

九州大学 大型計算機センターニュース

No. 302

福岡市東区箱崎6丁目10番1号
九州大学大型計算機センター
広報教育室(TEL092-641-1101)
内線 2505

目 次

- | | |
|--|---|
| 1. データベースシステムGENASへの蛋白質データの追加…………… | 1 |
| 2. 統計データ処理パッケージANALYSTのバージョン
アップについて…………… | 6 |

1. データベースシステムGENASへの蛋白質データの追加

GENAS [1] のデータベースは、これまで European Molecular Biology Laboratory (西独) の EMBL Nucleotide Sequence Data Library から作られてきましたが、10月1日から、これに National Biomedical Research Foundation, Georgetown University Medical Center (米国) の Atlas of Protein Sequence and Structure のデータが加わり、核酸塩基配列と蛋白質のアミノ酸配列が一緒に検索・解析できるようになります。これに伴い、下記のように使用法が若干変る部分や使用上の制約が生じました。

情報検索型コマンド系を用いる場合

- 1) DISPコマンドのオペランドにSEQが指定できるようになります。これまで、塩基配列を表示するには、非情報検索型コマンド系あるいは応用プログラムSEQLIST及びSCANによってしかできませんでしたが、促進メッセージ". "のあと、

D SEQ

と入力することによって、塩基配列又はアミノ酸配列が表示されます。アミノ酸の略記法は、Table 1に、アミノ酸間の記号の意味はTable 2にあります。

- 2) DISPコマンドにオペランドBIとSEQを指定するときは、他のオペランドと一緒に置くことができなくなります。これまでの、DISPコマンドは、

D DEF BI

のように入力することができましたが、これが禁止されます。もちろん、BIとSEQ以外については、これまでと同じです。

- 3) DEF(配列の定義)に塩基配列かアミノ酸配列かを区別する文字列が入りました。例えば、

D DEF

と入力したとき

DEF=(NUCLEOTIDE)……

と出ますと、塩基配列であり、

DEF=(PROTEIN)……

と出ると、アミノ酸配列です。

非情報検索型コマンド系を用いる場合

非情報検索型コマンド系そのものは、変更がありませんが、このコマンド系を使うには、データベースの定義情報が必要なので、新しい定義情報を Fig. 1 に示します。

応用プログラムを用いる場合

GENAS の応用プログラムのうち、蛋白質に対して適用できるものは次の 9 個です。

AACOM, AAFIT, AASERK, FMIP3, HARR1, HOMO, SEARCHX, SEQFIT, SEQLIST.

なお、現在のデータ量は、核酸がリリース 3.0 からの 1,481 配列で、蛋白質はリリース 2.1 からの 2,784 配列です。

参考文献

1. 久原ほか 核酸塩基配列データベースシステム GENAS の使用法(1), 九大大型計算機センター広報, 16, 5, 1983, 497-521.

Table 1
One- and Three-letter Amino Acid Abbreviations

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
B	Asx	Asp or Asn, not distinguished
Z	Glx	Glu or Gln, not distinguished
X	X	Undetermined or atypical amino acid

These abbreviations conform to those suggested by the IUPAC-IUB Commission on Biochemical Nomenclature, J. Biol. Chem. 243, 3557-3559, 1968.

Table 2
Punctuation in Protein Sequences

- A blank between two adjacent amino acids indicates that they are connected, as determined experimentally.
- () Encloses a region, the composition but not the complete sequence of which has been determined experimentally, or encloses a single residue that has been tentatively identified.
 - = Indicates)(, the juxtaposition of two regions of indeterminate sequence, while preserving proper spacing between residues.
 - / Indicates that the adjacent amino acids are from different peptides, not necessarily connected. When the amino end of a protein has not been determined, / precedes the first residue. When the carboxyl end has not been determined, / follows the last residue. When)/, /(, or)/(are needed, only / is used.
 - . Outside of parentheses, indicates the ends of sequenced fragments. The relative order of these fragments was not determined experimentally but is clear from homology or other indirect evidence.
 - . Within parentheses, indicates that the amino acid to its left has been placed with at least 90% confidence by homology with known sequences.
 - , Indicates that the amino acid to its left could not be positioned with confidence by homology. If the structure of related proteins is not known, the position of the amino acids within parentheses is arbitrary.

```

*
DEF(REF < RF*ID)
  L('REFERENCE CODE')
  U(PUBDB)
*
DEF(RNA < RF)
  L('RNA OR CDNA')
  U(PUBDB)
*
DEF(FR < RF)
  L('FRAGMENT')
  U(PUBDB)
*
DEF(DE : RF -> DEF)
  L('DEFINITION')
  U(PUBDB)
*
DEF(OS < RF*GEN*SPEC)
  L('ORGANISM')
  U(PUBDB)
*
DEF(OC < RF*LVL*TAX)
  L('ORGANISM CLASSIFICATION')
  U(PUBDB)
*
DEF(HS < RF*GEN*SPEC)
  L('THE MOST COMMON HOST SPECIES')
  U(PUBDB)
*
DEF(HC < RF*LVL*TAX)
  L('HOST CLASSIFICATION')
  U(PUBDB)
*
DEF(BIB : RF*SN -> BI)
  L('BIBLIOGRAPHIC INFORMATION')
  U(PUBDB)
*
DEF(KY < RF*KW)
  L('KEYWORD')
  U(PUBDB)
*
DEF(AUTH < RF*AU)
  L('AUTHOR')
  U(PUBDB)
*
DEF(YEAR < RF*YR)
  L('YEAR')
  U(PUBDB)
*
DEF(FN : RF -> NFTE)
  L('NUMBER OF FEATURE TABLE ENTRIES')
  U(PUBDB)
*

