
A DNA sequence analysis program for the Apple Macintosh

Robert H. Gross

Department of Biological Sciences, Dartmouth College, Hanover, NH 03755, USA

Received 29 August 1985

ABSTRACT

This paper describes a new set of programs* for analyzing DNA sequences using the Apple Macintosh computer, a computer ideally suited for this kind of analysis. Because of the Macintosh interface and the availability of high quality software-only speech synthesis, these programs are truly easy to use. Instead of typing in commands, the user directs the program by making selections with the mouse, thereby eliminating most typographical and syntax errors. Output options are selected by "pressing buttons" and then clicking "OK" with the mouse. DNA sequences are confirmed by having the program speak them. The high resolution graphics on the Macintosh not only allow for explanatory diagrams to be used to aid in deciding on input parameters, but can be used to produce slides for presentations and figures for papers. Because of the clipboard and the ability of the Macintosh to readily share data among different applications, data can be saved for use directly in word processing documents (e.g. manuscripts).

INTRODUCTION

A number of DNA sequence analysis programs are available for microcomputers (1). For the routine analyses, these programs have several advantages over the mainframe programs. Since there are no connect time charges for running the programs on your own microcomputer, the exploration of different kinds of analyses is encouraged. Except for the large number crunching routines, the actual time a microcomputer takes to carry out an analysis will be comparable to that of most mainframes since the microcomputer is devoted to the single task, while the mainframe is busy doing other things as well.

The unique user interface on the Apple Macintosh (2) eliminates the confusion many molecular biologists not intimately familiar with computers have when trying to run DNA analysis programs on other computers. In most programs, the user is given a

*The DNA Inspector II is available for \$245 from TEXTCO, 27 Gilson Road, West Lebanon, New Hampshire 03784 (phone: 603-641-1471). It is supplied as a 2 disk set along with a manual. The programs require a 512k Macintosh. A second disk drive is suggested but not required. An Imagewriter printer is needed to obtain printouts of the data analyses.

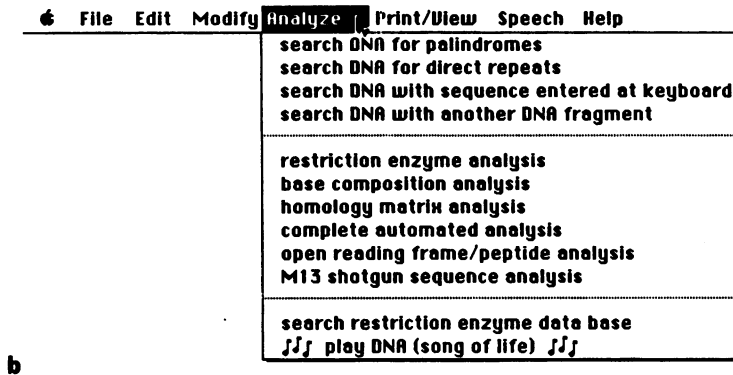
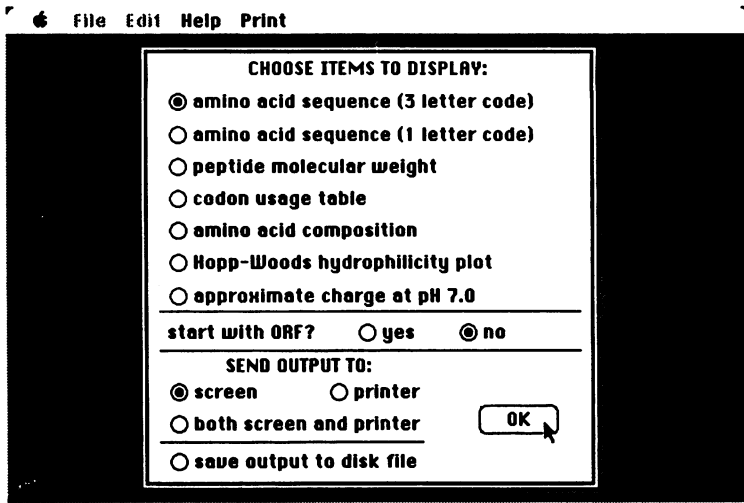


Figure 1a. This screen image that illustrates how choices can be specified with the mouse. By moving the mouse on the desktop, the cursor is positioned over one of the round circles on the screen (buttons). When the the button on the mouse is then depressed, the circle will change from filled <---> empty. Filled buttons indicate that the corresponding option is active. When all choices are made, the "OK" button is selected with the mouse and the program proceeds. Note that all of the information is passed to the program without any typing.

