## PDBのデータとその見方・探し方

金城玲 大阪大学蛋白質研究所 日本蛋白質構造データバンク (PDBj) http://www.pdbj.org

Mar. 3, 2008

### PDBjでできること

生体高分子(蛋白質・核酸)立体構造データの・・・

- 登録:構造決定の論文を発表する時には必須.
- 検索: 配列,構造そのものや構造に基づく機能情報など.



## PDBjのサービス一覧

- PDB ID によるエントリの取得
- キーワード検索 / より進んだ検索 / XMLを利用した検索 (xPSSS)
- アミノ酸配列検索 (Sequence-Navigator)
- 立体構造検索 (Structure-Navigator)
- 立体構造比較 (ASH)
- 蛋白質の動きのデータベース (ProMode)
- 蛋白質の分子表面データベース (eF-site)
- 電子顕微鏡構造データベース (EM-Navigator) New!
- 立体構造空間 GUI (Protein Globe) New!
- 蛋白質構造百科事典 (eProtS)
- 分子グラフィクスソフト (jV)

## PDBのデータ形式

**PDB format** 伝統的な PDB のファイル形式.いわゆるフラットファイル.コラム指向(昔の FORTRAN で用いられていたようなテキスト形式).

**PDBML** XMLファイル. wwPDB で策定した XML Schema によって形式が厳密に定義されて いる.

Westbrook et al. (2005), Bioinformatics 21(7):988-992.

mmCIF (省略)

http://www.wwpdb.org/docs.html



## PDBの可視化ソフト: jV3



- 無料.
- RasMol互換のコマンド体系.
- 分子表面も表示できる.
- PDBMLに対応.
- アニメーションもできます.
- OpenGLできれいな表示.
- Mac, Win, Linux で動作可. (要 Java 環境)

## エントリを取得してデータを眺める



### エントリ取得





「PDB ID」のフォームに(たとえば) 「1gof」と入力して「GO」ボタンをク 1gofのデータの要約が表示される. リック.

### PDB

## エントリの情報いろいろ: Structural Details



### 構造の詳細.

- 分子の名前
- 分子量
- 分子機能
- などなど……

## エントリの情報いろいろ: Experimental Details

▶ C + Ohttp://p	dbjs3.protein.osa	ika-u.ac.jp/xPSSS	/DetailServlet				- Q- ep	rots	C
Globe(ml) ARK Matras	blog アップル	Amazon.co.jp	ProtS ==== ス (32)	6) - GTOP	PubMed	PDBj Members' pag	e Gmail	Scopus	
PDBj oterin Data Bank Japan	xPSSS (xml-b	ased Pro	tein Struc	ture S	earch	Service)			
	Experia	nental Deital	s Page (X-RAY	Details)					
		[1	GOF]						
Summary Structural Det	ails Experime	ntal Details	Functional Detail	Sequenc	e Neighbo	r Download/Disp	ay Lin	k	
[Back] [xPSSS Top] [PDE	ij Top]				(se	arch • PDB ID	keywo	rd	
Cell axes [Å]	98.000	89.400	86.700						
Cell angles [degree]	90.00	117.80	90.00						
Spacegroup		C121							
Resolution limits [Å]		10.0 *1 - 1.7 *1							
R-factor		0.177							
R-free									
RMSD bond length [Å]		0.018							
RMSD bond angle [Å]		0.05 *1							
Data Collection Statis Resolution limits [Å] Number of reflections	stics 10.0 *2 - 1.7 ( 57092 *2 (5	1.8 - 1.7) 908) *2							
Number of measurement	s 117980 *2 (1	1365) *2							
Rmerge_I_obs	0.039	2							
Completeness [%]	79.1 *2 (52	6) <sup>*2</sup>							
Redundancy									
Values in parentheses refer Crystallization Condi	to the highest re i <b>tions</b>	solution shell.							
method pH temparatur	e								

実験情報の詳細.

