

**KLONER; a computer program to simulate recombinant DNA strategies by restriction map manipulation**

---

Paul R. Caron

---

Department of Biochemistry, The Johns Hopkins University, School of Hygiene and Public Health, Baltimore, MD 21205, USA

---

Received 15 August 1983

---

**ABSTRACT**

A computer program is described which allows for the manipulation of restriction maps of various DNA fragments to demonstrate techniques used in DNA cloning and to predict and/or confirm experimental results. This program is capable of reading in restriction enzyme cleavage sites for several different DNA molecules of interest. This information is then compiled in order to form restriction maps which can then be processed by digestion with restriction endonucleases and treatment with other common DNA modifying enzymes. Ligation can then be simulated by joining fragments with complementary ends in all possible orientations, producing restriction maps of the products. The resulting recombinants can then be further analyzed by physical mapping with appropriate restriction endonucleases. This program was written in Pascal on an Apple II computer.

**INTRODUCTION**

In recent years the utilization of microcomputers as research tools in the laboratory has been widespread, mainly due to relatively low cost in investment and to the large resource of software available. In particular, microcomputer programs for DNA sequence analysis have been extremely helpful in processing the large amounts of data generated by a laboratory actively engaged in sequencing, a task which would otherwise be quite time consuming.<sup>1,2,3</sup> One of the end products of many of these programs are lists of restriction enzyme cutting sites which are valuable in planning future sequencing strategies or cloning experiments. The program described here, KLONER, can take lists of restriction enzyme cutting sites for several DNA molecules and simulate the action of various enzymes on these molecules. Included are procedures for generating complete and partial restriction endonuclease digestions, and for simulating the

action of Bal31 nuclease, S<sub>1</sub> nuclease, terminal transferase, Klenow fragment of DNA polymerase, and ligase. KLONER is written as a single program with many procedures rather than a package of smaller programs to allow rapid communication of data from one section of the program to the next.

## METHODS

### Hardware

This program was written using the Apple UCSD Pascal 1.1 compiler on an Apple II plus computer with 64 K of RAM. Other equipment included two disk drives, an 80 column card, and a printer.

### Datafile Format

The input datafile consists of a list of restriction enzyme names, their recognition sequence, cutting positions within this sequence, and the sites within the DNA molecule. The DNA molecule may be circular or linear. In order to make this program applicable to a large group of users, this datafile was set up in the same format as the output of the BACHREST and INTREST programs described by Fristensky, Lis, and Wu.<sup>1</sup> Therefore, a restriction map can be generated from a known sequence using the latter programs and used as input to KLONER, or a file of similar format may be created manually if the sequence is not available, Fig. 1.

INSERT Name	Configuration: Recognition Sequence	Cut	CIRCULAR # of Sites	Length: 1086 bp Sites
AVA2	GGQCC	1	1	4
BAMH1	GGATCC	1	2	2 150
HPA2	CCGG	1	1	113
PVU2	CAGCTG	3	1	74

Fig. 1. Input datafile which was created manually for a DNA molecule from which a fragment will be isolated and inserted into pBR322.

---

### Data Structures

The data within the program is arranged in order to resemble actual laboratory parameters. The user begins with a series of empty tubes which can be filled with DNA, represented by restriction maps, from the data files. The DNA molecules are stored as linked lists of fragments, each consisting of a double-linked list of subfragments which contain all of the restriction enzyme cleavage sites within the fragment. Each fragment record contains information on the size and the origin of the fragment as well as the left and right 'sticky' ends. Each subfragment includes the start and stop sites and the delimiting restriction enzymes. Each restriction enzyme is read into a sequential list along with all of the necessary information to determine which type of ends that particular enzyme generates after digestion. Enzymes which cleave at multiple sites can either be specified by A and B forms of the enzyme or by specifying the ambiguity following the conventions of Dayhoff, et al.<sup>4</sup>

### PROGRAM OPTIONS

A list of the main options in this program, is shown in Fig. 2. A flowchart of a sample execution of this program, Fig 3., demonstrates how these options can be combined to simulate an actual experiment.

