Volume 9 Number 24 1981 Nucleic Acids Research

Computer programs to assist in high resolution thermal denaturation and circular dichroism studies on nucleic acids

Thomas C.Goodman*, Stephen C.Hardies**, Carlos Cortez+ and Wolfgang Hillen+

University of Wisconsin, Department of Biochemistry, College of Agricultural and Life Sciences, Madison, WI 53706, USA

Received 15 September 1981

ABSTRACT

Computer programs are described that direct the collection, processing, and graphical display of numerical data obtained from high resolution thermal denaturation (1-3) and circular dichroism (4) studies. Besides these specific applications, the programs may also be useful, either directly or as programming models, in other types of spectrophotometric studies employing computers, programming languages, or instruments similar to those described here (see Materials and Methods).

INTRODUCTION

The work in our laboratory centers on studying the physical and biochemical properties of DNA. The goal is to understand the role of DNA structure in gene regulation (reviewed in 5). In addition to using computer programs for DNA sequence handling (6,7), a number of programs were written in our laboratory to facilitate various physical studies on homogeneous DNA fragments. These programs have proved invaluable in gathering, storing, and processing large amounts of numerical data obtained from high resolution thermal denaturation (1-3) and circular dichroism (4) studies.

This paper describes the most generally applicable programs developed in our laboratory. These programs direct the collection, processing, and graphical display of experimental data gathered from specific instruments (see Materials and Methods). In addition to their use in high resolution thermal denaturation and circular dichroism studies, some of the programs may also be adaptable to other types of spectrophotometric studies and serve as models in applications requiring the rapid collection of large amounts of numerical data. Finally, this paper discusses some of our approaches in dividing complex data processing problems into units that are manageable by a small computer and in devising program and data formats that are simple and provide compatibility over a wide

range of data processing tasks.

MATERIALS AND METHODS

The programs described in this paper were developed for a Digital Equipment Corp. (DEC) PDP 11/03 microcomputer equipped with an 8-inch RX11 dual drive single density floppy disk system. An RT-11SJ (V02C-02G) operating system is currently used in this system. The computer language employed in writing these programs is either Macro Assembler machine language (designated as Macro or machine language in the text) or PDP-11 Fortran IV (refered to as Fortran).

The programs A2D and A2DCON require that the computer system have an analog-to-digital interface circuit board (DEC model ADV11-A). A parallel input/output circuit board (DEC model DRV11) is needed in the operation of the programs VARIN, VARCMD, and VARSTA. The interfacing of the Hewlett-Packard instruments mentioned below requires an IBV11A interface circuit board.

In addition to the computer and interfacing circuit boards, other equipment associated with the overall system and several of the programs mentioned in this paper include a Hewlett-Packard 7245A Plotter/Printer, a Varian-Cary 219 UV-VIS spectrophotometer equipped with a Digital Interface Port (DIP), and a Hewlett-Packard 2804A quartz thermometer.

Finally, references 8 and 9 provide useful definitions and examples for some of the specialized terms (e.g. "file", "subroutine", "ASCII") that are used in the text.

RESULTS

Computers provide speed, accuracy, and convenience in the repetitious collection of experimental data generated by various types of instruments. In interfacing a computer with some instrument, two components are necessary: first, a hardware interface to physically connect the computer and the instrument producing the data, and second, a software, or program, interface to regulate the flow of data and commands between the computer and the peripheral instrument.

In addition to controlling an experiment and receiving the data generated, this data must also be stored and processed by the computer. Programming is necessary to direct these operations. Finally, it is also desirable that the computer be programmed to produce plots and figures from the processed data by means of printing and graphics programs.

Considering these tasks, the programs described below are grouped on the basis of their functions: interfacing programs, programs for the acquisition and processing of data, and graphics programs for data display. The particular hardware requirements for these programs are enumerated in Materials and Methods. The coordination and overall operation of the individual programs is described in the Discussion.

A. Interfacing Programs

Three machine language subroutines were developed as the software interface to allow data, commands, and instrument status information to be transfered between the computer and a Varian spectrophotometer.

VARIN This subroutine returns 18 4-bit binary coded decimal words containing photometric, wavelength, and (when the Varian supplied temperature accessory is installed) temperature data from the spectrophotometer.

VARCMD This subroutine allows commands to be given to the spectrophotometer (such as wavelength scan and chart control) by the program in operation.

VARSTA This subroutine allows the status (such as the automatic changing of source lamps) of the spectrophotometer to be monitered by the computer.