Figure 1b. This is an example of a pull down menu. After the DNA sequence has been loaded into the Macintosh, you can choose any of these options by pointing to it with the mouse. With this kind of interface, there is no need to have complicated commands to choose a specific analysis. One simply points with the mouse.

prompt (that may change in different parts of the program) to which he/she must respond. The available commands are usually listed in a "help" file so the user can determine what options are available. Different programs will handle the inevitable typographical or syntactical errors in the typed responses in different ways, but they

may cause some difficulties.

These problems do not exist on the Macintosh, since no typing is required for these operations. Files are opened by pointing to the file name with the mouse. Instead of typing a list of commands to specify parameters to use in a routine or the kind of output desired from a given routine, selections are made by "pressing buttons" on the screen and then selecting "OK", as shown in Figure 1a for selecting the kinds of output desired from the open reading frame analysis. Restriction enzyme names are selected in a similar fashion by pointing to names in a table on the screen. All available commands are listed in plain English text in pull down menus as shown in Figure 1b, which shows some of the analyses that can be carried out by The DNA Inspector II.

The availability of high quality speech without any extra hardware (3) enhances the ability of the program to communicate with the user. Thus, if the speech option is turned on, errors can be addressed with a spoken explanation rather than a "beep". The main use of the speech, however, is in confirming DNA sequences. Instead of requiring two people to confirm a sequence (one to read and one to check), a DNA sequence can be easily confirmed by having the computer speak the sequence while the operator follows along on a piece of paper.

The high resolution graphics on the Macintosh can be utilized to make slides for presentations (simply by photographing the screen or a high quality printout), or can be saved to disk and then modified later with MacPaint (2). MacPaint pictures can be printed with Apple's LaserWriter printer to produce publication quality figures (the figures in this paper were produced on the LaserWriter).

Finally, the ability of the Macintosh to share data among different applications means that MacPaint pictures, or data written to disk by the program can be used directly in manuscripts by word processing programs such as MacWrite (2) or Microsoft Word (4).

PROGRAM DESCRIPTION

The DNA Inspector II programs are written in Microsoft BASIC (version 2.1) with assembly code embedded to handle the number analysis routines. This combination provides a complete Macintosh interface (through BASIC) along with the maximum possible speed (through the assembly code). The disks come with a run time interpreter, so the user does not have to supply BASIC.

The program will handle either linear or circular DNA sequences of up to 32,767 nucleotides. To analyze a DNA, the DNA file is loaded by selecting from a list of files with the mouse (Figure 2). Comments and other information about the DNA will appear on the screen and a number of menu choices will become available (they are

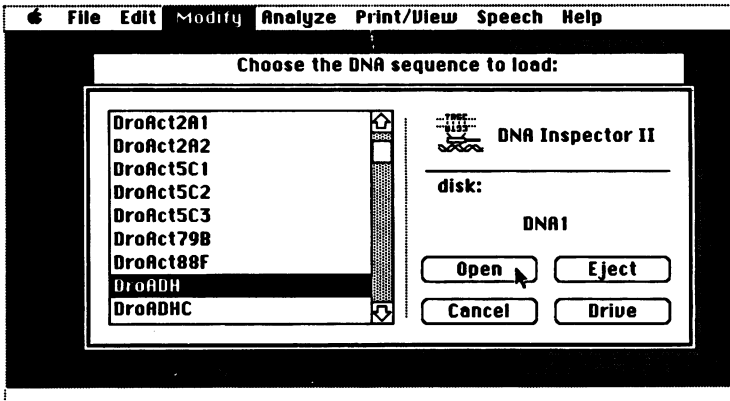


Figure 2. Selection of a DNA file to load is made by pointing to the file name to load and then pressing the mouse button. This causes the file name to invert, as shown. Next, by using the mouse to press the "OPEN" button, the file is automatically loaded. Selecting files to load in this way eliminates the need to type in names and also provides you with a list of all available files.