- A-Aliquot tube contents to other tube(s)
- B-Bal31 nuclease
- C-Change label on tube
- D-Digest with a restriction enzyme
- E-Empty current tube
- F-Fill in sticky ends with pol1 and NTP(S)
- G-Get another tube
- H-Help
- I-Isolate fragment(s)
- L-Ligate
- M-Mix two tubes together
- N-reNumber fragments in tube
- P-Partial digestion with restriction enzyme
- R-Restriction map of a given fragment
- S-S1nuclease
- T-Terminal transferase
- V-saVe tube contents on disk
- W-Write output somewhere else
- X-eXamine contents of all tubes

Fig. 2. Options available from main menu in the KLONER program.

---

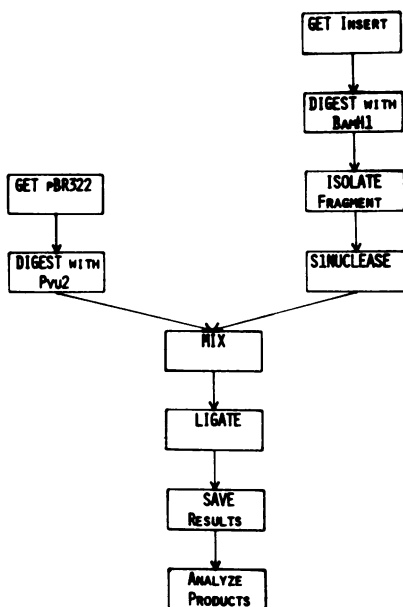


Fig. 3. Diagram indicating the process of inserting a fragment of DNA into the PvuII site of pBR322. The input datafile for the molecule which originally contains the insert is shown in Fig. 1. The datafile for pBR322 was created using the BACHREST program.<sup>4,5</sup> The insert fragment is isolated after digestion with BamHI and then treated with S<sub>1</sub> nuclease to produce blunt ends. This is mixed with PvuII digested pBR322 and ligated together using T<sub>4</sub> ligase, Fig. 4. The desired recombinants are then stored in tubes or on disk to be further analyzed or processed.

#### Input, Output, and Transfer Procedures

The KLONER program is set up to allow the handling of one tube at a time. The GETTUBE procedure transfers the execution from the current tube to the contents of another tube and saves the contents of the current tube in memory until it is needed again. If the tube is empty the user is prompted for the name of a input datafile. The user is advised when the amount of available memory begins to run low, thus avoiding serious errors.

By using the SAVETUBE procedure, the contents of a tube may be stored on disk as a list of restriction sites in a datafile having the same format as the input datafile already described. Additional information which is known about a particular DNA fragment, such as the base numbering at the beginning and the end

of the fragment and the sticky ends, is stored in this same file in locations which do not alter the main structure of this file. This includes additional information stored in the heading and at the end of the file. If a tube contains more than one fragment, then all of the fragments are stored in the same datafile. All files created by this procedure may be read back into KLONER. In addition, files which contain a single circular molecule may be read into other programs such as MAP and DIGEST which are part of the program package described by Fristensky, et al,<sup>1</sup> without alteration. The contents of a tube may also be transferred while still in memory by either the MIX or ALIQUOT procedures. The MIX procedure empties one of the two tubes mixed, while the ALIQUOT procedure simply makes another copy of the contents of one tube and adds it to another.

Individual DNA fragments within a tube may be isolated by the ISOLATE procedure. This is analogous to isolation of fragments from a gel or column and all undesired fragments can be eliminated.

The output from the program which describes the contents of a tube can be sent to either the console, printer, or to a file during the execution of the program.

#### Restriction Enzyme Digestion

The COMPLETEDIGEST procedure digests the contents of a tube with a restriction endonuclease, leaving the fragments generated by this action in the original tube, Fig 3. These fragments now have the sticky ends which that enzyme creates. PARTIALDIGEST acts in a similar manner except that the number of sites is determined first and the user is then allowed to abandon the procedure if the number of sites, and thus the number of fragments which would be generated, is too large.

#### Nucleases

BAL31 will remove a given number of base pairs from each end of all linear fragments in a tube. The resulting sizes are recalculated and the corresponding subfragments are removed with all sticky ends being converted to blunt ends. The ends of all fragments in a tube can also be converted to blunt ends through the execution of the S1NUCLEASE procedure.

