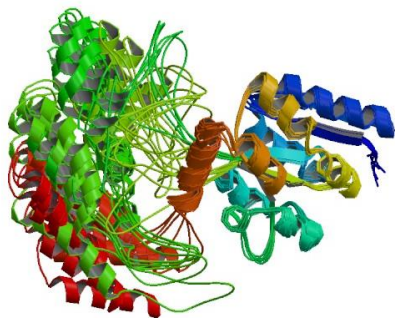


NMR構造公開件数 (12,290: 2018年6月)

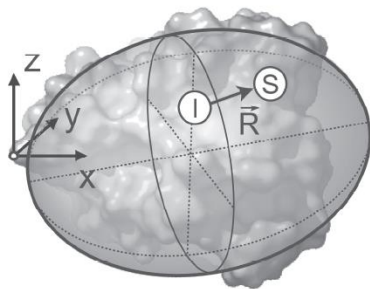
BMRB公開件数(12,318: 2018年6月)



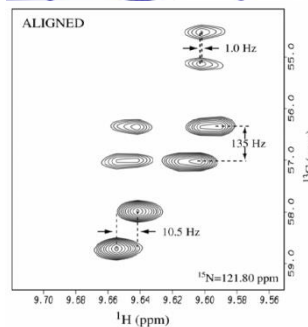
# NMRシグナル帰属データの応用:なにが分かるのか？



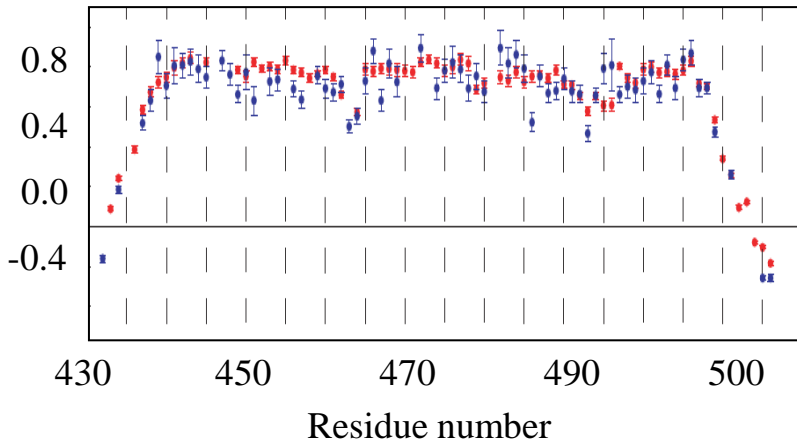
立体構造解析



ドメイン配向

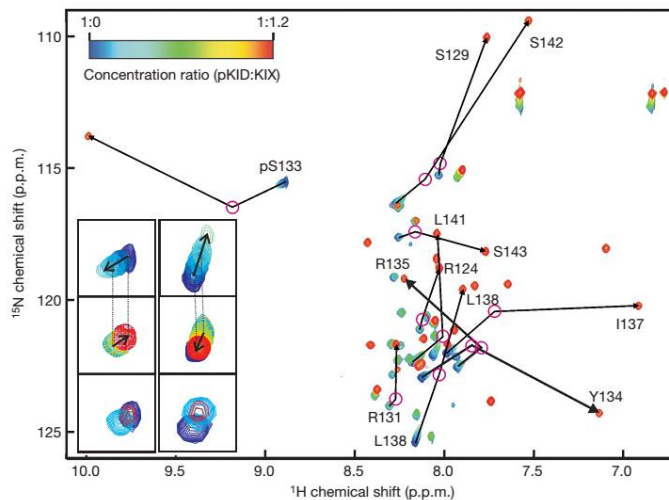


NOE



動的状態の解析

Sugase et al., *Nature*, 2007



リガンド相互作用





# BMRBデータのXML化: 辞書に基づいた完全な記述

**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 F1  
53-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 .
```

bmr15400.str

```
<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 F15  
5-flurotryptophan mutant of human cardiac troponin C</BMRBx:title>  
    <BMRBx:version_type>original</BMRBx:version_type>  
  </BMRBx:entry>  
</BMRBx:entryCategory>
```

bmr15400.xml



# XML化されたデータのRDF化: 計算機が理解できるデータへ

## 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_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
5-fluorotryptophan mutant of human cardiac troponin C</BMRBx:title>
    <BMRBx:version_type>original</BMRBx:version_type>
  </BMRBx:entry>
</BMRBx:entryCategory>
```