```

Fig. 1. Definition of GENEDB.

```

DEF(FT : RF*FTEN -> KEY*FROM*TO*DESCR)
  L('FEATURE TABLE')
  I(KEY, FROM, TO)
  U(PUBDB)
*
DEF(SH : RF -> A*C*T*G*U*Y*R*AB)
  L('SEQUENCE HEADER')
  U(PUBDB)
*
DEF(LN : RF -> LEN)
  L('LENGTH OF SEQUENCE')
  U(PUBDB)
*
DEF(SQ : RF*SN -> SEQ)
  L('SEQUENCE')
  U(PUBDB)
*
DEF(RF)      KEY
              T(I4)      L('REFERENCE CODE')
*
*** PROTEIN IF RF > 1,000,000,000; OTHERWISE NUCLEOTIDE.
*
DEF(ID)      T(C8)      L('SHORT IDENTIFICATION')
*
DEF(GEN)     T(C100)   L('GENUS')
DEF(SPEC)    T(C50)    L('SPECIES')
*
DEF(LVL)     T(I2)     L('LEVEL IN TAXONOMIC TREE')
DEF(TAX)     T(C50)    L('ENTRY IN TAXONOMIC CLASSIFICATION')
*
DEF(DEF)     T(C600)   L('DEFINITION WRITTEN IN ENGLISH')
DEF(BI)      T(C2000)  L('BIBLIOGRAPHIC INFORMATION')
DEF(KW)      T(C50)    L('KEY WORD')
DEF(AU)      T(C50)    L('AUTHOR')
DEF(YR)      T(I2)     L('YEAR')
*
DEF(NFTE)    T(I2)     L('NUMBER OF FEATURE TABLE ENTRIES')
DEF(FTEN)    T(I2)     L('FEATURE TABLE ENTRY NUMBER')
DEF(KEY)     T(C8)     L('FEATURE TABLE KEY NAME')
DEF(FROM)    T(I4)     L('ENDPOINT SPECIFICATION')
DEF(TO)      T(I4)     L('ENDPOINT SPECIFICATION')
DEF(DESCR)   T(C200)   L('MORE INFORMATION ABOUT FEATURE')
*
DEF(A)       T(I4)     L('NUMBER OF ADENINE BASES')
DEF(C)       T(I4)     L('NUMBER OF CYTOSINE BASES')
DEF(T)       T(I4)     L('NUMBER OF THYMINE BASES')
DEF(G)       T(I4)     L('NUMBER OF GUANINES')
DEF(U)       T(I4)     L('NUMBER OF URACIL BASES')
DEF(Y)       T(I4)     L('NUMBER OF PYRIMIDINE BASES')
DEF(R)       T(I4)     L('NUMBER OF PURINE BASES')
DEF(AB)      T(I4)     L('NUMBER OF ANY BASES')
*
DEF(LN)      T(I4)     L('LENGTH OF SEQUENCE')
*
DEF(SN)      T(I4)     L('SEQUENTIAL NUMBER')
DEF(SEQ)     T(C500)   L('NUCLEOTIDE/PROTEIN SEQUENCE')

```

Fig. 1. (Continued)

(データベース室 電話(内線) 2510)

2. 統計データ処理パッケージANALYSTのバージョンアップについて

9月26日(水)のANALYST講習会の日より、このパッケージをV10/L10にバージョンアップします。これに伴ない、起動コマンドがANALYSTから%ANALYSTへ変更されますのでご注意ください。また、今回のバージョンアップでマニュアルも全部新版となりましたので、以下の機能追加その他の詳細については、文献[1]またはマニュアル[2, 3, 4, 5, 6, 7]を参照して下さい。

主な変更点

- マクロの呼出し……………%マクロ名 から /マクロ名 に変更。
- ANALYSTの中から特定のTSSコマンドを直接呼出しできる。
ALLOCATE, ATTRIB, DSPRINT, FREE, LIST, PFD, LISTALC, LISTBC, LISTCAT, LISTDS, LOGOFF, PROFILE, SEND, SORP, TERMINAL, TIME
- 追加されたコマンド及び機能がサポートされたコマンド
データ入力/定義……MERGE, COMBINE, JOIN, PUTLABEL, GETLABEL, CONTENTS, DISPLAY, WAIT, ERASE, ADDVAR, ADDCASE, CHECK, CODE, MCODE, NCODE, RENAMEVAR
集計/作表……………FREQ, TFREQ
統計解析……………GTEST
- 廃止されたコマンド
集計/作表コマンドのうちCROSSコマンドが廃止され、その機能はMCROSSコマンドに統合。

参考文献

1. 石水, 桜井, 平野, 池田, 菅崎, 武富 統計データ処理パッケージANALYSTの使用について, 九大大型計算機センター広報, 17, 4-5, 1984.
2. 計算機マニュアル FACOM ANALYST解説書(99SG-7400-1), 富士通株.
3. 計算機マニュアル FACOM ANALYST/BASEコマンド説明書(99SP-7420-1), 富士通株.
4. 計算機マニュアル FACOM ANALYST/TABULATEコマンド説明書(99SP-7430-1), 富士通株.
5. 計算機マニュアル FACOM ANALYST/ANALYZEコマンド説明書(99SP-7440-1), 富士通株.
6. 計算機マニュアル FACOM ANALYSTメッセージ説明書(99SP-7480-1), 富士通株.
7. 計算機マニュアル FACOM OS IV ANALYST運用手引書(70SP-7410-1), 富士通株.

(ライブラリ室 電(内)2523)