

蛋白質構造データバンクの国際的な構築と統合化

(Protein Data Bank japan: PDBj, <http://pdbj.org>)

代表研究者: 中村春木(大阪大学蛋白質研究所・教授)

共同研究者: 藤原敏道(大阪大学蛋白質研究所・教授)



PDB (Protein Data Bank):

X線結晶解析、核磁気共鳴法 (NMR)、電子顕微鏡観測によって得られた8万件近い蛋白質・核酸・糖などの生体高分子3次元立体構造の国際的データベース。原子種とその座標、アミノ酸残基、実験手法、実験時の情報、実験観測データを含む。

1. 本研究開発課題の目的・要旨

国際協力のもと、蛋白質構造データバンク(PDB)の構築・運営を実施するとともに、NMR実験情報のデータベース(BMRB)、電子顕微鏡やX線小角散乱による構造情報データベース(DB)を構築し公開する。オントロジー、データ管理システム、新たなデータ書式、利用者インターフェースや二次的DBおよび種々のサービスも開発して公開し、PDBの高度化・統合化を図り、人材育成も実施して、他の生命DBとの統合化による高次生命機能の理解に資する研究開発を行う。

2. 3年間の研究開発のスケジュール

	H23 年度	H24 年度	H25年度
(1) PDBデータおよびNMR実験情報の登録・編纂の国際連携による実施とデータ検証法の開発	蛋白質構造情報とNMR実験情報の登録 登録業務の標準化・自動化の開発 wwPDB活動の推進 電子顕微鏡およびX線小角散乱の構造情報DBの構築と公開 NBDCとの統合的データベース運営の仕組みを構築し実施		
(2) 統合化に向けたデータベースの高度化	構造情報のRDF形式等による高度化と配布 構造データのオントロジー化 構造データの新フォーマットの開発・改良 ユーザインターフェースの改良と高度化 NMR化学シフトデータの管理・配布システムの高度化 (STAR形式への変換) (XML形式への変換) 新たなNMR実験データのDB化と登録システムの開発		
(3) 機能情報と構造情報との統合化システムの開発	蛋白質相互作用部位データベースの改良とPDBデータとの統合化 配列情報を入力とした構造情報に基づく機能の解析・推定パイプラインの構築 (パイプライン化と公開) (GIRAF, eF-seekの組込) (システムの確立と公開)		
(4) 人材養成の実施	データ寄託者・登録者に対する教育・講習会の実施 データ利用者に対する教育・講習会の実施 キュレータ・アナレータの人材育成		

3. 本研究開発課題のH23年度の目標、当初計画

(3-1) PDBデータおよびNMR実験情報の登録・編纂の国際連携による実施とデータ検証法の開発

- Primary Annotatorsによる欧米との協力したデータ登録の継続的实施。
- 登録業務の標準化・自動化 (Common Annotation & Deposition) システム開発の国際協力 (データ検証部分を担当)。
- wwPDB活動を推進。
- 電子顕微鏡およびX線小角散乱の実験による構造情報データベースの構築と公開
- NBDCとの統合的データベース運営の仕組みを構築し実施。

(3-2) 統合化に向けたデータベースの高度化

- 蛋白質構造データの管理・配布システムの高度化: 他のDBとの関連性メタデータとRDF化したPDBデータを提供し、DB統合化環境を整備する。
- 構造データのオントロジーの改良・開発: mmCIFの記述に基づき、OWLに準拠したRDF/XMLの確立と、他のRDF化がなされたDBとの統合的利用法の確立。
- データ記載のフォーマットの改良・開発: 新たなPDBフォーマットの開発と公開、およびマスターフォーマットの確立。NBDCとの連携による構造データのフォーマットの標準化。
- ユーザーインターフェースの改良: 多言語でのweb pageや、jVを用いたGUIの高度化
- NMR実験データの管理・配布システムの高度化: オントロジー工学に基づく標準的なXML記述法を開発し、外部DBとの統合的利用を可能とし、新規実験データを収集する。

(3-3) 機能情報と構造情報との統合化システムの開発

- ・**蛋白質相互作用部位データベース**の改良とPDBデータとの統合化: 原子レベルでの蛋白質相互作用を記述するオントロジーを定義し、相互作用部位をRDFで記述する。これにより、原子レベル→残基レベル→分子レベル→分子複合体レベルへつなげて、高次生命機能情報へ、原子レベルの情報を統合化できる環境を整える。
- ・**配列情報を入力とした、構造情報に基づく機能の解析・推定パイプライン**の構築: PDBjおよびその他の研究者が開発した、リモートホモログの推定や局所的蛋白質相互作用部位や機能的分子表面部位の推定のツールを基に、ゲノム解析等によって得られた配列情報から、構造情報あるいは複合体を含むモデル構造情報を一度経由し、高次生命機能情報を推定できる高速なパイプラインを開発し、提供する。
(*PreMAFFTash*, *MAFFTash*, *MSThread*, *Spanner*, *SeSAW*, *surFit* 等を改良・成熟させそれらをつないだ検索システムとする)

