

メニュー画面方式による蛋白質・核酸データベースの利用

医療短大 中島 廣志・情報処理センター 高田 良宏

メニュー画面方式による蛋白質・核酸データベースの利用システムを作成したので紹介する。現在、金沢大学情報処理センターの計算機に、(1)蛋白質のアミノ酸配列、(2)核酸塩基配列、(3)生体高分子立体構造座標データの3種の生体高分子データベースが登録されている。これらのデータベースは、計算機により初めて利用可能となる。利用に際しての説明書^{1,2)}もあり、慣れると便利であるが慣れるまでが大変であり、計算機に不慣れな場合、何となく抵抗を感じ利用をためらうのではと思う(筆者の一人H. Nがそうであった)。そこで、計算機を初めて使う人でも簡単に利用できるようにメニュー画面方式を作成したので利用していただきたい。

1. 使用方法

端末のスイッチを入れ計算機を利用可能状態にする。このとき

LOGON TSS AB9999/ パスワード(1)

とタイプし、**ENTER**キーを押せばよい(これを入力という)。ここで、AB9999 とパスワードは利用者の課題番号とパスワードである。(1)が正常に働くとREADYと表示され、利用可能状態となる(これをREADYモードという)。何かおかしいときは(1)のタイプをやり直す。

図1にメニュー画面システムの概略を示す。このシステムを起動させるとき、READYモードで

EX 'AB2535.IDENSI.CLIST(MENU)'(2)

と入力すると蛋白質・核酸データベース利用メニュー画面(画面1)が表示される。

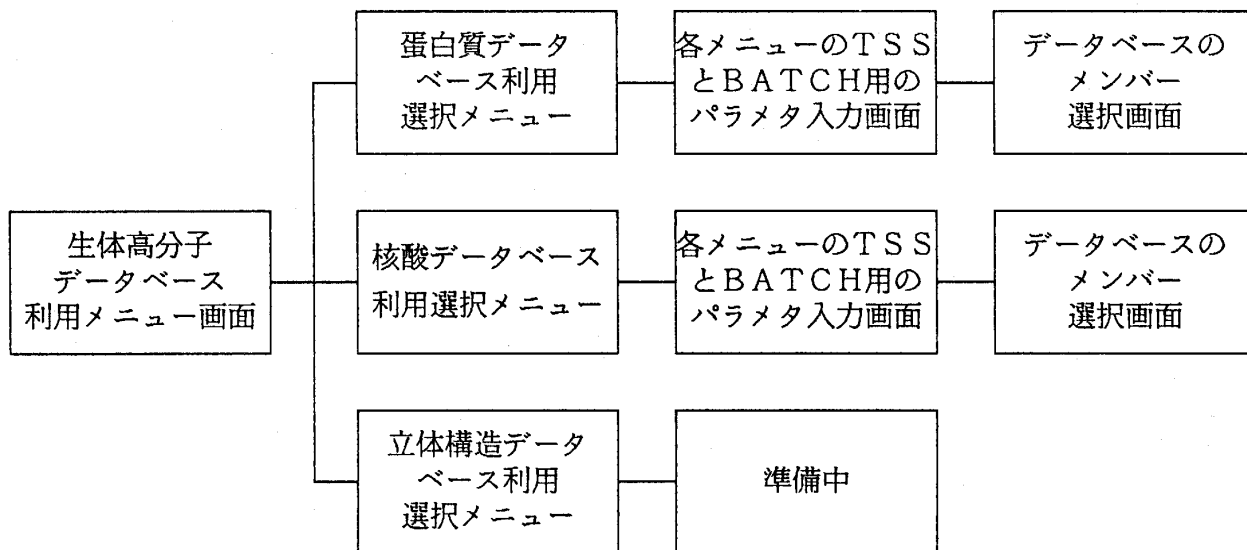


図1 メニュー画面による生体高分子利用システムの概略

-----<蛋白質・核酸データベース利用メニュー 画面-1 >----- MAIN -----
オプション ==> A

-----<データベース利用 オプションメニュー>-----

- A PROTEIN - 蛋白質アミノ酸配列データベース
- B DNA - 核酸塩基配列データベース
- C PDB - PROTEIN DATA BANK (立体構造)