bmr15400.xml

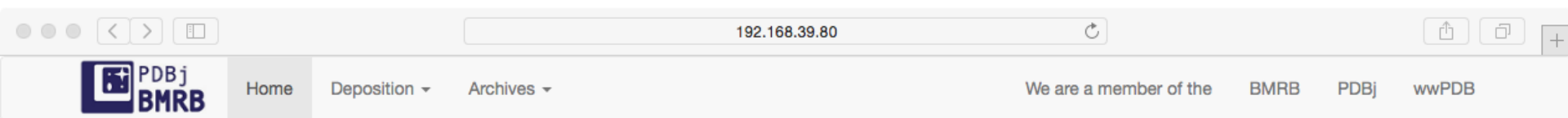
BMRBoTool

## BMRB/RDF

```
<BMRBo:has_entryCategory>
  <BMRBo:entryCategory rdf:about="http://bmrpub.protein.osaka-u.ac.jp/rdf/bmr15400">
    <BMRBo:has_entry>
      <BMRBo:entry rdf:about="http://bmrpub.protein.osaka-u.ac.jp/rdf/bmr15400/entry/15400">
        <BMRBo:of_datablock rdf:resource="http://bmrpub.protein.osaka-u.ac.jp/rdf/bmr15400/datablock/15400">
          <BMRBo:entry.id>15400</BMRBo:entry.id>
          <BMRBo:entry.accession_date>2006-12-07+09:00</BMRBo:entry.accession_date>
          <BMRBo:entry.experimental_method>NMR</BMRBo:entry.experimental_method>
          <BMRBo:entry.experimental_method_subtype>SOLUTION</BMRBo:entry.experimental_method_subtype>
          <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.original_nmr_star_version>
          <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_date>
          <BMRBo:entry.title>Backbone and side chain chemical shift assignments of
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>
```

bmr15400.rdf

# マルチデータベースサーチ機能



Search filter

Universal search box

Search button

# XML化されたデータベースを有効活用

Found 1574 Documents (0.295 - 0.885 seconds)

Order by ▾

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](#)

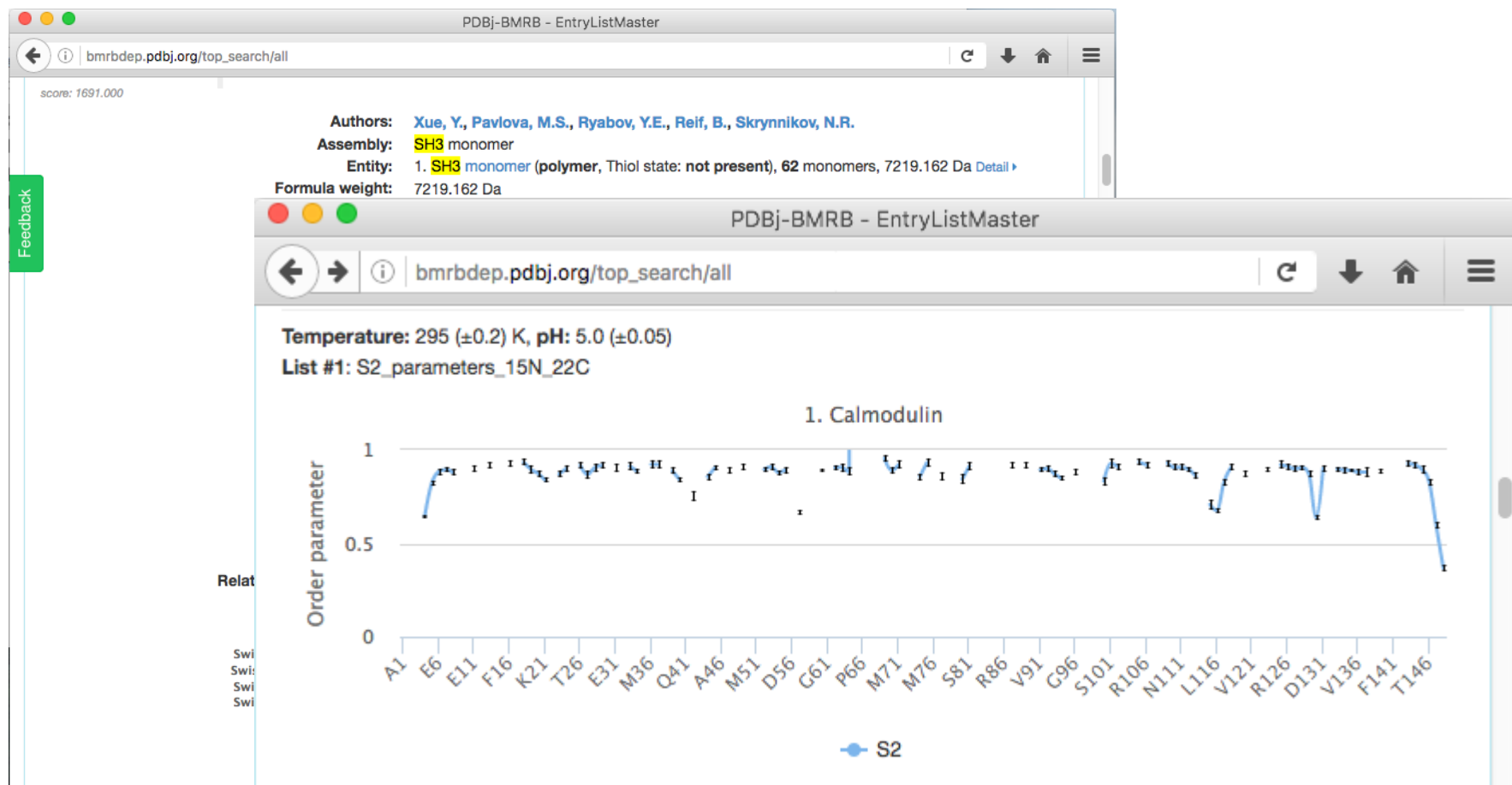
Polymer type: polypeptide(L)