- 結晶格子のパラメータ
- 実験手法 (X線, NMR, 電顕, 中性子など)
- 試薬などなど……

## エントリの情報いろいろ: Functional Details

© X

200		my) 2220v	-based Prot	ain Structure Sear	ch Service)	Pasult		
4 1 6	+ Ahtto:/	/odbis3 protein osaka-u ac ii	/xPSSS/Deta	ilServlet	en service)	resure		Qv eprots
M Globe	(ml) ARK Matra	s blog ZyZik Amazona	o.in eProtS	= 1 = 2 (3246) T	GTOP Pul	bMed	PDRi Members' page	Gmail Scopus
Protein Data	DBj Bask Japan	xPSSS (xml-based	Protei	n Structu	re Sea	rch	Service)	
	Functional Datails Page [IGOP] [Sumary] (Structural Datails) [Experimental Datails) [Functional Datails] (Seguence Heighber ] [Demload/Display ] Link, [Back] IAESSS Top] (PDB) Top] [Control of the second s							
			[1GOF	1				
Summa	ry Structural D	Details Experimental Det	ils Funct	ional Details	Sequence Ne	ighbor	Download/Displa	y Link
[Back]	[xPSSS Top] [PI	DBj Topl					arch  • PDB ID (	keyword
shout B	unctional Details	Page						
about P	uncuonal Detalis	rago						
Funct	ional Informa	tion from DDP atom	acordina	too for the "H	ETATM		ding citor	
runci	ional informa	tuon from PDB atom	coordina	les for the "h	ETAIM	- DIII	ung sues	
site_id:	ACY_1gof_0_7	03						
type	bir	nding site						
Numbe	er of Residues 11							
Details	A	CETIC ACID binding site						
Chain	Residue	ligand						
A	PHE194	ACY: ACETIC ACID						
A	PHE227-CYS22	28 ACY: ACETIC ACID						
A	TYR272	ACY: ACETIC ACID						
A	TRP290	ACY: ACETIC ACID						
A	TYR329-ARG3	30 ACY: ACETIC ACID						
A	PHE464	ACY: ACETIC ACID						
A	TYR495-HIS49	6 ACY: ACETIC ACID						
Α	HIS581	ACY: ACETIC ACID						
site id:	ACY 1gof 0 7	01						
type	bi	nding site						
Numbe	r of Residues 11	-						
Details	A	CETIC ACID binding site						
Chain	Residue	ligand						
A	TRP336	ACY: ACETIC ACID						
A	ARG371	ACY: ACETIC ACID						
A	ALA378-PRO37	79 ACY: ACETIC ACID						
A	ALA381	ACY: ACETIC ACID						
	THR 398.GI Y40	ACY: ACETIC ACID						

