

PDBのデータとその見方・探し方

ライフサイエンス・データベース講習会 in 名大

金城玲
大阪大学蛋白質研究所
2010年8月10日

この講義の内容

- PDBjを使った基本的な検索法
- PDBデータの読み方
 - PDB format
 - mmCIF format
 - PDBML format
- jVを使った立体構造の眺め方(時間があれば)

基本的な検索法

- エントリーの取得
- キーワード検索
- Advanced search (細かな条件を指定した検索)

<http://www.pdbj.org/>

を開いてください。

PDBjのウェブインターフェイス



- データの登録(今回は扱いません)。
- データの検索(エントリー／キーワード)
- その他のサービス(ページ左側)。
- ページの右上に現在のエントリー数が表示されている。(クリックすると「今週追加／更新されたエントリー」のリストが表示される。

エントリーの取得



- 「PDB ID」がチェックされていることを確認する。
- 箱に「1gof」と入力する（半角英数字）。
- 「Go」ボタンをクリック。
- 「1gof」はあるエントリーのPDB ID。

エントリーのサマリー(要約)ページ

- エントリー「1gof」のサマリーページが表示されます。
- サマリーとはこの蛋白質に関する記述、引用文献、実験手法、他のデータベースの該当するエントリーへのリンク、などをまとめたものの。
- 図もついています。

xPSSS (xml-based Protein Structure Search Service) Summary Page

http://pdjbs8.pdbj.org/xPSSS/DetailServlet?PDBID=1gof&PAGEID=Summary

PDBj (Protein Data Bank Japan) maintains a centralized archive of macromolecular structures and provides integrated tools, in collaboration with the RCSB in USA and the PDB in EU. PDBj is supported by JST-BIRD.

Summary [1gof]

Summary | Structural Details | Experimental Details | Functional Details | Sequence Neighbor | Download/Display | Link

Structure Viewers (JV3 / Jmol) JV and Jmol require Java(TM) Plug-in 5 or later.

rotated about x by 90° 250X250 500X500

rotated about y by 90° 250X250 500X500

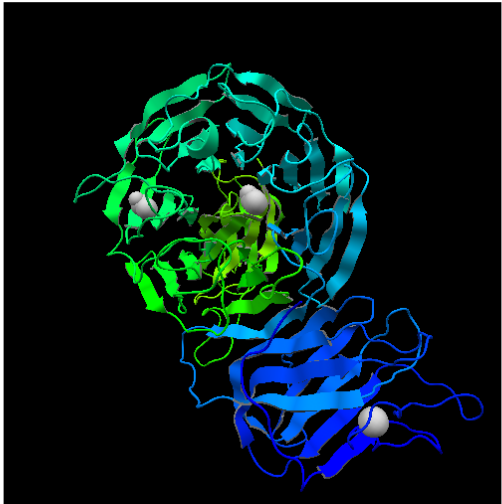
Electron density map is available to be displayed from the Experimental Details page.

PDB ID	1gof sequence information (FASTA format) download PDB format file
Descriptor	GALACTOSE OXIDASE (E.C.1.1.3.9) (PH 4.5)
Title	NOVEL THIOETHER BOND REVEALED BY A 1.7 ANGSTROMS CRYSTAL STRUCTURE OF GALACTOSE OXIDASE
Functional Keywords	OXIDOREDUCTASE(OXYGEN(A))
Biological source	Hypomyces rosellus
Cellular location	[UNP - GAOA_DACDE] Secreted
Total number of polymer chains	1
Total molecular weight	68785.9 (the details in Structural Details Page)
Authors	Ito, N., Phillips, S.E.V., Knowles, P.F. (deposition date: 1993-09-30, release date: 1994-01-31)
Primary citation	Ito, N., Phillips, S.E., Stevens, C., Ogel, Z.B., McPherson, M.J., Keen, J.N., Yadav, K.D., Knowles, P.F. Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. <i>Nature</i> , 350:87 - 90, 1991. (PubMed: 2002850) (DOI: 10.1038/350087a0)
Experimental method	X-RAY DIFFRACTION (1.7[Å])
Other Database Information	CATH, CE, FSSP, SCOP, VAST, UniProt (UNP-Q01745), eF-site, KEGG (EC 1.1.3.9), EzCatDB

© 2007 PDBj. All Rights Reserved.

分子グラフィクスを眺めてみる

xPSSS (xml-based Protein Structure Search Service) jV3 (Applet-Launcher)



jV3

Statistics Help Contact Us

molecular structures and provides integrated tools, by JST-BIRD.

About xPSSS
Update Information
wwPDB Snapshot Mirror

Sequence Neighbor Download/Display Link

search PDB ID keyword

nce information (FASTA format)
format file

SE OVIDASE (5C1120) (P415)

jV version 3 Applet Launcher

Style: ☒ Default ☐ Cartoon ☐ Wireframe
Color: ☒ Default ☐ group ☐ chain

Protein Globe
ASH
MAFFTash
jV: Graphic Viewer

Derived database >>
eF-site/eF-
seek/eF-surf
eProtS
ProMode
Molecule of the
Month

Download >>
PDB
Archive/Snapshot
Archive

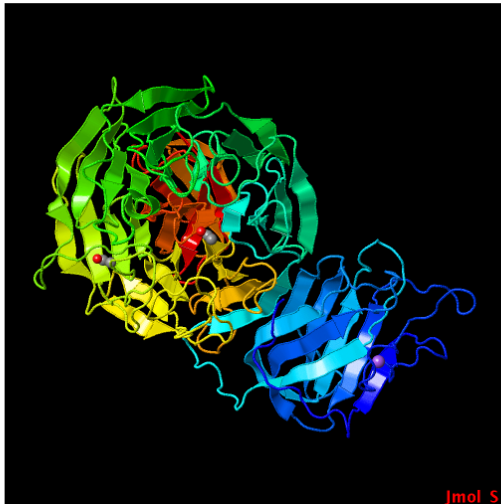
About Remediation
Data
Links

jV and Jmol require
Java(TM) Plug-in 1
later.

rotated about x by
250X250 500X500

rotated about y by
250X250 500X500

Electron density map is



Jmol version 11.3

Style: ☒ Default ☐ Cartoon ☐ Rocket ☐ Wireframe ☐ CPK_without_water ☐ CPK
Color: ☒ Default ☐ Group ☐ Chain ☐ Atom

- サマリーページの一番大きな図の下に「jV3」または「Jmol」というボタンをクリックすると、分子グラフィクスソフトのアプリットが起動します。
- マウスでぐるぐる回してしばし眺めてください。

Jmol

エントリーのその他の情報

Summary Page

English Statistics Help Contact Us

Home

Data Deposition >>

ADIT: PDB Deposition

ADIT-NMR

Search >>

Search PDB (xPSSS)

Latest Released Search

Sequence-Navigator

Structure-Navigator

SeSAW

Ligand Binding Sites (GIRAF)

EM Navigator

Search NMR Data (BMRB)

Status Search

Service and Software >>

Protein Globe

ASH

MAFFTash

JV: Graphic Viewer

Derived database >>

eF-site/eF-seek/eF-surf

eProtS

ProMode

Molecule of the Month

Download >>

PDB Archive/Snapshot Archive

About Remediation Data

Links

PDBj (Protein Data Bank Japan) maintains a centralized archive of macromolecular structures and provides integrated tools, in collaboration with the RCSB in USA and the PDBe in EU. PDBj is supported by JST-BIRD.

xPSSS (xml-based Protein Structure Search Service)

Summary [1gof]

About xPSSS Update information wwPDB Snapshot Mirror

Summary Structural Details Experimental Details Functional Details Sequence Neighbor Download/Display Link

search PDB ID keyword

Structure Viewers

jv3 / Jmol

jv and Jmol require Java(TM) Plug-in 1.5 or later.

rotated about x by 90°

250X250 500X500

rotated about y by 90°

250X250 500X500

Electron density map is available to be displayed from the Experimental Details page.

PDB ID	1gof sequence information (FASTA format)
Descriptor	download PDB format file
Title	GALACTOSE OXIDASE (E.C.1.1.3.9) (PH 4.5)
Functional Keywords	NOVEL THIOETHER BOND REVEALED BY A 1.7 ANGSTROMS CRYSTAL STRUCTURE OF GALACTOSE OXIDASE
Biological source	OXIDOREDUCTASE(OXYGEN(A))
Cellular location	Hypomyces rosellus
Total number of polymer chains	[UNP - GAOA_DACDE] Secreted
Total molecular weight	1
Authors	68785.9 (the details in Structural Details Page)
Primary citation	Ito, N., Phillips, S.E.V., Knowles, P.F. (deposition date : 1993-09-30, release date : 1994-01-31)
Experimental method	Ito, N., Phillips, S.E., Stevens, C., Ogel, Z.B., McPherson, M.J., Keen, J.N., Yadav, K.D., Knowles, P.F.
Other Database Information	Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. <i>Nature</i> , 350:87 - 90, 1991.(PubMed : 2002850) (DOI: 10.1038/350087a0)
	X-RAY DIFFRACTION (1.7Å)
	CATH , CE , FSSP , SCOP , VAST , UniProt (UNP - Q01745), eF-site , KEGG (EC 1.1.3.9), EzCatDB

© 2007 PDBj. All Rights Reserved.