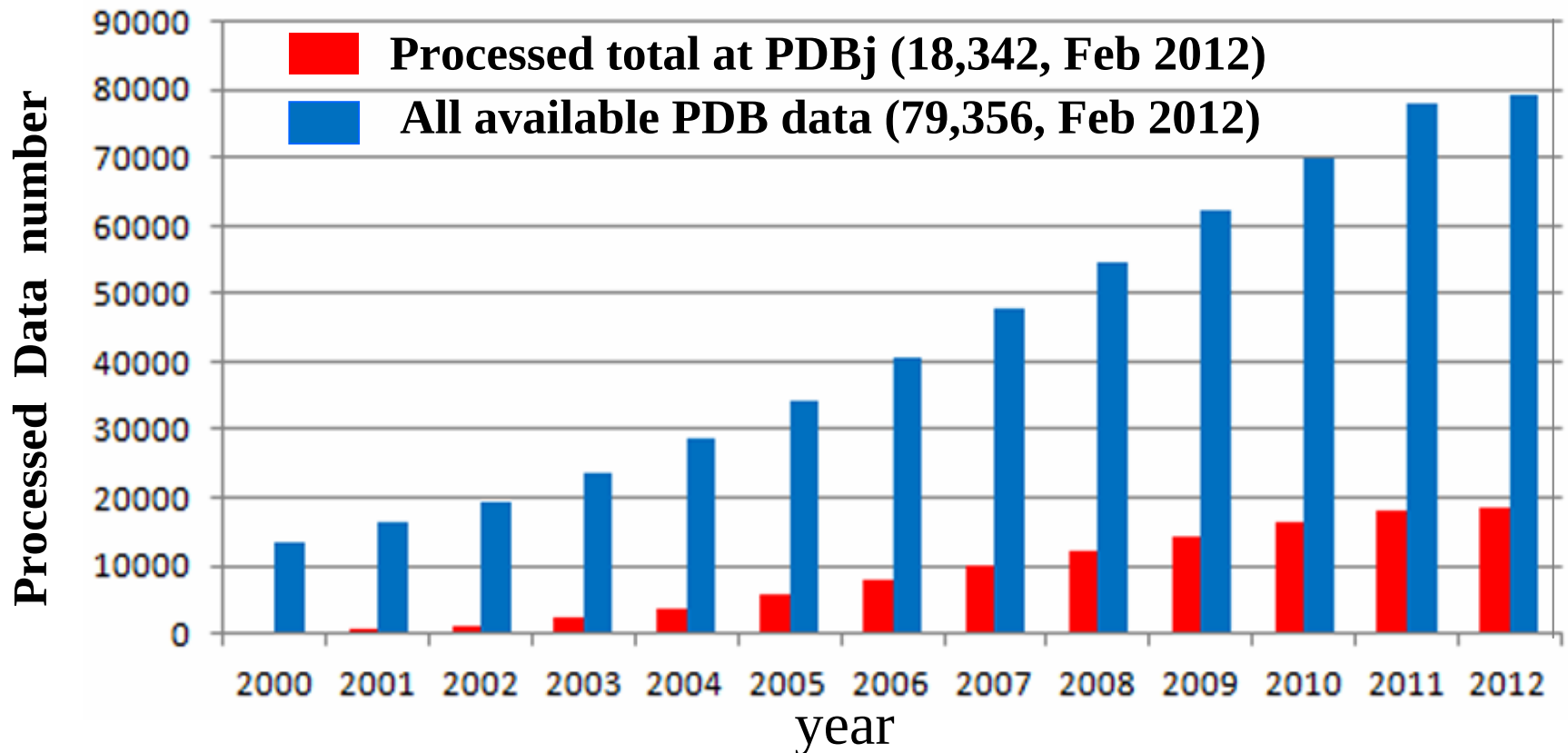
(3-4) 人材養成の実施

- ・**データ寄託・登録者に対する高精度のデータ整理と登録**についての教育: 構造生物学の研究者・学生に対し、データの品質管理について講習会やセミナーで紹介する。
- ・**データ利用者に対する初歩的および高度な利用法**についての教育: 初歩的な利用法と、高度なweb serviceの利用法の双方について、講習会やセミナーで紹介する。
- ・**キュレータ、アノテータの人材養成**: 国際的なon-the-job trainingを実施し、国際感覚を身に付けたキュレータ、アノテータを育成する。

4. 本研究開発課題のH23年度の進捗状況

(4-1) PDBデータおよびNMR実験情報の登録・編纂の国際連携による実施とデータ検証法の開発

- Primary Annotatorsによる欧米との協力したデータ登録の継続的实施。



PDBj curates and processes about a Quarter of the deposited data, mainly from Asian and Oceania regions.

- ・登録業務の標準化・自動化 (Common Annotation & Deposition) システム開発の国際協力 (データ検証部分を担当)。



Validation Tool for Master Format File

Upload Master Format file (PDBx/mmCIF)

C:\Users\harukin\Desktop\rcsb067524.cif 参照...

Run Validation

reset



mmcif file "rcsb067524.cif" has been uploaded

Return to Top page Display uploaded data

```
*** mmCIF_VALIDATE: dictionary= ./mmcif_pdbx_v40.dic
% dictionary.title mmcif_pdbx.dic
% dictionary.datablock_id mmcif_pdbx.dic
% dictionary.version 4.022
```

number of categories: 144 -> 76

Options for Error messages:

Errors for public cif (default) All errors

Category name	data type errors	mandatory errors	primary key errors	enumeration errors	alias errors
Error type	item name	error message			

Data check results

	data type: 0	mandatory: 0	primary key: 0	enumeration: 0	alias: 6
reflns	0	0	0	0	6
struct_conf_type	0	0	0	0	0
struct_biol	0	0	0	0	1
database_PDB_rev	0	1	1	0	4
refine	0	0	0	0	18
refine_hist	0	0	0	0	3

▪ **wwPDB活動を推進。**

wwPDB Advisory Committee meeting at EBI-PDBe, Hinxton, UK on Sept 30, 2011



IUCr 2011 at Madrid, Booth



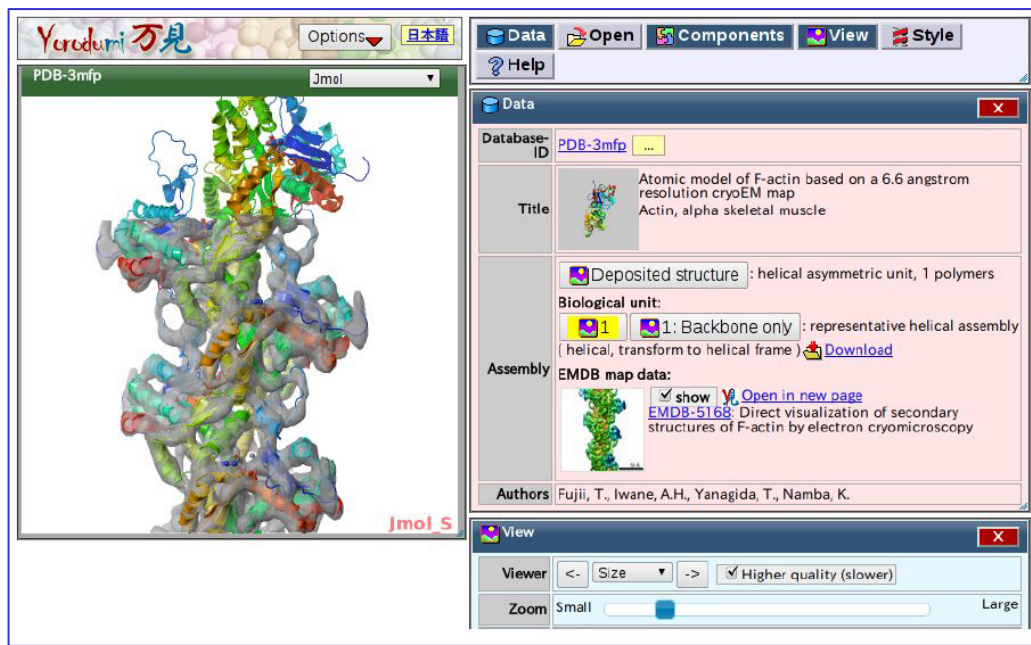
PDB 40th Anniv. Symposium at CSHL, New York



・電子顕微鏡およびX線小角散乱の実験による構造情報データベースの構築と公開



EM Navigator: Viewer of Images of EMDB



Yorodumi: Viewer of both Image and Atomic Structure

- ・PDB データと EMDB の統合化が2012年3月から開始予定。
それに伴い、統合的な検索を可能とした。

- **NBDCとの統合的データベース運営の仕組みを構築し実施。**
 - i) NBDCのデータセンターに、ポータルを設置し、ミラーリングする(構築中)
 - ii) データ配布サービスを、情報・システム研究機構 DBCLS と共同で構築

GIRAFの相互作用部位のstructural motif dataの統合化

<http://pdbj.org/giraf/cmotif/>

Supporting Information for *Composite structural motifs of interaction interfaces for delineating biological functions of proteins* by Akira R. Kinjo and Haruki Nakamura

List of data at TogoDB/compo_motif.