機能情報の詳細。

- Gene Ontology による分類
- リガンド結合部位
- 活性部位

### エントリの情報いろいろ: Sequence neighbor

		PSSS (xml-based	Protein Structure Sean	ch Service) Resi	lt	6	_
	http://pdbjs3.protein.os	aka-u.ac.jp/seqna	wix/SeqHomXServlet			• Q+ eprots	6
PPDR rotein Data Bank Japan	xPSSS (xml-t	ased Pro	otein Structu	re Searc	h Service)		
		Sequence N	leighbor Page				
		[1	GOF]				
Summery Structu [Back] [xPSSS Top	ral Details Experim [PDBjTop]	ntal Details)	Functional Details	Sequence Neigh	oor Download/Displa search • PDB ID (	ay Link	
Results (1-22)	/ 22						
<u>.GOF</u> 1 639 ASAPI . <u>T2XA</u> 1 639 ASAPI	GSAISRNNWAVTCDSAQS GSAISRNNWAVTCDSAQS	NECNKAIDGNKD	TFWHTFYGANGDPKPPHTY	TIDHKTTONVNO	LSMLPRODGNONGWIGH	HEVYLSSDOTNWGS	PVASGS PVASGS
Req. Identity: 9! New Search [1K3IA] GOF 1639 ASAPI	N Seq. Positives: Structural Superpositions SSAISRNNWANTCDSAQS	99% E-value: osition NECNKAIDGNKD	0.0 Score: 1315 Con	npound: GALA	LISHLINGDONQHONION	CURSOR RHEVYLSSDGTNWGS	PVASGS
Seq. Identity: 9! New Search [1K3IA] GOF 1 639 ASAPI [K3IA 13 639 ASAPI Beg. Identity: 35	N Seq. Positives: <u>Structural Superp</u> GSAISRNNHAVTCDSAQS GSAISRNNHAVTCDSAQS N Seq. Positives:	99% E-value: psition NECNKAIDGNKD NECNKAIDGNKD 55% E-value:	0.0 Score: 1315 Con FFWETFYGANGDPKPPETY FFWETFYGANGDPKPPETY 2e-19 Score: 94 Con	npound: GALA TIDMKTTQNVNG TIDMKTTQNVNG	CTOSE OXIDASE PRE LSMLPRQDGNQNGWIGP LSMLPRQDGNQNGWIGP ERIAL SIALIDASE	CURSOR RHEVYLSSDGTNWGS RHEVYLSSDGTNWGS	PVASGS PVASGS
Geq. Identity: 9!           New Search [IK3IA]           LGOP         1 639 ASAPI           KSIA         13 639 ASAPI           Geq. Identity: 35           New Search [IW80A]           LGOP         19 127           LMBOA         472 127           W80A Exact         4520 Exact	V Seq. Positives: <u>Structural Superp</u> GSAISRNNWAVTCDSAQSG GSATSRNWAVTCDSAQSG V Seq. Positives: <u>Structural Super</u> DSAQSGNECNKAIDGNN <u>BYAREGGASNVIDGNN</u> <u>149NA</u>	99% E-value: <u>sstion</u> NNECNKAIDGNKD NNECNKAIDGNKD 55% E-value: <u>cosition</u> LDTFWETFYGANGI STFWETEWSRAD	0.0 Score: 1315 Co CFWHEFYGANGDPKPPHT FWHEFYGANGDPKPPHT 2e-19 Score: 94 Co DPKPPHTYTIDMKTTQNVN PGYPHRISLDLGGTHTS	npound: GALA TIDHKTTQNVNC TIDHKTTQNVNC npound: BACT IGLSKLPRQDGNQ GLQYTRRQN-SJ	TOSE OXIDASE PRE LSMLPRQDGNQNGWIGR LSMLPRQDGNQNGWIGR ERIAL SIALIDASE NGWIGRHEVYLSSDGTM NEQVADYEIYTSLNGTT	CURSOR REEVYLSSDGTNWGS REEVYLSSDGTNWGS REGSPVASGSMFADS RHGSPVASGRFTTSL	PVASGS PVASGS TTKYSN APQRAV
Seq. Identity: 9! New Search [1K31A] 1602 1 639 ASAPT 1K31A 13 639 ASAPT Seq. Identity: 32 New Search [1W80A] 1602 19 127 1W80A Exact 14tches: 1994. Identity: 35	<ul> <li>Seq. Positives: Structural Superpresentation of the second structural structura structural structural structura s</li></ul>	99% E-value: spition INECNKAIDGNKD: S5% E-value: cosition CDTFWETFYGANGI STFWETEWSRADJ 55% E-value:	0.0 Score: 1315 Co CFWHIFYGANGDPKPPHT FFWHIFYGANGDPKPPHT 2e-19 Score: 94 Co PFKPPHTYIDHKTYGNVS APGYPHRISLDLGGTHTIS 2e-19 Score: 94 Co	npound: GALA TIDMKTTQNVNG TIDMKTTQNVNG npound: BACT GLSMLPRQDGNG GLQYTRRQN-SJ npound: SIAL	LSMLPRQDONQNGWIGR LSMLPRQDONQNGWIGR LSMLPRQDONQNGWIGR ERIAL SIALIDASE NGWIGRHEVYLSSDGTN NEQVADYEIYTSLNGTT	CURSOR HEEVYLSSDOTNHOS HEEVYLSSDOTNHOS HNGSPVASGSHFADS HNGSPVASGSHFADS	PVASGS PVASGS TTKYSN APQRAV
Seq. Identity: 91 New Search [IKIA] [IG02] 1 639 ASAP1 [IG13] 13 639 ASAP1 Seq. Identity: 35 New Search [IWBOA] [IG02] 19 127 IMBOA 472 127 IMBOA 472 127 IMBOA 472 127 IMBOA 172 127 IMBOA 17	A Seq. Positives: <u>Bircitural Superpresent</u> <u>Structural Superpresent</u> <u>Bircitural Superpresent</u> <u>B</u>	99% E-value: ssition INECNKAIDGNKD: SS% E-value: sosition SS% E-value: SS% E-value: ssfwattewSRADJ SS% E-value: ssfwattewSRADJ	0.0 Score: 1315 Co erratriandopkopatri 2e-19 Score: 94 Co opkopatriidaktoover 12e-19 Score: 94 Co 2e-19 Score: 94 Co opkopatriidaktoover poopkissiloloottiis	npound: GALA TIDKKTYQNVNG TIDHKTYQNVNG npound: BACT GLSHLPRQDGNG GLQYTRRQN-SJ GLQYTRRQN-SJ GLQYTRRQN-SJ	LSNLPRODUNGNUT IS LSNLPRODUNGNUT IS LSNLPRODUNGNUT IS LSNLPRODUNGNUT IS REAL SIALIDASE MONIGREVILSSOFT ISASE	CURSOR HEEVILSSDOTNIGS HEEVILSSDOTNIGS INGSPVASGSHPADS INGSPVASGSHPADS INGSPVASGSHPADS	PVASGS PVASGS TTKYSN APQRAV TTKYSN APQRAV
ieq. Identity: 9/ New Sarch [LXIA] (GO7 1 639 AGART IKAIA 13 630 AGART	A Seq. Positives: Rinciumal Superpresent Sinskinkarvichsages A Seq. Positives: Bindival Super- Endestrations Endestrations Rinciumal Super- Statestantives: Sinskinkarvichs Sinskinkarv	994 E-value: sation NECKRAIDONKO SS& E-value: contion DIFMETFIGANG SS& E-value: contion SS& E-value: SS& E-value: SS& E-value:	0.0 Becrei 1315 Co errettraakoopkopat 22-19 Scorei 94 Co opkopattriokktegevo heorphristoldommis 22-19 Scorei 94 Co opkopattriisklodommis 22-19 Scorei 94 Co	apound: GALA TIDNKTTQNVNG TIDNKTTQNVNG TIDNKTTQNVNG apound: BACT apound: SIAL MGLSMLPRQDONG GAQTTRRQN-SJ apound: SIAL	LINE QUINGHI UN LINE PROMONICA BALLING DAMANIA REALING DALLIDASE NONIGRIEVYLSSDOTM MEQUADYEIYTSLMOTT IDASE RIGUADYEIYTSLMOTT IRIAL SIALIDASE	CURSOR HEVYLSSDOTNHOS HEVYLSSDOTNHOS NGSPVASGENYADS NGSPVASGENYADS NGSPVASGENYADS	PVASGS PVASGS TTKYSN APQRAV TTKYSN APQRAV

類似のアミノ酸配列をもつ他のPDBエントリの リスト.