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
All	92.6 % (1172 of 1266)	91.6 % (602 of 657)	92.2 % (450 of 488)	99.2 % (120 of 121)
Backbone	98.2 % (612 of 623)	99.5 % (208 of 209)	97.1 % (302 of 311)	99.0 % (102 of 103)
Sidechain	88.8 % (659 of 742)	87.9 % (394 of 448)	89.5 % (247 of 276)	100.0 % (18 of 18)
Aromatic	54.1 % (66 of 122)	54.1 % (33 of 61)	53.3 % (32 of 60)	100.0 % (1 of 1)
Methyl	98.0 % (100 of 102)	98.0 % (50 of 51)	98.0 % (50 of 51)	

## 詳細なNMR実験情報を取得可能

viral protein Src [↗](#)

# 高度化かつ視認性の良い検索結果を表示



## 運動性の高い残基などを表示

# 他データベースとの自動リンク

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

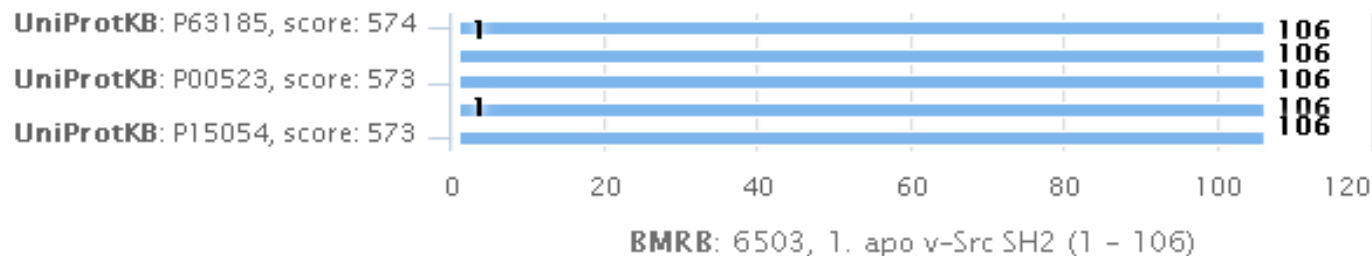
**Citation:** NMR assignment of the apo and peptide-bound SH2 domain from the Rous sarcoma

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

配列相同性、相互作用ターゲットなどの情報とも多様にリンク: 創薬研究への応用



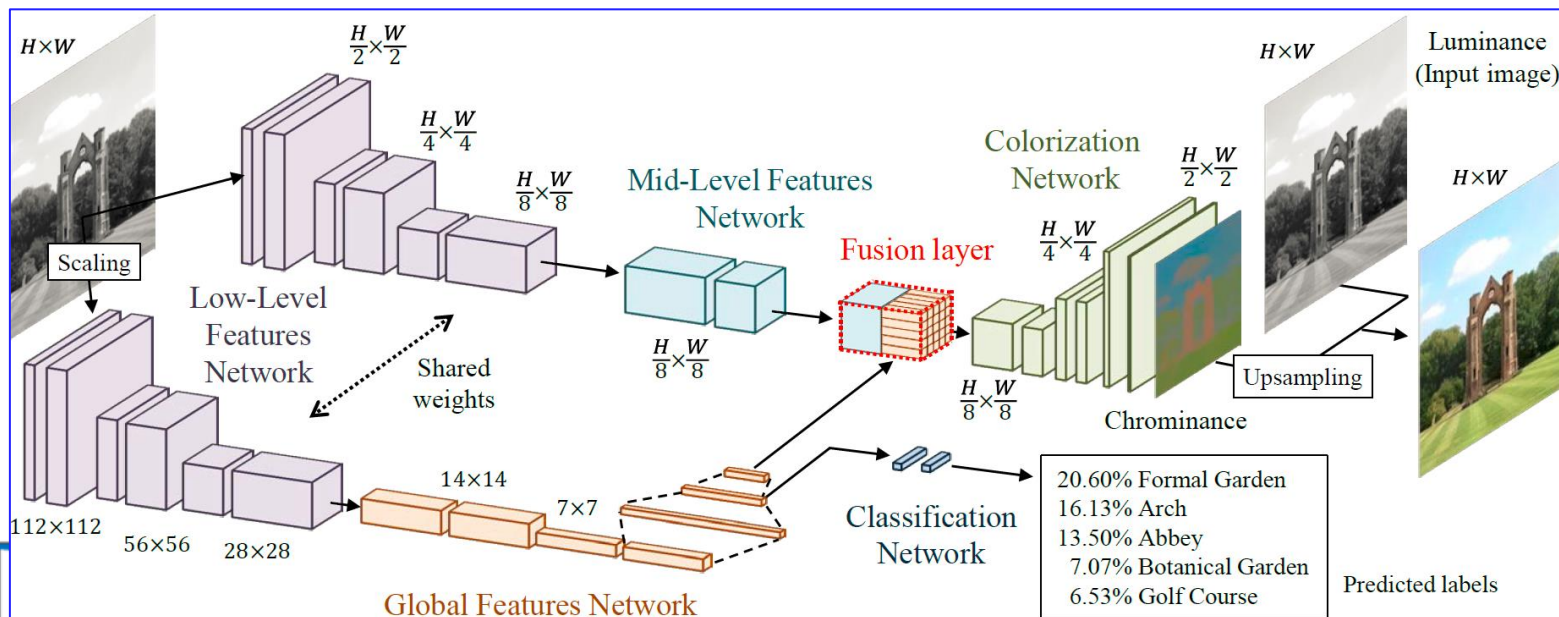
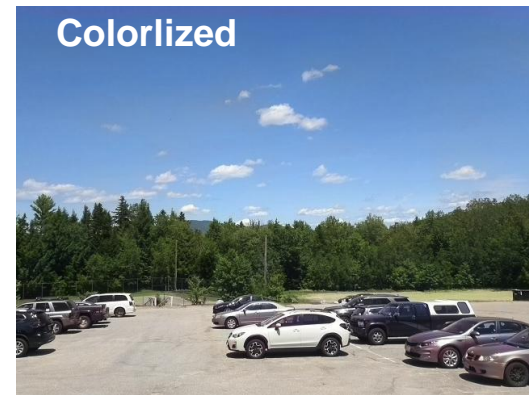
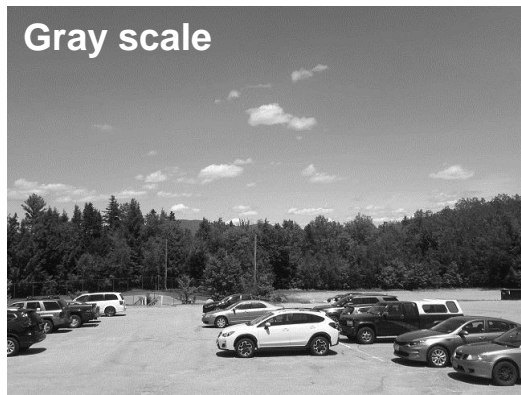
# 深層学習への応用：近年のAIブームによる技術の発展