Notes:

- All data for reproducing the results of the paper are provided below.
- All files are gzipped, comma-separated value (CSV) text files.
- In the following, chain ID's are in the form of "[struct_asym.label_asym_id](#)" - "[pdbx_struct_oper_list_id](#)" (e.g., A-1) if [pdbx_struct_assembly](#) category elements are provided in the corresponding PDBML file, or [struct_asym.label_asym_id](#) otherwise. They are different from chain ID's in the PDB format files!
- Most data can be browsed at TogoDB/compo_motif.

List of tables

table name	download file	file size (bytes)	TogoDB	description
bp_top	bp_top.csv.gz	2.6K	compo_motif_bp_top	The upper-most UniProt keywords in the Biological process category
cmotif	cmotif.csv.gz	432K	compo_motif_cmotif	List of composite motifs
emotif_nonpolymer	emotif_nonpolymer.csv.gz	557K	compo_motif_emotif_nonpolymer	List of elementary motifs for non-polymer binding interfaces
emotif_nucleicacid	emotif_nucleicacid.csv.gz	38K	compo_motif_emotif_nucleicacid	List of elementary motifs for nucleic acid binding interfaces
emotif_protein	emotif_protein.csv.gz	886K	compo_motif_emotif_protein	List of elementary motifs for protein binding interfaces
func_all_chain	func_all_chain.csv.gz	501K	compo_motif_func_all_chain	List of subunits associated with all UniProt functions.

(4-2) 統合化に向けたデータベースの高度化

- ・蛋白質構造データの管理・配布システムの高度化: 他のDBとの関連性メタデータとRDF化したPDBデータを提供し、DB統合化環境を整備する。
- ・構造データのオントロジーの改良・開発: mmCIFの記述に基づき、OWLに準拠したRDF/XMLの確立と、他のRDF化がなされたDBとの統合的利用法の確立。

Published online 5 October 2011

*Nucleic Acids Research, 2012, Vol. 40, Database issue D453–D460
doi:10.1093/nar/gkr811*

Protein Data Bank Japan (PDBj): maintaining a structural data archive and resource description framework format

Akira R. Kinjo¹, Hirofumi Suzuki¹, Reiko Yamashita¹, Yasuyo Ikegawa¹,
Takahiro Kudou¹, Reiko Igarashi¹, Yumiko Kengaku¹, Hasumi Cho¹,
Daron M. Standley², Atsushi Nakagawa¹ and Haruki Nakamura^{1,*}

¹Institute for Protein Research and ²Immunology Frontier Research Center, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan

Received August 8, 2011; Accepted September 15, 2011

PDB/RDF for Semantic Web

(Developed by PDBj: Akira R. Kinjo et al.)

<http://pdbj.org/rdf>

PDB/RDF

About PDB/RDF
chem_comp/RDF Search

PDB ID

(e.g., '7RSA')

property:

(e.g.,
'PDBo:entity.pdbx_description')

keywords:

(e.g., 'alcohol')

submit

reset

Download XSLT stylesheet for converting PDBML to RDF: [PDBML2rdf.xsl.gz](#)
(gzipped 22KB)

*) Some web browsers such as Internet Explorer will display the plain XML file. Please use other
browsers that can handle XSLT, e.g. Firefox, Google Chrome, or Safari etc..

PDB/RDF is a collection of PDB data in the **Resource Description Framework** (RDF) format. The RDF format is the **standard format for the Semantic Web**.

An ontology defined in the **Web Ontology Language** (OWL) is also provided for the PDB/RDF, which is a straightforward translation of the **PDB mmCIF Exchange Dictionary** (Kinjo et al. *Nucl. Acids Res.* (2012) 40, D453-D460).

PDB/RDF example

By accessing <http://pd bj.org/rdf/1GOF>, a list of category holders for the PDB entry 1GOF can be retrieved in the RDF/XML format.

Then, a list of category elements can be retrieved (again in the RDF/XML format). Finally, the a particular category element, the list of properties of that element is retrieved.

Subject: <http://pd bj.org/rdf/1GOF>

Predicate	Object
PDBo:datablockName	1GOF-noatom
PDBo:has_atom_sitesCategory	PDBr:1GOF/atom_sitesCategory
PDBo:has_atom_sites_footnoteCategory	PDBr:1GOF/atom_sites_footnoteCategory
PDBo:has_atom_typeCategory	PDBr:1GOF/atom_typeCategory
PDBo:has_audit_authorCategory	PDBr:1GOF/audit_authorCategory
PDBo:has_audit_conformCategory	PDBr:1GOF/audit_conformCategory
PDBo:has_cellCategory	PDBr:1GOF/cellCategory
PDBo:has_chem_compCategory	PDBr:1GOF/chem_compCategory
PDBo:has_citationCategory	PDBr:1GOF/citationCategory
PDBo:has_citation_authorCategory	PDBr:1GOF/citation_authorCategory
PDBo:has_computingCategory	PDBr:1GOF/computingCategory
PDBo:has_database_2Category	PDBr:1GOF/database_2Category
PDBo:has_database_PDB_matrixCategory	PDBr:1GOF/database_PDB_matrixCategory
PDBo:has_database_PDB_revCategory	PDBr:1GOF/database_PDB_revCategory
PDBo:has_database_PDB_rev_recordCategory	PDBr:1GOF/database_PDB_rev_recordCategory
PDBo:has_diffnCategory	PDBr:1GOF/diffnCategory
PDBo:has_diffn_radiationCategory	PDBr:1GOF/diffn_radiationCategory
PDBo:has_diffn_radiation_wavelengthCategory	PDBr:1GOF/diffn_radiation_wavelengthCategory
PDBo:has_entityCategory	PDBr:1GOF/entityCategory
PDBo:has_entity_keywordsCategory	PDBr:1GOF/entity_keywordsCategory
PDBo:has_entity_polyCategory	PDBr:1GOF/entity_polyCategory
PDBo:has_entity_src_genCategory	PDBr:1GOF/entity_src_genCategory
PDBo:has_entity_struct_asymCategory	PDBr:1GOF/entity_struct_asymCategory
PDBo:has_entity_struct_refCategory	PDBr:1GOF/entity_struct_refCategory