- BLASTで検索した結果が表示される。
- 複数の配列を含むエントリの場合、ユーザが 一つの配列を選択する。
- 後述の Sequence Navigator を利用している.

### エントリの情報いろいろ: Download / Display



PDBの「生データ」を表示またはダウンロード する.

- 伝統的な PDB 形式.
- 伝統的な PDB 形式の原子座標を除く部分 (ヘッダ部分).
- mmClF ファイル(主に PDB 内部で使われる ファイル形式:すべてのファイルの元データ).
- PDBのXML形式 (PDBML).
- PDBMLの原子座標を除く部分 (no-atom).
- PDBMLの原子座標のみの部分の簡易形式 (ext-atom).
- 構造因子.

### エントリの情報いろいろ: Link

▶ C + O http://pdbjs3.p	rotein.osaka-u.ac.jp/xP	SSS/DetailServlet				Q. eprots	
Globe(ml) ARK Matras blog	アップル Amazon.co.jp	eProtS ===-7	(3246) * GTOP	PubMed	PD8j Members' page	Gmail Sco	pus
PDBj xE	SSS ml-based Pr	otein Str	ructure S	earch	Service)		
	Li	nks Page					
		[1GOF]					
Summary Structural Details	Experimental Details	Functional D	stails Sequenc	e Neighbor	Download/Display	Link	
[Back] [xPSSS Top] [PDBj Top]					arch    PDB ID (	keyword	
PCSR-PDR (Protein Data	Rank)						
MSD-EBI (Macromolecula	r Structure Databas	e)					
CATH(Class, Architecture	, Topology and Hom	-/ iologous superf:	amily)				
<u>CE</u> (Combinatorial Extens	on of the optimal par	h)	-				
<u>eF-site</u> (electrostatic surface	of Functional-site)						
FSSP(Fold classification background bac	used on Structure-Stu	ructure alignme	nt of Proteins)				
SCOP(Structural Classific	ation Of Proteins)						
<ul> <li><u>UniProt(</u>the universal prot</li> </ul>	ein resource) ( <u>UNP</u>	- Q01745)					
VAST(Vector Alignment S	iearch Tool)						
- FaCotDR (A database of F	www.Cotobutio.Mo	(ampione)					

該当するPDBエントリの2次データベースなど へのリンク.

- RCSB-PDB, MSD-EBI: 米国と欧州のPDB.
- CATH, SCOP, FSSP: 構造の分類データベース.
- UniProt: アミノ酸配列データベース.
- KEGG: 代謝経路など様々な生物データベー スの複合体.
- EzCatDB: 酵素反応のデータベース.
- などなど……

### キーワード検索





「Keyword」のフォームに(たとえば) 「hemoglobin」と入力して「GO」ボタ ンをクリック.

ヒットしたエントリのリストが表示さ れる.

## より進んだ検索

000	xPSSS A	dvanced Search	
+ ttp://pdbjs3.protein	osaka-u.ac.jp/xPSSS/ad	vanced.html	😡 ^ Qr Google
[]] Globe(ml) ARK Matras blog アップ	Il Amazon.co.jp ePro	S ニュース (3252) ▼ GTOP	PubMed PDBj Members' page Gmail 🔊
PERFORMENT REPORT REPORT	S based Prote	in Structure S	Search Service)
	Advan	ced Search	
[PDBj Top] [xPSSS Top]			
search sorted by PDB ID	: Perform a	in exact word match	set
PDB ID:		Search theoretical n	odel
Release Date:	after month month before +	day year t day year	
Citation Author:	1 2 3 c.g.: Ito, N.		and and Authors of primary citation only!
Journal:	Title Name Year Volume		
Contains Chain Type:	Yes No ignore O O O polyp O O O polyp O O O polyd O O Polyd O O Polyd O O Polyd O O Polyd	eptide(L) eptide(D) eoxyribonucleotide bonucleotide uccharide(D)	
Compound Information:			
Title:			
Other DB :	; ID:		
Ligands and Prosthetic groups:			
Number of Chains and Chain longth	Number	Length	
rumber of Chains and Chain length:			

トップページの「Advanced Search」の矢印をク リック.