**Polymerases**

Polymerases are often helpful in modifying the sticky ends of DNA fragments before ligation. FILLIN is a procedure which simulates the Klenow fragment of DNA polymerase in filling in 5' sticky ends by the addition of one or more nucleotides. TERMTRAN is a procedure which allows the addition of a homopolymer of nucleotides by terminal transferase reactions. This tailing is possible with any type of fragment terminus.<sup>6</sup>

**Ligases**

The LIGATE procedure in this program is designed to first calculate all of the possible circular DNA molecules which can result from ligation of one, two, or three DNA fragments in all possible orientations while limiting the number of products by simulating selection and screening of resulting plasmids. The option of using T<sub>4</sub> ligase or E. coli ligase allows one to avoid blunt end ligation by using the latter.<sup>7</sup> The LIGATE procedure allows one of the fragments which is being ligated to be referred to as the vector fragment. Any other fragment can be identified as having a selectable marker. A successful ligation results in

TUBE NO 1 PBR322+INS

Number	Size	Start	Finish	Sticky Ends
1)	4362 BP	2068 PVU2	2067 PVU2	BLUNT BLUNT
2)	147 BP	2 BAMH1	149 BAMH1	BLUNT BLUNT

LIGATION RESULT

PLASMID 1	2068 PVU2 ....2067 PVU2	+ 2 BAMH1....149 BAMH1	
PLASMID 2	2068 PVU2 ....2067 PVU2	+ 2 BAMH1....149 BAMH1	+ 2 BAMH1....149 BAMH1
PLASMID 3	2068 PVU2 ....2067 PVU2	+ 2 BAMH1....149 BAMH1	+ 149 BAMH1....2 BAMH1
PLASMID 4	2068 PVU2 ....2067 PVU2	+ 149 BAMH1....2 BAMH1	
PLASMID 5	2068 PVU2 ....2067 PVU2	+ 149 BAMH1....2 BAMH1	+ 2 BAMH1....149 BAMH1
PLASMID 6	2068 PVU2 ....2067 PVU2	+ 149 BAMH1....2 BAMH1	+ 149 BAMH1....2 BAMH1

Fig. 4. The contents of tube 1 are shown along with the products of ligation. Fragment 1, pBR322, was designated as the vector and fragment 2 as an insert with a selectable marker. This program is designed such that each recombinant molecule contains one copy of the vector fragment (if there is one indicated) and one or more copies of any fragment with a selectable marker.

the ligation of a vector fragment with one or more fragments with selectable markers. This allows multiple copies of an insert to be ligated into a new plasmid while disregarding all ligation products which do not contain the vector fragment. The desired molecules then may be stored into separate tubes to be further analyzed or saved on disk for future reference, Fig 4. The new plasmids can also be renumbered starting with any restriction enzyme site.

### DISCUSSION

The program described in this paper allows the user to apply many currently used techniques in recombinant DNA to restriction maps of DNA molecules of known or unknown sequence and to predict the restriction map of the product. This program is especially useful when there are many steps involved in the cloning or when the process required in order to obtain the desired end product involves multiple digestions, partial digestions, or multiple inserts.

This program should be adaptable to all Pascal systems with possible minor changes. Copies of this program and further documentation may be obtained by Apple Pascal users by sending one blank 5 1/4 inch diskette, return postage paid.

### ACKNOWLEDGEMENTS

I would like to thank Dr. P.C. Huang for his encouragement and helpful discussions and for the use of his computer in writing this program.

### REFERENCES

1. Fristensky, B., Lis, J., and Wu, R. (1982) Nucl. Acids. Res., 10, 6451-6463.
2. Larson, R. and Messing, J. (1982) Nucl. Acids. Res., 10, 39-49.
3. Pustell, J. and Kafatos, F.C. (1982) Nucl. Acids. Res., 10, 51-59.
4. Dayhoff, M.O., Schwartz, R.M., Chen, H.R., Hunt, L.T., Barker, W.C., and Orcutt, B.C. (1981) Nucleic Acid Sequence Database, Vol. 1, National Biomedical Research Foundation, Washington, D.C.
5. Sutcliffe, J.G. (1979) Cold Spring Harbor Symp. Quant. Bio., 43, 77-90.
6. Deng, G. and Wu, R. (1983) Meth. Enzymol., 100, 96-116.
7. Higgins, N.P. and Cozzarelli, N.R. (1979) Meth. Enzymol., 68, 50-71.