

よくある質問から(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。

よくある質問から(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 ！

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「アルデヒド脱水素酵素」(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？
- 公開日 (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 formatとmmCIFを比較してみる。

- 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。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

詳細検索 (Advanced Search) を用いて以下のことを調べる。

- 「公開日」(Release date) が2000年1月1日以降のエントリーは何件あるか？
- 「公開日」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

詳細検索 (Advanced Search) を用いて、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？
- 「引用文献」(Primary) 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

詳細検索 (Advanced Search) を用いて、以下のことを調べる。

- 「ポリペプチド(L体)」(通常の蛋白質)を含むエントリーは何件あるか？
- 「ポリペプチド(D体)」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「多糖(D体)」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

詳細検索 (Advanced Search) を用いて、蛋白質の名前で検索する。

- 「化合物情報」(Compound information) に「prion」を入れて検索する。
- ヒットしたエントリーの概要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

詳細検索 (Advanced Search) を用いて、

- 「エントリーのタイトル」(Title) に「crystal」と入力して検索。
- 前問と同じように、概要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。
- 詳細検索 (Advanced Search) の「外部データベース」 (Other DB) からPubMedを選び、「ID:」の箱にPMIDを貼り付けて、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？

ヒント: citation

演習課題12

- 詳細検索 (Advanced Search) の「外部データベース」(Other DB) から「EC Number」(酵素番号) を選択する。
- 「ID:」の入力ボックスに「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

詳細検索 (Advanced Search) の「外部データベース」(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「機能情報」(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

詳細検索 (Advanced Search) の「外部データベース」(Other DB)から「UniProt」を選択して、「ID:」には何も入れずに検索してみる。

- ヒットしたエントリー(例: 101m)の各概要ページを見ると、ページ下の方に UniProtへのリンクがある。このリンク先を見て、この蛋白質に関連する情報を調べてみる。
- PDBj各エントリーの「機能情報」(Functional Details) ページに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のみ)、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のみ)、
struct_ref

演習課題16

詳細検索 (Advanced Search) の「リガンドと補欠分子族」 (Ligands and Prosthetic groups) に「ATP」と入力して検索する。

- 検索結果からいくつかのエントリーについて概要ページを確認し、3次元構造ビューアページでどこに ATP が結合しているか確認する。
- 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

詳細検索 (Advanced Search) の「ポリマー鎖の数/長さ」 (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

- 詳細検索 (Advanced Search) の「実験手法」(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。
 - 検索結果の分子静止画像をぱっと見て、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: entity, entity_poly, entity_poly_seq, pdbx_poly_seq_scheme

演習課題20

- 詳細検索 (Advanced Search) の「由来する生物種」(Source、その分子の遺伝子源生物) に生物種名を入れて検索。
例: human, bos taurus, pyrococcus など
- 「宿主生物種」(Host species、その分子を発現させた生物) に生物種名を入れて検索する。
- 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