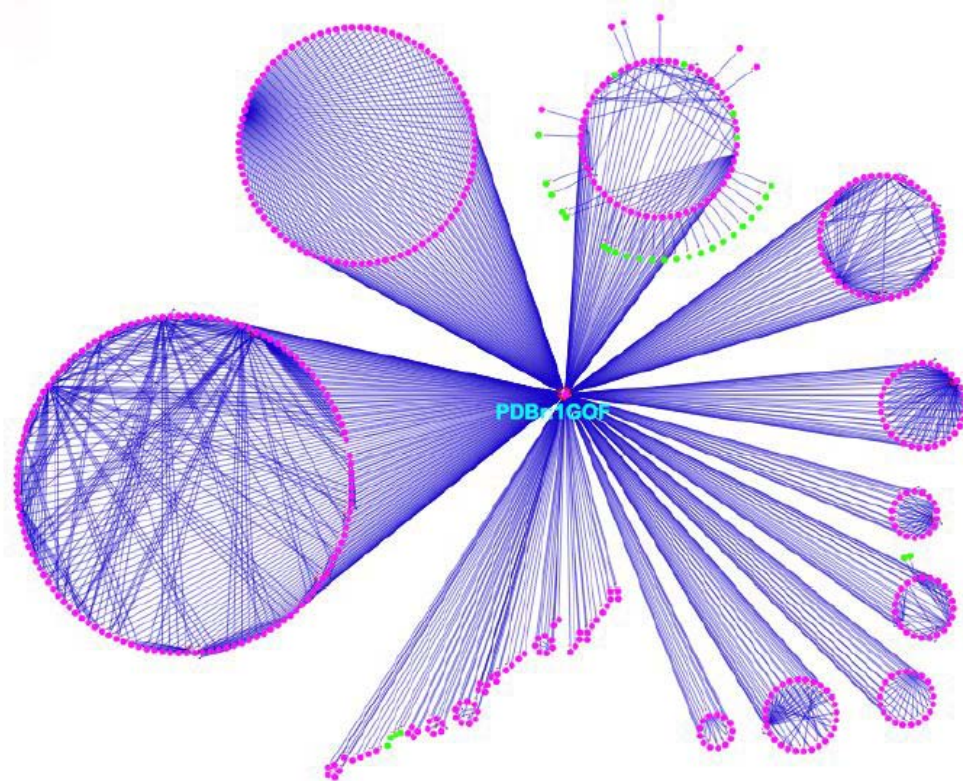
Subject: <http://pd bj.org/rdf/1GOF/entityCategory>

Predicate	Object
PDBo:of_datablock	PDBr:1GOF
PDBo:has_entity	PDBr:1GOF/entity/1
PDBo:has_entity	PDBr:1GOF/entity/2
PDBo:has_entity	PDBr:1GOF/entity/3
PDBo:has_entity	PDBr:1GOF/entity/4
PDBo:has_entity	PDBr:1GOF/entity/5
rdf:type	http://pd bj.org/schema/pdbx-v40

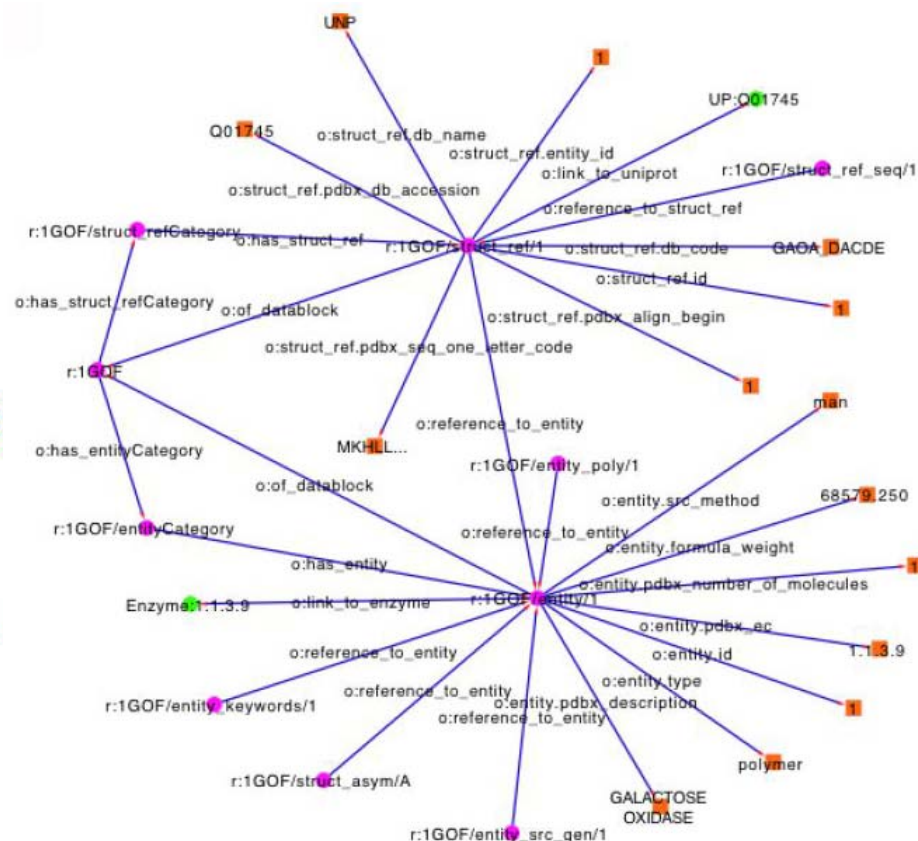
Subject: <http://pd bj.org/rdf/1GOF/entity/1>

Predicate	Object
PDBo:of_datablock	PDBr:1GOF
PDBo:of_category	PDBr:1GOF/entityCategory
PDBo:entity.formula_weight	68579.250
PDBo:entity.id	1
PDBo:entity.pdbx_description	GALACTOSE OXIDASE
PDBo:entity.pdbx_ec	1.1.3.9
PDBo:entity.pdbx_number_of_molecules	1
PDBo:entity.src_method	man
PDBo:entity.type	polymer
PDBo:link_to_enzyme	http://purl.uniprot.org/enzyme/1.1.3.9
PDBo:referenced_by_entity_keywords	PDBr:1GOF/entity_keywords/1
PDBo:referenced_by_entity_poly	PDBr:1GOF/entity_poly/1
PDBo:referenced_by_entity_src_gen	PDBr:1GOF/entity_src_gen/1
PDBo:referenced_by_struct_asym	PDBr:1GOF/struct_asym/A
PDBo:referenced_by_struct_ref	PDBr:1GOF/struct_ref/1
rdf:type	http://pd bj.org/schema/pdbx-v40.owl#entity

Example of an RDF graph



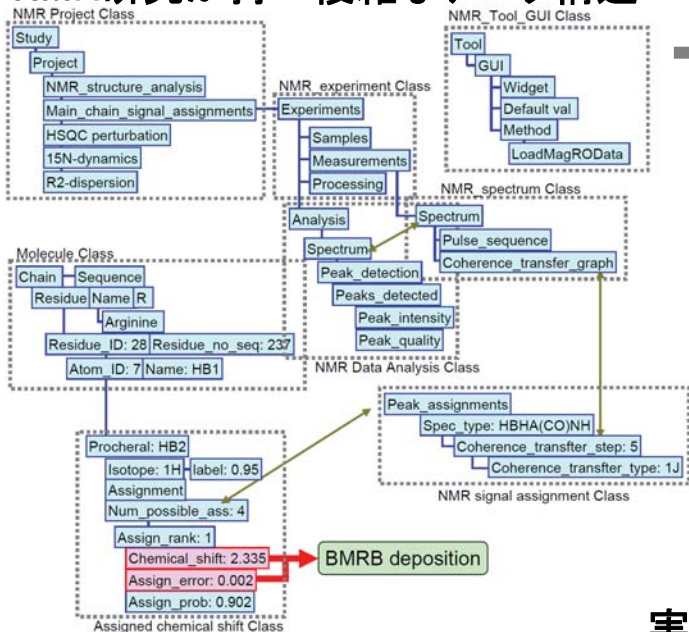
The network of RDF resources for the PDB entry 1GOF.



