蛋白質立体構造データの使い方

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)

概略

- PDBj & wwPDB
- PDBj をウェブサイトの基本
- PDBデータフォーマットの解説
- PDBjのサービス
- 演習問題

立体構造データベース:PDB

- wwPDB (Worldwide Protein Data Bank)
 - 以下PDBと略

http://www.wwpdb.or

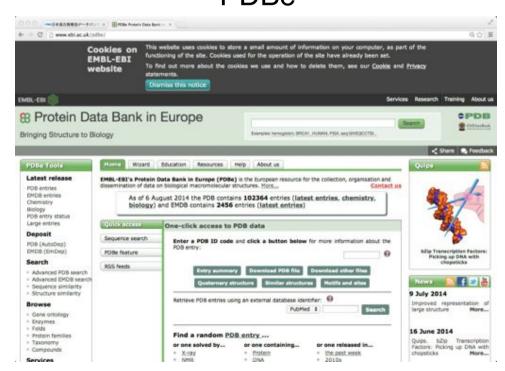
G

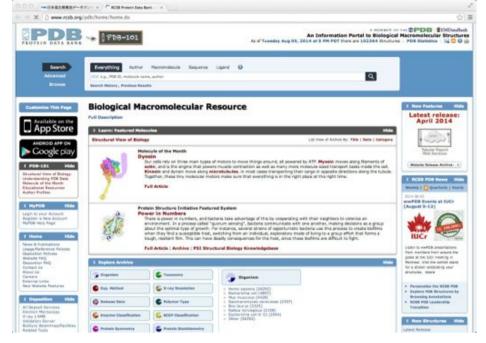
- 日米欧で共同運営
- 立体構造解析を行った研究者がPDBにデータを登録
- そのデータをPDBが全世界に無償で公開