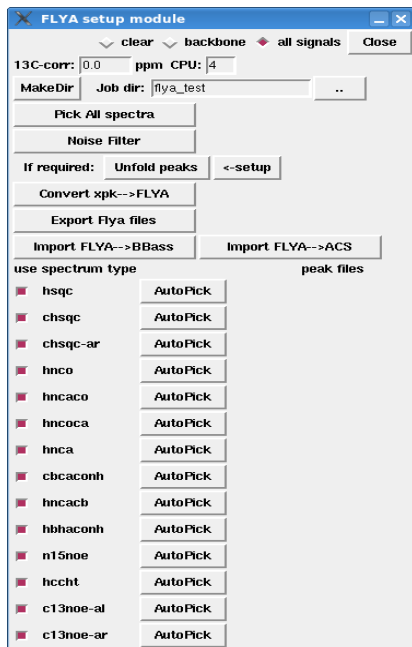
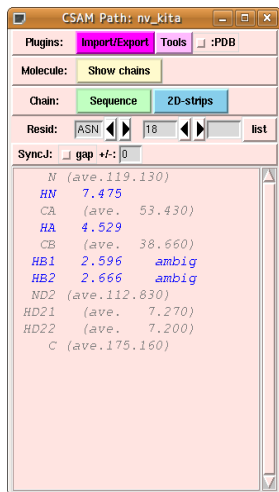


# 深層学習による自動着色の例: 物体認識を利用

早稲田大学: 飯島ら



# 深層学習技術を取り入れた統合的解析ツール Deep-MagRO



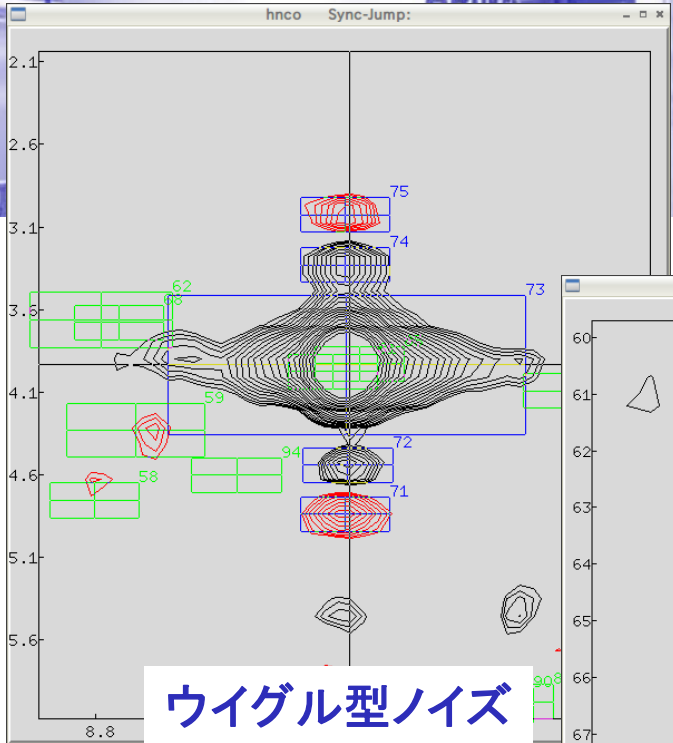
- 完全自動による閾値決定
- 完全自動によるNMR信号検出とノイズ除去機能
- FLYA, CYANAと連携した完全NMRシグナル自動解析と構造計算



