

Title	蛋白質および核酸データベースの使用法 : FROG system
Author(s)	三木, 直正
Citation	大阪大学大型計算機センターニュース. 1992, 85, p. 2-7
Version Type	VoR
URL	https://hdl.handle.net/11094/65963
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

蛋白質および核酸データベースの使用法

— F R O G s y s t e m —

三木直正
阪大医学部第一薬理
06-875-7330

I、データベースについて

1970年代後半にDNAの塩基配列決定法が開発されて以来、今日まで、膨大な量の核酸および蛋白質の一次配列のデータが蓄積されている。これら核酸と蛋白質にデータは、本として出版しても役にたたず、コンピュータにデータベースとして構築保存され、かつこれを利用できるソフトが開発され、はじめて有用となる。現在、利用できる核酸のデータベースは、米国のGenBankと、ヨーロッパのEMBLおよび我国のDDBJなどがある。一方、蛋白質のデータベースは、米国のNBRF(PIR)、ヨーロッパのSwissProtや蛋白質奨励会のPRFなどがある。エントリー数は毎年増加し続けている。核酸および蛋白質のデータとも巻物の様に順番(順ファイル)に並んでおり、検索はデータベースの頭の部分から、一行ずつ最後まで順次行うようになっており、検索には時間がかかる。したがって、ホモロジー検索には高速コンピュータが必要である。

阪大大型計算機センターのACOS2000に入っているデータベースは

- 1) Genbank (核酸データベース) ----- 14項目に分かれている。
PRI. SEQ, ヒト、サル; ROD. SEQ, マウス、ラット; MAM. SEQ, ウシ、ウサギ;
VRT. SEQ, トリ ; INV. SEQ, ハエ ; PLT. SEQ, 植物 ;
ORG. SEQ, ミトコン ; BCT. SEQ, 細菌 ; RNA. SEQ, R N A ;
VRL. SEQ, ウイルス ; PHG. SEQ, ファージ ; SYN. SEQ, プラスミド ;
UNA. SEQ; 未分類 ; ALL, 上記全部
- 2) PIR1、PIR2 (蛋白質データベース) --- NBRFでPIR1とPIR2に分かれている。
PIR1, ファミリーに分類済 ; PIR2, 未整理
- 3) Swissprot (蛋白質データベース) -- EMBO作成
- 4) Consensus Sequences (核酸、蛋白質の共通配列データベース) ---三木作成中ですが全てを採取 できませんので、皆さんの情報提供をお願いします。
NUCL. CNS, 核酸コンセンサス配列 ; PEPT. CNS, 蛋白質コンセンサス配列

上記データベースは約3ヶ月毎に更新されます。

II、自分のデータ形式(阪大形式)

ACOS2000では、次の様な形式を標準とする(国立遺伝研と同じ)。

要は、

- 1) >で始まるコメント行
- 2) 配列があり、
- 3) 最終行には、//を付ける。

一行は80字以内で、大文字入力する、スペースを入れてもよい。

[例]

```
> Visinin, 192 AA, Open reading frame from bovine cDNA
MGNSRSSAL SREVLQDCAPPL RYTEEELSRWYEGFQRQCS DG
```

```
-----
-- -- -- -- --

```

```
AEFIDGVMKNDAIMRLIQYEPKK
```

```
//
```

III、電話回線によるACOS2000への接続

センターでは「ASTER」というソフトを無料配布しているので利用するとよい。
通信ソフトを立ち上げ、通信ポートを次のように設定する。

通信速度：手持ちのモデムに合わせ、1200から9600ボアのどれかにセットする。

7ビット、偶数 (EVEN)

XON/XOFF：あり (ON)

エコー：あり

センターの電話番号は、内線2911または外線876-3145である。
吹田地区の場合は、次のように入力する。

AT

ATDT712911 または ATDT0,8763145
接続されると、CONNECTと出るので、

\$\$\$CON,TSS,,ASC (f・1キーで代用できる)

USER ID-XXXXXX;Y

PASSWARD--ABCDEF

SYSTEM? FRT7

OLD,NEWまたはSAME?

