DIGICALC: a restriction fragment analysis program

Armand MacMurray

Department of Medical Microbiology, Sherman Fairchild Science Building, Stanford University School of Medicine, Stanford, CA 94305, USA

Received 16 July 1985

ABSTRACT

DIGICALC is a program designed to aid in the acquisition, storage, and analysis of nucleic acid restriction fragment data. The chief considerations during program design were (i) ease of use for people with varying degrees of computer experience, (ii) minimal hardware requirements (e.g. an IBM PC), (iii) portability and ease of modification, and (iv) improved functionality in sizing and comparing restriction fragments over manual methods. The program accepts manual or digitizer input of nucleic acid fragment mobility, calculates the fragments' sizes, and provides the means to search the fragment database and to produce charts of fragment sizes.

DESCRIPTION OF THE PROGRAM

General Features

Background. DIGICALC was developed to manage restriction fragment length polymorphism (RFLP) data in our study of disease associations with the human major histocompatibility complex. Manual calculation and comparison of fragment sizes is time-consuming and prone to error. The program greatly reduces the time and effort required to produce comparisons of cell line restriction fragment patterns from our Southern blot data. Since most of its users have little previous computer experience, the user interface is designed to be easy to learn, but flexible enough so that experienced users are not hindered.

In order to minimize hardware requirements, DIGICALC is written in C for the IBM PC and compatibles. However, the use of the C language and the isolation of machine-dependent display and keyboard routines in a single source code module simplifies any

*DIGICALC is available from the author. The executable code is available for $200 U.S.; the source code is available in two modules (program and utility) at $250 U.S. each.
desired conversion of the program for a different software and/or hardware environment.

**Program Organization.** The program is organized as a blot database, with all functions available from the main menu. Each entry in the database contains the blot name, the name of the probe used for that blot, and information on each lane in the blot. This lane information consists of the name of the DNA sample in that lane, a user-defined description of the DNA sample, the name of the restriction enzyme used to cut the DNA, and a list of the fragment mobilities, intensities, and sizes.

The functions accessible from the main menu allow one to enter blot data into the database (either manually or with the aid of a digitizer), display or delete such data, calculate the fragment sizes, retrieve data on fragments with specific characteristics (e.g. those from a particular cell line), and produce charts showing the distribution of fragments from a set of cell lines. In a typical program session, one enters data from one or more blots, calculates the band sizes, and then produces a table showing the restriction fragment sizes for each DNA sample.

**Key Features.** The most important features of DIGICALC are its ability to accept digitized band mobilities and its fragment sizing procedure.

Although fragment mobilities may be measured by hand and manually entered into the computer, the computer is better suited to this task. Once the vertical orientation of a blot's lanes has been determined, one simply digitizes the center of each band to enter its mobility value into the computer. This method is manyfold faster than manual measurement, and eliminates the human error inherent in the ruler-reading and data-entry of the manual method.

Accurate and consistent fragment sizing is often difficult to achieve, especially with data from a number of different blots. In order to minimize this problem, DIGICALC first calculates the sizes of fragments on all blots with marker lanes; the fragment sizes on other blots are then calculated, using DNA samples from the already calculated lanes as internal standards. For example, if the fragment sizes of DNA sample AVL, digested
with EcoRI and probed with DR beta 1200bp, have already been calculated on one blot, any other blot containing a lane with the same sample/enzyme/probe combination can use that lane as a size standard for the calculation of its fragment sizes. This process is repeated until all possible calculations have been accomplished.

Thus, having chosen the appropriate DNA samples for the blot with marker lanes, all fragment sizes can be calculated with reference to a single blot. This greatly reduces the occurrence of size inconsistencies between lanes on different blots that are due to slight differences in the mobilities of marker fragments on the different blots.

**User Interface.** In order to lessen the time required to learn the program, efforts were made to provide a clear and consistent user interface.

An important facet of this interface is user input. Novice users tend to prefer to make a selection from a list of choices, while more advanced users tend to prefer minimal-keystroke input. The input routines were designed, whenever possible, as selections from a list of choices, in order to aid novice users. However, they may also be used in a minimal-keystroke manner by expert users.

For example, the restriction enzyme name-entry routine maintains a list of possible enzyme choices. At the "ENZYME NAME: " prompt, pressing RETURN produces the first enzyme in the list. Successive presses of the space bar or "P" key then display, respectively, the next or previous enzyme choices in the list. Once the correct enzyme is displayed, pressing RETURN enters that enzyme as the user's response.