- サマリーページ上部にその他の情報のページへのリンクがあります。
- Structural Details
- Experimental Details
- Functional Details
- Sequence Neighbor
- Download/Display
- Links
- これらを順に説明します。

Structural Details

The screenshot shows the xPSSS (xml-based Protein Structure Search Service) Structural Details page for PDB entry 1gof. The page is titled "xPSSS (xml-based Protein Structure Search Service) Structural Details" and has a URL of "http://pdbjs8.pdbj.org/xPSSS/DetailServlet". The page is in English and includes a navigation bar with links for Home, Statistics, Help, and Contact Us. The main content area is titled "Structural Details [1gof]" and includes a search bar and a table of entities. The table lists the following entities:

Chain ID	Description	Type	DB Name (Accession)	Formula weight	Number of molecules	Biological source	Descriptive key words
A	GALACTOSE OXIDASE sequence information (FASTA format)	polymer	UniProt (Q01745)	68579.2	1	Hypomyces rosellus	
	COPPER (II) ION	non-polymer	PDBChem (CU) Chem.Comp.(CU)	63.5	1		
	SODIUM ION	non-polymer	PDBChem (NA) Chem.Comp.(NA)	23.0	1		
	ACETIC ACID	non-polymer	PDBChem (ACY) Chem.Comp.(ACY)	60.1	2		
	water	water	PDBChem (HOH) Chem.Comp.(HOH)	18.0	316		

Below the table, there is a section titled "Contents of the asymmetric unit" which provides a summary of the contents:

Polymers	Number of Chains	Total molecular weight
	1	68579.2

Non-Polymers*	Number of Molecules	Total molecular weight
	4	206.7

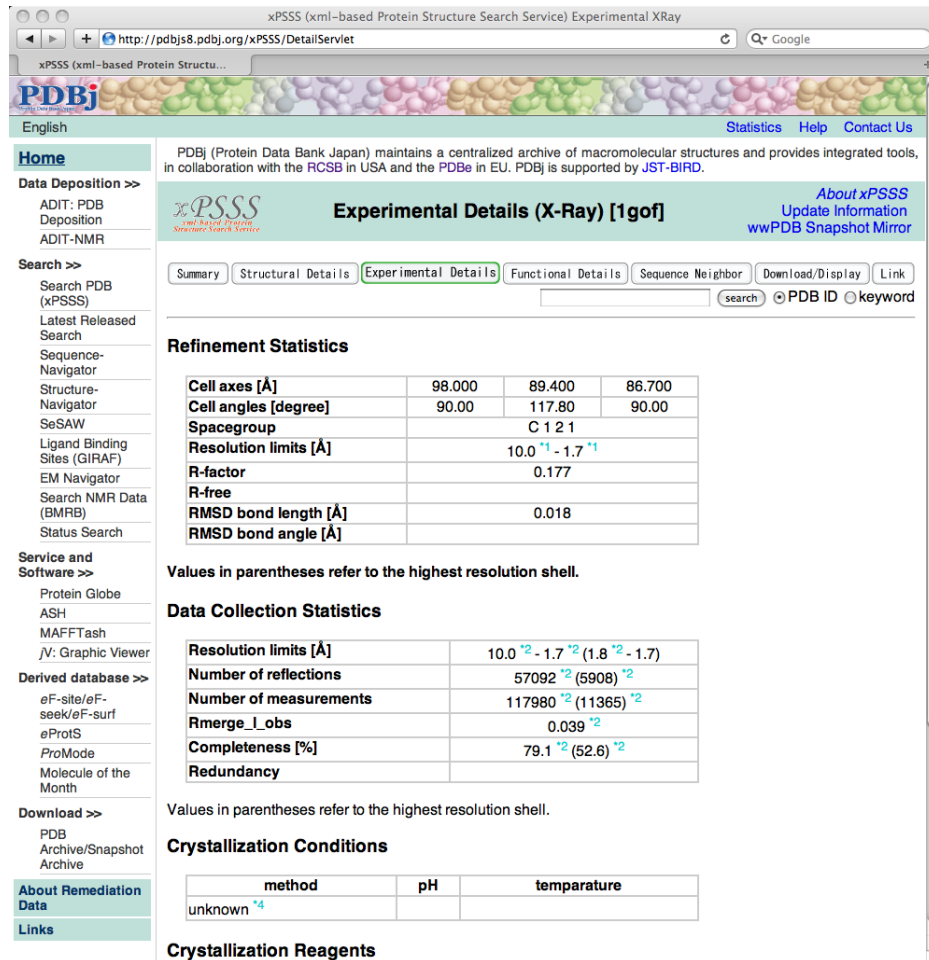
All*	Total molecular weight
	68785.9

*Water molecules are not included.

The page also includes a sidebar with links for Home, Data Deposition, Search, Service and Software, Derived database, Download, About Remediation Data, and Links.

- 解かれている構造の詳細情報です。
- 「どのようなモノがふくまれるか」が主な内容です。
- 1gofの場合はポリペプチド鎖が1つと銅イオン、ナトリウムイオン、酢酸、水、が含まれていることがわかります。

Experimental Details



The screenshot shows the xPSSS (xml-based Protein Structure Search Service) Experimental X-Ray page. The page is titled "Experimental Details (X-Ray) [1gof]" and includes a search bar and navigation links. The left sidebar contains links for Home, Data Deposition, Search, Service and Software, Derived database, Download, About Remediation Data, and Links. The main content area displays the following information:

Refinement Statistics

Cell axes [Å]	98.000	89.400	86.700
Cell angles [degree]	90.00	117.80	90.00
Spacegroup	C 1 2 1		
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]			

Values in parentheses refer to the highest resolution shell.

Data Collection Statistics

Resolution limits [Å]	10.0 ^{*2} - 1.7 ^{*2} (1.8 ^{*2} - 1.7)
Number of reflections	57092 ^{*2} (5908) ^{*2}
Number of measurements	117980 ^{*2} (11365) ^{*2}
Rmerge I_obs	0.039 ^{*2}
Completeness [%]	79.1 ^{*2} (52.6) ^{*2}
Redundancy	

Values in parentheses refer to the highest resolution shell.

Crystallization Conditions

method	pH	temperature
unknown ^{*4}		

Crystallization Reagents

- 構造が解かれた実験の条件に関する詳細です。
- 結晶構造の場合は結晶の各種パラメータや結晶化条件、溶液 NMR構造の場合は分光器やスペクトルの種類などが記載されています。

Functional Details

The screenshot shows the xPSSS (xml-based Protein Structure Search Service) Functional Details page. The page is titled 'Functional Details [1a24]' and includes a search bar with 'search' and 'PDB ID' options. The left sidebar contains navigation links such as 'Home', 'Data Deposition', 'Search', 'Service and Software', 'Derived database', and 'Download'. The main content area displays 'Functional Information from GO Data' and 'Functional Information from PROSITE/UniProt'.

Chain	GOid	namespace	contents
A	0015035	molecular_function	protein disulfide oxidoreductase activity
A	0030288	cellular_component	outer membrane-bounded periplasmic space
A	0045454	biological_process	cell redox homeostasis

(More GO information ...)

site_id	type	Number of Residues	Details
PS00194	prosite	19	Thioredoxin family active site. [LIVMF]-[LIVMSTA]-x-[LIVMFYC]-[FYWSTHE]-x(2)-[FYWGTN]-C-[GATPLVE]

(More PROSITE information ...)