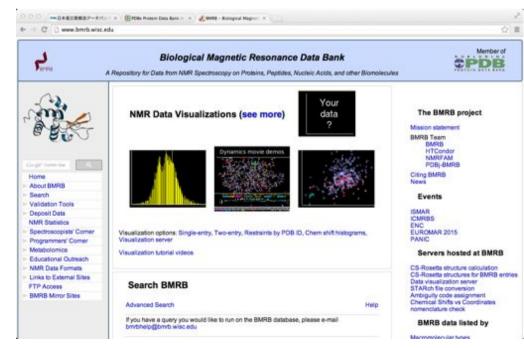
PDBe





RCSB PDB

BMRB



PDBのやっていること

- 登録
 - 日米欧で作業を分担

- 公開
 - データそのものは日米欧で同一
 - その他のサービスは日米欧が独自に提供

PDBjをつかってみる



http://pdbj.org/

基本的な検索法

- エントリーの取得
- キーワード検索
- 詳細条件検索

エントリーの取得



- 検索ボックスに "PDB ID"を入れる
 - 例:1gof
- リターン(エン ター)キーを押す/ または「虫眼鏡」を クリック

エントリーのサマリーページ

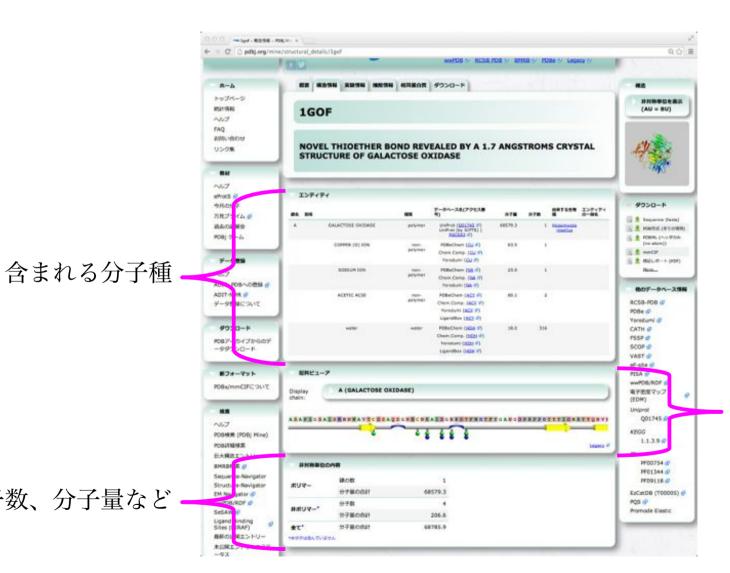


エントリーのその他の情報



- 構造情報
- 実験情報
- 機能情報
- 相同蛋白質
- ファイルダウンロード
 - などのページがあります。

エントリーの構造情報



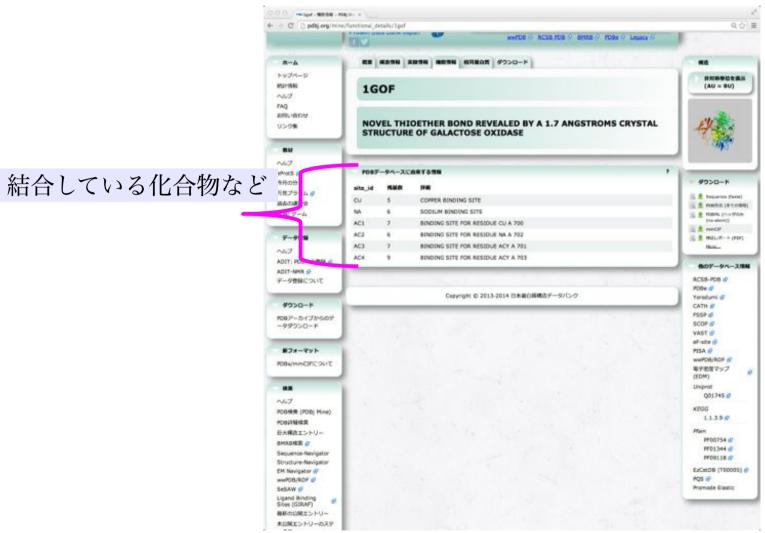
二次構造、SS結合、 リガンド結合サイト

分子数、分子量など -

エントリーの実験情報



エントリーの機能情報



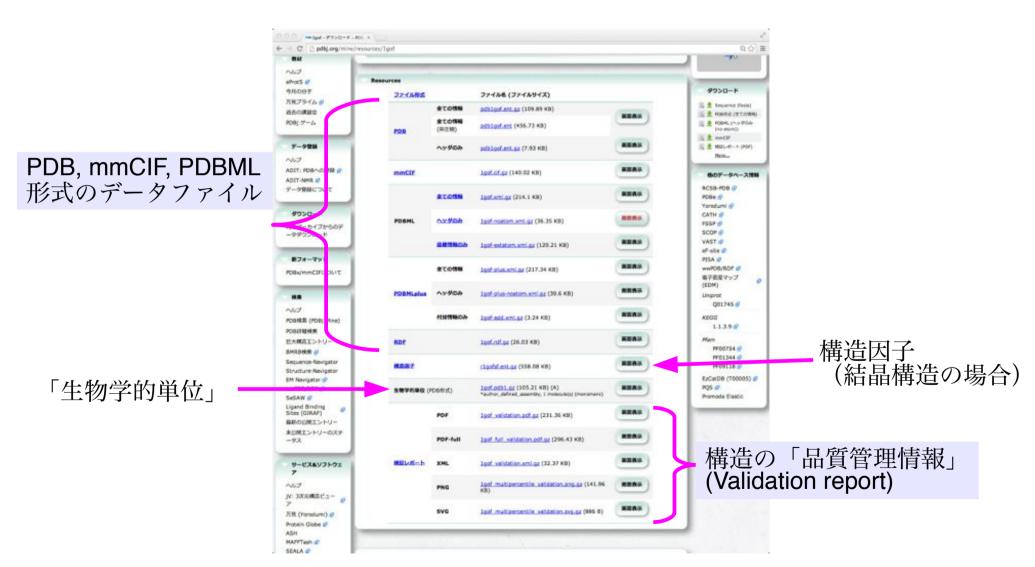
蛋白質の「機能」とは?

