Optimal Size Calling Methods for Electrophoretic Analysis Utilizing Internal Size Standards

K. Corcoran*, E. Mayrand*, J. Robertson*, T. Shaefer** and M. Kronick*

*Applied Biosystems, Inc., Foster City, CA, USA
**Applied Biosystems, GmbH, Weiterstadt, Germany

A technically demanding problem in analytical genetics is quantitative matching of electrophoretic bands. The number usually associated with a match in forensic science is the size of the nucleic acid molecule in base pairs. Unfortunately, there are many inherent difficulties in the accurate quantitative determination of the size of DNA fragments.

A significant problem associated with size determination is lane-to-lane variation in DNA mobility due to differences in such parameters as salt concentration, amount of DNA loaded, thermal gradients, band location on the gel, and ethidium bromide band shifting effects. By utilizing the fluorescence emission differences of two different dyes, it is possible to place the size ladder in the same lane as the unknown. The ladder fragments and the sample DNA are labeled with the two different fluorophores, respectively. To eliminate errors due to lane-to-lane differences, we have utilized such an in-lane size ladder, which fluoresces in red, to construct a standard curve with which we calculate the sizes of products from PCR reactions utilizing either blue, green, or yellow fluorescing 5'-end labeled primers.

Another problem of major concern is that DNA fragments may interact with the gel matrix in ways that are governed by their DNA sequence, so that mobility is no longer just a function of size. Thus, when such a scenario is operative, two fragments of the same size in base pairs may not co-migrate; whereas, two fragments of dissimilar size may migrate at the same rate. In this communication, we present two size ladders, each of which possesses a band that migrates in an anomalous manner.

Using red labeled fragments from an Alu I digest of pBR322, we constructed calibration curves and calculated the lengths of the Collagen 2A1 VNTR alleles using four different approaches. Calibration curves are obtained from either a cubic splines or second order curve fitting procedure. We also determined the sizes of the alleles using the reciprocal relationship between fragment length and mobility described by E. M. Southern (Edinburgh), using either three points from the size ladder (local form), or all the points (global form). We calculated the sizes of the alleles first with all the restriction fragments included in the size ladder and then calculated using the ladder lacking the 695 bp fragment, which runs as a 708 bp band on agarose gels. We examined DNA samples from 100 unrelated Caucasians for the VNTR locus Collagen 2A1 using the PCR and primers labeled at the 5'-end with fluorophores. The samples were run on 2% agarose gels for 6 h in the Applied Biosystems Model 362 Fluorescent Fragment Analyzer.

The data obtained from 8 samples are depicted in Table 1. It is apparent from the data in the Table that the cubic splines and second order least squares curve fitting procedures yield significantly different results when the size ladder contains the 695 fragment. The results obtained with the reciprocal method are quite different, too, depending on whether a local or global approach is utilized. Note that the sizes obtained with the cubic splines and local Southern methods are similar, as are the data obtained with the second order least squares and global Southern methods.

Advances in Forensic Haemogenetics 4
Edited by Ch. Rittner and P. M. Schneider

Table 1. Size Calls of the Collagen 2A1 VNTR Alleles Calculated with the Alu pBR322 Size Ladder either *Containing* the 695 bp Fragment or *Without* the 695 Fragment

lane	With the 695 Fragment				Without the 695 Fragment			
	cubic splines	2nd order least squares	Southern local	Southern global	cubic splines	2nd order least squares	Southern local	Southern global
1	684	668	679	667	671	670	671	668
2	685,696	670,705	680,696	669,703	673,709	672,707	673,709	670,705
3	685	670	681	668	673	672	673	670
4	685,695	669,702	680,695	668,700	672,706	672,704	672,705	669,701
5	685	669	681	668	673	672	673	669
6	619,685	602,668	641,680	601,667	602,672	605,671	603,672	660,669
7	686,710	671,744	682,705	670,743	675,750	674,747	675,750	671,745
8	686	671	681	669	674	674	674	671

The two repeats reported for Collagen 2A1 from sequencing work are 31 and 34 bp, respectively. When we look for a repeat size pattern in the data set obtained with the complete size ladder, we only see a reasonable correlation with the reported repeat length in the data set obtained with either the second order least squares or the global Southern methods: a repeat of about 30 bp is readily apparent. The data obtained by the other two methods, however, give no indication of what the repeat length might be. In fact, this set indicates there might be partial alleles! Examination of 80 other DNA samples from non-related individuals yielded results similar to those presented here. Also in the Table, we show the results obtained from the four calculations when the anomalously running 695 fragment was deleted from the size ladder by the software. Now, all four methods yield similar results. It is apparent that the repeat pattern is indeed about 30 base pairs.

