Improved Estimation of DNA Fragment Length from Gel Electrophoresis Data Using a Graphical Method

ABSTRACT

A computer program has been developed for computing DNA fragment size from its electrophoretic mobility using a graphical method. The program uses DNA marker data and selects the semilogarithmic linear range (sl-range), i.e., the linear portion of the semilogarithmic curve (mobility vs. log of DNA fragment length). Over this range a linear interpolation is derived for calculating the size of a DNA fragment whose mobility falls in the sl-range. The program also derives a hyperbolic interpolation formula that covers the entire range for determining the size of a DNA fragment whose mobility is beyond the semilogarithmic linear range. The method described in this paper is sensitive, accurate and reliable. This program can also be used to compute protein or polypeptide size from sodium dodecyl sulfate polyacrylamide gel electrophoresis data. The DOS™ version of the DNASIZE program is freely available from Netserver at EMBL or from BioTechNet® by EMail.

INTRODUCTION

Gel electrophoresis is the standard method for determining the size of various macromolecules such as DNA, proteins and polypeptides. In gel electrophoresis, it is customary to run the markers along with the experimental fragments. DNA marker data are used to construct a standard curve, which serves as the internal calibration for determining the length of the experimental fragments on the gel. Several methods have been proposed for approximating the relationship between migration distance and DNA size (2,4,5,11). Based on these approximations, a number of computer programs have been developed (5,7,10).

The standard method for measuring DNA fragment size is to plot the log of DNA size vs. mobility and to determine the values of the unknown from this plot [the semilogarithmic method (4)]. This method of plotting is useful because the graph is a straight line over a portion of the fragment size range. Unfortunately, it is nonlinear in the high molecular weight range, precluding accurate estimates (11). Southern (11) has shown that a reciprocal relationship [$(M - M_0) \times (L - L_0) = C$, where M is mobility, L is length and M_0, L_0 and C are constants] exists over a wider molecular weight range than for the semilogarithmic method. However, this method is limited to the use of three standards at a time for calculating the size of an unknown fragment. An alternative to this local method of estimation is to perform a least squares fit over the entire range, in order to calculate a single set of constants, M_0, L_0 and C (the global method) (10). Similarly, Studnicka (12) estimates a single set of constants using the hyperbolic regression method. The global method (10,12) yields a single equation, and the contribution of error to the estimates, due to faulty (or misestimated) standards, is less. However, this method is less accurate than the local method (2,3,5). On the other hand, the local method is less reliable because it fits even the aberrant points exactly (5,7).

In this paper, a graphical method for estimating the size of a DNA fragment has been described to overcome the limitations of the global and local methods described above. The graphical method, combining the hyperbolic regression and linear regression methods, derives (i) the linear interpolation formula over the sl-range for calculating fragment length and (ii) the hyperbolic interpolation over the entire range for estimating the size of a DNA fragment whose mobility is beyond the sl-range (9). The computer program can also be used for estimating DNA fragment size from gel electrophoresis data or protein/peptide size from sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) data using the graphical method.

ALGORITHM

Analysis of Standard Data

The graphical method was used to fit the standard data obtained by running the DNA marker with the experimental fragments. The detailed derivation of the graphical method, which is the combination of linear regression and hyperbolic regression, has been described previously by Raghava et al. (8,9). The linear interpolation over the semilogarithmic linear range and the hyperbolic interpolation formula over the entire range were derived from standard data as described below.

Linear Interpolation Formula

The semilogarithmic curve of the log of DNA fragment length vs. mobility of known standard was used for deriving the linear interpolation formula. In this curve, the sl-range, which is the linear portion of the curve, was depicted. The sl-range was selected by using a tunable algorithm (8). Data in the sl-range of the curve were fitted by using the linear equation:

$$M = A_0 + A_1 \times \log_{10}(L) \quad \text{(Eq. 1)}$$

where M, L, A_0 and A_1 represent the mobility, the size of the DNA fragment, the constant and the slope of the curve, respectively. The values of A_0 and A_1 were calculated by fitting the data in the sl-range using a least squares curve-fitting method (1).

Hyperbolic Interpolation Formula

The hyperbolic interpolation formula was derived from a known standard by using hyperbolic regression analysis. It has been shown that a reciprocal relationship between the mobility and the DNA fragment size exists over a wider DNA molecular weight range and can be expressed by the following equation:

$$(M - M_0)(L - L_0) = C_0 \quad \text{(Eq. 2)}$$

where M is mobility, L is length, M_0, L_0 and C_0 are constants. Equation 2 is a rectangular hyperbola whose asymptotes are M_0 and L_0. The constants were determined by fitting standard data to rectangular hyperbola using the least squares curve-fitting method. The detailed derivation of the hyperbolic interpolation formula has been described previously (9,12).