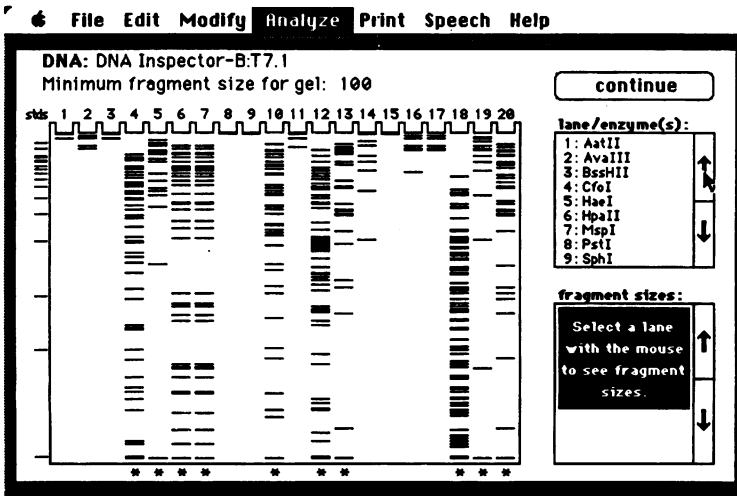


Figure 3. One of the options available following the restriction enzyme digestions is to display the predicted gel electrophoresis patterns of the digests. This figure illustrates the high resolution of the Macintosh screen. After specifying the size of the smallest fragment to remain on the gel, this display is generated. This particular pattern is the result of various digestions of the first 29960 bases of T7 DNA. The mouse can be used to scroll the information in the "lane/enzyme(s)" box (upper right) to see which enzymes correspond to each lane. By pointing to a lane with the mouse and then depressing the mouse button, the fragment sizes of the bands in any lane will appear in the "fragment sizes" box (lower right).

not available until a sequence is loaded). At this point, the program can do any of the following analyses on the DNA:

- edit DNA or comments with standard Macintosh cut/paste
- confirm DNA sequence either by having the Macintosh speak the sequence or by retyping the sequence and having the program compare what you type with the previously entered sequence
- restriction enzyme analysis can produce any or all of the following output: restriction map (linear or circular), list of sites, list of fragments by position, list of fragments by size, gel electrophoresis pattern for digest (see figure 3)
- base composition analysis produces number and percent of each nucleotide and map of GC content along length of DNA
- search for palindromes or direct repeats (user specified separation distance and number of allowed mismatches)
- search with typed sequence (user specified mismatching)
- search with another DNA fragment (user specified mismatching)
- open reading frame/peptide analysis with one or more of the following outputs: position of first coding nucleotide, reading frame of peptide, 3 letter amino acid sequence, 1 letter amino acid sequence, peptide molecular weight, codon usage table, amino acid composition table, Hopp-Woods hydrophilicity plot, charge at pH 7.0

Several other routines and analyses can also be done, which do not involve pre-loading a specific DNA sequence. These include:

- M13 shotgun sequence reconstruction
- recombining of up to 10 different DNA fragments with the option of inverting any fragment before joining.
- homology matrix analysis can be conducted at three levels of resolution. A region of interest can be selected with the mouse and enlarged by "zooming in". Mismatching can be specified. Self-homology analysis can be done with the option compare a sequence to itself, it's reverse sequence, it's complement, or it's inverse (reversed complement).
- downloaded sequences (e.g. GenBank) can be automatically converted to a format appropriate for The DNA Inspector II
- the restriction enzyme data base (over 350 enzymes) can be searched for isoschizomers or for those enzymes recognizing a specific sequence
- the restriction enzyme table can be edited to include any sequence (e.g. TATA box) or to modify enzyme names or sites for easy identification (e.g. all enzymes with * are available in lab)

As a whole these programs provide a very useful package of routines for analyzing DNA sequences. Their true value, however, is in their ease of use. People who have had no prior experience with DNA analysis programs and only brief exposure to the Macintosh (less than 30 minutes) can be doing DNA analyses in less than 5 minutes with this set of programs.

ACKNOWLEDGEMENTS

I would like to thank Dartmouth College for the use of a Macintosh computer and would especially like to thank my colleagues here for their many helpful suggestions and debugging time.

BIBLIOGRAPHY

1. Nucl. Acids Res., vol 12#1 (1984)
2. Apple, MacPaint, and MacWrite are registered trademarks of and Macintosh is a trademark licensed to Apple Computer, Inc., Cupertino, Ca.
3. SmoothTalker speech is licensed to TEXTCO for use in The DNA Inspector II. SmoothTalker is a registered trademark of First Byte, Long Beach, California.
4. Microsoft is a registered trademark of Microsoft Corporation, Bellevue, Washington.