- 蛋白質の機能に関する情報が記載されています。
- 多くは、PDBjが独自に収集したデータで元のPDBのデータにはない情報も含まれます。
 - Gene ontology
 - Functional site (残基)
 - リガンド結合部位
 - などなど

Sequence Neighbor

- PDBに対する配列類似性検索 (BLAST) の結果を表示します。
- 注意: 複数のポリペプチド鎖が含まれるエントリーの場合は最初のポリペプチド鎖に対して検索されます。

xPSSS (xml-based Protein Structure Search Service) Result

http://pdbjs3.protein.osaka-u.ac.jp/seqnavix/SeqHomXServlet

English Statistics Help Contact Us

Home

PDBj (Protein Data Bank Japan) maintains a centralized archive of macromolecular structures and provides integrated tools, in collaboration with the RCSB in USA and the PDBe in EU. PDBj is supported by JST-BIRD.

xPSSS (xml-based Protein Structure Search Service)

Summary Structural Details Experimental Details Functional Details Sequence Neighbor Download/Display Link

search PDB ID keyword

Results (1-36) / 36

1GOFA Exact Matches: 2JKXA 2VZ1A 2E1EA 1GOHA 1GOFA 1GOGA

Seq. Identity: 99% Seq. Positives: 99% E-value: 0.0 Score: 1320 Compound: GALACTOSE OXIDASE

New Search [1T2XA] Structural Superposition

2VZ3A 1 639 ASAPIGSAISRNNNAVTCDSAQSGNECNKAIDGNKDTFWETFGANGDPKPPPTTTIDMKTTQNVNGLSMLPRQDGNQNGWIGREEVYLLSSDGTNNGSPVASG

1T2XA 1 639 ASAPIGSAISRNNNAVTCDSAQSGNECNKAIDGNKDTFWETFGANGDPKPPPTTTIDMKTTQNVNGLSMLPRQDGNQNGWIGREEVYLLSSDGTNNGSPVASG

Seq. Identity: 99% Seq. Positives: 100% E-value: 0.0 Score: 1320 Compound: GALACTOSE OXIDASE

New Search [2E1EA] Structural Superposition

2VZ3A 1 639 ASAPIGSAISRNNNAVTCDSAQSGNECNKAIDGNKDTFWETFGANGDPKPPPTTTIDMKTTQNVNGLSMLPRQDGNQNGWIGREEVYLLSSDGTNNGSPVASG

2E1EA 1 639 ASAPIGSAISRNNNAVTCDSAQSGNECNKAIDGNKDTFWETFGANGDPKPPPTTTIDMKTTQNVNGLSMLPRQDGNQNGWIGREEVYLLSSDGTNNGSPVASG

Seq. Identity: 99% Seq. Positives: 99% E-value: 0.0 Score: 1318 Compound: GALACTOSE OXIDASE

New Search [2E1EA] Structural Superposition

2VZ3A 1 639 ASAPIGSAISRNNNAVTCDSAQSGNECNKAIDGNKDTFWETFGANGDPKPPPTTTIDMKTTQNVNGLSMLPRQDGNQNGWIGREEVYLLSSDGTNNGSPVASG

2E1EA 1 639 ASAPIGSAISRNNNAVTCDSAQSGNECNKAIDGNKDTFWETFGANGDPKPPPTTTIDMKTTQNVNGLSMLPRQDGNQNGWIGREEVYLLSSDGTNNGSPVASG

Seq. Identity: 99% Seq. Positives: 99% E-value: 0.0 Score: 1318 Compound: GALACTOSE OXIDASE

New Search [2E1EA] Structural Superposition

2VZ3A 1 639 ASAPIGSAISRNNNAVTCDSAQSGNECNKAIDGNKDTFWETFGANGDPKPPPTTTIDMKTTQNVNGLSMLPRQDGNQNGWIGREEVYLLSSDGTNNGSPVASG

2E1EA 1 639 ASAPIGSAISRNNNAVTCDSAQSGNECNKAIDGNKDTFWETFGANGDPKPPPTTTIDMKTTQNVNGLSMLPRQDGNQNGWIGREEVYLLSSDGTNNGSPVASG

Seq. Identity: 99% Seq. Positives: 99% E-value: 0.0 Score: 1315 Compound: GALACTOSE OXIDASE PRECURSOR

New Search [1K3IA] Structural Superposition

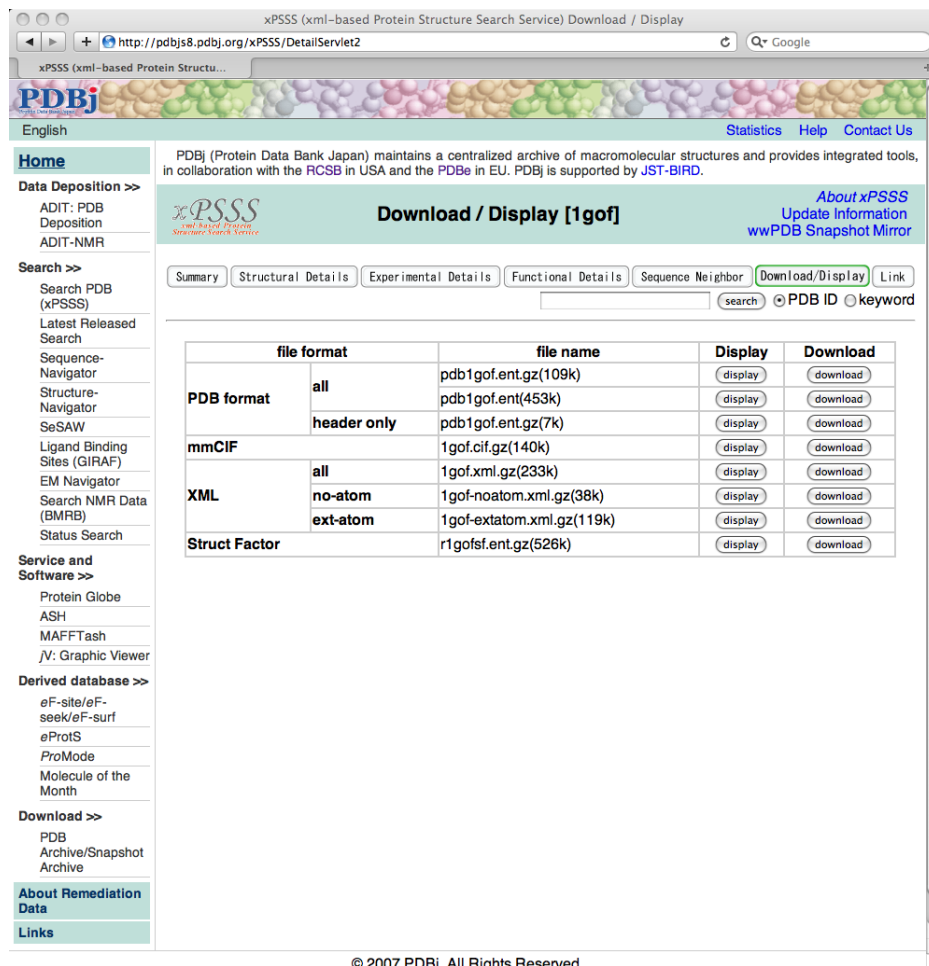
2VZ3A 1 639 ASAPIGSAISRNNNAVTCDSAQSGNECNKAIDGNKDTFWETFGANGDPKPPPTTTIDMKTTQNVNGLSMLPRQDGNQNGWIGREEVYLLSSDGTNNGSPVASG

1K3IA 1 639 ASAPIGSAISRNNNAVTCDSAQSGNECNKAIDGNKDTFWETFGANGDPKPPPTTTIDMKTTQNVNGLSMLPRQDGNQNGWIGREEVYLLSSDGTNNGSPVASG

Seq. Identity: 35% Seq. Positives: 55% E-value: 3e-19 Score: 94 Compound: SIALIDASE

New Search [1WCQA] Structural Superposition

Download/Display



The screenshot shows the xPSSS (xml-based Protein Structure Search Service) Download / Display page. The page has a navigation bar with links for Home, Data Deposition, Search, Service and Software, and Derived database. The main content area is titled "Download / Display [1gof]" and contains a table of available data formats and files.

file format	file name	Display	Download
PDB format	all	display	download
	header only	display	download
mmCIF	all	display	download
	no-atom	display	download
XML	all	display	download
	ext-atom	display	download
Struct Factor	r1gofsf.ent.gz(526k)	display	download

- 元データを表示またはダウンロードできます。
- 「生」のPDB format, mmCIF, PDBMLフォーマット
の他に、原子座標を除いたバージョンも用意されています。
 - アノテーションのみがみたいときに便利
- 結晶構造の場合は構造因子も用意されていることもあります。
- 後ほどじっくり眺めます。

Link



- 他のデータベースの該当するエントリーへのリンク集です。

キーワード検索

- トップページに戻って、キーワード検索を試みましょう。
- 「Keywords」ボタンをチェックする。
- 箱に適当なキーワードを入れる。
- 「Go」ボタンをクリック。
- 例:「awamori」



キーワード検索の結果

xPSSS (xml-based Protein Structure Search Service) Result Page

http://pdbs8.pdbj.org/xPSSS/xPSSSSearch?search_type=keyword_search&query=awamori&xpss_search_g...

English Statistics Help Contact Us

Home

Data Deposition >>

ADIT: PDB Deposition

ADIT-NMR

Search >>

Search PDB (xPSSS)

Latest Released Search

Sequence-Navigator

Structure-Navigator

SeSAW

Ligand Binding Sites (GIRAF)

EM Navigator

Search NMR Data (BMRB)

Status Search

Service and Software >>

Protein Globe

ASH

MAFFTash

JV: Graphic Viewer

Derived database >>

eF-site/eF-seek/eF-surf

eProtS

ProMode

Molecule of the Month

Download >>

PDB Archive/Snapshot Archive

About Remediation Data

Links

PDBj (Protein Data Bank Japan) maintains a centralized archive of macromolecular structures and provides integrated tools, in collaboration with the RCSB in USA and the PDB in EU. PDBj is supported by JST-BIRD.

xPSSS

Result Page

About xPSSS Update Information wwPDB Snapshot Mirror

(More information appears when each PDB-ID code is clicked.)