いろいろな条件で検索できます.

- 著者名および論文情報
- 実験手法
- リガンド名
- 残基数
- 解像度
- 生物種
- などなど……

## 他にもいろいろできますが……

配列検索、構造検索、などなどは D. M. Standley の講義で。



## PDB のファイルの中身と仕組み

- PDB ファイル
- PDBML ファイル

これらを実際に開いて読んでみましょう。



### 復習:エントリの取得





「PDB ID」のフォームに(たとえば) 「1gof」と入力して「GO」ボタンをク リック.



PDB format ファイルを眺める.



PDB	xPS	SS I-based	Protein Struch	ire Sea	urch Se	ervice)	"
Procent Lina Base Japan		D	ownload Page	and a	CU.		
			[1GOF]				
Summary Structural	Details Expe PDBjTop]	rimental Deta	ils Functional Details	Sequence Ne	ighbor Do search	wnload/Display Link ⊙PDB ID ⊜keyword	
	<b>6</b> 10 <b>6</b>	mat	fle nome	Display	Download	กไ	
	PDB format	all	pdb1gof.ent(442k)	(display)	download download		
		header only	pdb1gof.ent.gz(6k)	display	download		
	mmCIF	all	1gof.cm.gz(134k) 1gof.xml.gz(199k)	display	download		
	XML	no-atom	1gof-noatom.xml.gz(28k)	display	download		
		ext-atom	1gof-extatom.xml.gz(118k)	display	download		
	Struct Facto	r	r1gofsf.ent.gz(525k)	display	download		
Back (xPSS-lop) PL	<u>ні тор</u> і						

000	http://pdbjs3.protein.osaka-u.ac.jp/xPSSS/DetailServlet3?PDBID=1GOF&ARCHIVE=LegacyAll
HEADER	OXIDOREDUCTASE(OXYGEN(A)) 30-SEP-93 1GOF
TITLE	NOVEL THIOETHER BOND REVEALED BY A 1.7 ANGSTROMS CRYSTAL
TITLE	2 STRUCTURE OF GALACTOSE OXIDASE
COMPND	MOL ID: 1;
COMPND	2 MOLECULE: GALACTOSE OXIDASE;
COMPND	3 CHAIN: A:
COMPND	4 EC: 1.1.3.9:
COMPND	5 ENGINEERE, VEC
CONTRO	NOT TO. 1.
COURCE	ADD_10. 1, 2 OBCANICH SCIENMIETC, DACMVIIIM DENDROIDES
VEVHDO	2 OKGANISH_SCIENTIFIC: DECILION DENDROIDES
REIWDS	
DIFUT	A-RAI DIFFRACTION
AUTHOR	N.ITO, S.E.V.PHILLIPS, P.F.KNOWLES
REVDAT	2 01-AFR-03 IGOF 1 JRNL
REVDAT	1 31-JAN-94 IGOF U
JENL	AUTH N.ITO,S.E.PHILLIPS,C.STEVENS,Z.B.OGEL,
JRNL	AUTH 2 M.J.MCPHERSON, J.N.KEEN, K.D.YADAV, P.F.KNOWLES
JRNL	TITL NOVEL THIOETHER BOND REVEALED BY A 1.7 A CRYSTAL
JRNL	TITL 2 STRUCTURE OF GALACTOSE OXIDASE.
JRNL	REF NATURE V. 350 87 1991
JRNL	REFN ASTM NATUAS UK ISSN 0028-0836
REMARK	1
REMARK	1 REFERENCE 1
REMARK	1 AUTH N.ITO,S.E.V.PHILLIPS,K.K.S.YADAV,P.F.KNOWLES
REMARK	1 TITL THE CRYSTAL STRUCTURE OF A FREE RADICAL ENZYME,
REMARK	1 TITL 2 GALACTOSE OXIDASE
REMARK	1 REF TO BE PUBLISHED
REMARK	1 REFN
REMARK	1 REFERENCE 2
REMARK	1 AUTH M.J.MCPHERSON, Z.B.OGEL, C.STEVENS, K.D.S.YADAV,
REMARK	1 AUTH 2 J.M.KEEN, P.F.KNOWLES
REMARK	1 TITL GALACTOSE OXIDASE OF DACTYLIUM DENDROIDES: GENE
REMARK	1 TITL 2 CLONING AND SEQUENCE ANALYSIS
REMARK	1 REF J.BIOL.CHEM. V. 267 8146 1992
REMARK	1 REFN ASTM JBCHA3 US ISSN 0021-9258
REMARK	2
REMARK	2 RESOLUTION. 1.70 ANGSTROMS.
REMARK	3
REMARK	3 REFINEMENT.
REMARK	3 PROGRAM : PROLSQ
REMARK	3 AUTHORS : KONNERT, HENDRICKSON
REMARK	3
REMARK	3 DATA USED IN REFINEMENT.
REMARK	3 RESOLUTION RANGE HIGH (ANGSTROMS) : 1.70
REMARK	3 RESOLUTION RANGE LOW (ANGSTROMS) : 10.00
REMARK	3 DATA CUTOFF (SIGMA(F)) : 0.000
REMARK	3 COMPLETENESS FOR RANGE (%) : NULL
REMARK	3 NUMBER OF REFLECTIONS : NULL
REMARK	3
REMARK	3 FIT TO DATA USED IN REFINEMENT.
REMARK	3 CROSS-VALIDATION METHOD : NULL
REMARK	3 FREE R VALUE TEST SET SELECTION : NULL
REMARK	3 R VALUE (WORKING + TEST SET) : 0.177
REMARK	3 R VALUE (WORKING SET) : NULL
REMARK	3 FREE R VALUE : NULL
REMARK	3 FREE R VALUE TEST SET SIZE (%) : NULL
REMARK	3 FREE R VALUE TEST SET COUNT : NULL
REMARK	3
REMARK	3 FIT/AGREEMENT OF MODEL WITH ALL DATA.
REMARK	3 R VALUE (WORKING + TEST SET, NO CUTOFF) : NULL
REMARK	3 R VALUE (WORKING SET, NO CUTOFF) : NULL
REMARK	3 FREE R VALUE (NO CUTOFF) : NULL
REMARK	3 FREE R VALUE TEST SET SIZE (%, NO CUTOFF) : NULL
REMARK	3 FREE R VALUE TEST SET COUNT (NO CUTOFF) : NULL
REMARK	3 TOTAL NUMBER OF REFLECTIONS (NO CUTOFF) : NULL
DEMADE	3