エントリーの相同蛋白質

PDBに存在するこのエントリーの配列の相同蛋白質をBLASTで検索



各種ファイルのダウンロード



PDBのファイル形式

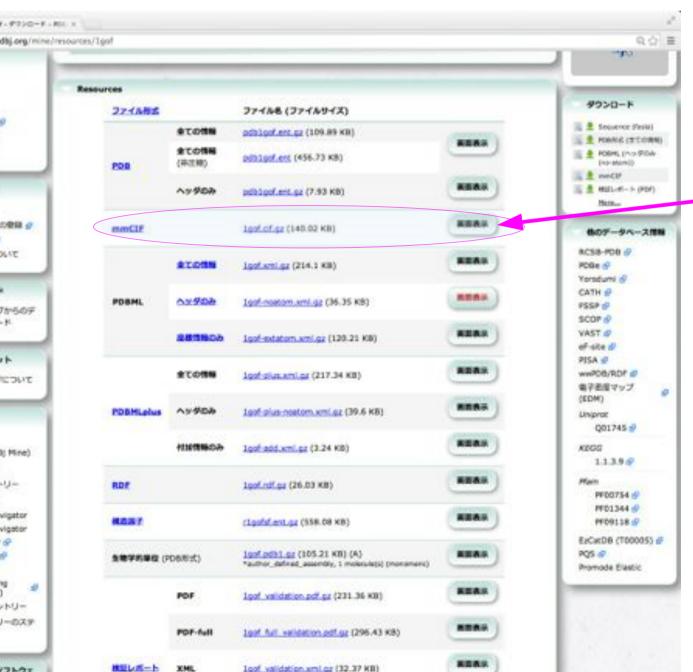
PDBのファイル形式の種類

- mmCIF
 - wwPDBのもっとも基本的なフォーマット
- PDBML
 - mmCIFをXMLに「直訳」したもの
- PDBフォーマット
 - 昔ながらのplain textフォーマット
- PDB/RDF
 - PDBMLをセマンティックウェブの標準形式である RDFに変換したもの

「その他」のデータ

- 検証レポート(Validation Report)
 - 構造データの品質チェック
- 構造因子 (X線結晶構造解析の場合)
- NMR距離制限情報(NMR構造の場合)

mmCIFの例



エントリーの 「ダウンロード」 ページでここをクリック!

mmCIFの基本

- データはいろいろなカテゴリに分類されている
 - _category.item
 - 例:_entry.id "entry"はカテゴリ名、"id"はその項目(item)
 - 「_entry.id 1GOF」はentryカテゴリのid項目の値が「1GOF」である、という意味。
- データの記述法は2通り
 - key-value: 1 つのカテゴリに 1 つの値 (の組) しかない場合。
 - loop: 1 つのカテゴリに複数の値(の組)がある場 合。

20

mmCIF: 追加情報

- 文脈自由文法 (STAR[Self-def ning Text Archive and Retrieval]形式)で記述。
- タグの意味はPDBx/mmCIF dictionaryで定義されている。
- 詳しくは http://mmcif.wwpdb.org/

mmCIFをもう少し見てみる

```
このタグでエントリーが始まる("datablock"がエントリーの単位)
data_1GOF
#
                       entry ID (PDB ID)は"1GOF"
_entry.id
#
        1GOF
audit conform.dict name
                          mmcif_pdbx.dic
_audit_conform.dict_version
                          4.007
__audit_conform.dict_location
#
                         http://mmcif.pdb.org/dictionaries/ascii/mmcif_pdbx.dic
                        PDB
database 2.database id
_database_2.database_code #
                         1GOF
                                                         「管理情報」も含まれる。
```

key-valueの例

```
cell.entry id
                          1GOF
cell.length a
                          98.000
cell.length b
                          89.400
cell.length c
                          86.700
cell.angle alpha
                          90.00
cell.angle beta
                          117.80
cell.angle gamma
                          90.00
cell.Z PDB
cell.pdbx unique axis
```

最後の「#」はそのカテゴリの記述の終わりを示す慣習 (convention)

loopの例

```
ループの開
始
loop
entity.id
entity.type
entity.src method
entity.pdbx description
entity.formula weight
                                      項目のリスト(「1行1項目」
entity.pdbx number of molecules
                                      は慣習)
entity.details
entity.pdbx mutation
entity.pdbx fragment
 entity.pdbx ec
                                        68579.250 1 ? ? ? 1.1.3.9
63.546 1 ? ? ? ?
22.990 1 ? ? ? ?
60.052 2 ? ? ?
1 polymer
              man 'GALACTOSE OXIDASE'
2 non-polymer syn 'COPPER (II) ION'
3 non-polymer syn
                   'SODIUM ION'
4 non-polymer syn 'ACETIC ACID'
                                                   316 ? ? ? ?
5 water
            nat water
                                        18.015
                                   • 各項目は空白で区切られる
```

最後の「#」はループの終わりを示す慣習囲む

(convention)