1 - 16/ 18

1 2

keyword : awamori search sorted by: PDB ID

next

1agm	descriptor	title	authors	exp.method	deposition date	release date
	: GLUCOAMYLASE-471 (GLUCAN 1,4-ALPHA-GLUCOSIDASE, RESIDUES 1 - 471) (E.C.3.2.1.3) COMPLEXED WITH ACARBOSE	: Refined structure for the complex of acarbose with glucoamylase from Aspergillus awamori var. x100 to 2.4 angstroms resolution	: Aleshin, A.E., Firsov, L.M., Honzatko, R.B.,	: X-RAY DIFFRACTION	: 1994-05-13	: 1994-09-30
download						
1b0l	descriptor	title	authors	exp.method	deposition date	release date
	: LACTOFERRIN	: RECOMBINANT HUMAN DIFERRIC LACTOFERRIN	: Baker, E.N., Jameson, G.B., Sun, X.,	: X-RAY DIFFRACTION	: 1998-11-11	: 1999-11-18
download						
1bk1	descriptor	title	authors	exp.method	deposition date	release date
	: ENDO-1,4-B-XYLANASE C	: ENDO-1,4-BETA-XYLANASE C	: Fushinobu, S., Ito, K., Konno, M., Wakagi, T., Matsuzawa, H.,	: X-RAY DIFFRACTION	: 1998-07-14	: 1999-01-13
download						
1dog	descriptor	title	authors	exp.method	deposition date	release date
	: GLUCOAMYLASE-471 (GLUCAN 1,4-ALPHA-GLUCOSIDASE) (E.C.3.2.1.3) COMPLEXED WITH 1-DEOXYNOJIRIMYCIN	: REFINED STRUCTURE FOR THE COMPLEX OF 1-DEOXYNOJIRIMYCIN WITH GLUCOAMYLASE FROM (ASPERGILLUS AWAMORI) VAR. X100 TO 2.4 ANGSTROMS RESOLUTION				

Electron density map is available to be displayed from the Experimental Details page.

- 検索結果はPDB Idの辞書順で表示されます。
- ただし、「sorted by」のメニューから別のオプションを選んでsearchボタンを再び押せば並べ替えができます。
- 複数のキーワードも指定可能。
- (注意: Googleのようなものは期待しないこと。)

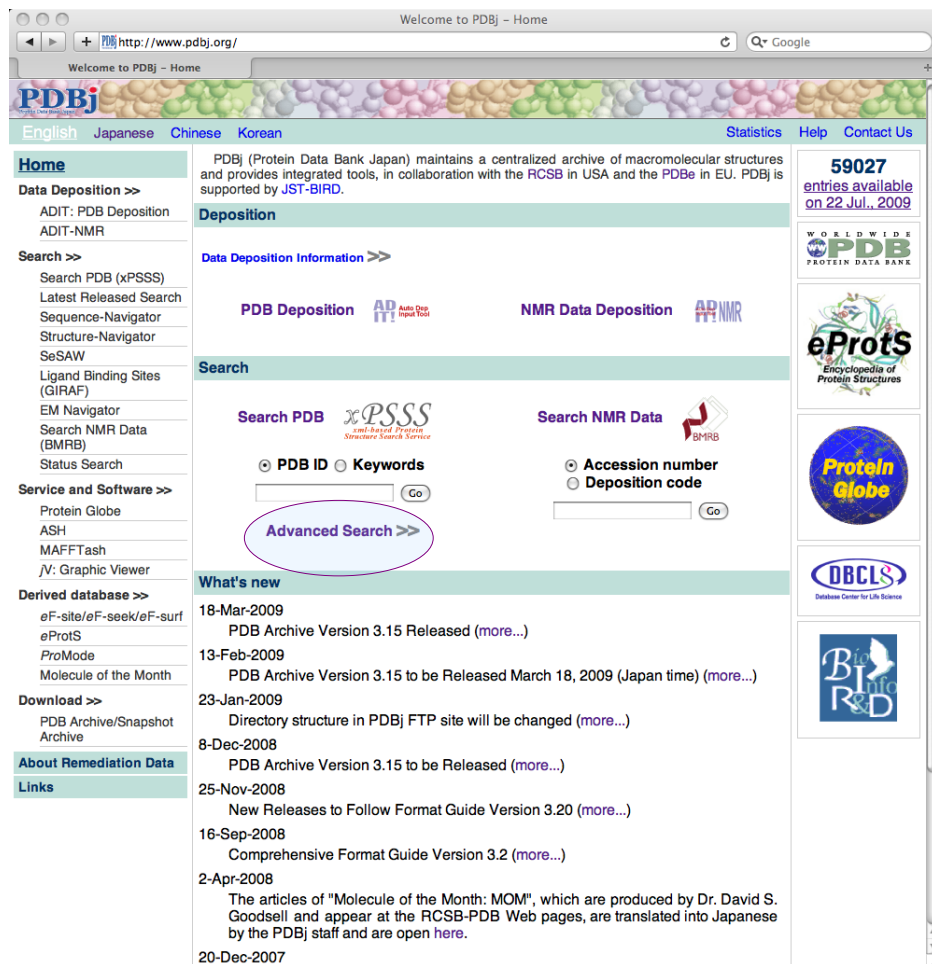
キーワード検索の結果を見る

Search results for keyword: awamori

PDB ID	Descriptor	Title	Authors	exp.method	deposition date	release date
1b0l	LACTOFERRIN	RECOMBINANT HUMAN DIFERRIC LACTOFERRIN	Baker, E.N., Jameson, G.B., Sun, X.,	X-RAY DIFFRACTION	1998-11-11	1999-11-18
1bk1	ENDO-1,4-B-XYLANASE C	ENDO-1,4-BETA-XYLANASE C	Fushinobu, S., Ito, K., Konno, M., Wakagi, T., Matsuzawa, H.,	X-RAY DIFFRACTION	1998-07-14	1999-01-13
1dog	GLUCOAMYLASE-471 (GLUCAN 1,4-ALPHA-GLUCOSIDASE) (E.C.3.2.1.3) COMPLEXED WITH 1-DEOXYNOJIRIMYCIN	REFINED STRUCTURE FOR THE COMPLEX OF 1-DEOXYNOJIRIMYCIN WITH GLUCOAMYLASE FROM (ASPERGILLUS AWAMORI) VAR. X100 TO 2.4 ANGSTROMS RESOLUTION				

- 検索結果リストのPDB ID または図をクリックすると該当するエントリーのサマリーページがポップアップします。

Advanced Search



- 再びトップページに戻り、「Advanced Search」というリンクをクリック。
- より詳細な条件を指定した検索ができます。

Advanced Searchを使う

Welcome to PDBj - xPSSS Advanced Search

http://pdbjs8.pdbj.org/xpss/xpss_advanced.html

Welcome to PDBj - xPSSS Advance...

English Statistics Help Contact Us

Home

Data Deposition >>

ADIT: PDB Deposition

ADIT-NMR

Search >>

Search PDB (xPSSS)

Latest Released Search

Sequence-Navigator

Structure-Navigator

SeSAW

Ligand Binding Sites (GIRAF)

EM Navigator

Search NMR Data (BMRB)

Status Search

Service and Software >>

Protein Globe

ASH

MAFFTash

JV: Graphic Viewer

Derived database >>

eF-site/eF-seek/eF-surf

eProtS

ProMode

Molecule of the Month

Download >>

PDB Archive/Snapshot Archive

About Remediation Data

Links

PDBj (Protein Data Bank Japan) maintains a centralized archive of macromolecular structures and provides integrated tools, in collaboration with the RCSB in USA and the PDB in EU. PDBj is supported by JST-BIRD.

xPSSS
x-ray Protein Structure Search Service

Advanced Search

About xPSSS
Update Information
wwPDB Snapshot Mirror

sorted by PDB ID

☐ Perform an exact word match

search reset

PDB ID:

☐ Search theoretical model

Release Date:

after month day year before month day year

Deposition Date:

after month day year before month day year

Citation Author:

1 and

2 and

3 ☐ primary citation only!

e.g.: Ito, N.