-----< PFD オプションメニュー > -----

- 0 ATTRIBUTES, 1 BROWSE, 2 EDIT, 3 UTILITY,
- 4 FOREGROUND, 5 BACKGROUND, 6 TSS, 7 TEST,
- H HELP, X EXIT

ユーザID - AB9999
パスワード -
時刻 - 10:08
PFキー - 24
第一修飾子 - AB9999

終了する時、ENDキー (PF3) を押して下さい。

画面1 蛋白質・核酸データベース利用メニュー画面

-----<蛋白質アミノ酸配列データベース >----- MPRO -----

番号選択 ==> 3
実行モード ==> T T: TSS, B: BATCH

-----<データベース利用 選択メニュー>-----

- 1 AFIND1 - 指定したアミノ酸配列の検索 1
- 2 AHOMO1 - アミノ酸配列のホモロジー検索 1
- 3 AHOMO2 - アミノ酸配列のホモロジー検索 2 (高速)
- 4 ALI - アミノ酸配列データの蛋白質名のリスト
- * HARR - HARR PLOT 1
- * HPATHY - HYDROPATHY PLOT
- * PLOT1 - アミノ酸配列に対するデータのPLOT 1
- * PP2ND - 蛋白質2次構造の予測
- * PFTYPE - アミノ酸組成より蛋白質構造型の予測
- * 印 - 準備中

終了する時、ENDキー (PF3) を押して下さい。

画面2 蛋白質データベース利用選択メニュー画面

-----< ホモロジー検索2, (高速) AHOMO2 >----- TPRO03 -----
コマンド=> GO GO,RUN:実行依頼 SAVE:コマンドの保存
U:データセット編集 B:データセット検索

ホモロジー検索データ名 => 'AB0408.KEEP.DATA(AHOMO2)' DISK

(例 : 'AB0408.KEEP.DATA(AHOMO2)')

得点 => -150 100残基で-60ぐらい

検索するアミノ酸配列の位置

最初の残基番号 => 1

省略可能: 省略時は1

最後の残基番号 => 100

省略可能: 省略時は残基数

出力(FT06P001) =>

出力クラス(C,A), プリンター名, データセット名, 省略時は端末

次はデータベースからメンバー名の選択画面です

コマンド保存データ・セット名==>

メンバ名==> TPRO03

終了時は END キーを押す

画面3 TSS用ホモロジー検索2のメニュー画面

この画面の ==> の所に PROTEIN の A を入力すると、蛋白質データベース利用選択メニュー画面 (画面2) に進む。このシステムでは、メニュー画面上の入力はすべて ==> の所で行う。もし、入力が適切でない場合は、その旨のメッセージが表示され先に進まない。

いま、アミノ酸配列のホモロジー検索2 (高速) を行うとき、画面2の番号選択欄と実行モード欄に、3とTを入力するとTSS用ホモロジー検索2のメニュー画面 (画面3) に進む。この画面で、必要なパラメータを指定する。検索データとして AB0408.KEEP.DATA (AHOMO2) の様なデータ (図2) を用意する。この様なデータの準備が一番大変な仕事になるかと思う。

図2のデータは、データベースからコピーしたものであるが、必要なのは1行目の残基数と蛋白質名、2行目のSOURCE、20行目のSEQUENCEと21~24行のアミノ酸配列データである。

ここで、画面3の様にパラメータをタイプし、**ENTER**キーを押すと、蛋白質データベースからメンバー名の選択画面 (画面4) へと進む。いま画面4において@ NBRFの前にSと入力すると計算が始まり、しばらくすると結果が端末に出力される (図3)。