PDB format」のall, displayのボタン をクリックする.

PDB format の中身が見える.

4 4

## PDB formatの良いところ・悪いところ

良いところ:

- 人にとって読みやすい.
- 長年の蓄積がある.(慣れ)
- ファイルサイズが(比較的)小さい.

悪いところ:

- コラム数が決まっている. (1行80文字まで)
- 例外が多い。
- 最近の超巨大分子に対応できない.

### PDBML ファイルを眺める



### 「なま」のPDBMLファイル

"download" ボタンを押してダウンロードしたあと、ファイルを開いてみると、

```
<?xml version="1.0" encoding="UTF-8" ?>
<PDBx:datablock datablockName="1GOF-noatom"
           xmlns:PDBx="http://deposit.pdb.org/pdbML/pdbx.xsd"
           xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
           xsi:schemaLocation="http://deposit.pdb.org/pdbML/pdbx.xsd pdbx.xsd">
  <PDBx:atom_sites_footnoteCategory>
      <PDBx:atom_sites_footnote id="1">
         <PDBx:text>CIS PROLINE - PRO
                                           52</PDBx:text>
      </PDBx:atom_sites_footnote>
      <PDBx:atom_sites_footnote id="2">
         <PDBx:text>CIS PROLINE - PRO
                                          163</PDBx:text>
      </PDBx:atom_sites_footnote>
      <PDBx:atom_sites_footnote id="3">
         <PDBx:text>CIS PROLINE - PRO
                                          350</PDBx:text>
      </PDBx:atom_sites_footnote>
  </PDBx:atom_sites_footnoteCategory>
  <PDBx:audit_authorCategory>
      <PDBx:audit_author name="Ito, N."></PDBx:audit_author>
```

<PDBx:audit\_author name="Phillips, S.E.V."></PDBx:audit\_author> <PDBx:audit\_author name="Knowles, P.F."></PDBx:audit\_author> </PDBx:audit\_authorCategory> <PDBx:cellCategory> <PDBx:cell entry\_id="1GOF"> <PDBx:length\_a>98.000</PDBx:length\_a> <PDBx:length\_b>89.400</PDBx:length\_b> <PDBx:length\_c>86.700</PDBx:length\_c> <PDBx:angle\_alpha>90.00</PDBx:angle\_alpha> <PDBx:angle\_beta>117.80</PDBx:angle\_beta> <PDBx:angle\_gamma>90.00</PDBx:angle\_gamma> <PDBx:Z\_PDB>4</PDBx:Z\_PDB> </PDBx:cell> </PDBx:cellCategory> <PDBx:database\_PDB\_matrixCategory> <PDBx:database\_PDB\_matrix entry\_id="1GOF"> <PDBx:origx11>1.000000</PDBx:origx11>

. . . . . . . . .

と、XMLのタグだらけの形式になっている.

