GELYSIS: Pascal-implemented analysis of one-dimensional electrophoresis gels

Peter J. Russell¹, Richard E. Crandall² and Rhonda Feinbaum¹

¹Biology Department, and ²Physics Department, Reed College, Portland, OR 97202, USA

Received 1 July 1983

ABSTRACT

GELYSIS (GEL ANALYSIS) is a Pascal implementation of an algorithm that takes data on the distances migrated by DNA fragments in a one-dimension electrophoretic gel and calculates the sizes of the fragments based on the known sizes of marker fragments. A key step in the overall algorithm is a cubic-spline best-fit of marker calibration data. An option of the program is the generation of graphic representations of the gel and of the calibration spline that are convenient for primary data storage or for publication. Provision is made for specifying different band intensities in the graphic output of the gel.

INTRODUCTION

The use of computers for the storage and analysis of research data has increased significantly in recent years, particularly in the area of nucleic acids research (1). One routine type of experiment is the separation of DNA species, or of DNA fragments generated by restriction endonuclease digestion, by agarose-gel, acrylamide-gel, or agarose-acrylamide composite-gel electrophoresis. Typically this involves a one-dimensional separation in a slab gel (2). This paper describes a Pascal implementation that generates a best-fit cubic spline of size (in kilobase pairs (kb)) vs. distance for marker DNA fragments and, using that curve, computes the size of DNA fragments in lanes of experimental samples. As an option, the program produces a graphic representation of the banding pattern and of the marker calibration curve.

PROGRAM DESCRIPTION**System requirements**

The program described is written in Standard Pascal and was developed on a DEC (Digital Equipment Corporation) PDP 11/70 computer running under the UNIX operating system (UNIX is a trade mark of Bell Telephone Laboratories, Murray Hill, New Jersey). The graphics output is designed either for a Hewlett-Packard 7470A plotter, or for a Tektronix 4012 graphics terminal with

associated hard copy unit (e.g. Tektronix 4631). In the former case, all options can be controlled from a standard terminal that, along with the plotter, is interfaced with the computer.

The graphing program, GELYSIS

The program GELYSIS (GEL analysis) takes input from a computer file containing data for the positions of bands of DNA fragments in a one-dimension electrophoresis gel. Each data file should contain one set of marker data (i.e. that for which DNA fragment sizes are known), and the sets of experimental data for which the marker data are appropriate to compute DNA fragment sizes. The marker data set must be at the beginning of the file. A maximum of 20 sets of data, including the marker set, is allowed in one data file. A set of data begins with a title (up to 25 characters) which will appear in the GELYSIS outputs. The general format of the data file is:

```
[name of marker]
1 [distance] [size] [intensity]
2 [distance] [size] [intensity]
.
.
[name of first experimental]
1 [distance] [intensity]
2 [distance] [intensity]
.
.
[name of second experimental]
1 [distance] [intensity]
.
.
etc.
```

The first column is the number of the band, counting from the origin. The definitions of the data entries are:

```
name:      characters (up to 25); must not start with a numeral.
distance:  band distance from origin, in millimeters.
size:      marker band (DNA fragment) size in kilobase pairs (kb).
intensity: subjective intensity of band, with intensity increasing as
           values are selected from 0 up.
```

The number of spaces between entries in any one row is unimportant, but each line of data must be left-justified.

EXAMPLE OF PROGRAM USE

The use of GELYSIS is demonstrated through the analysis of experimentally-derived data. FIGURE 1 shows a physical map of the recombinant

FIGURE 1. Physical map of the recombinant plasmid pKD002 showing some of the known restriction enzyme sites. This plasmid consists of approximately one-half of the ribosomal DNA repeat unit of the *Neurospora crassa* Oak-Ridge wild-type strain inserted into the HindIII site of pBR322. pKD002 contains two HindIII sites and three EcoRI sites.

plasmid pKD002 showing the cleavage sites for the restriction enzymes EcoRI and HindIII.

An experiment was done in which samples of pKD002 were cut with HindIII, with EcoRI, and with HindIII + EcoRI. The resulting fragments were subjected to electrophoresis in a 1% agarose gel. The marker was HindIII + EcoRI cut lambda DNA. After the gel was stained with ethidium bromide, it was photographed with long-wave ultra-violet light subillumination through red and yellow filters using Polaroid Type 57 film. Transparent plastic rulers wrapped with Parafilm laboratory film (which fluoresces with UV illumination) were placed next to the gel for part of the exposure time. From the photograph, the distance each fragment had migrated was recorded, and the values were entered into a computer data file (called GEL.1 in this example) in the required format as illustrated in FIGURE 2.