端末のブラウン管が一杯になったとき左下に***が表示される、このとき**ENTER**キーを押すと次のページに進む。また、計算を途中で中止したいとき、**DUP PA1**のキー (ブレイクキーとも言う) を押せばよい。画面4で、@ NBRFを選んだが、個々のメンバーについての簡単な説明をメンバー名

日本語EDIT --- ABO408.KEEP.DATA(AHOMO2) - 01.00 ----- 表示欄 001 072
コマンド ==> 移動量 ==> HALF

```
***** ***** データの先頭 *****V10L30*****
000001 NAME 186 RAT PREPRORELAXIN UPDATED 11/17/84
000002 SOURCE RAT (RATTUS NORVEGICUS)
000003 NBRF RXRT
000004 DDBJ RTPRLAX
000005 GENBANK RATRELAX
000006 ACCESSION J00780
000007 REFERENCE 1
000008 JOURNAL FEBS LETT. 129,80(1981)
000009 REFERENCE 2
000010 AUTHORS HUDSON,P., HALEY,J., CRONK,M., SHINE,J. AND NIALL,H.
000011 TITLE MOLECULAR CLONING AND CHARACTERIZATION OF CDNA SEQUENCES
000012 CODING FOR RAT RELAXIN
000013 JOURNAL NATURE: 291, 127-131 (1981)
000014 COMMENT N-TERM; (GLN-163) IS PYPROLIDONE CARBOXYLIC ACID
000015 SITES FORM TO DESCRIPTION
000016 PEPT 1 186 PRECURSOR
000017 SIGP 1 22 SIGNAL PEPTIDE
000018 MATP 1 23 57 RELAXIN B-CHHIN
000019 MATP 2 163 186 RELAXIN A-CHHIN
000020 SEQUENCE 186 AA
000021 MSSRLLQLLGFWLFSLQPCRRARVSEBWMQVIQVCGRGYARAWIEVCGASVGRLLALAQE 60
000022 EPAPLARQATAEVVPSFINKDAEPPDMLKCLPNLSEERKAALSEGRAPPPELQQHAPAL 120
000023 SDSVVSLEGFKKTFHNQLGEAEDGGPPELKYLGSDAQSRKKRQSGALLSEQCCHIGCTRR 180
000024 SIKLC 186
000025 //
```

図2 蛋白質ホモロジー検索データ

MENU #に準備してあるので、このファイルを一度見ておくことをおすすめする。

図3は、ブラウン管のハードコピー（**PRINT**キーを押し印刷させたもの）であるが、データ@NBRF の5251個の蛋白質において、ホモロジーの条件-150点以下の蛋白質2個の存在とその相同性を表示している。このときの計算時間は38秒であった。前の画面3において、出力欄の所にC又はAを指定すると、情報処理センターの1階又は2階の日本語ラインプリンタからの出力となり、分室のプリンタ名を入力すると、そのプリンタからの出力となる。データセットに出力したいときは、データセット名（例えば A.OUTLIST の様な名称）を指定すればよい。画面3のコマンド欄にU又はBとすると、それぞれ EDIT と BROWSE に相当し、データの編集、検索等が行える。

ENTER キーを押すと実行されます。とり止める時は、PF3 キーを押す。
出力したいMEMBER名の前にSと入力してください。いくつでもかまいません。
ただし次のメンバーは配列データではありません。

MENU	ADDED	CLNBRF	FEATURE	KEYWORD	REVISED
#MENU#	@ADDED	@CLNBRF	@DBIP	@DBIPN	
@FEATURE	@KEYWORD	<u>S</u> @NBRF	@NEWNBRF	@PGTRANS	
@REVISED	@SEQ2ND	ANTIGEN	AZUPLA	BASE1	
BASE2	BASE3	COAT	COLLA	CYT	
DNABIND	DOXIN	EC11	EC16	EC21	
EC27	EC31	EC32118	EC34	EC35	
EC456	HAGLU	HB	HIST	HORMONE1	
HORMONE2	IG	INHIB	INSULIN	INTER	
LENS	LIGHT	LIPO	MB	MUSCLE	
ONCO	PHAGE	POLYPEP	RIBO	TOXIN	
UNDEF1	UNDEF2	UNDEF3	UNDEF4	VIRUS1	
VIRUS2	VIRUS3				

