

## Protein Data Bank データベース

医療技術短期大学部 中島 廣志

金沢大学計算機に、CC0892.PDB.DATAという区分データセット名で公開登録している Protein Data Bank (PDB) データベースについて紹介する。

PDB データは、X線結晶構造解析に基づく生体高分子(蛋白質、核酸、糖など)の立体構造データを集録したデータベースであり、約250個の蛋白質三次元座標データが主体となっている。このデータベースと、以前に登録した蛋白質1次構造データベース、核酸塩基配列データベースとの関連についてもふれる。

### 1. PDB データ

PDB データは、米国ブルックヘブン国立研究所で、集収、編集され、世界に配布、利用されている。日本においては、大阪大学蛋白質研究所の安岡則武先生がPDBデータの窓口となっております<sup>1)</sup>。データは80字固定長のカードイメージで、一定FORMATのもとに記載されている<sup>2)</sup>。利用法としては、印字出力による利用や、プログラムによるデータ処理利用法があり、PDBデータ解析用として、すでに多くのプログラムが報告されている<sup>3), 4)</sup>。PDBデータには、解析プログラム、データ目録、立体座標データが含まれているが、ここでは立体座標データのみ登録をおこなっている。

個々のデータには、数字で始まる4文字のPDB識別名がつけられている。ここでは、この識別名を@@の記号ではさみメンバー名に用いている。例えば、蛋白質アズリンの立体座標データには、1AZUの識別名がついているので、このデータはメンバ名@1AZU@に登録されている。PDBデータ利用の便宜上、メンバー名とデータ内容の一覧を#MENU#に示した。図1に、CC0892.PDB.DATA(#MENU#)の一部を示す。PDBデータ利用の際、このメニューが役立つものと思う。

### 2. 関連データベースについて

金沢大学計算機にはすでに、蛋白質1次構造データベース<sup>5)</sup>、核酸塩基配列データベース<sup>6)</sup>が登録公開されており、ここに蛋白質立体構造データベースが追加されたことにより、蛋白質関係のデータがかなり充実したと思われる。蛋白質関係データベースの名称と内容を図2に示す。

MENU OF CC0892.PDB.DATA  
PDB REPRESENTS PROTEIN DATA BANK

UPDATED 04/25/85

```

*****
NO CODE   PROTEIN NAME
          AUTHORS (DEPOSITORS)          UPDATED DATE STATUS
*****
1 2APE   ACID PROTEINASE (ENDOTHIA PARASITICA)
          T. BLUNDELL                      9/81
2 2APP   PROTEINASE (PENICILLIUM JANTHINELLUM)
          A. SIELECKI, M. JAMES           1/83
3 1APR   ACID PROTEASE (PHIZOPUS CHIMENSIS)
          D. DAVIES                        8/79
4 2ACT   ACTINIDIN
          E. BAKER                          11/79 R
5 1ACX   ACTINOXANTHIN
          V. PLETNEY, A. KUZIN            12/82 R
6 2ADK   ADENYLATE KINASE (PORCINE MUSCLE)
          G.E. SCHULZ AND CO- WORKER      3/77 R
          :
          :
  
```

図1 PDBデータリストの一部

データ内容	データベース名
DNA Sequence	CC0892.KAKUSAN.DATA
RNA Sequence	CC0892.KAKUSAN.DATA
Protein Sequence	CC0892.AMINO.DATA
Protein Structure	CC0892.PDB.DATA
DNA Structure	CC0892.PDB.DATA

図2 蛋白質に関する登録データベースの名称と内容

AMINO.DATA, KAKUSAN.DATAにおいて、以前と比べデータFORMATを変更したので以下に記す。KAKUSAN.DATAに関しては、塩基配列記述において右に4カラム移動をおこなった。このFORMATは京都大学化学研究所の大井龍夫教授によって設定されDDBJ (DNA Data Bank of JAPAN) FORMATと呼ばれている。

AMINO.DATAについて、DDBJ FORMATを採用し、全面的に書き直した。また、他のデータベースとの関係を表わす目的で、NBRF, PDB, DDBJ等の項目を追加し、それぞれのデー