Journal:

Title

Name

Year

Volume

Yes No ignore

☐ ☐ ☐ polypeptide(L)

☐ ☐ ☐ polypeptide(D)

☐ ☐ ☐ polydeoxyribonucleotide

☐ ☐ ☐ polyribonucleotide

☐ ☐ ☐ polysaccharide(D)

Contains Chain Type:

Compound Information:

Title

Other DB :

ID :

Ligands and Prosthetic groups:

Number of Chains and Chain length:

Number Length

- -

min max min max

Experimental Technique:

Resolution:

smaller than or equal to Å larger than or equal to Å

- 以下の条件などで検索できます
 - 登録/リリース年月日
 - 登録者(論文著者)名
 - 論文のタイトル、雑誌名
 - 含まれるポリマーの種類(ポリペプチド、DNA, RNA, etc.)
 - 解像度
 - リガンド分子名
 - (後でみっちり練習します)

データバンクとウェブサイト

- 「ウェブサイト(例:www.pdbj.org)がデータバンク(例:PDB)」ではない。
- データバンクとは1次データが集積されいる場所(組織)のこと(PDBやINSD[核酸配列])。
- PDBの基本はデータバンク(≠データベース)。
- PDBj内に構築されたデータベースはウェブサイト経由でしか使えない。
- ウェブサイトで表示されるものがデータバンクに含まれるデータのすべてではない。
- 詳しい情報はエントリーのファイルを直接みる必要がある。

エントリーとエントリーファイル

- 「エントリー」とは一つの実験で決まったひとまとまりの立体構造の情報のこと。
- 「PDB ID」は各エントリーに固有の識別子で英数字4文字からなる。例: 1a00, 1gof (大文字・小文字の区別はない)
- 各エントリーのファイルはwww.pdbj.org などからダウンロードできる。
- エントリーファイルにはそのエントリーに関する(ほとんど)すべての情報が含まれる。(ただし、各種 convention の定義などは除く。)

PDBのファイル形式

各PDBエントリーは一つのファイルに収められている。

- PDB フォーマット
 - 伝統的なフラットファイル形式
 - 情報に欠落があるので厳密な解析には向かない(簡便フォーマット)。
- mmCIF
 - 国際結晶学連合(IUCr)が定めたCIF(Crystallographic Information File)を拡張したもの
 - Hall, S.R., et al. (1991), *Acta Cryst. A*, **47**:655-685.
- PDBML
 - mmCIFに基づくXMLフォーマット
 - Westbrook, J. et al. (2005), *Bioinformatics* **21**:988-992

「Download/Display」をもう一度開く

The screenshot shows the xPSSS (xml-based Protein Structure Search Service) interface. The main content area displays a table of available files for download or display. The table has four columns: file format, file name, Display, and Download. The 'Download/Display' button is highlighted in the navigation bar.

file format	file name	Display	Download
PDB format	all	display	download
		display	download
	header only	display	download
mmCIF	1gof.cif.gz(140k)	display	download
	1gof.xml.gz(233k)	display	download
XML	all	display	download
	no-atom	display	download
	ext-atom	display	download
Struct Factor	1gofsf.ent.gz(526k)	display	download

- 「Display」ボタンを押すとブラウザでファイルの中身が表示されます。
- ただし、XML(no-atom)だけは例外。

PDBフォーマット

```
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 ENGINEERED: YES
SOURCE      MOL_ID: 1;
SOURCE      2 ORGANISM_SCIENTIFIC: HYPOMYCES ROSELLUS;
SOURCE      3 ORGANISM_TAXID: 5132
KEYWDS      OXIDOREDUCTASE(OXYGEN(A))
EXPDTA      X-RAY DIFFRACTION
AUTHOR      N. ITO, S.E.V. PHILLIPS, P.F. KNOWLES
REVDAT      3   24-FEB-09 1GOF    1      VERSN
REVDAT      2   01-APR-03 1GOF    1      JRNL
REVDAT      1   31-JAN-94 1GOF    0
JRNL         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                               ISSN 0028-0836
JRNL         PMID    2002850
JRNL         DOI     10.1038/350087A0
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                               ISSN 0021-9258
REMARK      2
REMARK      2 RESOLUTION.      1.70 ANGSTROMS.
```

PDBフォーマット(つづき)

ATOM	1	N	ALA	A	1	38.840	0.236	1.012	1.00	34.65	N
ATOM	2	CA	ALA	A	1	38.356	-0.999	0.357	1.00	42.26	C
ATOM	3	C	ALA	A	1	37.098	-1.547	1.056	1.00	41.25	C
ATOM	4	O	ALA	A	1	36.619	-0.946	2.028	1.00	29.44	O
ATOM	5	CB	ALA	A	1	39.398	-2.114	0.379	1.00	40.70	C
ATOM	6	N	SER	A	2	36.610	-2.666	0.495	1.00	32.67	N
ATOM	7	CA	SER	A	2	35.411	-3.244	1.202	1.00	34.90	C
ATOM	8	C	SER	A	2	35.683	-4.740	1.081	1.00	38.30	C
ATOM	9	O	SER	A	2	36.827	-5.147	0.747	1.00	28.59	O
ATOM	10	CB	SER	A	2	34.063	-2.660	0.823	1.00	24.49	C
ATOM	11	OG	SER	A	2	33.031	-3.308	1.686	1.00	20.37	O
ATOM	12	N	ALA	A	3	34.660	-5.537	1.334	1.00	35.91	N
ATOM	13	CA	ALA	A	3	34.815	-6.995	1.246	1.00	33.38	C
ATOM	14	C	ALA	A	3	33.416	-7.594	1.259	1.00	21.71	C
ATOM	15	O	ALA	A	3	32.529	-6.833	1.679	1.00	30.82	O
ATOM	16	CB	ALA	A	3	35.687	-7.414	2.433	1.00	25.44	C
ATOM	17	N	PRO	A	4	33.335	-8.786	0.733	1.00	36.18	N
ATOM	18	CA	PRO	A	4	32.068	-9.552	0.674	1.00	40.82	C
ATOM	19	C	PRO	A	4	32.067	-10.418	1.934	1.00	37.76	C
ATOM	20	O	PRO	A	4	33.145	-10.698	2.466	1.00	42.84	O
ATOM	21	CB	PRO	A	4	32.222	-10.479	-0.536	1.00	40.12	C
ATOM	22	CG	PRO	A	4	33.729	-10.691	-0.634	1.00	48.00	C
ATOM	23	CD	PRO	A	4	34.452	-9.579	0.148	1.00	34.36	C

PDBフォーマット超入門

- 行指向(1行80コラム)(cf. FORTRAN77)
- ヘッダ(行頭6コラム)とアノテーションの組。
 - 長いアノテーションは次の行に継続。(7～10コラム目に継続を示す数字がある)
- 代表的なヘッダは: HEADER, TITLE, COMPND, SOURCE, EXPDTA, AUTHOR, JRNL, REMARK (番号つき), ATOM, HETATM
- 詳しくは <http://www.wwpdb.org/docs.html>

mmCIF

```
data_1GOF
#
_entry.id      1GOF
#
_audit_conform.dict_name      mmcif_pdbx.dic
_audit_conform.dict_version   1.0670
_audit_conform.dict_location  http://mmcif.pdb.org/dictionaries/ascii/mmcif_pdbx.dic
#
_database_2.database_id      PDB
_database_2.database_code    1GOF
#
loop_
_database_PDB_rev.num
_database_PDB_rev.date
_database_PDB_rev.date_original
_database_PDB_rev.status
_database_PDB_rev.replaces
_database_PDB_rev.mod_type
1 1994-01-31 1993-09-30 ? 1GOF 0
2 2003-04-01 ? ? 1GOF 1
3 2009-02-24 ? ? 1GOF 1
#
loop_
_database_PDB_rev_record.rev_num
_database_PDB_rev_record.type
_database_PDB_rev_record.details
2 JRNL ?
3 VERSN ?
#
_pdbx_database_status.status_code    REL
_pdbx_database_status.entry_id       1GOF
_pdbx_database_status.deposit_site   ?
_pdbx_database_status.process_site   ?
_pdbx_database_status.SG_entry       .
#
loop_
_audit_author.name
_audit_author.pdbx_ordinal
'Ito, N.' 1
'Phillips, S.E.V.' 2
'Knowles, P.F.' 3
#
```

mmCIF (つづき)

loop_
_atom_site.group_PDB
_atom_site.id
_atom_site.type_symbol
_atom_site.label_atom_id
_atom_site.label_alt_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_entity_id
_atom_site.label_seq_id
_atom_site.pdbx_PDB_ins_code
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.occupancy
_atom_site.B_iso_or_equiv
_atom_site.Cartn_x_esd
_atom_site.Cartn_y_esd
_atom_site.Cartn_z_esd
_atom_site.occupancy_esd
_atom_site.B_iso_or_equiv_esd
_atom_site.pdbx_formal_charge
_atom_site.auth_seq_id
_atom_site.auth_comp_id
_atom_site.auth_asym_id
_atom_site.auth_atom_id
_atom_site.pdbx_PDB_model_num