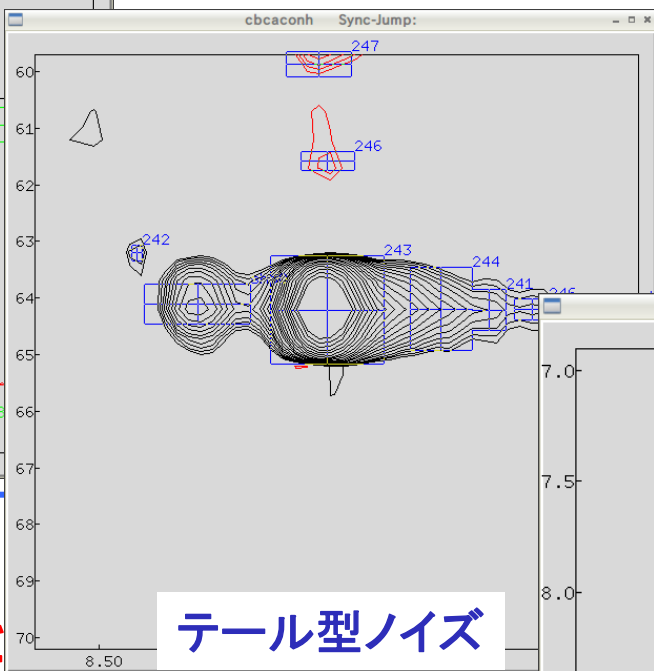
現状として、ピーク検出から開始し、  
かつ高精度に連鎖帰属を達成できる  
プログラムは存在しない・・・ようである

[理由はおそらく]