The program is run in the UNIX operating system by:

```
GELYSIS <return>
```

The program responds:

```
Input file? :
```

The name of the data file is entered, i.e. GEL.1 here, and GELYSIS echos the file contents to the standard output and calculates the relevant calibration curve for the marker as a best-fit cubic spline in the form:

$$\ln(\text{size}) = aD^3 + bD^2 + cD + d$$

```
HindIII+EcoRI x lambda
1 22 21.7 3
2 48 5.24 2
3 49.5 5.05 2
4 56 4.21 2
5 65 3.41 2
6 93.5 1.98 2
7 96.5 1.9 2
8 106 1.57 2
9 113 1.32 2
10 130.5 0.93 1
11 136.5 0.84 1
HindIII x pKD002
1 54.5 2
2 55.5 2
EcoRI x pKD002
1 45 2
2 83 2
3 158 1
HindIII+EcoRI x pKD002
1 54.5 2
2 83 2
3 97 2
4 158 1
```

FIGURE 2. Data file in format suitable for GELYSIS program. File name is GEL.1 for this discussion.

where size is in kb and D is the distance in millimeters. The equation for the best-fit cubic spline is then displayed in the output. This is exemplified for the data in file GEL.1 as follows:

BEST-FIT CUBIC IS:

$\ln(\text{size in kb}) =$

$$-0.00000326 * D^3 + 0.00093897 * D^2 + -0.10531511 * D^1 + 4.95585043$$

where "ln" is natural log, D is in mm, and ^ indicates "to the power of".

A menu of options is then displayed:

```
[c]alibration - make a graph of the spline curve for marker
[g]el         - make a graph of the gel banding pattern
[t]able       - print out a table of calculated fragment sizes
[q]uit        - exit the program
```

An option is selected by typing the character enclosed in the square brackets. As examples of the capabilities of the program, FIGURE 3 shows the spline curve for the lambda marker data in the GEL.1 file produced by the [c] option, and FIGURE 4 shows the graphic representation of the experimental data generated by the [g] option. In the latter, each lane is labeled with the title entered in the computer file. The origin is at the end of the lane nearest to the title.

FIGURE 3. Cubic spline calibration curve generated by the GELYSIS program for the lambda marker fragment data of FIGURE 2.

Selection of the [t] option prints out a table of calculated fragment sizes to the standard output - usually a video display terminal. The most efficient way to obtain a hard copy of the table along with the echo of the original data file and the best-fit cubic spline information, is to channel the program output to a file that can subsequently be printed out. The command sequence for this is:

```
GELYSIS > output filename <return>
gel data filename <return>
t <return>
q <return>
```

FIGURE 4. GELYSIS generated graphic output of the DNA fragment data in FIGURE 2.

TABLE 1.

FINAL TABLE:			
BAND	DIST (mm)	SIZE (kb)	INTENSITY
HindIII+EcoRI x lambda			
1	22.00	21.70	3
2	48.00	5.24	2
3	49.50	5.05	2
4	56.00	4.21	2
5	65.00	3.41	2
6	93.50	1.98	2
7	96.50	1.90	2
8	106.00	1.57	2
9	113.00	1.32	2
10	130.50	0.93	1
11	136.50	0.84	1
HindIII x pKD002			
1	54.50	4.38	2
2	55.50	4.25	2
EcoRI x pKD002			
1	45.00	6.18	2
2	83.00	2.27	2
3	158.00	0.33	1
HindIII+EcoRI x pKD002			
1	54.50	4.38	2
2	83.00	2.27	2
3	97.00	1.82	2
4	158.00	0.33	1

which places the non-graphics program output in a file with the chosen name. TABLE 1 shows the output of calculated fragment sizes for the GEL.1 data of FIGURE 2.

ACKNOWLEDGEMENTS

We thank R. L. Metzenberg, University of Wisconsin, Madison, for providing the plasmid pKD002. This work was supported by grant GM22488 from the National Institutes of General Medical Sciences, NIH, and by Biomedical Research Support Grant RR07168 from NIH. Some (non-computer) equipment used in the experiments described in this paper was purchased with funds from the National Science Foundation, Special Equipment Program Grants PRM-7919955, CDP-8007801, and PRM-8114794. A listing of the program may be obtained from P. J. Russell upon request. The non-graphics parts of the program are available in UCSD Pascal for implementation on an Apple II computer; they can be copied onto a 5.25" disk if one is supplied.

REFERENCES

1. Soll, D. and Roberts, R. J. (1982) The Applications of Computers to Research on Nucleic Acids. IRL Press, Oxford, England.
2. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.