データベースの識別コードをつけた。NBRF は、National Biomedical Research Foundation を表わし、米国 NIH の蛋白質1次構造データベースを意味する。AMINO.DATA には、NBRF のデータベースも1つのメンバーとして登録されている。AMINO.DATAのデータ例を図3に示す。

KAKUSAN.DATA , AMINO.DATAそれぞれ、PDB.DATA 同様メニューによりデータ内容を簡単に知ることができるよう整備をおこなった。

```

NAME 311 ASPARTATE TRANSCARBAMOYLASE C CHAIN UPDATED 10/17/84
SOURCE ESCHERICHIA COLI, PYR B GENE PRODUCT
NBRF DTECC
PDB DATA 2ATC
DDBJ ECPYRBI
GENBANK ECPYRBI
ACCESSION J01668
REFERENCE 1
AUTHORS PAUSA,C.D., KARELS,M.J., NAVRE,M. AND SCHACHMAN,H.K.
TITLE GENES ENCODING ESCHERICHIA COLI ASPARTATE TRANSCARBAMOYLASE:
THE PYR B-PYR I OPERON
JOURNAL PROC NAT ACAD SCI USA: 79, 4020-4024 (1982)
REFERENCE 2
AUTHORS HOOVER,T.A., ROOF,W.D., FOLTERMANN,K.F., O'DONOVAN,G.A.,
BENCINI,D.A., AND WILD,J.R.
TITLE NUCLEOTIDE SEQUENCE OF THE STRUCTURAL GENE (PYRB) THAT
ENCODES THE CATALYTIC POLYPEPTIDE OF ASPARTATE
TRANSCARBAMOYLASE OF ESCHERICHIA COLI
JOURNAL PROC. NATL. ACAD. SCI. USA 80, 2462-2466 (1983)
REFERENCE 3
AUTHORS KONIGSBERG,W.H. AND HENDERSON,L.
JOURNAL PROC. NATL. ACAD. SCI. USA 80, 2467-2471 (1983)
COMMENT CATALYTIC SUBUNIT(C CHAIN)
ASPARTATE TRANSCARBAMYLASE, EC 2.1.3.2
DODECAMER, TWO CATALYTIC TRIMERS(C3) AND THREE REGULATORY
DIMER(R2)
SITES FROM TO DESCRIPTION
MATP 2 311 MATURE PROTEIN
SEQUENCE 311 AA
MANPLYQKHIISINDLSRDDLNVLATAAKLKANPPELLKHKVIASCFFEASTRTRLSF 60
QTSMHRGASVVGFSANTSGLGKGETLADTISVISTYDAIVMRHPQEGAARLATEFS 120
GNVPVLNAGDGSNQHTQTLLDLFTIQETQGRDLNLHVAMVGDLLKTRTVHSLTQALAKF 180
DGNRFYFIAPDALAMPQYILDMLDEKGIAWSLHSSIEEVMVEVDILYMRVQKERLDPSE 240
YANVKAQFVLRASDLHNAKANMKVHLHPLPRVDEIATDVDKTPHAWYFQQACNGIFARQAL 300
LALVLRDLVL 311
//

```

図3 蛋白質1次構造データの例

## おわりに

数種のコマンドをマスターすれば、簡単にデータベースが利用できます。データベースが多くの人に利用されることを期待する。

## 文 献

- 1) 安岡則武：生物物理, 23, 248-252 (1983) .
- 2) Bernstein, F.C., Koetzle, T.F., Willams, G.J.B., Meyer, E.F. Jr., Brice, M.D., Rogers, J.R., Kennand, O., Shimanouchi, T. and Tasumi, M. : J.Mol.Biol., 112, 535-542 (1977) .
- 3) 郷 信広, 水野裕重, 郷 通子, 宮澤三造, 武富 敬：蛋白質・核酸・酵素, 23, 1336-1344 (1978) .
- 4) 中島広志, 岡本克稔：金沢大学医療技術短期大学部紀要, 5, 79-84 (1981) .
- 5) 中島広志：金沢大学計算機センター広報, 7, 55-60 (1983) .
- 6) 中島広志：金沢大学計算機センター広報, 8, 123-132 (1984) .