The combination of the above subroutines with a Fortran main program allows virtually any type of spectrophotometric task to be automated with respect to control of the process and data collection.

Many instruments produce an analog voltage as their final output. This voltage is usually sensed by a chart recorder and may be directly used by a computer only when it is converted to a digital number. The hardware component of this type of interface was described in Materials and Methods. The software component is embodied in the following two subroutines:

A2D This Macro subroutine determines the number of analog channels that will be read and calls A2DCON to do the conversion.

A2DCON This Macro subroutine directs the actual analog-to-digital conversion and returns the digital value obtained to the main (calling) program.

These programs allow up to 16 analog channels (0-15) to be read sequentially by the computer. Repetitive calling of these subroutines allows data averaging and effectively reduces the problems of instrument and interface noise.

Two machine language device handlers implement the interfacing of a Hewlett-Packard 7245A Plotter/Printer with the computer by means of an IBV11A interface circuit board.

HP.SYS This program selects the printer function of the instrument and controls the transfer of ASCII characters to the printer.

PL.SYS This program selects the plotting mode of the instrument and allows the transfer of plotting commands and data.

An additional machine language program, QSETUP, is callable as a Fortran subroutine and implements, according to a preset mode, the transfer of data from the quartz thermometer (Hewlett-Packard 2804A) to the computer.

B. Data Processing Programs

The following Fortran programs were written to process high resolution thermal denaturation data. Each program accepts an input file with a set data structure and ordering (see Discussion). The program then creates a new output file for the numerical results of the processing and any additional comments that the operator wishes to add to the file from the terminal.

CALHYP This program receives a file containing temperature and absorbance data pairs recorded during the experiment. It locates the initial and final absorbances (the minimum and maximum), calculates the fractional hyperchromicity for each point and writes the temperature and hyperchromicity data pairs to the output file.

CTHETA This program accepts a file with hyperchromicity and temperature data pairs and produces a file with theta ($\theta = 1.0 -$ hyperchromicity) and temperature data pairs.

PRUNE1 This program accepts a file containing ordered data pairs. It fits the first 21 data points to a cubic equation by a least squares method and displays the standard deviation of the points. It then accepts a cut-off level (e.g. $3.0 = 3\sigma$) from the terminal keyboard. This cut-off level is then used as the program successively fits the additional points in the file to delete any point having a larger than specified deviation. A separate file is produced containing only those points within the stipulated limits. Deleted points and their deviation are displayed on the terminal screen.

SMDIF1 and SMDIF4 These programs accept an input file of ordered data points and smooth these points by fitting successive

groups of points (odd numbers between 9 and 21 for SMDIF1; odd numbers between 21 and 41 for SMDIF4) to a cubic equation. The derivative of the curve at the midpoint is also calculated. Two separate output files are created for the smoothed and differentiated data.

In addition to the above programs for the processing of high resolution thermal denaturation data, other programs were written to handle circular dichroism data. These programs allow digitized data points from a chart paper record to be entered manually into the computer (CDINPT) and later plotted (see below) for comparison (SIGMAF) with various types of calculated spectra (4).

C. Graphics Programs

Computer systems provide speed and relative ease for plotting numerical data. Not only are such plots useful analytical tools in interpreting the data, but in many cases photo-ready figures can be produced for publication. The following graphics programs were developed for these applications:

HPPLOT Employing the PL.SYS device handler described above, this program accepts a data file, determines the maximum and minimum values of the X and Y data pairs in that file, and scales these coordinates onto a defined plotting area. The data points are then plotted and tic marks and labels are also put on the final graph. A variety of line types can be drawn between points.

LONGAX This program is similar to the program HPPLOT described above. The major difference is that it allows long axis plots (up to 1 m) to be produced and contains more extensive provisions for labeling the final plot (e.g. Sample:, Date:, File:, Comments:, etc.).

DISCUSSION

In a typical high resolution thermal denaturation experiment, a Fortran main program repeatedly calls the subroutine VARIN to obtain absorbance measurements on the DNA sample in the Varian spectrophotometer. A2D and A2DCON are repeatedly called to read the voltage from a thermistor placed in the heated cuvette. This data is stored on a floppy disk as a file containing absorbance - voltage data pairs. The voltage readings are then converted to temperature readings after the experiment is fin-

ished by means of a calibration curve for the particular thermistor. (In our laboratory the calibration is done by means of a quartz thermometer using the subroutines QSETUP, A2D, and A2DCON.) The resulting file, containing absorbance - temperature data pairs is then used as input to the program CALHYP which produces a new output file containing hyperchromicity - temperature data pairs. In the same way, the programs CTHETA, PRUNE1, and SMDIF1 or SMDIF4 are then used sequentially to produce the final smoothed $d\theta/dT$ - temperature data pair file. This file, or any intermediate file, may then be plotted using the program HPPLOT or LONGAX.

