An integrated software system for microcomputer management of recombinant DNA data

José Campione-Piccardo* and Martha Ruben¹

Health and Welfare Canada, Laboratory Centre for Disease Control, Tunney’s Pasture, Virus Building, Ottawa, Ontario K1A 0L2, and ¹Department of Medicine, University of Ottawa, Health Sciences Centre, Ottawa, Ontario, K1H 8L5, Canada

Received 23 May 1985

ABSTRACT

A database-oriented system (pCP123) is described for the manipulation of recombinant DNA data. This system was developed within the context of an integrated software package with spreadsheet, database, graphing and programming capabilities. The system includes two databases, one of sites and another of regions, coordinately handled by a series of macro-programs operated from four user-defined menus. A distinctive feature of the system is the possibility of handling both ends of defined functional or structural regions in situations of simulated deletions or insertions.

INTRODUCTION

Several powerful software packages have been described for the analysis and storage of nucleic acid sequences in microcomputers (1-14). Nevertheless, a large part of recombinant DNA work is not done at the nucleotide sequence level, which in many cases remains unknown. Usually, ligations, insertions and deletions can be more meaningfully modeled with restriction enzyme cleavage maps which can provide a better idea of the architecture of a genome than the picture that can be immediately derived from the nucleotide sequence. The pCP123 system described in this communication allows the simultaneous handling of two databases, one of sites as defined by single positions in a genome, and another one of regions defined by two boundary positions. Three other systems developed along similar concepts but unable to handle boundaries of regions have been reported (15-17). The pCP123 system was developed within the infrastructure provided by Lotus 1-2-3, a powerful integrated software package with spreadsheet, database, graphing and programming capabilities. The integrated software provides a compatible computer environment regardless of the hardware configuration. The macro language-like programming tool of Lotus 1-2-3 is well adapted to its database

The pCP123 system will be made available upon receipt of a 5 1/4 inch two sided diskette in a self-addressed diskette-mailer envelope.

Nucleic Acids Research

and spreadsheet functions. On the other hand, it was found to be extremely limited in its syntax, editing and graphing possibilities. These limitations forced the development of many ad-hoc programming solutions.

Some knowledge of the use of Lotus 1-2-3 is essential for the implementation of pCPl23 but no programming abilities are required as the entire system is operated through four user-defined menus. A minimum of 128 K (preferably 256 K) in RAM as well as the Lotus 1-2-3 System Diskette are essential requirements for operation. The pCPl23 system was developed using a Hyperion microcomputer (Dynalogic Info-Tech Bytec-Comterm, Montreal, Quebec, Canada) with 2 disk drives, 256 K in RAM, and running version 1.25H of MS-DOS (Microsoft Corporation, Bellevue, WA, U.S.A.) and version 1.AH of Lotus 1-2-3 (Lotus Development Corporation, Cambridge, MA, U.S.A.). Minor adjustments may be required to run the pCPl23 macro-programs in computers with a different keyboard.

DESCRIPTION OF THE SYSTEM

Active screens

At the top left corner of the pCPl23 spreadsheet there are three "pages", each of which corresponds to an entire screen. Page 1 is dedicated to the database of site positions, page 2 is used by the region database and page 3 provides a range for comments on the peculiarities of each particular construct. The two databases are independent of each other and can be accessed and manipulated using any of the many commands of Lotus 1-2-3. Their only functional link is provided by the simultaneous modifications they experience during the operation of the different commands of pCPl23.

Menus

The entire pCPl23 system is operated by four menus. The Master-menu is the one which provides the routing to the other A-, B- and C-menus containing the functions implemented by the system. The A-menu contains those functions dedicated to the handling of the site database, the B-menu contains commands for operating deletions, insertions and ligations, and the C-menu has been reserved for utility commands. Lotus 1-2-3 programming only allows one user-defined menu with 8 commands. This limitation was overcome in pCPl23 by internally redefining the range operated by the single user-defined menu.

Variables

The screen contiguous to page 2 contains a series of single and double ranges used as variables or criteria for several of the functions of the system.
The macro-programs and subroutines responsible for the different pCP123 commands are listed at the right of the screens already described. Sitesort and Enzsort sort the site database by ascending site position or by alphabetical order of the restriction enzyme name. Xtract allows the extraction from the database of the sites cleaved by one or more defined restriction enzymes. Fragment displays the fragment sized in circular and linear conformations originated by one or more defined restriction enzymes. Origin redefines position 1 while retaining the orientation of the fragments within the plasmid. Vector, on the other hand changes the orientation of the sites while maintaining the same origin (site 1). These last two programs also modify the position limits of the genomic regions defined in the region database. Master causes the return of the Master-menu, and Quit allows the exit from the user-defined system of menus. Save creates files with the database content of the worksheet currently on display, and Retrieve recalls within the pCP123 system the database content of files previously created. Zavelig redefines the database regions and saves the information they contain in file LIGATE.WKS which is automatically recalled during simulation of ligations by Ligate. Delete, Insert and Ligate allow the described operations in the site database with the corresponding updating of the region database. Insert allows the simulation of an insertion while Ligate performs the reciprocal insertion of two files into each other. For this operation to take place one of the files should be in the worksheet and the other in file LIGATE.WKS in the default disk. These features, and particularly its ability to update the region database, allow pCP123 to become a powerful tool for handling recombinant DNA data. Printreg allows the printing of the region database. The printing conditions can be adapted to different printers supported by Lotus 1-2-3. Circmap provides a rudimentary but effective and extremely fast circular representation of the plasmid sites and fragments as determined in the output of a previously called Fragment command. Linemap allows the linear representation of enzyme cleavage maps by up to six different restriction enzymes. The outputs of Circmap and Linemap can be directed to a printer or a plotter using Lotus 1-2-3 PrintGraph program. Klear clears the system worksheet and prepares it for the retrieval of a new pCP123 file. Justify justifies the text of comments entered on the third page of the spreadsheet and Sizereg sorts the region database and adjusts the range of the column for sizes of regions.
CONCLUSION

The convenience of the system described in this communication stems from two sources. First, it uses the software environment provided by a powerful package of integrated software accounting for the high computation speed as well as for additional editing capabilities and hardware compatibility. This compatibility is achieved in a way similar to operating systems providing compatible environments created by software in otherwise different computers. Second, the system is database-oriented and allows the handling of large numbers of records describing the limits and names of structural and functional regions within recombinant DNA molecules. The ability to automatically manipulate and up-date both these limits is one of the most distinct features of the system.

ON NAMES AND TRADEMARKS

"pCP123" reproduces the usual nomenclature for plasmids whose data the system is designed to handle, as represented by the letter P located at the centre of the name. At the same time it reminds that the system was developed for personal Computers running Lotus 1-2-3. Lotus and 1-2-3 are registered trademarks of Lotus Development Corporation, and MS-DOS is a registered trademark of Microsoft Corporation.

REFERENCES