**割とノイズが多い(アーティファクト)**  
シグナル強度の低いシグナルを拾いにくい  
若干重なっているシグナルなど

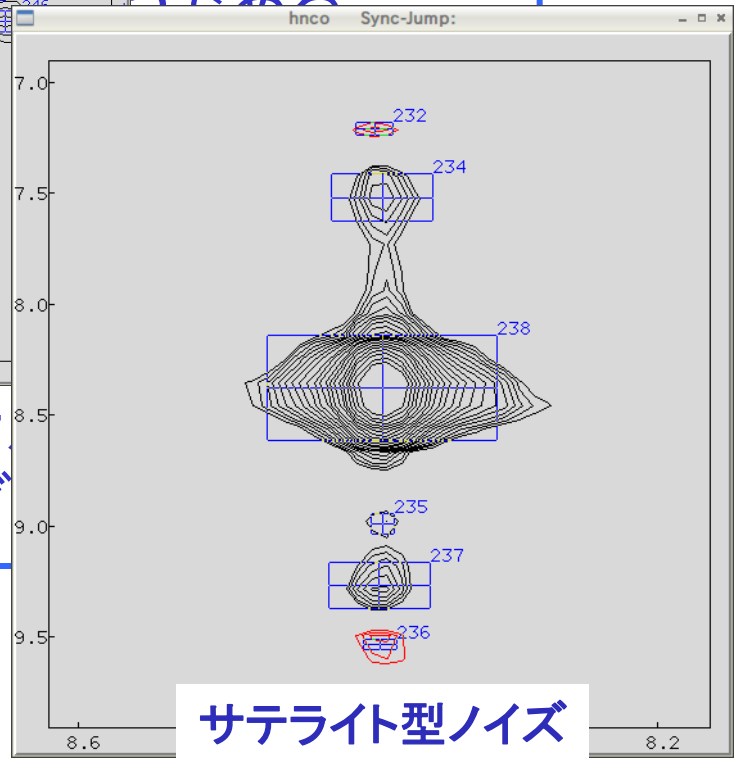


ウイグル型ノイズ



テール型ノイズ

開始し、  
成できる  
うである



サテライト型ノイズ

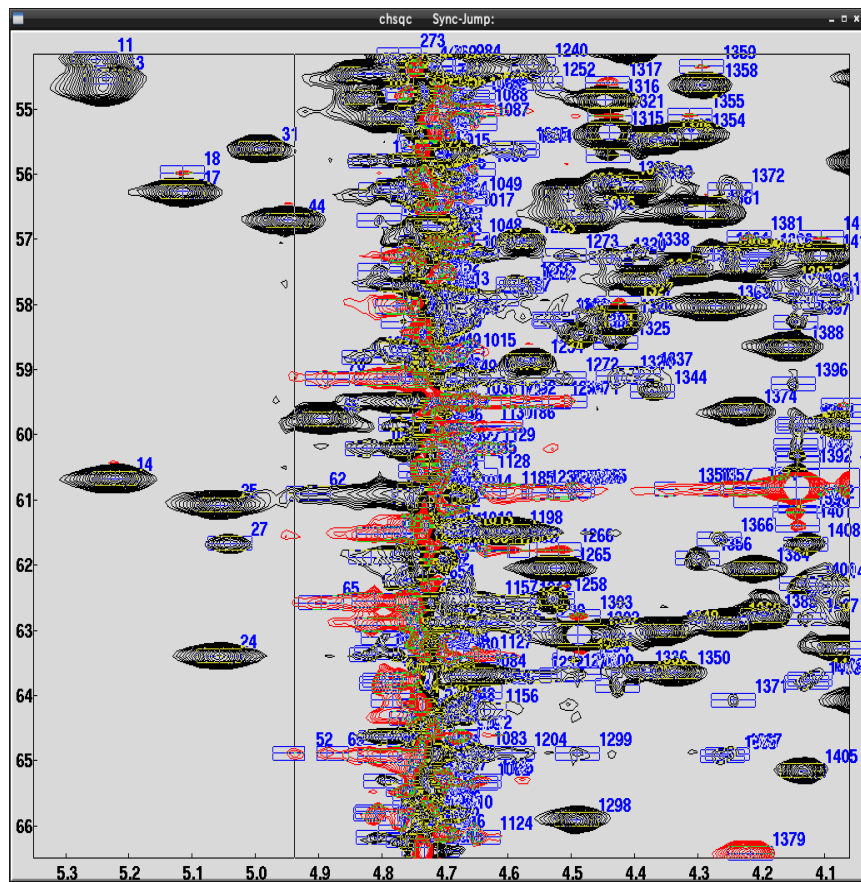
割と  
シグナル強度の低いシグナ  
若干重なっているシグ



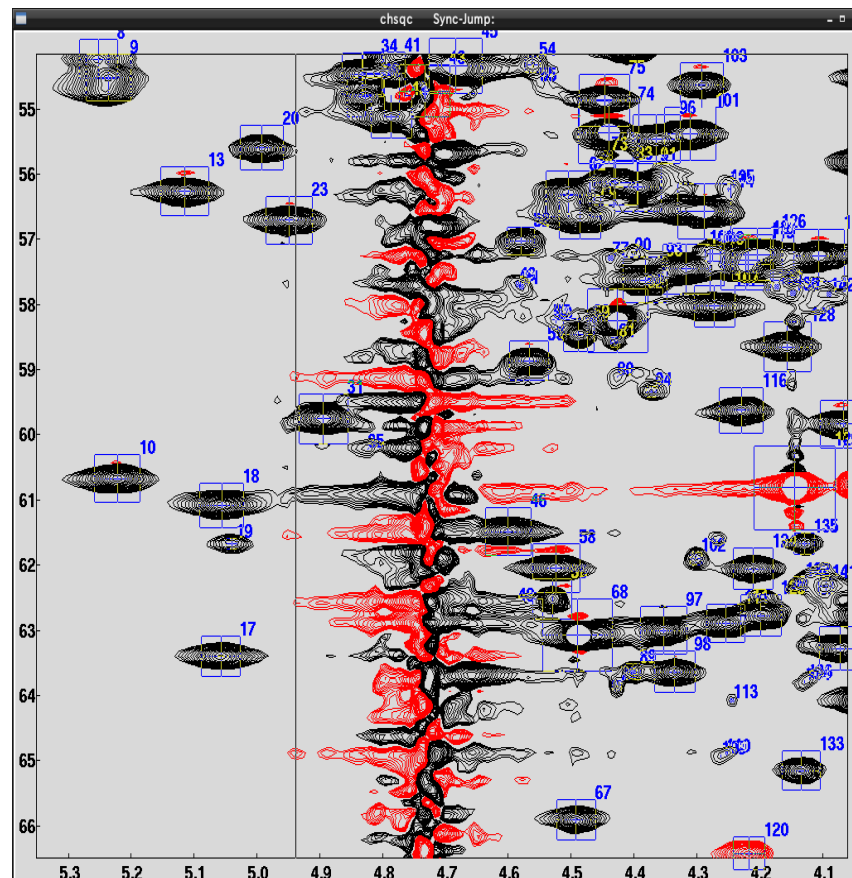
# 実際のノイズ除去の結果



2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC (Constant Time), 111aa ( $^{13}\text{C}/^{15}\text{N}$ ラベル体、メチル化FKBP2)