## 原子座標: PDB format

ATOM 2 CA ALA 1 38.356 -0.999 0.357 1.00 42.26 1GOF 220



### 原子座標: PDBML

<PDBx:atom\_site id="2"> <PDBx:group\_PDB>ATOM</PDBx:group\_PDB> <PDBx:type\_symbol>C</PDBx:type\_symbol> <PDBx:label\_atom\_id>CA</PDBx:label\_atom\_id> <PDBx:label\_alt\_id xsi:nil="true" /> <PDBx:label\_comp\_id>ALA</PDBx:label\_comp\_id> <PDBx:label\_asym\_id>A</PDBx:label\_asym\_id> <PDBx:label\_entity\_id>1</PDBx:label\_entity\_id> <PDBx:label\_seq\_id>1</PDBx:label\_seq\_id> <PDBx:Cartn\_x>38.356</PDBx:Cartn\_x> <PDBx:Cartn\_y>-0.999</PDBx:Cartn\_y> <PDBx:Cartn\_z>0.357</PDBx:Cartn\_z> <PDBx:occupancy>1.00</PDBx:occupancy> <PDBx:B\_iso\_or\_equiv>42.26</PDBx:B\_iso\_or\_equiv> <PDBx:auth\_seq\_id>1</PDBx:auth\_seq\_id> <PDBx:auth\_comp\_id>ALA</PDBx:auth\_comp\_id> <PDBx:auth\_asym\_id>A</PDBx:auth\_asym\_id> <PDBx:auth\_atom\_id>CA</PDBx:auth\_atom\_id> <PDBx:pdbx\_PDB\_model\_num>1</PDBx:pdbx\_PDB\_model\_num> </PDBx:atom\_site>

## 原子座標: PDBMLの簡易版(extatom)

<PDBx:atom\_record id="2"> ATOM 1 A A 1 1 ? . ALA ALA C CA CA 38.356 -0.999 0.357 1.00 42.26 </PDBx:atom\_record>



### PDBMLの良いところ・悪いところ

参考: Westbrook et al. Bioinformatics (2005) 21(7):988-992.

良いところ:

- 形式が厳密に定められている。
   XML Schema: http://pdbml.pdb.org/schema/pdbx.xsd
- 拡張性がある.
- プログラムで扱いやすい.(処理の自動化)

悪いところ:

- 煩雑. 人が直接読むのには適していない.
- ファイルサイズが大きくなる。
- スキーマ(タグの意味)の学習が(ある程度、かなり?)必要.

### XMLの基礎知識



XML 文書は木構造 (入れ子構造) をしている. 要素 (element) と属性 (attribute) からなる. 可能な木構造は XML Schema などで定義されている.

### XPath と XQuery

XMLの構造を利用すれば、非常に詳細な検索が可能になる。PDBj の XMLデータベースを 検索する方法は2通りある.

XPath XML文書の木構造の指定したノードで、ある条件を満たすものを返す。返せる値は一つだけ.

XQuery XPathと同様だが、より複雑な条件や複数の戻り値を指定できる.



## pdbMLplusの構造

XPath/XQueryを使うためにはpdbMLplusの構造(スキーマ)をある程度知っている必要がある. 参考:http://pdbjs3.protein.osaka-u.ac.jp/xPSSS/schema.html

datablock	すべてのエントリはdatablockで囲まれる.
<pre>@datablockName</pre>	datablockの属性= PDB ID
citation_authorCategory	論文の著者のカテゴリ
citation_author	
Qname	論文の著者名
citationCategory	論文のカテゴリ
citation	論文の情報
journal_abbrev	雑誌の略称(J. Biol. Chem.など)
journal_volume	雑誌の Volume番号
page_first	論文の最初のページ番号

タンパク質が採れた生物
生物の通称(Human)
学名(Homo sapiens)
実験手法
実験手法(NMR, X-ray, etc.)

#### PDB5 Protein Data Bank Japan

. . . . .

### xml-based Protein Structure Search Service (xPSSS)





xPSSSを開く.

ページの下部に XQuery と XPath のフ ォームがあります.

## XPath 検索の基本

# /datablock[検索条件]/出力アイテム

検索条件はなくても良い.



## XPathの例(1)

すべての PDB ID を列挙する.

datablock @datablockName

 $\Rightarrow$ 

/datablock/@datablockName



## XPathの例(2)

著者名を列挙する.

datablock
 citation\_authorCategory
 citation\_author
 @name

 $\Rightarrow$ 

/datablock/citation\_authorCategory/citation\_author/@name

すべてのエントリのすべての文献のすべての著者名が列挙される.

## XPathの例(3)

• あるエントリ(1gof)の著者名をすべて列挙する. ⇒

/datablock[@datablockName='1GOF-noatom']
 /citation\_authorCategory/citation\_author/@name

• 著者名が" Ito, N." を含むエントリを列挙する. ⇒

```
/datablock[
  citation_authorCategory/citation_author/@name='Ito, N.']
  /@datablockName
```



## XPathの例(4)

• 分子量が100000以上の分子を含むエントリを列挙する.

/datablock[entityCategory/entity/formula\_weight >= 100000]/@datablockName

• 30残基以上のαヘリックスを含むエントリを列挙する.