• 項目リストと同じ順番で並ぶ

• 空白を含むデータは引用府「'」で

mmCIFの原子座標情報

```
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
       1
            Ν
                Ν
                    . ALA A 1 1
                                              0.236
                                                       1.012
ATOM
                                   ? 38.840
                                                               1.00 34.65 ? ? ? ? ?
                                                                                             AT<sub>1</sub>A A N
                                                                                                        1
АТОМ
                CA
                                   ? 38.356
                                              -0.999
                                                       0.357
                                                               1.00 42.26 ?
                                                                             ?
                                                                                                        1
                    . ALA A 1 1
                                                                                             ALA A CA
ATOM
                    . ALA A 1 1
                                   ? 37.098
                                              -1.547
                                                       1.056
                                                               1.00 41.25 ?
                                                                                             ATIA A C
                                   ? 36.619
                                              -0.946
                                                       2.028
ATOM
                    . ALA A 1 1
                                                               1.00 29.44 ?
                                                                                             ALA A O
ATOM
                    . ALA A 1 1
                                   ? 39.398
                                              -2.114
                                                       0.379
                                                                                             ALA A CB
АТОМ
                    . SER A 1 2
                                   ? 36,610
                                              -2.666
                                                       0.495
                                                                                             SER A N
                N
                                                               1.00 32.67 ?
                CA
                    . SER A 1 2
                                   ? 35.411
                                              -3.244
                                                       1.202
                                                                                             SER A CA
ATOM
ATOM
                    . SER A 1 2
                                   ? 35.683
                                              -4.740
                                                      1.081
                                                               1.00 38.30 ?
                                                                                             SER A C
                                                                                                        1
ATOM
                    . SER A 1 2
                                   ? 36.827
                                              -5.147
                                                       0.747
                                                                                             SER A O
                                                                                                        1 25
                CB
                                                      0.823
ATOM
       10
                    . SER A 1 2
                                   ? 34.063
                                              -2.660
                                                               1.00 24.49 ? ? ? ? ?
                                                                                             SER A CB
ATOM
       11
                OG
                    . SER A 1 2
                                   ? 33.031
                                              -3.308
                                                       1.686
                                                               1.00 20.37 ? ?
                                                                                             SER A OG
```

主なカテゴリ(グループ)

- _entity 研究対象の分子情報
 - entity, entity_poly, pdbx_entity_nonpoly, ...
- _atom 各原子の情報(座標など)
 - atom_site
- _struct 構造の特色(分子全体、2次構造など)
 - struct, struct_conf, struct_sheet, struct_conn, pdbx_struct_assembly, ...
- _chem_comp 化合物データ
 - chem_comp
- - citation, citation_author, ...

カテゴリ間の関係

```
loop
struct asym.id
struct asym.pdbx blank PDB chainid flag
struct asym.pdbx modified
struct asym.entity id
                         「子(child)」
struct asym.details
A N N 1 ?
B N N 2 ?
C N N 3 ?
                                      「親子関係」もPDBxで定義されている
 N N 4 ?
E N N 4 ?
F N N 5 ?
                                             「親(parent)」
                                   entity.type
                                   entity.src method
                                   entity.pdbx description
                                   entity.formula weight
                                   entity.pdbx number of molecules
                                   entity.details
                                   entity.pdbx mutation
                                   entity.pdbx fragment
                                   entity.pdbx_ec
                                  1 polymer
                                                                                    ? ? ? 1.1.3.9
                                                                       68579.250 1
                                               man 'GALACTOSE OXIDASE'
                                  2 non-polymer syn 'COPPER (II) ION'
                                                                       63.546
                                  3 non-polymer syn 'SODIUM ION'
                                                                       22.990
                                  4 non-polymer syn 'ACETIC ACID'
                                                                       60.052
                                  5 water
                                                nat water
                                                                       18,015
                                                                                316 ? ? ? ?
```