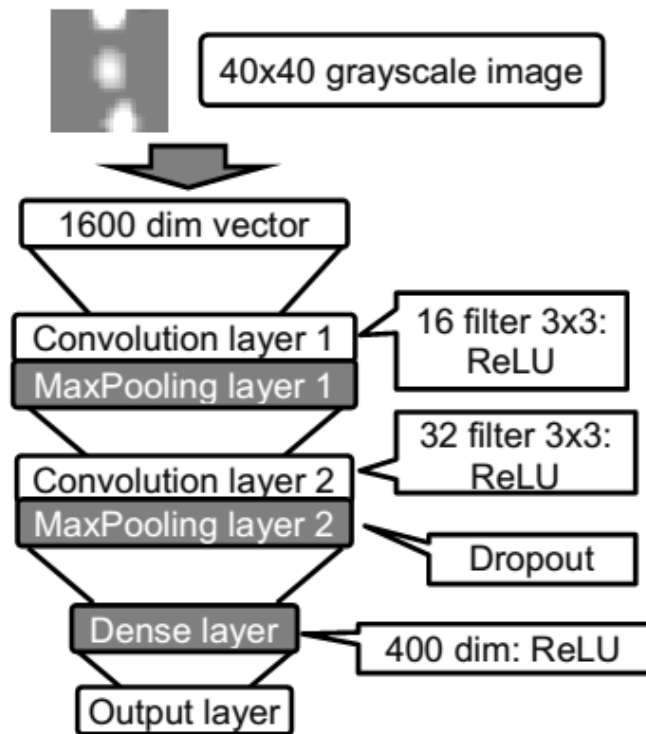


自動ピックによるピーク



CNNによるノイズフィルタ

# Deep Convolutional Neural Networksを応用したNMR信号・ノイズ自動判定器



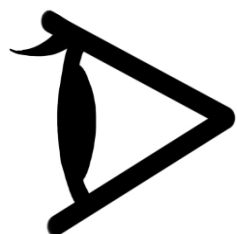
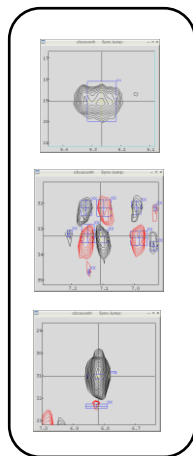
NMR experiment	Number of peaks		Accuracy		
	initial	filtered	Recall (%)	Precision (%)	F-value (%)
<b>bmr16647</b>					
2D $^1\text{H}$ - $^{15}\text{N}$ HSQC	323	87	<b>100.0</b>	90.8	95.2
2D $^1\text{H}$ - $^{13}\text{C}$ HSQC	1,332	334	<b>98.8</b>	73.4	84.8
3D CBCA(CO)NH	1,023	140	<b>100.0</b>	97.1	98.6
3D HNCACB	723	230	<b>97.0</b>	99.6	98.3
3D HCCH-TOCSY	17,760	845	<b>95.1</b>	85.6	90.1
3D $^{15}\text{N}$ -edited NOESY	2,880	806	<b>99.7</b>	95.4	97.5
3D $^{13}\text{C}$ -edited NOESY	20,243	1,997	<b>99.0</b>	84.1	91.0
<b>Lamin-G465D</b>					
2D $^1\text{H}$ - $^{15}\text{N}$ HSQC	364	173	<b>95.5</b>	98.8	97.2
2D $^1\text{H}$ - $^{13}\text{C}$ HSQC	2,144	815	<b>99.4</b>	86.7	92.7
3D CBCA(CO)NH	1,862	292	<b>100.0</b>	94.9	97.4
3D HNCACB	1,700	490	<b>98.9</b>	92.7	95.7
3D HCCH-TOCSY	35,280	1,290	<b>97.1</b>	90.8	93.8
3D $^{15}\text{N}$ -edited NOESY	5,776	2,497	<b>99.2</b>	97.9	98.5
3D $^{13}\text{C}$ -edited NOESY	54,016	4,273	<b>96.2</b>	97.0	96.6



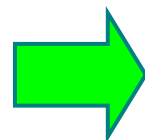
# 計算機が”目で見て・判断する”



## 2次元NMRデータ

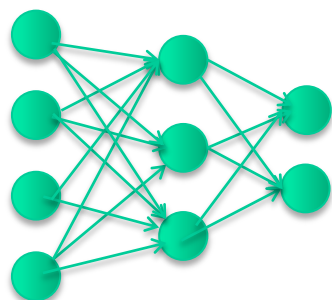
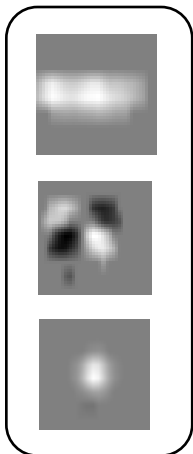


## 人間の場合



ノイズかシグナルかを判定

## 2次元画像



## 深層学習



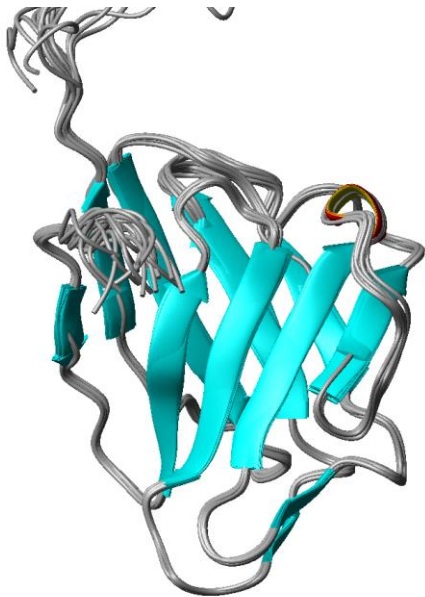
ノイズかシグナルかを判定

# Deep-MagRO, FLYA, CYANAを連携させた完全自動解析の事例 (Kobayashi, N. et al., 国際誌投稿中)

Lamin G465D 144アミノ酸長 (7種の3D NMRスペクトル)

完全自動 (左)

手動による (右)



1. Peak pick on MagRO

2. CNN filter

3. FLYA (20-Core 3.0 GHz)

4. TALOS+

5. CYANA (20 models)

6. RDC validation ( $^1\text{H}$ - $^{15}\text{N}$ )

**全工程: 163min**

(通常の半自動解析: 14~20days)