画面4 蛋白質データベースのメンバー選択画面

画面2の蛋白質データベース利用選択メニューにおいて、3とBを入力すると画面5のBATCH用ホモロジー検索2のメニュー画面となる。この画面5の入力は先のTSS用の画面3と同様に行う。画面5のコマンド欄にEDITと入力すると、メンバー名選択画面(画面4)を経て図4のJOB制御文の編集画面となる。図4において、検索データ(25行目)、データベース(27行目)、検索条件(22行目)の確認を行うことができる。実行するとき、図4のコマンド欄の所にSUBと入力する。

```

FORTRAN 77 COMPILER ENTERED
END OF COMPILATION
QUERY PROTEIN SEQUENCE
186  RAT PREPRORELAXIN                                RAT (RATTUS NORVEGICUS)
INPUT SEQUENCE
  *  +  *  +  *  +  *  +  *  +  *  +  *  +  *
+  *  +  *  +
MSSRLLLQLLGFWLFSLQPCRARVSEEWMDQVIQVCGRGYARAWIEVCGASVGRALALSQEEPAPLARQATAEVVPSFIN
KDAEPFDMTLKCLPNLSEERK 100
AALSEGRAPPELQQHAPALSDSVSLEGFKKTFFHNQLGAEDEGGPELKYLGSDAQSRKKRQSGALLSEQCCHIGCTR
RSIAKLC 186
ANALYSIS FROM 1 TO 100 RESIDUES
  *  +  *  +  *  +  *  +  *  +  *  +  *  +  *
+  *  +  *  +
MSSRLLLQLLGFWLFSLQPCRARVSEEWMDQVIQVCGRGYARAWIEVCGASVGRALALSQEEPAPLARQATAEVVPSFIN
KDAEPFDMTLKCLPNLSEERK 100
ANALYSIS OF LOCAL HOMOLOGOUS REGIONS
CUT OFF SCORE -150
1) RXRT 186 PRORELAXIN PRECURSOR - RAT
HOMOLOGOUS REGION  QUERY SEQUENCE FROM 1  LIBRARY SEQUENCE FROM 1 RES
IDUE
HOMOLOGY SCORE (UNCORRECTED) -526 ( -532)
ALIGNED LENGTH 100 MATCH 100.0 % ( 100/ 100)
***
  +  *  +  *  +  *  +  *  +  *  +  *  +  *
*  +  *  +  *
MSSRLLLQLLGFWLFSLQPCRARVSEEWMDQVIQVCGRGYARAWIEVCGASVGRALALSQEEPAPLARQATAEVVPSFIN
KDAEPFDMTLKCLPNLSEERK 100
:
:
MSSRLLLQLLGFWLFSLQPCRARVSEEWMDQVIQVCGRGYARAWIEVCGASVGRALALSQEEPAPLARQATAEVVPSFIN
KDAEPFDMTLKCLPNLSEERK 100

2) RXPG 182 PRORELAXIN PRECURSOR - PIG
HOMOLOGOUS REGION  QUERY SEQUENCE FROM 3  LIBRARY SEQUENCE FROM 2 RES
IDUE
HOMOLOGY SCORE (UNCORRECTED) -222 ( -173)
ALIGNED LENGTH 99 MATCH 45.5 % ( 45/ 99)
  +  *  +  *  +  *  +  *  +  *  +  *  +  *
*  +  *  +  *
SRRLLLQLLGFWLFSLQPCRARVSEEWMDQVIQVCGRGYARAWIEVCGASV-GRLALSQEEPAPLARQATAEVVPSFINK
DAEPFDMTLKCLPNLSEERK 99
:
:
:
PRLFSYLLGVWLLLSQLPRE-IPGSTNDFIKACGRELVRLWVEICGSVSWGRTALSLEEQ-LETGPPAETMPSSITK
DAEILKMMLEFVPLNPQELK 99

***