"label" 2 "auth"

```
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
                                  "label ...": wwPDB が内部的に付与したラベル
 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 seg id
 atom site.auth comp id
                                 "auth …": 登録者が任意に付与したラベル
 atom site.auth asym id
 atom site.auth atom id
 atom site.pdbx PDB model num
            N N
АТОМ
                    . ALA A 1 4
                                  ? 11.751 37.846 29.016
                                                           1.00 44.65 ?
                                                                                      AT<sub>1</sub>A A N
ATOM
            C CA
                    . ALA A 1 4
                                  ? 12.501 39.048 28.539
                                                           1.00 30.68 ?
                                                                                      ALA A CA
                                                                                                  1
            C
              C
                                                                                      ALA A C
                                                                                                  1
ATOM
                    . ALA A 1 4
                                  ? 13.740 38.628 27.754
                                                           1.00 24.74 ?
ATOM
            0 0
                    . ALA A 1 4
                                  ? 14.207 37.495 27.890
                                                          1.00 25.59 ?
                                                                                      ALA A O
                                                                                                  1
                                                                                                  1
ATOM
            C CB
                    . ALA A 1 4
                                  ? 12.902 39.919 29.730
                                                          1.00 16.77 ?
                                                                                      ALA A CB
            N N
                                  ? 14.235 39.531 26.906
                                                           1.00 19.29
                                                                                      TYR A N
                                                                                                  1
ATOM
                    . TYR A 1 5
                                                                                      TYR A CA
            C CA
                    . TYR A 1 5
                                  ? 15.552 39.410 26.282
                                                          1.00 8.51
                                                                                                  1
ATOM
АТОМ
            C
              C
                    . TYR A 1 5
                                  ? 16.616 38.913 27.263
                                                          1.00 6.11
                                                                                      TYR A C
                                                                                                  1
       9
                                                                                                  1
ATOM
            0 0
                    . TYR A 1 5
                                  ? 17.187 37.844 27.068
                                                          1.00 17.99 ?
                                                                                      TYR A O
ATOM
       10
            C CB
                      TYR A 1 5
                                  ? 15.988 40.762 25.702
                                                          1.00 2.00
                                                                                                  1
                                                                                      TYR A CB
```

ファイルの解読演習

PDBj 左のタブ:新フォーマット「PDBx/mmClFについて」
H23年度大阪大学蛋白研セミナー「PDBj講習会」-in 福岡- (演習問題と回答)

http://pdbj.org/workshop/20120207/ito2.pdf

少し凝った検索

- キーワード検索を試してみる。
- 「詳細条件検索」を試してみる。
 - 指定した条件が、mmCIFまたはPDBMLまたはPDBファイルのどこに書かれているかを確認する。

その他の検索

「PDB検索(PDBj Mine)」



詳細条件検索

「PDB詳細検索」



いろいろなサービス

万見(Yorodumi)

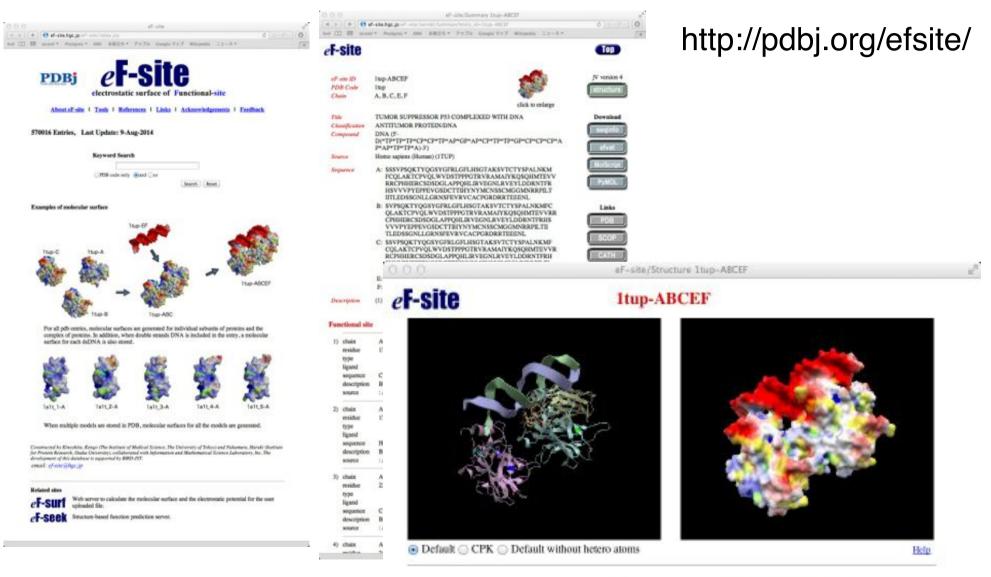
高機能&対話的な構造ブラウザ



万見で機能部位を見る



eF-siteで分子表面を見る



Java環境が必要です!

details of a Functional site

相同蛋白質をさがす

Sequence Navigator (PDBに対するBLAST検索)

PDB IDとChain IDを指定

アミノ酸配列(1文字コード)を入力