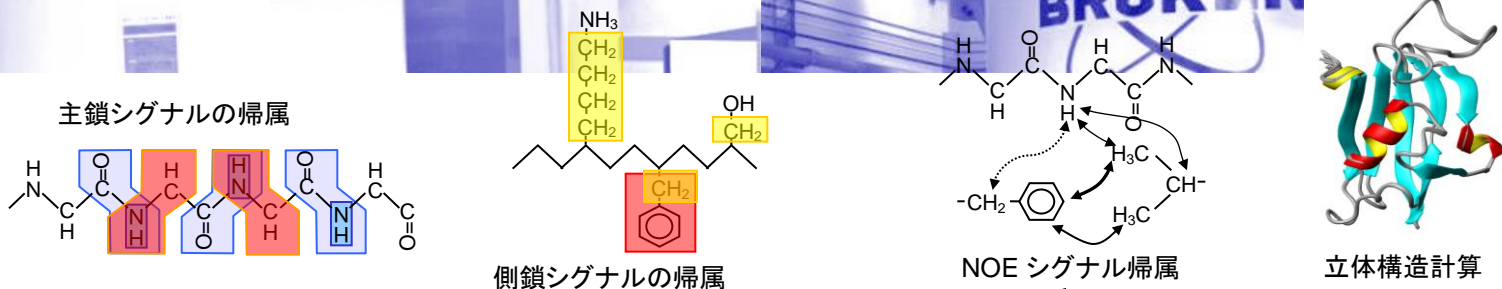
Assignment completeness : 92.5%

accuracy: 97.2%

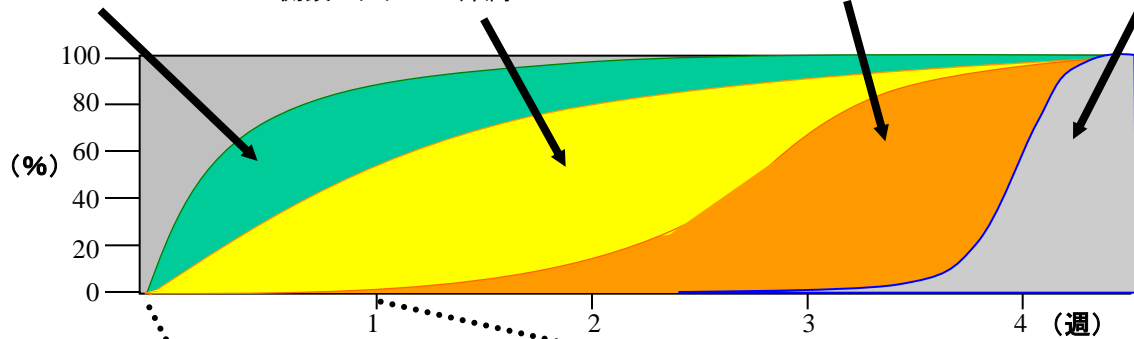
RMSD bias (CO,N,Ca):  $1.10\text{\AA} \pm 0.03\text{\AA}$

RDC ( $^1\text{H}$ - $^{15}\text{N}$  with Phase) corr =  $0.85 \pm 0.04$

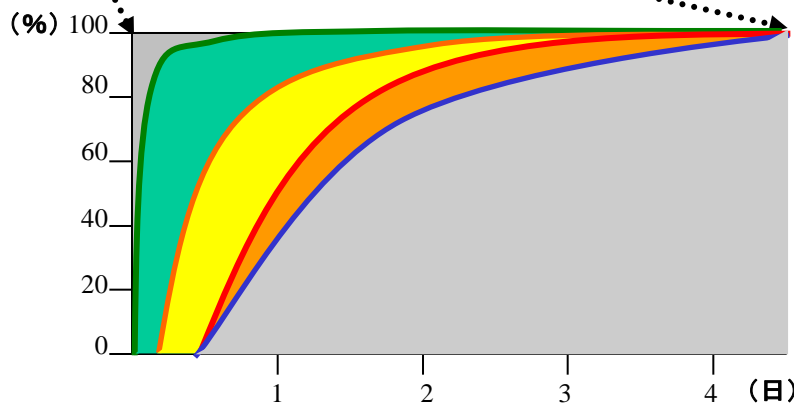
# 深層学習により強化された解析作業時間の改善



半自動的な  
手動解析



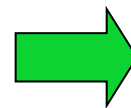
旧来の  
MagRO/FLYA/  
CYANAを用いる  
手法



深層学習により強化  
された解析法

およそ:4-5時間

(ただし、データの質  
に強く依存)



解析初期	中間期	解析終盤
------	-----	------

## まとめ

800 US

BRUKER

- PDB, BMRBにおけるデータは自動的にXML化、RDF化され、高度な検索、WEB表示機能を獲得した
- 深層学習の画像認識技術を応用したDeep-MagROにより、小型のタンパク質ならば少数の3D NMRデータセット+残余双極子 (RDC) を用いる事で信頼性が高くなおかつ高速に自動解析が可能になった。

# 大阪大学PDBj-BMRBグループ

藤原敏道（阪大・教授、グループリーダー）

宮ノ入洋平（阪大・准教授）

児嶋長次郎（阪大・横浜国大）

小林直宏（阪大・特任准教授）

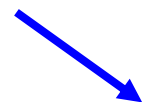
横地政志（阪大・特任研究員）

岩田武史（阪大・特任研究員）

# 各年におけるNMR構造登録件数とBMRB登録件数の推移



NMR構造登録件数 (11,901: 2017年6月)



BMRB登録件数(12,318: 2018年4月)



All_PDB	NMR struc	EM									
138994	12290	2179									
136414	12205	1888									
125488	11668	1323									
109821	11028	916									
105397	10818	701									
95747	10264	515									
86368	9758	397									
78161	9228	332									
70068	8705	279									
62165	8181	226									
54779	7612	186									
47815	6958	145									
40616	5993	126									
34141	5130	100									
28779	4254	77									
23596	3525	65									
19429	2994	36									
13200	2550	18									



■ A  
■ N  
■ E