```

図3 ホモロジー検索の結果

-----< HOMOLOGY 検索2(FAST), AHOMO2 >----- BPR003 ----
コマンド=> EDIT SUB:バッチ依頼 SAVE:制御文の保存 EDIT:制御文の編集

ホモロジー検索データ名 => AB0408.KEEP.DATA(AHOMO2) DISK

(例 : 'AB0408.KEEP.DATA(AHOMO2)')

得点 => -50 100残基で-60ぐらい

検索するアミノ酸配列の位置

最初の残基番号 =>

省略可能: 省略時は1

最後の残基番号 =>

省略可能: 省略時は残基数

//AB9999? JOB ,CLASS=B,MSGCLASS=X,REGION=2048K

//*

//*

ジョブ制御文保存データ・セット名==>

メンバ名==> BPR003

終了時は END キーを押す

画面5 BATCH用ホモロジー検索2のメニュー画面

実行終了後、結果を印刷させる。

以上、このシステム利用の一例を示したが、他のメニューについての利用も同様である。

ここで整備したメニューは登録されているプログラムの一部であり、メニュー以外の多くの項目についてプログラムが登録され利用可能となっている²⁾。立体構造データベース利用については、データベースからメンバー選択に困っている所である。

メニュー画面方式を作動させるソフトは、課題番号AB2535のファイルに、実行に必要なプログラムやデータはAB0408のファイルに登録されているので、興味のある人は利用していただきたい。

```

EDIT --- AB9999.KPFD.OUTLIST(BPRO03) ----- 表示欄 001 072
コマンド ==> SUB                               移動量 ==> CUR
***** ***** データの先頭 *****V10L30*****
000100 //AB9999B JOB ,CLASS=B,MSGCLASS=X,REGION=2048K,
000200 //      PASS=パスワード
000300 //*
000400 //*
000500 //*      KANAZAWA UNIV. KPFD SUBMITO V10L11
000600 //*
000700 //      EXEC FORT7CLG
000800 /*JOBPARM L=200
000900 /** AHOMO2
001000 /** PROTEIN HOMOLOGY SEARCH 2 (FAST)
001100 //FORT.SYSIN DD DISP=SHR,DSN=AB0408.MY.FORT77(AHOMO2)
001200 //      DD DISP=SHR,DSN=AB0408.MY.FORT77(ISEQ1)
001300 //      DD DISP=SHR,DSN=AB0408.MY.FORT77(SEQP)
001400 //      DD DISP=SHR,DSN=AB0408.MY.FORT77(DIF23)
001500 //      DD DISP=SHR,DSN=AB0408.MY.FORT77(PSDB3)
001600 //      DD DISP=SHR,DSN=AB0408.MY.FORT77(INTEG2)
001700 //      DD DISP=SHR,DSN=AB0408.MY.FORT77(PSEQ2)
001800 //      DD DISP=SHR,DSN=AB0408.MY.FORT77(ALIGN2)
001900 //      DD DISP=SHR,DSN=AB0408.MY.FORT77(HIST1)
002000 /** SCORE, SPAN (FROM, TO)
002100 //GO.FT05F001 DD *
002200 -50 1 50
002300 /*
002400 /** QUERY SEQUENCE **
002500 //GO.FT01F001 DD DISP=SHR,DSN=AB0408.KEEP.DATA(AHOMO2)
002600 /** PROTEIN DATABASE **
002700 //GO.FT02F001 DD DISP=SHR,DSN=CC0892.AMINO.DATA(@SEQ2ND)
002800 //

***** ***** データの末尾 *****

```

図4 JOB制御文

文 献

- 1) 利用の手引き, 核酸及び蛋白質のデータベースの使用法, MNLW01, 中島廣志, 三木直正
- 2) 核酸及び蛋白質のデータ処理 (TSS用), 三木直正