A subgraph of the left network augmented with literal objects

・NMRデータ構造の管理配布システムの高度化

NMR研究が持つ複雑なデータ構造



STAR形式として表現

```
#
_cyana_input.id 201110101
_cyana_input.user 'Naohiro Kobayashi'
_cyana_input.mag ?

loop_
_hierachy_cyana_input.tag
_hierachy_cyana_input.item
_hierachy_cyana_input.type
_hierachy_cyana_input.item_up

cyana_input_gui_mother_top item top none
cyana_input_gui_mother_top widget norm item
cyana_input_gui_mother_top geom end widget
cyana_input_gui_mother_top color end widget
cyana_input_gui_mother_frame item top none

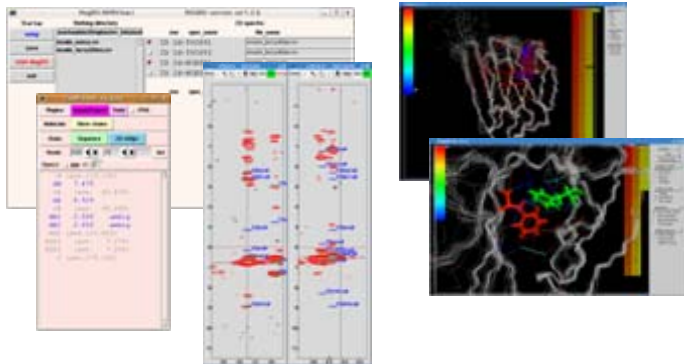
cyana_input_gui_mother_frame widget norm item
cyana_input_gui_mother_frame row end widget
cyana_input_gui_mother_frame col end widget
cyana_input_gui_mother_frame anchor end widget
cyana_input_gui_mother_frame parent end widget
#
```