Experienced users, however, need only type sufficient initial characters for unique identification of the desired enzyme's name at the original prompt, then press RETURN. The enzyme name will then be displayed (if no match is found, the user may either try again or enter a new enzyme name). As above, the space bar and "P" key then allow movement along the list of names, while RETURN key enters the displayed choice.

Two additional interface design features are the elimination of unfamiliar keystroke combinations (e.g. control characters)
Nucleic Acids Research

Blot: D56  Probe: DR BETA 300BP

8  |  |  |
7   |  ]  |
6   ]  |
5   ]  |
4   :  |
3   ]  |
2   ]  |
1   ]

8 MARKER LAMBD A HINDIII
7 DR7 LBF PVUII
6 DR6 WVD PVUII
5 DR5 THR PVUII
4 DR4 BSN PVUII
3 DR3 WT49 PVUII
2 DR2 IWB PVUII
1 DR1 IBW4 PVUII

Fig. 1: Screen display of blot fragment mobilities. "J" represents heavy intensity, "I" represents medium intensity, and ":" represents light intensity.

and the presence of dedicated HELP and CANCEL keys. The HELP key provides context-sensitive descriptive information at any user input, while the CANCEL key allows one to cancel the current operation and return to the previous one. For example, the first press of CANCEL during the entry of a fragment's mobility allows one to re-enter the mobility value; the second press of CANCEL allows one to re-enter the entire lane, and so forth.

When using the digitizer for data entry, it is often inconvenient to alternate between the digitizer for the entry of fragment mobility values and the keyboard for the entry of other pertinent data. Therefore, a user-specified "Digi-Key" area of the digitizing surface is dedicated to simulating keyboard entry of the digits and other frequently-used keys. Each box in the 3x6 rectangular grid represents a key; digitizing the box is equivalent to pressing the associated key on the keyboard.

The final significant feature of the user interface is the display of blot fragment mobilities in analog form (Fig. 1). Up to 24 lanes may be displayed on the screen at one time. This facilitates quick visual inspection and comparison with the actual blots.

Program Functions

Below are brief descriptions of the program functions available from the main menu:
Rechecking gel D55 DR BETA 1200BP
calculated D55 DR BETA 1200BP based on D55MARK PGF HINCI; dev = 97.038612
Rechecking gel JB2 DR BETA 1200BP
calculated JB2 DR BETA 1200BP based on JB2MARK PGF BGLII; dev = 91.912590
Rechecking gel JB2.2 DR BETA 300BP
Rechecking gel JB3 DR BETA 300BP
Rechecking gel JB4 DR BETA 300BP
OVER LIMIT: JB4 DR BETA 300BP based on Alll LBF ECO R; dev = 412.156891
Rechecking gel JB5.2 DR BETA 300BP

Making another pass to try to calculate any remaining gel(s)

Fig. 2: Part of an "audit trail" from the fragment sizing procedure.

Entering Data. Blot information may be entered either manually or with the aid of a digitizer. Information such as the names of the blot and probe used are entered at the keyboard, while fragment mobilities may be entered by digitizing the fragment position once the blot has been properly oriented in the digitizing area. Fragment intensities are recorded on a three-level scale (light, medium, and heavy) for increased discrimination between distinct fragments of similar mobilities. As fragment data is entered, a screen display similar to that of Figure 1 is continually updated to provide visual feedback on the current state of data entry.

Calculating Sizes. Figure 2 shows a portion of an "audit trail" produced by the fragment sizing procedure. This information is written to a disk file, and shows which DNA sample served as a marker for each blot calculated. The "deviation" value is a measure of the internal consistency of the marker fragment sizes. This measure and the sizing algorithm used are described in reference 1.

Predicting Band Mobilities. This function allows one to predict the mobility of a band, given the mobility of marker bands on the same gel or blot. Such prediction is useful in determining which region of a gel to cut out in order to isolate a specific nucleic acid fragment.