ATOM	1	N	N	.	ALA	A	1	1	?	38.840	0.236	1.012	1.00	34.65	?	?	?	?	?	?	?	1	ALA	A	N	1
ATOM	2	C	CA	.	ALA	A	1	1	?	38.356	-0.999	0.357	1.00	42.26	?	?	?	?	?	?	?	1	ALA	A	CA	1
ATOM	3	C	C	.	ALA	A	1	1	?	37.098	-1.547	1.056	1.00	41.25	?	?	?	?	?	?	?	1	ALA	A	C	1
ATOM	4	O	O	.	ALA	A	1	1	?	36.619	-0.946	2.028	1.00	29.44	?	?	?	?	?	?	?	1	ALA	A	O	1
ATOM	5	C	CB	.	ALA	A	1	1	?	39.398	-2.114	0.379	1.00	40.70	?	?	?	?	?	?	?	1	ALA	A	CB	1
ATOM	6	N	N	.	SER	A	1	2	?	36.610	-2.666	0.495	1.00	32.67	?	?	?	?	?	?	?	2	SER	A	N	1
ATOM	7	C	CA	.	SER	A	1	2	?	35.411	-3.244	1.202	1.00	34.90	?	?	?	?	?	?	?	2	SER	A	CA	1
ATOM	8	C	C	.	SER	A	1	2	?	35.683	-4.740	1.081	1.00	38.30	?	?	?	?	?	?	?	2	SER	A	C	1
ATOM	9	O	O	.	SER	A	1	2	?	36.827	-5.147	0.747	1.00	28.59	?	?	?	?	?	?	?	2	SER	A	O	1
ATOM	10	C	CB	.	SER	A	1	2	?	34.063	-2.660	0.823	1.00	24.49	?	?	?	?	?	?	?	2	SER	A	CB	1
ATOM	11	O	OG	.	SER	A	1	2	?	33.031	-3.308	1.686	1.00	20.37	?	?	?	?	?	?	?	2	SER	A	OG	1
ATOM	12	N	N	.	ALA	A	1	3	?	34.660	-5.537	1.334	1.00	35.91	?	?	?	?	?	?	?	3	ALA	A	N	1
ATOM	13	C	CA	.	ALA	A	1	3	?	34.815	-6.995	1.246	1.00	33.38	?	?	?	?	?	?	?	3	ALA	A	CA	1
ATOM	14	C	C	.	ALA	A	1	3	?	33.416	-7.594	1.259	1.00	21.71	?	?	?	?	?	?	?	3	ALA	A	C	1
ATOM	15	O	O	.	ALA	A	1	3	?	32.529	-6.833	1.679	1.00	30.82	?	?	?	?	?	?	?	3	ALA	A	O	1
ATOM	16	C	CB	.	ALA	A	1	3	?	35.687	-7.414	2.433	1.00	25.44	?	?	?	?	?	?	?	3	ALA	A	CB	1
ATOM	17	N	N	.	PRO	A	1	4	?	33.335	-8.786	0.733	1.00	36.18	?	?	?	?	?	?	?	4	PRO	A	N	1
ATOM	18	C	CA	.	PRO	A	1	4	?	32.068	-9.552	0.674	1.00	40.82	?	?	?	?	?	?	?	4	PRO	A	CA	1
ATOM	19	C	C	.	PRO	A	1	4	?	32.067	-10.418	1.934	1.00	37.76	?	?	?	?	?	?	?	4	PRO	A	C	1
ATOM	20	O	O	.	PRO	A	1	4	?	33.145	-10.698	2.466	1.00	42.84	?	?	?	?	?	?	?	4	PRO	A	O	1
ATOM	21	C	CB	.	PRO	A	1	4	?	32.222	-10.479	-0.536	1.00	40.12	?	?	?	?	?	?	?	4	PRO	A	CB	1
ATOM	22	C	CG	.	PRO	A	1	4	?	33.729	-10.691	-0.634	1.00	48.00	?	?	?	?	?	?	?	4	PRO	A	CG	1
ATOM	23	C	CD	.	PRO	A	1	4	?	34.452	-9.579	0.148	1.00	34.36	?	?	?	?	?	?	?	4	PRO	A	CD	1

mmCIF超入門

- 文脈自由文法(STAR[Self-defining Text Archive and Retrieval]形式)
- タグとアノテーションの組。
- タグは「_」で始まる。
- タグとアノテーションの組が複数ある場合には「loop_」構文で繰り返しを指定する。
- タグの意味はmmCIF dictionary(STAR形式)で定義されている。
- 色々な種類のデータは「カテゴリー」によって分類・整理されている。
 - カテゴリーの例: entity, struct, citation, citation_author, atom_site, chem_comp, cell, ...
- 詳しくは <http://mmcif.pdb.org/>

PDBML

```
<?xml version="1.0" encoding="UTF-8" ?>
<PDBx:datablock datablockName="1GOF-noatom"
  xmlns:PDBx="http://pdbml.pdb.org/schema/pdbx-v32.xsd"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://pdbml.pdb.org/schema/pdbx-v32.xsd pdbx-v32.xsd">
  <PDBx:atom_sitesCategory>
    <PDBx:atom_sites entry_id="1GOF">
      <PDBx:Cartn_transform_axes xsi:nil="true" />
      <PDBx:fract_transf_matrix11>0.011535</PDBx:fract_transf_matrix11>
      <PDBx:fract_transf_matrix12>0.000000</PDBx:fract_transf_matrix12>
      <PDBx:fract_transf_matrix13>0.000000</PDBx:fract_transf_matrix13>
      <PDBx:fract_transf_matrix21>0.000000</PDBx:fract_transf_matrix21>
      <PDBx:fract_transf_matrix22>0.011186</PDBx:fract_transf_matrix22>
      <PDBx:fract_transf_matrix23>0.000000</PDBx:fract_transf_matrix23>
      <PDBx:fract_transf_matrix31>0.006081</PDBx:fract_transf_matrix31>
      <PDBx:fract_transf_matrix32>0.000000</PDBx:fract_transf_matrix32>
      <PDBx:fract_transf_matrix33>0.011534</PDBx:fract_transf_matrix33>
      <PDBx:fract_transf_vector1>0.00000</PDBx:fract_transf_vector1>
      <PDBx:fract_transf_vector2>0.00000</PDBx:fract_transf_vector2>
      <PDBx:fract_transf_vector3>0.00000</PDBx:fract_transf_vector3>
    </PDBx:atom_sites>
  </PDBx:atom_sitesCategory>
  <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>
```

PDBML (つづき)

```
<PDBx:atom_siteCategory>
  <PDBx:atom_site id="1">
    <PDBx:B_iso_or_equiv>34.65</PDBx:B_iso_or_equiv>
    <PDBx:B_iso_or_equiv_esd xsi:nil="true" />
    <PDBx:Cartn_x>38.840</PDBx:Cartn_x>
    <PDBx:Cartn_x_esd xsi:nil="true" />
    <PDBx:Cartn_y>0.236</PDBx:Cartn_y>
    <PDBx:Cartn_y_esd xsi:nil="true" />
    <PDBx:Cartn_z>1.012</PDBx:Cartn_z>
    <PDBx:Cartn_z_esd xsi:nil="true" />
    <PDBx:auth_asym_id>A</PDBx:auth_asym_id>
    <PDBx:auth_atom_id>N</PDBx:auth_atom_id>
    <PDBx:auth_comp_id>ALA</PDBx:auth_comp_id>
    <PDBx:auth_seq_id>1</PDBx:auth_seq_id>
    <PDBx:group_PDB>ATOM</PDBx:group_PDB>
    <PDBx:label_alt_id></PDBx:label_alt_id>
    <PDBx:label_asym_id>A</PDBx:label_asym_id>
    <PDBx:label_atom_id>N</PDBx:label_atom_id>
    <PDBx:label_comp_id>ALA</PDBx:label_comp_id>
    <PDBx:label_entity_id>1</PDBx:label_entity_id>
    <PDBx:label_seq_id>1</PDBx:label_seq_id>
    <PDBx:occupancy>1.00</PDBx:occupancy>
    <PDBx:occupancy_esd xsi:nil="true" />
    <PDBx:pdxb_PDB_ins_code xsi:nil="true" />
    <PDBx:pdxb_PDB_model_num>1</PDBx:pdxb_PDB_model_num>
    <PDBx:pdxb_formal_charge xsi:nil="true" />
    <PDBx:type_symbol>N</PDBx:type_symbol>
  </PDBx:atom_site>
```

```
<PDBx:atom_site id="2">
  <PDBx:B_iso_or_equiv>42.26</PDBx:B_iso_or_equiv>
  <PDBx:B_iso_or_equiv_esd xsi:nil="true" />
  <PDBx:Cartn_x>38.356</PDBx:Cartn_x>
  <PDBx:Cartn_x_esd xsi:nil="true" />
  <PDBx:Cartn_y>-0.999</PDBx:Cartn_y>
  <PDBx:Cartn_y_esd xsi:nil="true" />
  <PDBx:Cartn_z>0.357</PDBx:Cartn_z>
  <PDBx:Cartn_z_esd xsi:nil="true" />
  <PDBx:auth_asym_id>A</PDBx:auth_asym_id>
  <PDBx:auth_atom_id>CA</PDBx:auth_atom_id>
  <PDBx:auth_comp_id>ALA</PDBx:auth_comp_id>
  <PDBx:auth_seq_id>1</PDBx:auth_seq_id>
  <PDBx:group_PDB>ATOM</PDBx:group_PDB>
  <PDBx:label_alt_id></PDBx:label_alt_id>
  <PDBx:label_asym_id>A</PDBx:label_asym_id>
  <PDBx:label_atom_id>CA</PDBx:label_atom_id>
  <PDBx:label_comp_id>ALA</PDBx:label_comp_id>
  <PDBx:label_entity_id>1</PDBx:label_entity_id>
  <PDBx:label_seq_id>1</PDBx:label_seq_id>
  <PDBx:occupancy>1.00</PDBx:occupancy>
  <PDBx:occupancy_esd xsi:nil="true" />
  <PDBx:pdxb_PDB_ins_code xsi:nil="true" />
  <PDBx:pdxb_PDB_model_num>1</PDBx:pdxb_PDB_model_num>
  <PDBx:pdxb_formal_charge xsi:nil="true" />
  <PDBx:type_symbol>C</PDBx:type_symbol>
</PDBx:atom_site>
```