The overall task of producing the smoothed $d\theta/dT$ versus temperature plot is divided into a number of smaller tasks, not associated with the initial data collection. The advantages of this procedure are the following: a.) During an experiment, a maximum rate of data collection can be achieved. None of the computer's time during the course of the experiment is used for processing or smoothing the data. b.) Intermediate results that are sometimes useful in the analysis of the data are preserved. If necessary, these files may be deleted to save disk storage space and recalculated from the raw data file only as needed. c.) Greater flexibility is achieved in the processing of the data because, for example, the degree of smoothing (how many points will be fitted to the cubic equation) does not have to be determined prior to the start of the experiment. Differing degrees of smoothing may be tested on data obtained from a single experiment (2). d.) Because each program is smaller and requires less computer memory space than would the combination of them all, a less powerful and less expensive computer is sufficient.

In addition to the logical division of tasks in a complicated process of data processing, each of the series of programs allows a section at the beginning of the output file for written comments. In the case of the high resolution thermal denaturation data processing programs, the programs themselves automatically add a message recording the particular processing step in addition to allowing the operator to enter comments from the keyboard. This is a particularly helpful feature in that it allows processing steps to be traced for any file at any time.

A number of problems may be encountered in transferring computer programs between different laboratories. Many of these arise from differences in the computer hardware and different computer languages used in various labs. Subtle differences in the problems being addressed and

the differing data formats and processing needs may also require that programs received from outside sources be modified (sometimes extensively) in order to meet the particular needs of the research. In spite of these drawbacks, however, the publication and exchange of scientific programming may be helpful in providing researchers with models on which to base their own programming efforts and ideas as to helpful features that can be included.

ACKNOWLEDGEMENTS

This work was supported by grants to R.D. Wells from the National Institute of Health (CA-20279), the National Science Foundation (PCM-15033), and the Graduate School of the University of Wisconsin, Madison. W.H. was supported, in part, by the Max Kade Foundation and the Deutsche Forschungsgemeinschaft. We thank Drs. R.D. Wells and R.B. Inman for many helpful discussions.

FOOTNOTES

*To whom correspondence about these programs should be addressed. In requesting programs, individuals should specify which programs they would like to receive and whether they require a source listing or floppy disk.

**Present Address: University of North Carolina, School of Medicine, Department of Bacteriology and Immunology, Chapel Hill, N.C. 27514

+Present Address: Shinar Digital, Inc., 128 W. Main Street, Evansville, W.I. 53536

†Present Address: Institut für Organische Chemie und Biochemie, Technische Hochschule Darmstadt, Petersenstr. 22, D-6100 Darmstadt, FRG.

REFERENCES

1. Hardies, S.C., Hillen, W., Goodman, T.C., and Wells, R.D. (1979) J. Biol. Chem. 254, 10128-10134.
2. Hillen, W., Goodman, T.C., and Wells, R.D. (1981) Nucleic Acids Res. 9, 415-436.
3. Hillen, W., Goodman, T.C., Benight, A.S., Wartell, R.M., and Wells, R.D. (1981) J. Biol. Chem. 256, 2761-2766.
4. Hillen, W., Goodman, T.C., and Wells, R.D. (1981) Nucleic Acids Res. 9, 3029-3045.
5. Wells, R.D., Goodman, T.C., Hillen, W., Horn, G.T., Klein, R.D., Larson, J.E., Müller, U.R., Neuendorf, S.K., Panayotatos, N., and Stirdivant, S.M. (1980) Progress in Nucleic Acids Research and Molecular Biology 24, 167-267.
6. Staden, R. (1977) Nucleic Acids Res. 4, 4037-4051.
7. Staden, R. (1978) Nucleic Acids Res. 5, 1013-1015.

8. Cooper, J.W. (1977) The Minicomputer in the Laboratory, John Wiley & Sons, New York.
9. Murrill, P.W. and Smith, C.L. (1973) Fortran Programming for Engineers and Scientists, 2nd Ed., Intext Educational Publishers, New York.