Size Estimation of DNA Fragments

The sizes of experimental fragments were calculated using either equation

Table 1. Computation of Size of Restriction Fragments Using the Graphical and Hyperbolic Regression Methods

Mobility (mm)	Actual Size (bp)	Hyperbolic Method		Graphical Method		
		Computed Size (bp)	Percentage Variation	Computed Size (bp)	Percentage Variation	
20.0	23129	23340	0.9	23340	0.9	
21.9	9416	9451	0.4	9451	0.4	
23.3	6557	6426	-2.0	6426	-2.0	
25.5	4362	4159	-4.7	4159	-4.7	
29.8	2322	2302	-0.8	2302	-0.8	
30.8	2027	2059	1.6	2059	1.6	
32.9	1632	1658	1.6	1658	1.6	
37.8	1030	1063	3.2	1063	3.2	
44.3	602	625	3.8	625	3.8	
46.2	517	536	3.6	536	3.6	
46.4	506	527	4.1	518	2.4	sl-range
49.5	396	407	2.8	396	0.0	sl-range
51.3	344	348	1.1	339	-1.5	sl-range
53.2	298	292	-2.1	288	-3.4	sl-range
56.2	221	215	-2.7	222	0.4	sl-range
56.2	220	215	-2.2	222	0.9	sl-range
60.3	154	128	-16.8	156	1.3	sl-range

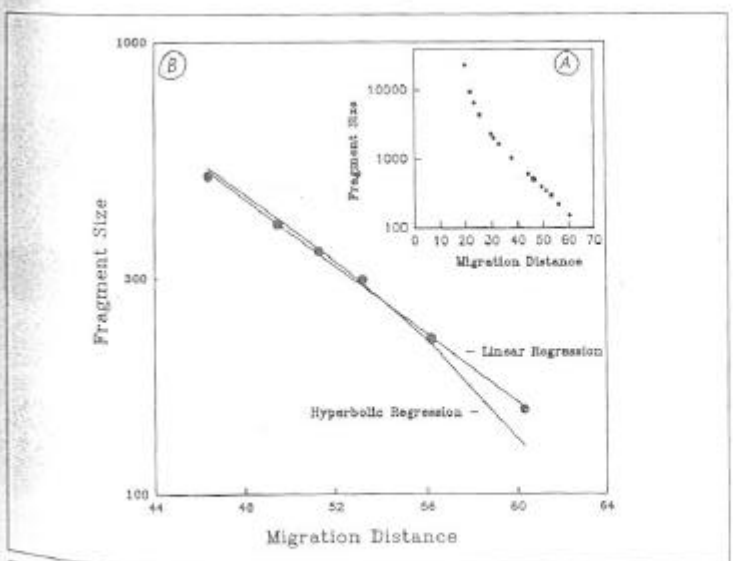

Figure 1. Restriction fragment of λDNA and pBR322 separated by 1% agarose gel (12). A) Semilogarithmic plot (migration distance vs. log of fragment size) for the entire range of actual data. B) Semilogarithmic plot for the sl-range of actual data, data fitted by linear regression and data fitted by hyperbolic regression.

or equation 2 depending on the mobility of the fragments. The program determines whether the mobility of the DNA fragment is in the sl-range. If the mobility is found in the sl-range, equation 1 is used to calculate the size of the DNA fragment; if the mobility is outside the sl-range, equation 2 is used.

COMPUTER PROGRAM

The DNASIZE computer program was written in C language and compiled under Turbo-C for running on an IBM®-compatible computer under the DOS™ environment. The program is fully menu driven and can be run by users with little knowledge of computers. The main options of the program are:

(i) **Input data.** The program allows the user to enter and update the standard (DNA marker) data and unknown data.

(ii) **Size calculation of experimental fragments.** This option allows the user to calculate the size of the experimental fragment using standard data. The size of the experimental protein/peptide fragment can also be calculated by this program using SDS-PAGE data of known markers.

(iii) **Standard curve.** This option allows on-screen display of the standard curve and fit the data using the hyperbolic regression method. The standard curve can also be plotted using an HP-GL-compatible plotter.

(iv) **Miscellaneous.** This option allows the user to display/print the standard data/experimental data and also to exit from the program.