解析用GUIの自動構築

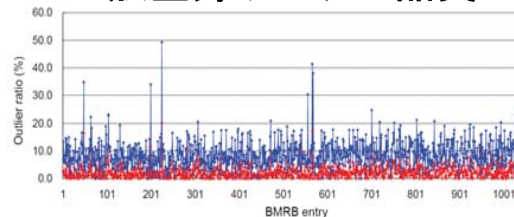


実験データとの整合性評価ツール

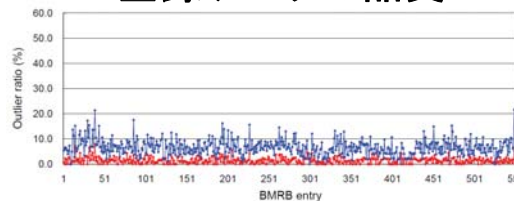
NMRデータの解析と登録準備の支援



一般登録データの品質

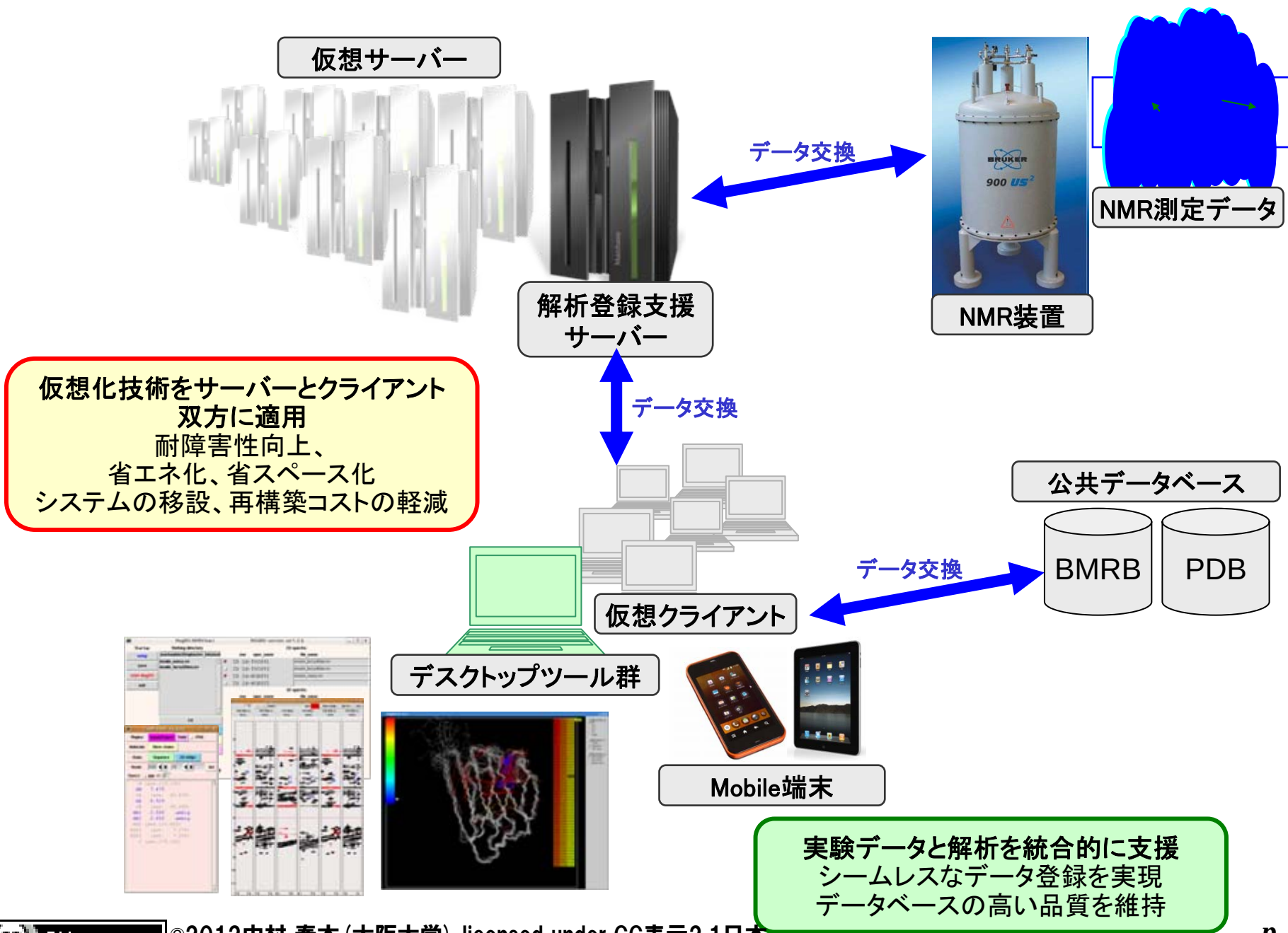


支援システムによる登録データの品質



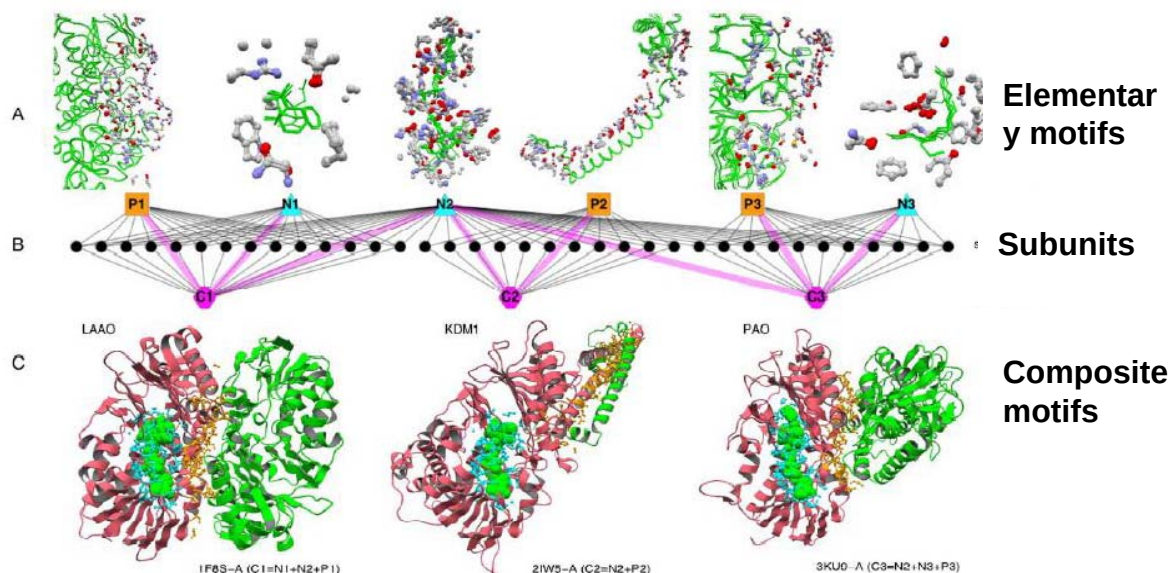
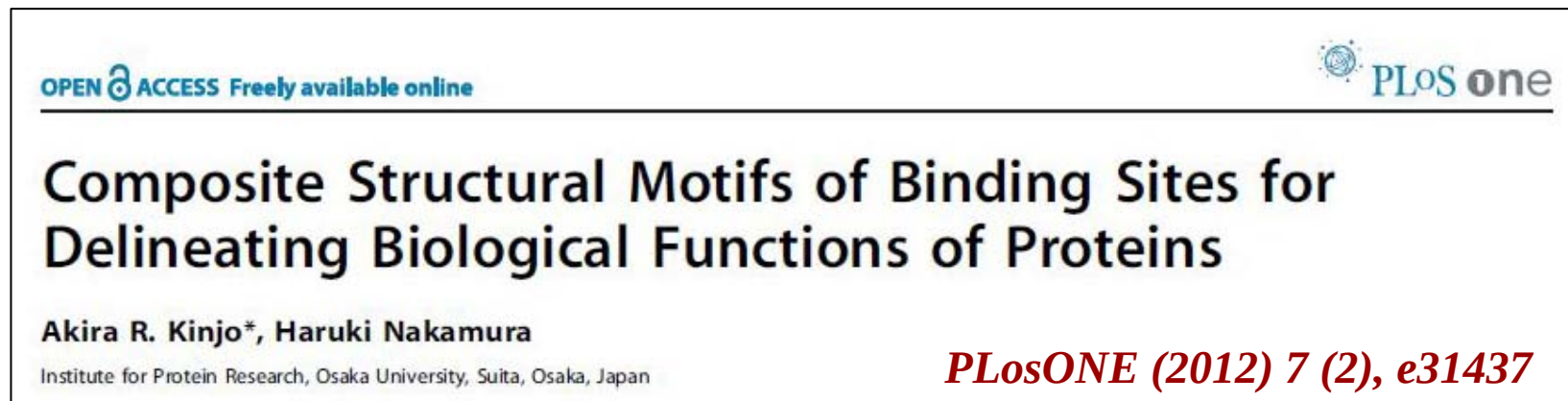
複雑なNMRデータ構造をオントロジーにより記述し、解析と登録作業を統合した支援システムを開発、配布することで高品質なデータベースを実現

・統合を推進する生体系NMRデータ群の網羅的登録



(4-3) 機能情報と構造情報との統合化システムの開発

・**蛋白質相互作用部位データベース**の改良とPDBデータとの統合化: 原子レベルでの蛋白質相互作用を記述するオントロジーを定義し、相互作用部位をRDFで記述する。これにより、原子レベル→残基レベル→分子レベル→分子複合体レベルへつなげて、高次生命機能情報へ、原子レベルの情報を統合化できる環境を整える。



Atom レベルでのstructural motifとして、低分子リガンド結合モチーフ(Ligand binding)と蛋白質間相互作用モチーフ(PPI)があり、それら**複数のモチーフ (elementary motifs) の固有の組み合わせ (composite motif)**によって、蛋白質の固有の高次生命機能が、より明確に整理される。

・配列情報を入力とした、構造情報に基づく機能の解析・推定パイプラインの構築:

- Multiple Sequence Alignment (MSA) for query (MAFFT) and templates (MAFFTash)
- MSA-profile alignment (MSThread)
- Fragment-based structural modeling (Spanner)
- Function prediction (SeSAW) <http://sysimm.ifrec.osaka-u.ac.jp/pipeline5/> (β 版)

Structure & Function Prediction Pipeline

Input Progress Alignments Spanner Sesaw About

Query Sequence

```
>sp|Q5D1E7|ZC12A_MOUSE Ribonuclease ZC3H12A OS=Mus musculus GN=Zc3h12a PE=1
SV=2MSDPCGTPKPVQESNPTMSLWSLEDHRHSSQGRPQDPQDPVAKEAPTSELQMKVDFFRKLGYSSSEIHSV LQKLGVQADTNTVLGELVKHGSATER
ECQALTAPSPQPPLVPRGGSTPKPSTLEPSLPEEDREGSDLRPVVIDGSNVAMSHGNKEVFSCRGILLAVNWFLERGHDTITVFVPSWRKEQRPDPVIT
DQHILRELEKKKILVFTPSRRVGGKRVVCYDDRFIVKLAFESDGVVVSNDTYRDLQGERQEWKRFIEERLLMYSFVNDKFMPPDDPLGRHGPSLDNFL
RKKPLPSEHRKQPCPYGKKCTYGIKCRFFHPERPSRPQRSVADELARANALLSPPTPVKDKSSQRPSPASQSSSVSLEAEPGLDGGKLGARSSPGPHRE
GSPQTCAPAGRSLPVSGGSFGPTEWLAHTQDSLPTYTSQECLDSGIGSLESQMSSELWGVRGGSPGESGPTRGYPYAGYHSYGSKVPAAPSFSPFRPAMGA
GHFSVPTDYVPPPTYPSPREYWSEPYPLPPPTPVLPQEPQRPSPGAGGGPWGRVGD LAKERAGVYTKLCGVFPPLVEAVMRRFPQLLDPQLAAEILSY
KSQHLSE
```

Structure & Function Prediction Pipeline

Input Progress Alignments Spanner Sesaw About

Alignment methods

- ☒ hhpred options
- ☐ ffas options
- ☐ forte options
- ☐ mstthread options
- ☐ psiblast
- ☐ blast

E-Mail

Job id (optional)

Run!