* が出る。

パソコンからデータを転送する場合は、まずVII章をお読み下さい。

データの転送が済んだら、

* F R O G と入力すると、メニューがでる。データ形式が阪大形式でない場合は、メニュー番号6でデータ形式の変換をして下さい。

IV、メニュー画面について

メニュー画面が次のように出るので番号を選ぶ。

MENU (by N.MIKI)
データベース更新の日付表示

1. FindGet
2. Protein Homology
3. DNA Homology
4. Consensus Sequences
5. Input Data
6. FORM Conversion
7. Complement
8. 1st Step Protein Search
9. 2nd Step Protein Search
10. End

注： 8と9は、より詳細な蛋白質ホモロジー検索であるが、大幅に時間がかかる。

V、各プログラムの説明

プログラムは、

- 1)ホモロジー検索
 - 2)ベータベースのデータ抽出、のみに限った。
- その他はパソコンで処理してほしい。

1、データベースからのデータの抽出

Select Database Name [データベースの項目一覧]

I*? と出るので、[]内の細目の中から、一つを選び、入力する。

* I*? PIR1

ただし、NUCL DBは核酸のデータベース、PEPT DBは蛋白質のデータベース、CONS DBはコンセンサス配列のデータベースである。

次にInput a Wordと出るので入力する。ただし、二語にわたる場合は単語を=でつなぐ。

* I*? MYELIN=BASIC

検索が始まり、検索文字がデータベース上にあれば、そのデータを表示する。Do You Want to Take This Data ? (Y/N) と出るので、データを保存する場合は'Y'を入力する。Input Filename for Resultsと聞いてくるので、結果保存用のファイル名を入力する。

* I*? RES.DAT ----- ACOSではファイル名は、を含め8字以内

注：ASTERの場合は、ROLLUP およびROLLDOWN キーにより、消失した画面を再び出すことができるので、COPY キーでプリンターにハードコピーを取ることができる。

2、蛋白質のホモロジー検索

京大化研の金久實氏作成のプログラム (AHOMO) で、これIDEASの中の一つであ

る。

Input Your File Name?と聞いてくるので、自分のデータ名を入れる。

* I*? MY.DAT

次にデータベースの[細目一覧名]があるので、そのなかの一つを選ぶ。

* I*? PIR1

次にInput Cut Score (Standard -70) と、ホモロジー検索の緩さをきいてくるので、標準では-70とする。-100はきつく、-50は緩いホモロジー検索をする。

* I*? -70

まず、自分のファイルが表示される。

Do You Want Save Results ?(Y/N)と聞いてくる。もし、'Y'を選んだ場合は、続いて、保存するファイル名を聞いてくるので入力する。

* I*? RESTDAT

ただし、この場合は結果が画面に現れない。従って、

1) 'N'でどの程度ホモロジーのあるものが出るかを調べておき、次に、結果を保存するか、

2) ASTERを使用の場合は、ROLLUP およびROLLDOWN キーにより、消失した画面を再び出すことができるので、COPY キーでプリンターにハードコピーを取る。

3、核酸のホモロジー検索

京大化研の金久實氏がIDEAS用に作成されたものを移植した。

Input Your File Name?と聞いてくるので、自分のデータ名を入れる。

次にデータベースの細目一覧名があるので、そのなかの一つを選ぶ。ALLを選んだ時は、GenBankのすべての細目を検索するが、時間が30分位かかる。

次にホモロジー検索の緩さをきいてくるので、標準では-15とする。-20はきつく、-10は緩いホモロジー検索をする。

結果をファイルに保存する時は、'Y'とする。この場合は、結果は画面に結果が現れない。従って、1) 'N'でどの程度ホモロジーのあるものが出るかを調べておき、次に、結果を保存したり、ダウンロードするか、2) ASTERを使用の場合は、ROLLUP およびROLLDOWN キーにより、消失した画面を再び出すことができるので、COPY キーでプリンターにハードコピーを取る。

4、核酸および蛋白質のコンセンサス配列の検索

現在、DNAとアミノ酸のコンセンサス配列のデータベースを構築中であり、これを利用する。ただし、データベースの内容には責任を持たないので、出た結果は自分で確かめてほしい。