PDBML超入門

- 文脈自由文法(XML)
- mmCIFのカテゴリに対応したXMLのタグ。
- XML要素の型はPDBMLのXML Schemaで定義されている。(タグの意味はmmCIFから継承されている。)

まとめ:フォーマットの比較

	形式	データの完全性	人が読みやすい	プログラムで処理しやすい	標準化
PDB	フラットファイル	×	◎	×	wwPDB
mmCIF	STAR	○	○	○	IUCr, wwPDB
PDBML	XML	○	×	◎	W3C, wwPDB

- PDBフォーマットは基本的なデータを目で確認するには適しているが、その他の用途には信頼性の面で問題がある。プログラムで一貫して扱うのはほとんど不可能。
- mmCIFは完全なデータが比較的目に易しい形で閲覧できるのが良い。パーザライブラリ(C++)はRCSBから用意されている。(あるいは lex/yacc などで作る)
- PDBMLは適当なブラウザ(PDBjの「Display」機能、Firefoxなど)がないと直接目で読むに耐えない。一般的なXML処理ライブラリを用いることができる。

アンケートから(1)

- 「PDB fileを見ると、1つのchainに複数の低分子リガンドが入っているものがあるが、全てべつのchainとして登録してもらえるとありがたいです。」

→PDB ファイルではなく、mmCIFまたはPDBMLを参照すればよい。

- PDBファイルの “chain ID”は auth_asym_id
- mmCIF/PDBMLにはそれとは別に label_asym_idがある。これは異なる分子には異なるIDが割り当てられる。

分子の「chain ID」

1A00を例にとると……

PDB file

```
ATOM      1  N      VAL  A      1      101.601  38.534  -1.962   1.00  53.29      N
HETATM  4387  FE      HEM  A     142      93.518  58.468  -3.212   1.00  18.42      FE
```

mmCIF

```
ATOM      1      N  N      . VAL  A  1  1      ? 101.601  38.534  -1.962   1.00  53.29      ? ? ? ? ? ? 1 VAL  A  N      1
HETATM  4383  FE  FE      . HEM  E  3  .      ? 93.518  58.468  -3.212   1.00  18.42      ? ? ? ? ? ? 142 HEM  A  FE      1
```

PDBML

<PDBx:atom_site id="1">

```
...
<PDBx:auth_asym_id>A</PDBx:auth_asym_id>
<PDBx:auth_atom_id>N</PDBx:auth_atom_id>
<PDBx:auth_comp_id>VAL</PDBx:auth_comp_id>
<PDBx:auth_seq_id>1</PDBx:auth_seq_id>
<PDBx:group_PDB>ATOM</PDBx:group_PDB>
<PDBx:label_asym_id>A</PDBx:label_asym_id>
<PDBx:label_atom_id>N</PDBx:label_atom_id>
<PDBx:label_comp_id>VAL</PDBx:label_comp_id>
<PDBx:label_entity_id>1</PDBx:label_entity_id>
<PDBx:label_seq_id>1</PDBx:label_seq_id>
<PDBx:occupancy>1.00</PDBx:occupancy>
<PDBx:pd_bx_PDB_model_num>1
  </PDBx:pd_bx_PDB_model_num>
<PDBx:type_symbol>N</PDBx:type_symbol>
</PDBx:atom_site>
```

<PDBx:atom_site id="4383">

```
...
<PDBx:auth_asym_id>A</PDBx:auth_asym_id>
<PDBx:auth_atom_id>FE</PDBx:auth_atom_id>
<PDBx:auth_comp_id>HEM</PDBx:auth_comp_id>
<PDBx:auth_seq_id>142</PDBx:auth_seq_id>
<PDBx:group_PDB>HETATM</PDBx:group_PDB>
<PDBx:label_asym_id>E</PDBx:label_asym_id>
<PDBx:label_atom_id>FE</PDBx:label_atom_id>
<PDBx:label_comp_id>HEM</PDBx:label_comp_id>
<PDBx:label_entity_id>3</PDBx:label_entity_id>
<PDBx:label_seq_id></PDBx:label_seq_id>
<PDBx:occupancy>1.00</PDBx:occupancy>
<PDBx:pd_bx_PDB_model_num>1
  </PDBx:pd_bx_PDB_model_num>
<PDBx:type_symbol>FE</PDBx:type_symbol>
</PDBx:atom_site>
```

chainと分子の対応

- struct_asymCategoryと entityCategoryの対応

```
<PDBx:entityCategory>
  <PDBx:entity id="1">
    <PDBx:formula_weight>15150.497</PDBx:formula_weight>
    <PDBx:pdxb_description>HEMOGLOBIN (ALPHA CHAIN)</PDBx:pdxb_description>
    <PDBx:pdxb_number_of_molecules>2.</PDBx:pdxb_number_of_molecules>
    <PDBx:src_method>man</PDBx:src_method>
    <PDBx:type>polymer</PDBx:type>
  </PDBx:entity>
  <PDBx:entity id="2">
    <PDBx:formula_weight>15899.374</PDBx:formula_weight>
    <PDBx:pdxb_description>HEMOGLOBIN (BETA CHAIN)</PDBx:pdxb_description>
    <PDBx:pdxb_mutation>V1M, W37Y</PDBx:pdxb_mutation>
    <PDBx:pdxb_number_of_molecules>2.</PDBx:pdxb_number_of_molecules>
    <PDBx:type>polymer</PDBx:type>
  </PDBx:entity>
  ...
</PDBx:entityCategory>

<PDBx:struct_asymCategory>
  <PDBx:struct_asym id="A">
    <PDBx:entity_id>1</PDBx:entity_id>
  </PDBx:struct_asym>
  <PDBx:struct_asym id="B">
    <PDBx:entity_id>2</PDBx:entity_id>
  </PDBx:struct_asym>
  <PDBx:struct_asym id="C">
    <PDBx:entity_id>1</PDBx:entity_id>
  </PDBx:struct_asym>
  <PDBx:struct_asym id="D">
    <PDBx:entity_id>2</PDBx:entity_id>
  </PDBx:struct_asym>
```



アンケートから(2)

- 「“MSE”といったアミノ酸表記を“MET”に直した場合のXMLファイルや、少なくとも対応付けリストがあると助かります。」

→ mmCIF/PDBMLのpdbx_struct_mod_residueカテゴリに対応付けリストがある！

```
<PDBx:pdbx_struct_mod_residueCategory>
  <PDBx:pdbx_struct_mod_residue id="1">
    <PDBx:PDB_ins_code xsi:nil="true" />
    <PDBx:auth_asym_id>A</PDBx:auth_asym_id>
    <PDBx:auth_comp_id>MSE</PDBx:auth_comp_id>
    <PDBx:auth_seq_id>45</PDBx:auth_seq_id>
    <PDBx:details>SELENOMETHIONINE</PDBx:details>
    <PDBx:label_asym_id>A</PDBx:label_asym_id>
    <PDBx:label_comp_id>MSE</PDBx:label_comp_id>
    <PDBx:label_seq_id>45</PDBx:label_seq_id>
    <PDBx:parent_comp_id>MET</PDBx:parent_comp_id>
  </PDBx:pdbx_struct_mod_residue>
```

演習課題1

- エントリー「101d」を検索して、以下のことを調べる。
 - 発表された論文は？
 - 実験手法は？
 - 含まれる分子種は？
 - 解像度(もしあれば)は？