RESULTS

Estimation of DNA Fragment Size

The size and mobility of DNA restriction fragments, separated by 1% agarose gel electrophoresis, were taken from Figure 2B of Studnicka et al. (12). The fragments are a mixture of lambda/*Hin*dIII, pBR322/*Hin*fI and pBR322/*Hin*dIII+*Hin*fI. The actual data, the data fitted to the linear equation and the data fitted to the hyperbola for the sl-range are shown in Figure 1. Figure 1 shows that linear regression fits better than hyperbolic regression in

Table 3. Computation of Molecular Weight of Proteins from SDS-PAGE

Mobility (mm)	Actual Size (kDa)	Hyperbolic Method		Graphical Method		
		Computed Size (kDa)	Percentage Variation	Computed Size (kDa)	Percentage Variation	
2.500	205.000	205.941	0.459	205.941	0.459	
6.200	116.000	115.646	0.306	116.491	0.432	sl-range
8.000	97.400	95.031	2.432	96.790	0.626	sl-range
11.700	66.000	69.230	4.894	66.136	0.206	sl-range
19.500	45.000	43.448	3.450	43.448	3.450	
29.000	29.000	29.367	1.267	29.367	1.267	

the relationship between mobility and DNA length for the entire molecular weight range. The graphical method, combining the two approaches, takes advantage of the increased sensitivity of the linear regression method when in the sl-range and can be applied to the entire range when using the hyperbolic regression method.

Protein/Peptide Size Estimation

It has been shown previously that the size of a protein can be estimated from SDS-PAGE data using the hyperbolic regression method (10,12). Studnicka (12) has shown that the protein molecular weight standard data obtained from SDS-PAGE under different experimental conditions of polyacrylamide gel concentration, temperature, width, etc., fits the hyperbolic curve very well, so this method can be applied to the estimation of protein or polypeptide size.

A mixture of proteins (carbonic anhydrase, 29 kDa; egg albumin, 45 kDa; bovine serum albumin, 66 kDa; phosphorylase B, 97.4 kDa; β-glactosidase, 116 kDa; and myosin, 205 kDa) was run in 10% acrylamide gel containing 0.4% SDS at 100 V for 90 min, and the data were analyzed using the graphical and hyperbolic regression methods. The comparison of both methods is shown in Table 3. As seen in Table 3, the percentage of error in the case of the graphical method (1.07% for entire range and 0.42% for the sl-range) is much lower than that for the hyperbolic regression method (2.13% for the entire range and 2.54% in the sl-range). This shows that the graphical method can be used for estimating protein or peptide size.

the sl-range. The percentage of error between known size and computed size using both methods, i.e., the hyperbolic regression and graphical methods, are shown in Table 1. As seen in this table, the mean of absolute error for the entire range using the graphical method is 1.9%, which is much lower than that obtained by the hyperbolic regression method (3.2%). This significant reduction in error is due to the use of the linear regression method in the sl-range. In the sl-range, the mean of absolute error using linear regression is 1.4% (graphical method) and when using hyperbolic regression it is 4.5%. The maximum variation is 3.4% in the case of linear regression, and 16.8% with the hyperbolic method.

The DNA gel electrophoresis data obtained from a published report (6) of pBR322 DNA (0.4 μg) digested with different restriction enzymes at an agarose concentration of 1.2% were also analyzed by the graphical and hyperbolic regression methods. A comparison of both methods is shown in Table 2. It can be seen that the mean of absolute error for the entire range in the case of the graphical method is 0.47%, which is less than that for the hyperbolic method (0.7%). This indicates that the linear regression method is more effective than the hyperbolic regression method for estimating fragment size in the sl-range. In conclusion, the hyperbolic regression method is a fast and reliable method for obtaining a simple analytical equation that represents

Table 2. Estimation of Restriction Fragment Sizes of pBR322 Using the Graphical and Hyperbolic Regression Methods

		Hyperbolic Method		Graphical Method	
Mobility (mm)	Actual Size (bp)	Computed Size (bp)	Percentage Variation	Computed Size (bp)	Percentage Variation
35.0	4362	4381	0.4	4363	0.0
39.0	3799	3798	0.0	3812	0.3
44.0	3237	3204	1.0	3220	0.5
54.0	2293	2322	1.2	2297	0.2
58.0	2067	2048	0.9	2048	0.9
76.5	1123	1132	0.8	1132	0.8
95.0	561	558	0.6	558	0.6

DISCUSSION

The objective of this study was to develop a sensitive, accurate and reliable method for estimating DNA fragment size and estimating the molecular weight of proteins or polypeptides from SDS-PAGE data. A computer program was also developed to automate the estimation using the method described above. Presently, there are two types of methods, i.e., local and global, generally used to estimate DNA fragment size. Generally, the local method is more accurate in comparison to the global method, but the error contribution is high due to (i) faulty or misestimated standards and (ii) the unknown that lies outside the range of standards. The global method, by using a unified equation, provides an alternative for overcoming these limitations; however, the global method is less accurate because none of the relationships between size and mobility exists over a wide range (2,3).