まず、DNAか蛋白質かを聞いてくるので、DNA->DまたはProtein->Pで答える。

次に、Input File Nameと聞いてくるので、自分のファイル名を入力する。

結果を保存するかどうかを聞いてくるので、Yes->YまたはNo->Nで答えると、実行される。

5、データの入力

短い配列の場合は、このプログラムを用いるとよい。

まず、データについてのコメントを聞いてくるので、入力する。次に配列を入力する。以下の点に注意！

一行は80字以内で改行する。一度改行するとその行は、修正は出来ないので注意すること。入力が終わると、/を入れると、入力したデータが表示される。ファイル名を聞いてくるので、入力する。

*LIST ファイル名 で保存されているかどうかを確かめる。

注：入力後のデータ修正は、このプログラムでは出来ないので、ダウンロードして、パソコン上で修正し、再びアップロードしてほしい。

6、データ形式変換

データが阪大形式でない場合は、阪大形式にデータを変換する。

Input File Nameと聞いてくるので、ファイル名を入力する。

変換が終わると表示される。データは大文字に変換される。

Input New File Nameと聞いてくるので入力する。

7、相補鎖を作る

Input File Nameと聞いてくるので、入力する。

変換後、Input New File Nameと聞いてくるので入力する。

8と9、蛋白質の詳細なホモロジー検索

蛋白質工学研の西川建氏により作成されたプログラムである。これは、2ステップ検索をするもので、初めに8のプログラムでホモロジーのあるエントリーを大まかに抽出し、次に、9のプログラムで、これらについてさらに詳細に調べるものである。ただし「2」のプログラムに比べて3倍以上の時間がかかる（約30分）。時間と経費の節約のために、まず、「2」のプログラムで検索し、ホモロジーのあるものがあれば、このプログラムで実行してほしい。

操作法：

まず、8を実行し終える。結果はOUT.DATAに入っているので、

*LIST OUT.DATA とし、エントリー項目が抽出されているのを確かめる。

次に、9のプログラムを実行する。

8、9とも、ファイル名入力を聞いてくるので、自分のファイル名およびデータベースとも同じ名前のものを入力すること。

ASTERを使用の場合は、ROLLUP およびROLLDOWN キーにより、消失した画面を再び出すことができるので、COPY キーでプリンターにハードコピーを取る。

VI、エラーについて

- 1、ファイル名が存在しないか、誤って入力した時は、
「装置識別子0?は接続されていない」と出るので、FROGから再起動する。
- 2、パソコンから転送したデータが阪大形式になっているか？
- 3、その他のエラーがある場合は遠慮なく三木まで連絡下さい。

VII、ACOSへのデータの転送（アップロード）

1) ワープロ等で予め入力したデータをモデムで転送する。
センターが配布しているプログラム「ASTER」が便利である。

ASTERでの例を述べる。

パソコンのディスクがCドライブにある場合：

- * ^DTLOAD C:DOSファイル名 （^DTLOADは、SHIFT+f・6を押すと出る）
ACOSのカレントファイルに入るので、
- * SAVE ACOSのファイル名 で保存する。
- * LIST ACOSのファイル名 で保存されているかを調べる。

VIII、ACOSのデータをパソコンに転送（ダウンロード）

ホモロジー検索やデータ抽出で得られた結果を、自分のディスクに落とすには、
通信ソフトでダウンロードの操作を行う。

ASTERでは、

- * CATA, A, S でファイルの存在を確かめてから、
- * OLD ACOSのファイル名 を実行してから、
- * ^DTSAVE C:DOSファイル名 （^DTSAVEは、SHIFT+f・7を押すと出る）
これでパソコンに転送される。

IX、謝辞

ACOS2000でのプログラムの公開使用を快く承諾いただいた
京大化研の金久實先生および蛋白工学研の西川建先生に感謝いたします。
論文を作成されるときは、以下の文献を引用していただければ幸いです。
また、プログラム作成にご指導いただいた、薬学部の高木達也先生に感謝いたし
ます。

X、文献

1. Goad, E. G. and Kanehisa, M: Nucleic Acid Res., 10, 247-263, 1992.
2. Nishikawa, K: Prot. Seq. Data Anal., 1, 107-116, 1987.

XI. 付記

本データベースの開発は、主として大阪大学大型計算機センター研究開発計画の一
環として行われ、平成4年5月1日よりサービスを開始しています。