1

1ixs_A	1exn_A	2cqe_A
2w84_A	3h71_A	2cqe_A
3oq9_A	3ix7_A	1m9o_A
1tte_A	1o4w_A	2rhk_A
1cuk_A	2qip_A	1m9o_A
2dna_A	3i8_A	2d9n_A
3a1y_A	3cq5_A	1axh_A
3d0w_A	3e3n_A	2d9n_A
1fad_A	1byk_A	2e5s_A
1wxp_A	3fdb_A	2e5s_A
2ztd_A	2x5d_A	2r_A
2lbf_A	1c7n_A	1_A
3fff_A	2hsg_A	



Spanner Model

(4-4) 人材養成の実施

・データ寄託・登録者に対する高精度のデータ整理と登録についての教育: 構造生物学の研究者・学生に対し、データの品質管理について講習会やセミナーで紹介する。

2011年日本結晶学会年会 ランチョンセミナー –in 北海道–

(2011年11月25日(金)、北海道大学 学術交流会館: 参加者 95名)

「PDBj (PDB Japan) 日本蛋白質構造データバンク」講師: 中村 春木

「PDB (Protein Data Bank) への登録について」講師: 中川 敦史

・データ利用者に対する初歩的および高度な利用法についての教育: 初歩的な利用法と、高度なweb serviceの利用法の双方について、講習会やセミナーで紹介する。

第11回日本蛋白質科学会年会「PDBjサテライトワークショップ」–in 大阪–

(2011年6月6日(月)、大阪大学・吹田キャンパス: 参加者 23名)

(入門編)「PDBj (PDB Japan) 日本蛋白質構造データバンク」講師: 中村 春木

(初級編)「『万見』でPDBのデータを体験する」講師: 鈴木 博文

(中級編)「PDBj Mine のSQL検索とREST API」講師: 金城 玲

第49回生物物理学会年会 ランチョンセミナー –in 兵庫–

(2011年9月16日(金)、兵庫県立大学姫路書写キャンパス: 参加者 110名)

「PDBj (PDB Japan)の活動とNBDCおよびwwPDBとの連携」講師: 中村 春木

「PDBj のRESTful ウェブサービスとセマンティックウェブサービス」講師: 金城 玲

H23年度大阪大学蛋白研セミナー「PDBj講習会」ーin 福岡ー (2012年2月7日(火)、九州大学・病院キャンパス: 参加者 34名)

H23 年度大阪大学蛋白質研究所セミナー
PDBj 講習会 in 福岡

日時: 2012年2月7日(火) 10:00-17:20 (9:30受付開始)
場所: 九州大学・病院キャンパス
(〒812-8682福岡市東区馬出 3-1-1)
総合研究棟 IT ルーム (2 階)
主催: 大阪大学蛋白質研究所 日本蛋白質構造データバンク
対象: PDBj を利用される方などでも歓迎します (定員 30 名)
内容: PDBj の web サービスなどの利用についての講義および
PC を用いた実習
費用: 無料
2012 年 1 月 10 日より参加登録受付を開始します。
http://www.pdbj.org/pdbj_workshop_j_20120207.html

2月7日(火) 10:00-17:20

10:00-10:40	PDBj の紹介と PDBj Mine の使い方 金城 玲 (大阪大学蛋白質研究所)
10:50-12:30	PDBML/mmCIF に関すること 伊藤暢聡 (東京医科歯科大学)
13:30-14:50	複雑な構造の見方 鈴木博文 (大阪大学蛋白質研究所)
15:00-16:20	ProMode-Elastic: 弾性ネットワーク モデルによる基準振動解析データベース 輪湖 博 (早稲田大学)
16:30-17:20	構造からの機能予測入門 木下賢吾 (東北大学)

PDBj
Protein Data Bank Japan

〒565-0871 大阪府吹田市山田丘 3-2
大阪大学蛋白質研究所 日本蛋白質構造データバンク
TEL: 06-6789-4311 <http://pdbj.org/>



- ・**キュレータ、アノテータの人材養成**: 国際的なon-the-job trainingを実施し、国際感覚を身に付けたキュレータ、アノテータを育成 (**5名のPrimary Annotators**)

5. 本研究開発課題のH24年度の実施計画

(5-1) PDBデータおよびNMR実験情報の登録・編纂の国際連携による実施とデータ検証法の開発

- ・Primary Annotatorsによる欧米との協力したデータ登録の継続的实施。
- ・登録業務の標準化・自動化(Common Annotation & Deposition)システム開発の国際協力(データ検証部分を担当)。
- ・wwPDB活動を推進。(2012年10月12日にwwPDBACを大阪で開催。wwPDBとしての教育面を重要視した第1回目のアウトリーチ活動も行う。)
- ・クラウドコンピューティング技術を用いたNMR実験データの統合的管理システムを構築。
- ・NBDCとの統合的データベース運営の仕組みを構築し実施。

(5-2) 統合化に向けたデータベースの高度化

- ・蛋白質構造データの管理・配布システムの高度化: RDF化したPDBデータをさらに高度化し、他のRDF化がなされたDBとの統合化環境を整備。
- ・wwPDBとしての新たなフォーマット開発とその記述検証法の開発
- ・ユーザインターフェースの改良: 継続的にポータルの改良を行い、多言語でのweb page記述とグラフィックスビューアによるGUIの高度化を実施。
- ・NMR実験データの管理・配布システムの高度化: オントロジー工学に基づく標準的なXML記述法を開発し、外部DBとの統合的利用を可能として、薬剤や蛋白質間の相互作用、動力学解析データ等の登録を支援する仕組みを開発。H24年度はプロトタイプを作成