/datablock[struct\_confCategory/struct\_conf/pdbx\_PDB\_helix\_length >= 30]
/@datablockName

 分子量が100000以上の分子を含み、かつ30残基以上のαヘリックスを含むエントリを列挙 する。

/datablock[ entityCategory/entity/formula\_weight >= 100000
and struct\_confCategory/struct\_conf/pdbx\_PDB\_helix\_length >= 30]
/@datablockName

## XQuery検索の基本

"for ... where ... return" が基本形.

for \$b in input()XPath where \$b の満たすべき条件 return <datablock>

{返すアイテム1} {返すアイテム2} {返すアイテム3}

</datablock>

where節はなくても良い.

### XQuad: XQuery Advisor for PDBML

- 1. XQueryは一種のプログラミング言語なので、マスターするのは大変です。
- → 簡便に XQuery プログラムを組み立てたい
- $\rightarrow$  XQuad



## XQuadその1

ulope()	all ABY Maters blas Zerfil		Bulling BORINS	and Consil
P J ein Data	Anticipan and a subject of the second	otein Structure Se	earch Service) <sup>8</sup>	wPDB inapshot Mirror
Qu	ick Search			
	Enter a PDB ID :	eerch Display size per page 16	sorted by PDB ID	;
	Enter a keyword :	Display size per page 16	sorted by PDB ID	:
Ad	click here			
	What is the structure of pdbML2 <u>An explan</u> What do categories and items mean? <u>The p</u> About the structure of pdbMLplus <u>pdbMLp</u> xPSSS Soap Service <u>Example Page</u> Sample programs to handle pdbML <u>Downl</u> XQuery	iation is here! nmCIF dictionary is here! ilus schema file load Page		
	How can you make an XQuery?Samples a	are here! Ne ! PDBML X	Query Advisor (XQuad)	
	Send Query Reset			
	Send Query Peset XPath How can you make an XPath?Samples an	beret	splay size per page 16	
	3ani Query Peset XPath How can you make an XPath?Samples an	Di e here]	splay size per page 16	

				XQuad (PD8N	IL XQuen	Advisor	r)						
<ul> <li>C + Ohttp:/</li> </ul>	/pdbjs7.protein.os	aka-u.ac.jp/xqu	ad/							<b>O</b> 1	Q- maria	k pages	0
[]] Globe(ml) ARK Matra	blog アップル	Amazon.co.jp	eProts :	==⊼ (3274) ¥	GTOP I	PubMed	PDBj Members' pa	ge Gmail	Scopus	Lang ¥	アップル	Google マップ	х
PDBB Batt Japan	PDBM	IL XQue	ry Ad	lvisor (X	Qua	d)	Operation E: About XQua	cample d	1550				
Search Category:	:	)				Sear	rch History						
Search Keyword:	SEARCH												
Result Display Max:	30 \$								-				

				Display size per pa	ge 16	SUENIT	RESET	XQuen
no	XPath	type	exp	value	and/or	match	return	clear
1		String :			and 🗘	8	8	clear
2		String :	- :		and 0	8	8	clear
3		String 0	- :		and 0	8	8	clear
4		String 0	- •		and 🗘	8	8	clear
5		String :			and	8	8	(clear
6		String :	- :		and 0	8		clear
7		String 2	- •		and 0	8	8	clear
8		String :	- •		and	8	8	clear
9		String :	- :		and	8	8	dear
10		String 1	- •		and 0	8	8	clear
11		String :	- •		and 0	8	8	clear
12		String :	- •		and	8	8	(dear
13		String :			and 0	8	8	clear
14		(Textus	— — — — — — — — — — — — — — — — — — —		1			6000

xPSSS のページから XQuad のリンク を選ぶ。

XQuad の初期画面。

## XQueryその2:使用例

#### 1) Category / Keyword



## その他



## 蛋白質構造百科事典 (eProtS)

初心者向けの解説. http://eprots.pdbj.org/jp.cgi





## Protein Globe: 蛋白質空間に基づいた PDBj へのGUI



球面上の各点がタンパク質の構造ドメインを表す。

### jV3をインストールする

PDBjで立体構造の対話的にグラフィクスを表示するには, jV3 (PDBjViewer version 3)が便利です.

- 1. Java 環境 (JRE 1.4.2以降)が必要です.
- 2. www.pdbj.org の 「jV version3」をクリック.
- そのページ (www.pdbj.org/PDBjViewer/index.html)の下の方にある JOGL JSR-231 1.0.0 をダウンロードしてインストールする. 以上で Web ブラウザ上で jV3が使えるようになる (はず). デスクトップで利用する場合には、以下も必要.
- 「jVのダウンロード」から binary ファイルをダウンロードする (ライセンスに同意する必要あり). 興味のある方はソースコードもどうぞ. 適当なディレクトリに jV をインストールする.
- 5. マニュアル類もダウンロードしておきましょう.