The graphical method described in this paper combines a global method, i.e., the hyperbolic method, with the linear regression method, thus significantly increasing the accuracy of DNA fragment size estimation in the sl-range. The linear regression method used in this paper is different from earlier linear regression methods where the entire curve was fitted by the linear regression and the least squares curve-fitting methods; in our linear regression method, only the set of data points in the linear portion of the curve is selected and fitted to a straight line. For calculating the size of a fragment having mobility beyond the sl-range, the hyperbolic regression method is used. Thus, the graphical method described here is accurate, sensitive and applicable for a wider range.

A computer program has been developed that is based on the graphical method for estimating DNA fragment size. In addition to this, the program allows one to present a comparison of standard curves obtained from gel electrophoresis data by running the gel under different sets of conditions. The program is written in C and can be run on an IBM PC or compatible computer under MS-DOS® version 2.0 or higher and requires a CGA, EGA or VGA card. It runs without a math coprocessor, and a hard copy of a graph can be obtained. A copy of the executive version of the program DNASIZE can be obtained from the author (Internet: raghava@imtech.ernet.in), from BioTechNet® or from Netserver at EMBL.

REFERENCES

1. **Agrewala, J.N., G.P.S. Raghava and G.C. Mishra.** 1993. Measurement and computation of murine interleukin-4 and interferon-γ by exploiting the unique abilities of these lymphokines to induce the secretion of IgG1 and IgG2a. J. Immunoassay *14*:83-97.
2. **Elder J.K. and E.M. Southern.** 1983. Measurement of DNA length by gel electrophoresis II: Comparison of methods for relating mobility to fragment length. Anal. Biochem. *128*:227-231.
3. **Elder J.K. and E.M. Southern.** 1987. Computer-aided analysis of one dimensional restriction fragment gels, p. 165-172. *In* M.J. Bishop and C.J. Rawlings (Eds.), Nucleic Acid and Protein Sequence Analysis - A Practical Approach. IRL Press, Oxford.
4. **Fisher, M.P. and C.W. Dingman.** 1971. Role of Molecular conformation in determining the electrophoretic properties of polynucleotides in agarose-acrylamide composite gels. Biochemistry *10*:1895-1899.
5. **Garlepy, C.E., M.J. Lomax and L.I. Grossman.** 1986. SPLINT: a cubic spline interpolation program for the analysis of fragment sizes in one-dimensional electrophoresis gels. Nucleic Acids Res. *14*:575-581.
6. **Gough, E.J. and N.M. Gough.** 1984. Direct calculation of the size of DNA fragments separated by gel electrophoresis using programmes written for a pocket calculator. Nucleic Acids Res. *12*:845-853.
7. **Grabner, M. and R. Hofbauer.** 1991. A computer program for molecular weight determination of DNA fragments (HOWBIG). Comput. Appl. Biosci. *7*:317-319.
8. **Raghava, G.P.S. and J.N. Agrewala.** 1994. Method for determining the affinity of monoclonal antibody using non-competitive ELISA: A computer program. J. Immunoassay *15*:115-128.
9. **Raghava, G.P.S., A.K. Joshi and J.N. Agrewala.** 1992. Calculation of antibody and antigen concentrations from ELISA data using a graphical method. J. Immunol. Methods *153*:263-264.
10. **Schaffer, H.E. and R.R. Sederoff.** 1981. Improved estimation of DNA fragment lengths from agarose gels, Anal. Biochem. *115*:113-122.
11. **Southern, E.M.** 1979. Measurement of DNA length by gel electrophoresis. Anal. Biochem. *100*:319-323.
12. **Studnicka, G.M.** 1987. Hyperbolic regression analysis for kinetics, electrophoresis, ELISA, RIA, Bradford, Lowry, and other applications. Comput. Appl. Biosci. *3*:9-16.

I thank Mr. Amit Choudhary for performing SDS-PAGE of protein samples and Dr. G. Sahni and Dr. G.C. Varshney for critically reading the manuscript. This is communication 23/93 from the Institute of Microbial Technology, Chandigarh, and supported by grants from the Council of Scientific and Industrial Research, and the Department of Biotechnology, Government of India. Address correspondence to G.P.S. Raghava, Institute of Microbial Technology, Sector 39A, Chandigarh-160 014, India.

Received 30 November 1993; accepted 22 April 1994.

G.P.S. Raghava
Institute of Microbial Technology
Chandigarh, India

CATRIMOX-14™

. . .a simple solution for isolating RNA

Catrimox-14™ reagent is a novel cationic surfactant which lyses cells and simultaneously precipitates RNA (*Nature* 362: 186-188, 1993). RNA can be recovered by one of several simple methods. A kit for isolating RNA from twenty samples of whole blood or other cell suspensions is offered. Catrimox-14™ reagent can also be purchased separately with instructions.

Iowa Biotechnology Corp

211 TECHNOLOGY INNOVATION CENTER, OAKDALE IA 52319
(319)335-4125 (800)392-2810
FAX(319)335-4127

Circle Reader Service No. 184