A Web Server for Computing the Size of DNA/Protein Fragments using a Graphical Method

GPS Raghava
Institute of Microbial Technology, Chandigarh, India

Abstract

A web server has been developed for computing the size of DNA/protein fragments from their electrophoretic mobilities using a graphical method. This server is based on the computer program DNASIZE (Raghava 1994). It uses DNA marker data and selects the semilogarithmic linear range (sl-range), i.e., linear portion of the semilogarithmic curve (mobility vs. log of DNA fragment length). Over this range, a least-squares interpolation is derived for calculating the size of DNA fragments whose mobility falls in sl-range. The program derives hyperbolic interpolation formula over the entire range for determining the size of DNA fragments whose mobility is beyond the semilogarithmic linear range. This server can also be used to compute the size of proteins or polypeptides from SDSPAGE. This web server utilizes a CGI program written in Perl and Javascript, which makes the server live and interactive (imtech.res.inf raghavaf dnasizef).

Introduction

Gel electrophoresis (GE) is the standard method which is commonly used in biological labs for determining the size of DNA, proteins, and polypeptides. In GE, one runs a marker, whose size is known, along with the experimental fragments. A standard curve is drawn from the DNA marker data, which serves as internal calibration for determining the length of the experimental fragments on the gel. Numerous methods have been developed in the past to estimate the size of DNA fragments. There are two types of methods (local and global) that are generally used to estimate the size of DNA fragments. Generally, local methods are more accurate in comparison to global methods, but the error contribution is high due to either faulty (or misestimated) standards or if the unknown is outside the range of the standards. On the other hand, global methods provide an alternative to overcome the above limitations using a unified equation; but the global methods are less accurate because none of the relationships between size and mobility hold accurately over a wide range (Elder & Southern 1983 and 1987).

Raghava (1994) described a method for estimating the size of DNA or protein fragments with high accuracy using a graphical method. The graphical method combines both global (hyperbolic regression) and local (linear regression) methods. The graphical method derives 1) the linear interpolation formula over the sl-range (a linear portion in the standard curve) for calculating fragment length and 2) the hyperbolic interpolation over the entire range for estimating the size of DNA fragments whose mobility is beyond the sl-range. Raghava (1994) also developed a program called DNASIZE for estimating the size of DNA fragments from gel elec trophoresis or the size of proteins or peptides from SDS-PAGE data,

using the above graphical method. However, although DNASIZE is a powerful program written in C for MS-DOS/Windows, one needs to load this program onto their computer in order to use it. DNASIZE is written for MS-DOS/Windows, so it cannot be implemented on UNIX or Mac machines. In order to provide wide accessibility of the program to users, the author developed a web server based on DNASIZE.

Algorithm

The web server uses the graphical method to fit the standard data obtained by running the DNA marker along with the experi mental fragments. The detailed derivation of the graphical method, using a combination of linear regression and hyperbolic regression, has been described previously (Raghava 1992, 1994, and 1995).

Linear Interpolation: The semilogarithmic curve of the log of DNA fragment length vs. the mobility of known fragments was used to fit a standard curve using a linear equation. In the first step, the linear range in the standard curve was detected; this is called the semilogarithmic linear (sl) range. In the second step, the data was fitted by using a linear equation in the sl-range as given below.

Equation 1:

$$M = Ao + A_1 Log_{10}(L)$$

where M, L, Ao, and Al represent the mobility, size of the DNA fragment, constant, and the slope of the curve, respectively.

Hyperbolic Interpolation: The hyperbolic interpolation formula was derived from the known standard by using hyperbolic regression analysis. In hyperbolic regression analysis, the curve was fitted using the following equation

Equation 2:

$$(M- Mo)(L- Lo) = Co$$

where M is mobility, L is length, and Mo, Lo, and Co are constants. The constants were determined by fitting standard data to rectangular hyperbolas using the least-squares curve-fitting method. The detailed derivation of the hyperbolie interpolation formula has been described previously.

Graphical Approach: The size of the experimental fragments was calculated using either Equation 1 or Equation 2 depending on the mobility of the fragments. For experimental fragments, the server first checks to determine whether they fall within the sl-range. If they are within the sl-range, then the server computes the DNA size using the linear interpolation formula (Equation 1), otherwise it computes the size by using the hyperbolic interpolation formula (Equation 2).

Web Server

The web site is written in the Perl language under the Digital UNIX environment on a DEC Alpha server (DS-lO). The common gateway interface (CGI) is also written in the programming language Perl. Javascript is used to provide the interactive graphical interface. The main options of the server are as below.

1) Input data: The program allows the user to enter and update the standard (DNA marker) data and unknown data. In the case of a protein, the user should enter the SDS-PAGE data of a known protein fragment and mobility distance. Mter entering the data, the user needs to submit the data to the server.

2) Size calculation of experimental fragments: The server computes the parameter from the input data using the graphical approach for internal calibration. The server opens a page where the user can submit the mobility distance of their experimental fragment. The server will then compute the size of the DNA or protein corresponding to the mobility distance in an interactive fashion using Javascript.

Hardware/Software

Requi rements

Use of the server (imtech.res.in/raghava/dnasize/) requires the user to have access to the Internet and a web browser. These web pages can be loaded onto any computer that can run the web server and the Perl language interpreter. The web pages use Javascript and CGI script, written in Perl.

References

Elder J.K. & Southern E.M. (1983) Measurement of DNA length by gel electrophoresis II: Comparison of methods for relating mobility to fragment length. *Anal. Biochem.* 128,227-231.

Elder J.K. & Southern E.M. (1987). Computer-aided analysis of one dimensional restriction fragment gels, p. 165-172. In: Bishop, M.J. & Rawlings, CJ. (Eds.), Nucleic acid and protein sequence analysis-a practical approach: IRL Press, Oxford.

Raghava, G.P.S., Joshi, A.K. & Agrewala, J.N. (1992) Calculation of antibody and antigen concentrations from ELISA data using a graphical

method.]. *Immunol. Methods* 153, 263-264.

Raghava, C.P.S. (1994) Improved estimation of DNA fragment lengths from gel electrophoresis. *Biotechniques* 17: 100-104.

Raghava, C.P.S. DNAOPT: A computer program to aid optimization of gel conditions of DNA gel electrophoresis and SDS-PACE. *Biotechniques* 18:274-81.

Address for Correspondence

Dr. G.P.S. Raghava, Scientist
Bioinformatics Centre
Institute of Microbial Technology
Sector 39A, Chandigarh
India
E-mail: raghava@imtech.res.in
Web: imtech.res.inlraghava
Phone: +91-172-690557
Fax: +91-172-690632