Managing Fragment Database. This heading actually encompasses a number of functions for managing the database. These functions include listing the names of all blots in the database, deleting a blot from the database, saving the database to disk, and
Probe: DQ BETA 900BP  Restriction enzyme: HINDIII
Band intensity: -- light; I - medium; X - dark

Fragment sizes (kb)
1 1 6 5 4 4 3 3 3 2 1 1 1 1
2 1 . . . . . . . . . . . . .
9 3 9 2 6 4 2 1 7 6 5 3 1

DR 1
IBW4  I -  I  X  X  I  A82
DR 2
IWB   I   I  X  I  X  I  A82
DR 3
WT49  X  I  I  I  X  I  X  I  A82
DR 5
FFP   I   -  I  X  I  I  X  I  A39
THR   I   -  I  X  I  I  X  I  A39
DHI   I  I   -  I  X  I  I  X  I  A39
DHI   I  I   I  X  I  I  X  I  A82
DR 6
WVD   I  I   I  X  I  I  X  I  A39
APD   I  I   I  X  I  I  X  I  A39
HHK   I  -  I  X  I  I  X  I  A39
WVD   I  -  I  X  I  I  X  I  A82
DR 7
LBF   X  X   -  I  X  -  I  I  I  A39
LBF   X  I  I  -  I  X  -  I  I  I  A82

Fig. 3: Sample chart of fragment sizes.

displaying a blot on the screen in the manner of figure 1.

Searching the Database. This function displays all lanes in the
database matching the characteristics the user has specified.
For instance, one might request all lanes digested with EcoRI.
The search specification categories include DNA sample name,
probe name, enzyme name, and the contents of the user-specified
DNA sample description field.

Producing a Chart of Fragment Sizes. This function produces a
chart displaying the fragment sizes associated with all the DNA
samples sharing an enzyme/probe combination (Fig. 3). In our use
of the program, we utilize the user-specified DNA sample
description field to describe the DR type of the DNA sample. The
charting function sorts the DNA samples in alphabetical order on
the basis of their description field. Thus, in the figure the
samples are sorted by DR type.
Utility Functions. This menu selection allows the user to specify the serial port in use for the digitizer, and the location and size of the Digi-Key digitizer keyboard area.

Auxiliary Programs

The blot data is stored on disk in binary form in order to minimize the time required for the program to read and write the data. Using the programs BINTOASC and ASCTOBIN, the data may be converted from binary to ASCII form and back (e.g. for use by other programs).

The charts described above under "Producing a Chart..." are saved on disk as ASCII text files, and thus may easily be edited with any word processor in order to modify titles and legends, and consolidate data.

In addition, the lists of names used in the name-entry routines (e.g. the enzyme name-entry routine) can be displayed and modified outside the DIGICALC program by a set of auxiliary programs.

Hardware Requirements

DIGICALC runs on the IBM PC and compatibles. It requires at least 256K of RAM (640K recommended) and one floppy disk drive (two floppies or a hard disk recommended). A printer is also recommended.

Digitizer entry of blot data requires the Science Accessories Corporation (Southport, CT) Grafbar model GP-7 sonic digitizer (serial port version).

Technical Information

DIGICALC is written using the Computer Innovations C86 C compiler (for PC/MS-DOS) version 2.20I, except for one small display function written in 8086 assembly language (IBM Macro Assembler, version 1) for speed (it, too, may be written in C).

The source code is divided into three groups of modules: the program, utility (e.g. windowing), and machine-dependent code modules. Thus, the program may be modified relatively easily to add new features or to allow it to run on other computers.

The compiled program, when loaded, uses approximately 110K of RAM with no entered blot data. The blot data is also kept in RAM, due to speed considerations; RAM requirements for blot data are about 350 bytes/lane entered (assuming an average of 7
bands/lane). Thus, the 640K RAM of a fully expanded PC allows entry of over 1200 lanes of data; this should be sufficient for most applications.

DISCUSSION

The successful use of DIGICALC by scientists with both minimal and extensive prior computer experience indicates that this type of user interface works well for most users. The information provided by the HELP key was especially useful for novices. Typically, users became proficient in using the program after one session of use.

Use of a digitizer for data entry increased data entry speed manyfold. In addition, the rapid accessibility of visual displays of band patterns often eliminated the need to refer to the original blot data.

Currently planned program enhancements include providing a choice of fragment size determination algorithms, making use of graphics on systems that support it, and adding more "intelligence" to the fragment size determination procedure.

ACKNOWLEDGEMENTS

I thank Linda Foster, Deborah Hardy, and John Bell for their help in making the program "novice-proof." In addition, I would like to thank my colleagues in the Department of Medical Microbiology for their helpful comments.

REFERENCES

1. Schaffer, H.E. and Sederoff, R.R. (1981) Anal. Biochem. 115, 113-122.