An inexpensive semi-automated sequence reader for the Apple II computer

Michael Komaromy and Herman Govan

Molecular Biology Institute, University of California, Los Angeles, CA 90024, USA

Received 15 August 1983

ABSTRACT

We describe a relatively inexpensive digitizing pad, available from Houston Instruments, for which we have written software which allows the direct reading of DNA (or RNA) sequencing gel autoradiographs. This eliminates several sources of error in nucleic acid sequencing, and renders computer analysis less cumbersome.

INTRODUCTION

Several papers have previously described systems for semi-automated reading of nucleic acid sequencing gels (1,2). In addition, there are commercially available readers (BRL, Intelligenetics). The advantages of such systems are obvious--they reduce to a minimum the operator errors involved in transferring sequences from autoradiograph to paper to computer. Regions at the tops of gels, which often contain a large number of closely-spaced bands, can be more quickly and accurately read. Depending on the exact type of system, it can provide double-checks of the sequence. Help can be provided in reading sequences, by quantitation of distances over regions of compression or distortion, and provision of information on the expected number of bands in these regions.

We have chosen the Apple II Plus computer for a variety of reasons. Most important is the large amount of software and hardware available for the Apple. This includes many scientific packages, for both data analysis and data acquisition. The available hardware makes it easy to upgrade the system to a quite powerful machine, and most languages can be used.

MATERIALS AND METHODS**Hardware**

We are using an Apple II Plus computer with 16K extra memory (64K RAM total), with dual minifloppy disk drives, an Apple Monitor ///, an 80-column

card, a Silentye printer, a Novation Apple-Cat II modem, and a Houston Instruments (Austin, Texas) HiPad digitizing tablet with an Apple II interface card and crosshair cursor. The entire system can be assembled for about \$3500, which includes \$900 for the HiPad. Apple Computer is no longer making the Apple II Plus, but their Apple IIe is totally compatible with the hardware and software described here.

The HiPad has an effective digitizing surface of 28 x 28 cm, and a limit of resolution of .125 mm. It is translucent, and, for gel reading, is sufficiently illuminated from underneath by a light box (Fig. 1). The Apple II interface card is supplied by Houston Instruments as the DT-11A option. The instrument functions by returning X and Y coordinates to specified memory locations in the computer, which can be retrieved by calling machine subroutines. Either cursors or light pens can be used to activate the tablet, but the crosshair magnifying cursor which we chose allows extremely fine resolution. In addition, one- or four-button cursors are available, with the latter being independently programmable.

Software

The HiPad is most easily programmed in BASIC, because of the above-mentioned calls (it is, however, possible to do in Pascal by calling BASIC routines from Pascal). The program described here was written in BASIC, and reads the sequence into a BASIC textfile which is stored to disk. The sequence analysis software which we are using is that described in (3), and is written in Pascal. Thus, a file transfer from DOS BASIC to Pascal is necessary (DOS is the Apple Disk Operating System). This is accomplished by use of PUFFIN (4), a program which directly does the transfer. Other programs which do this are available from RCM Software (Kettering, Ohio) and Link Systems (Santa Monica, Ca.). These latter will also do conversions between DOS and CP/M files, which would make the sequence portable to most other microcomputer systems.

RESULTS

The program itself is relatively simple. The user is asked to define the tracks for A, C, G, and T, which is done by spotting the cursor at the left and right sides of each lane. An N, for nonspecified nucleotide, is defined as the region to the right of the rightmost track. Each band is then spotted, again using the cursor. This can be done in any manner convenient, since the X and Y coordinates of each cursor spot are stored. After all entries are made, the computer assigns values of A, C, G, T or N to each X

Figure 1. Top--The digitizing pad with four-button magnifying crosshair cursor. Bottom--The digitizing pad, lighted from underneath by a fluorescent light box, with a sequencing ladder.

coordinate and then sorts the Y coordinates, thus assembling the sequence, which is then read into a textfile and stored to disk. This enables one to, for example, first read the sequence in normal sequence order, and then read it in by spotting a track at a time. This provides a useful cross-

check, since the sequence handling software can then be used to compare the two sequences, and inspection of an area of question resolves conflicts.

DISCUSSION

The system described here is inexpensive and quite versatile. In addition to the sequence reading, we have programmed the computer and tablet to read bands from restriction digest photographs, fit curves to standards and calculate molecular weights. The sequence reading software is being modified to account for gel smiling, and to provide interactive help by calculating the expected number of bands in a region.

The HiPad can be used with other microcomputer systems. A variety of interfaces and data coding types are available. The software described here should be easily adaptable to other computers. The DOS-Pascal-CP/M file transfer utilities make the process of conversion to formats used by other microcomputers straightforward. The addition of a modem to the system allows sequence transfers to and from large data bases, such as that at Los Alamos. It is thus possible to read sequence ladders, analyze the sequences and put them in a data base without possibility of human transfer error.

The programs described here, which include the BASIC program for the actual reading process and the Apple Pascal PUFFIN program, are available upon request. The Apple Pascal system must be purchased from Apple (about \$150). Apple users should send two 5.25 inch diskettes. Listings are also available.

ACKNOWLEDGEMENTS

We would like to thank Dr. Randolph Wall, in whose lab this work was done, and Mayo Yerington for help with manuscript preparation. This work was supported by NIH grants AI 13410 and CA 12800. H.G. was a President's Fellow of UCLA.

REFERENCES

1. Lautenberger, J.A. (1982) Nucleic Acids Research 10, 27-30
2. Gingeras, T. R., Rice, R. and Roberts, R.J. (1982) Nucleic Acids Research 10, 103-114
3. Fristensky, B., Lis, J. and Wu, R. (1982) Nucleic Acids Research 10, 6451-6463
4. Wo, D. (1981) Call-A.P.P.L.E. 4/#9, 13-42