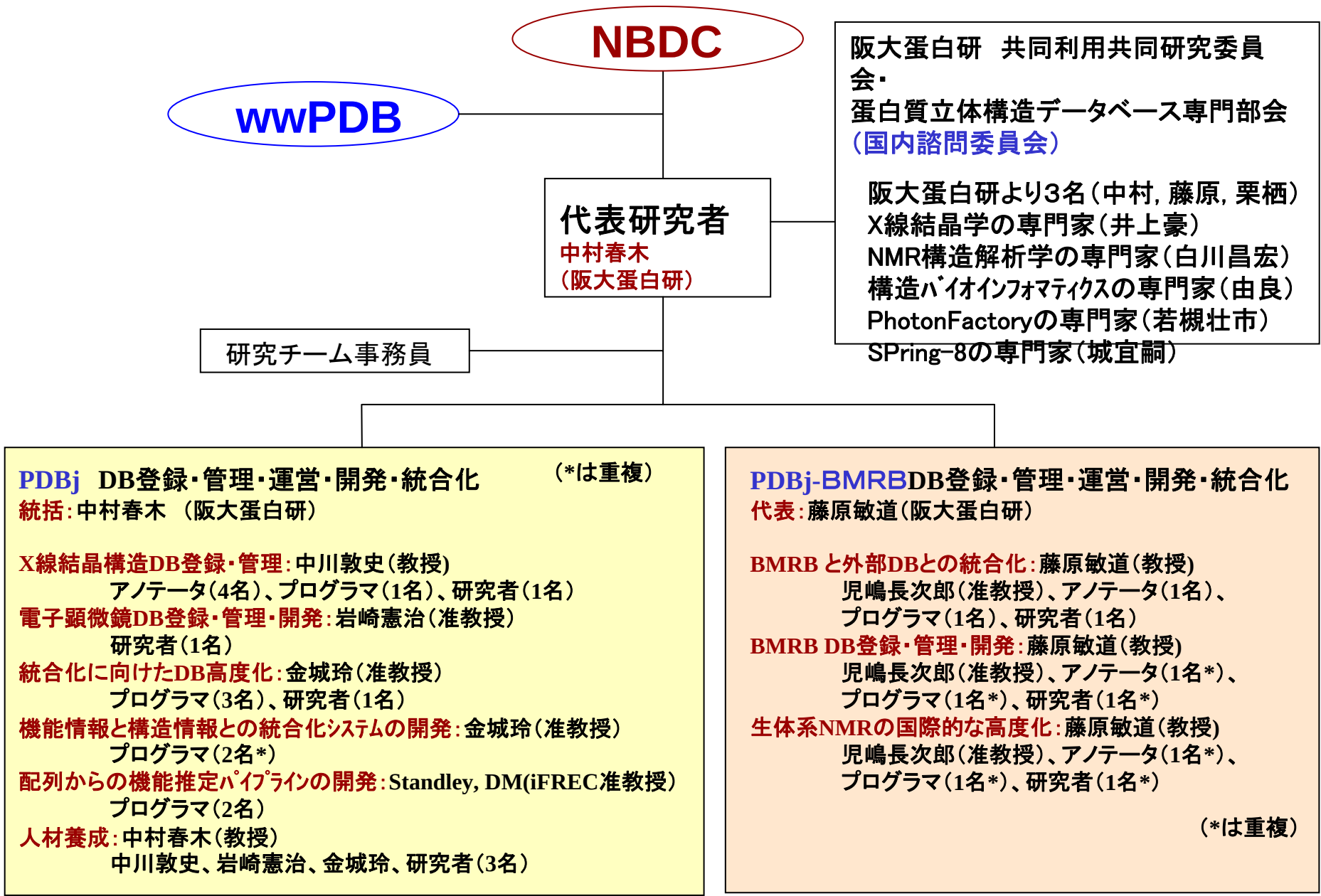
(5-3) 機能情報と構造情報との統合化システムの開発

- ・**蛋白質相互作用部位データベース**の改良とPDBデータとの統合化: 蛋白質相互作用を記述するGIRAF, eF-site/eF-seek等に対してもSemantic Web技術を適用したサービスとし、原子レベル→残基レベル→分子レベル→分子複合体レベルへつなげて、高次生命機能情報へ、原子レベルの情報を統合化できる環境を整える。
- ・**配列情報を入力とした、構造情報に基づく機能の解析・推定パイプライン**の構築: ゲノム解析等によって得られた配列情報から、構造情報あるいは複合体を含むモデル構造情報を経由し、高次生命機能情報を推定できる高速なパイプラインを改良・提供する。H23年度に、*PreMAFFTash*, *MAFFTash*, *MSThread*, *Spanner*, *SeSAW*, *surFit* 等で構築されたパイプラインを、複数のthreading手法(*HHPred*, *FFAS*, *FORTE*)が利用できるように改良して成熟させ、パイプラインによる機能推定が行えるシステムとする。

(5-4) 人材養成の継続的实施

- ・**データ寄託・登録者に対する高精度のデータ整理と登録**についての教育: 構造生物学の研究者・学生に対し、データの品質管理について講習会やセミナーで紹介する。
- ・**データ利用者に対する初歩的および高度な利用法**についての教育: 初歩的な利用法と、高度なweb serviceの利用法の双方について、講習会やセミナーで紹介する。
- ・**キュレータ、アノテータの人材養成**: 国際的なon-the-job trainingを実施し、国際感覚を身に付けたキュレータ、アノテータを育成する。

6. 実施体制



7. 保有しているあるいは今後保有する予定のデータベース

	名称	データ件数 (2012年2月)	アクセス件数 (2011年4月-2012.1月末)	概要
1	PDB : wwPDB (PDBjによる累積登録処理数)	79,356 (18,342)	PDBjMine:7,653,942 (file download数: 32,285,624)	蛋白質等の生体高分子の国際的立体構造データバンク。PDBjMineはその検索用RDBシステム
2	PDB/RDF	79,356	335,797	RDF化したPDB data(2011年9月から)
3	BMRB : wwPDB (PDBj-BMRBによる累積登録処理数)	7,595 (829)	5,195,244	生体高分子の化学シフト、緩和データ、相互作用データ等のNMRの実験データのデータバンク
4	EM-Navigator	1,280	4,500,734 (file download数 : 54,746)	生体高分子の電子顕微鏡画像データベース(EBIのEMDBデータベースを基に開発)
5	GIRAF	652,027	4,323	PDBデータに基づく相互作用部位データベース
6	eF-site	445,517	10,179	PDBデータに基づく分子表面データベース
7	ProMode	23,768	31648	PDBデータに基づく基準振動データベース(オリゴマーやElastic Network Modelも含む)
8	eProtS (英語および日本語)	353	338,808	高校生以上の学生・社会人向け蛋白質構造百科事典
9	MoM (日本語)	146	230,702	米国RCSBで提供されている月替わりの分子構造紹介記事を日本語に翻訳

公開できないデータ:なし

即時公開できないデータ:wwPDBの投稿時データ(理由:論文がpublishされるまで、または半年あるいは1年間の猶予期間をwwPDBで設定し、その間は即時公開しないでも良いルールとなっている)

8. 開発・運営しているデータベースの解析ツール/サービス

名称	アクセス件数 (2011年4月–2012.1月末)	概要
Spanner	3,719	既知の立体構造との類似性に基づいて、構造が分かっていないアミノ酸配列の立体構造を予測するプログラムおよびそれを用いたウェブサービス
SeSAW	31,648	機能未知の蛋白質構造データに対して、構造の類似性検索に基づき、局所的および大域的なアミノ酸配列の類似性も加味して、ファミリーあるいはスーパーファミリーを推定し、生化学的機能のアノテーションを行うサービス
Sequence Navigator	20,281	似た配列を持つ分子を検索するツール
Structure Navigator	2,565	類似の立体構造を持つ分子を検索するツール
Protein Globe	71,658	代表的な蛋白質を点で表し、構造の似ている分子ほど距離が近くなるよう球上に配置して蛋白質の世界を表現したもの。類似フォールドの検索も行える
MAFFTash	4,323	構造アライメントASHの出力を制約条件として配列・構造の統合マルチプルアライメントを構築するシステム
eF-surf/eF-seek	607/374	PDBフォーマットを入力として、表面の形状と物性(静電ポテンシャルと疎水性度)を計算するWeb サービス / PDBフォーマットのファイルを入力とした、表面の類似性検索による機能部位の予測法のWebサービス。
Yorodumi	7,655,506	生体分子の3次元構造(PDBおよびEMDB)を眺めて、学んで、楽しむためのウェブページ