ヒント: citation, exptl, entity, refine

演習課題2

- キーワード検索で「aldehyde dehydrogenase」を入力にして、以下のことを調べる。
 - 何件のエントリーがヒットするか？
 - Release date が最も古いエントリーはどれか？
 - その論文は何年に発表されたか？
 - そのエントリーは何年にPDBからリリースされたか？
 - その由来する生物種は何か？
 - その活性残基はどれか？

ヒント: citation, database_PDB_rev, entity_src_gen, struct_site

演習課題3

- PDBフォーマットとmmCIFを比べてみる。エントリー「1gof」を例にとる。
 - PDB, mmCIFで、release dateはそれぞれどこに書かれているか？
 - PDB, mmCIFで、登録者名はそれぞれどこに書かれているか？
 - PDB, mmCIFで、実験手法はそれぞれどこに書かれているか？
 - PDB, mmCIFで、由来する生物種(biological species)はそれぞれどこに書かれているか？
 - PDBの”ATOM”行はmmCIFのどこにどう対応するか？

ヒント: database_PDB_rev, audit_author, exptl, entity_src_gen, atom_site

演習課題4

- mmCIFとPDBMLを比較してみる。
 - 演習課題3と同じことをmmCIFとPDBMLで行う。
 - http://service.pdbj.org/mine/pdbml_image.html も参照。
 - PDBMLには原子座標の部分だけを簡略化した「extatom」と呼ばれる形式もある。PDBjのサイトで、これとmmCIFと対応する部分を比較せよ。

ヒント: database_PDB_rev, audit_author, exptl, entity_src_gen, atom_site

演習課題5

- PDBjのトップページから、今週の更新情報を調べる。
 - 今週新規に追加されたエントリーは何件あるか？
 - いくつか例を眺めてみる。
 - 今週更新されたエントリーは何件あるか？
 - いくつか例を眺めてみて、何度更新されたかを調べる。
 - “database_PDB_rev” カテゴリをみるとわかる。
 - どこが更新されたかを調べる。
 - “database_PDB_rev_record”カテゴリをみると(だいたい)わかる。

演習課題6

- Advanced Searchを用いて以下のことを調べる。
 - Release dateが2000年1月1日以降のエントリーは何件あるか？
 - Release dateが1999年12月31日以前のエントリーは何件あるか？

演習課題7

- Advanced Searchを使って、引用文献に関して調べてみる。
 - “Walker, J.E.” を著者に含むエントリーは何件あるか？
 - ヒットしたものの中には、一見して “Walker, J.E.” を含まないように見えるものもある。なぜか？
 - “Primary citation only!” ボタンをチェックして、同じ検索を試みる。結果はどう変わるか？
 - 雑誌 “Nature” に掲載されたエントリーは何件あるか？
 - 雑誌 “Science” に2005年に掲載されたエントリーはいくつあるか？

ヒント: audit_author, citation, citation_author

演習課題8

- Advanced Searchをつかう。
 - “polypeptide(L)”（要するに普通の蛋白質）を含むエントリーは何件あるか？
 - “polypeptide(D)”のみを含むエントリーを検索するにはどうすればよいか？
 - 蛋白質とDNAの複合体を検索するにはどうすればよいか？
 - 蛋白質、DNA、RNAを含むエントリーは何件あるか？
 - RNAと “polysaccharide(D)” の両方を含むエントリーはあるか？
 - その他色々な組み合わせを試してみる。

ヒント: entity_poly, entity

演習課題9

- Advanced Searchを使って蛋白質の名前で検索する。
 - “Compound information” に “prion” を入れて検索する。
 - ヒットしたエントリーのサマリーページをいくつか眺めて、“prion” という単語がどこに現れているかを観察する。
 - mmCIFまたはPDBML-noatomを読んで、prionという単語がどこに含まれているかを調べる。(どのタグか?)
 - “prion”の代わりに “oxidoreductase”として同じことを試してみる。

ヒント: entity, struct, pdbx_database_related, citation, entity_name_com, struct_keywords

演習課題10

- Advanced Search
 - “Title” に “crystal” と入力して検索。
 - 前問と同じように、サマリーページと元ファイルを確認する。
 - 他の単語でも試してみる。
 - たとえば “DNA complex”

演習課題 11

- PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>) にアクセスして、“Borhani DW [au] truncated human apolipoprotein 1997” を検索する。
 - PubMed ID (PMID) をコピーする。
- Advanced Searchで “Other DB” から PubMedを選び、“ID:” の箱にPMIDをペーストして、検索する。
 - PDB IDは何か？
 - Primary citation とpubMedでの検索結果を比較する。
 - PMIDはmmCIF/PDBMLのどこに記載されているか？

演習課題12

- Advanced Search の “Other DB” から “EC Number” (酵素番号) を選択する。
- “ID:” の箱に “1.1.1.1” (alcohol dehydrogenase) を入力する。
 - ヒットするエントリーは何件か？
 - EC番号は mmCIF/PDBML のどこに書かれているか？
 - ついでにいくつかのエントリーで活性残基も調べてみる。

ヒント: entity, struct_site

演習課題13

- Advanced Search の “Other DB” から “GO” (gene ontology) を選択して、“ID:” には何も入れずに検索してみる。
 - 何件ヒットするか？
 - いくつかのエントリーで “Functional Details” ページを眺めてみる。
 - GO の ID は mmCIF/PDBML のどこに書かれているか？

ヒント: gene_ontology (PDBMLplusのみ)

演習課題14

- Advanced Search の “Other DB” から “UniProt” を選択して、“ID:” には何も入れずに検索する。
 - これでヒットするエントリーの例えば 101m のサマリーページを見ると、ページ下の方に UniProt へのリンクがある。これをたどって、この蛋白質のアノテーションを調べてみる。
 - “Functional Details” ページ (PDBj) に gene ontology のアノテーションはあるか？
 - 対応する UniProt エントリー (リンク先) には gene ontology のアノテーションがあるか？
 - UniProt の ID (P02185 など) は mmCIF/PDBML のどこに書かれているか？

ヒント: gene_ontology(PDBMLplus only), struct_ref

演習課題15

- 前問と同様のことを“GB” (GenBank), “EMBL”および“PIR”でもやってみる。
 - それぞれで、最も最近リリースされたエントリーは何か？
 - (PIRは2002年にUniProtに併合された)

ヒント: `gene_ontology(PDBMLplus only)`, `struct_ref`

演習課題16

- Advanced Search の “Ligands and Prosthetic groups” に “ATP” と入力して検索する。
 - ヒットしたエントリーのいくつかをサマリーページで確認し、jV3でグラフィクス表示してどこに ATP が結合しているかを確認する。
 - mmCIF/PDBMLではどこに “ATP”に関する記述があるか？
 - PDBフォーマットファイルではどうか？
 - ATPの原子座標もみってみる。蛋白質の原子座標の表記とどのように異なるか？

ヒント: chem_comp, entity, pdbx_entity_nonpoly

演習課題17

- 前問と同様のことを色々なリガンド名で試してみる: FAD, NAD, HEM, PO4 など。
- PDB に含まれるリガンド化合物は chemical component dictionary で定義されている。詳しくは、
 - <http://www.wwpdb.org/ccd.html>

演習課題18

- Advanced Searchの“Numbers of chains and Chain length”でポリマー分子の数(上限、下限)とポリマー分子の長さ(上限、下限)をいろいろ指定して検索してみる。
 - ポリマーは蛋白質、核酸などを含む。
 - 例えば、最低(min)10分子のポリマーを含み、長さが最大20ユニット(残基、塩基)のポリマーを含むエントリーは？
 - ポリマーとしては蛋白質のみを含み、かつ二量体として解かれた構造はどうやって検索するか？

ヒント: entity, entity_poly, entity_poly_seq, pdbx_poly_seq_scheme

演習課題19

- Advanced Searchの“Experimental Technique”にはどのような選択肢があるか？
 - それぞれを選んで検索し、何件ヒットするか確認する。
 - この選択肢以外に登録されているデータは(今のところ)ない。
 - 検索結果の絵をぱっと見て、実験手法ごとにどのような傾向があるか考えてみる。
 - 実験手法の情報はmmCIF/PDBMLのどこに記載されているか？
 - 次に resolution (解像度)も指定してみる。
 - 実験手法によっては解像度がないものもあることに注意。
 - ところが、“SOLUTION NMR”で解像度 2.5Å以下と指定するとヒットするものがある。なぜか？

ヒント: exptl, refine, reflns

演習課題20

- Advanced Searchの“Source”に生物種名を入れて検索する。(その蛋白質などが由来する生物種)
 - 例: human, bos taurus, pyrococcusなど
- “Host species”に生物種名を入れて検索する。(その蛋白質などを精製するのに用いた生物種)
- これらの情報はmmCIF/PDBMLのどこに記載されているか？

ヒント: entity_src_gen, entity_src_nat, entity