Figure 1 illustrates two size ladders obtained from restriction digests of either plasmid or phage DNA. Shown are curve fits to the data by either the cubic splines or second order least squares methods. Upon examination of the cubic splines curves, one observes a point, which acts as an origin of irregularity in the flow of the line, caused by the fact that the cubic splines line must pass through each point. Now, if the DNA of unknown length falls within the boundary of the irregularity, it is obvious that the estimation of its size will be weighed by the effect of the anomalously running band of the size ladder on the form of the calibration curve. The second order curves form smoothly flowing lines, and no single point seems to influence the shape of the curve. Rather, one can see that the 695 fragment in the Alu I pBR322 curve and the 508 fragment in the Pst I Lambda curve migrate atypically during electrophoresis, because they lie somewhat off the line describing the best fit. The sequence of these fragments probably contain some subtle feature that causes an unusual interaction with the gel matrix.

Next, we let the software ignore the point corresponding to the fragment that seems to be irregular and see if the cubic splines curve becomes similar to that described by the second order least squares method. In the series of curves depicted in Fig. 2, the fragments 695 and 508 are deleted from the Alu I pBR322 and Pst I Lambda ladders, respectively. The cubic spline curve fits of these methods are shown in the figure and it is observed that theh approach the form of the second order curve fits illustrated in Fig. 1. This result explains why the sizes of the Collagen 2A1 alleles presented in Table 1 are dissimilar with the cubic splines and the second order least squares methods: the alleles fell in the region of irregularity on the calibration curve.

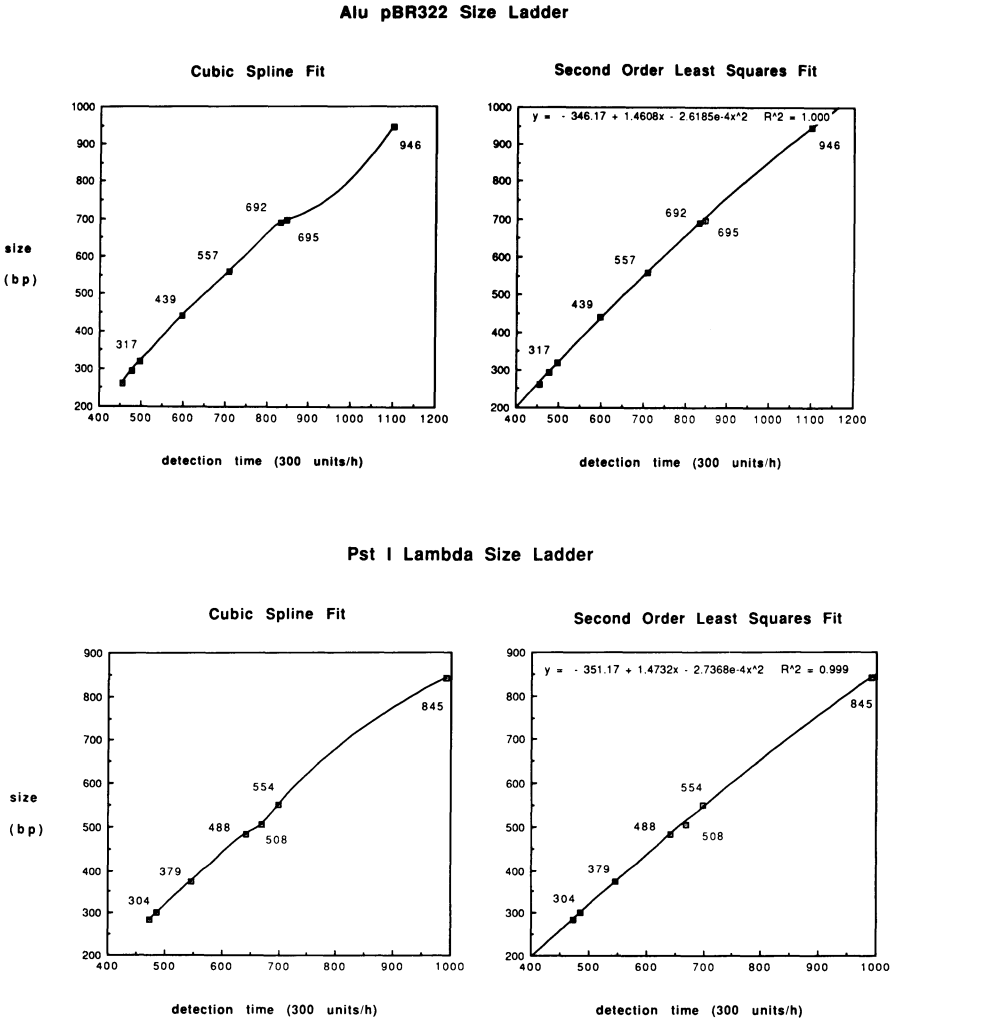

Fig. 1. Two examples showing different curve fitting procedures do not always yield similar results (shown above). Details are given in the text

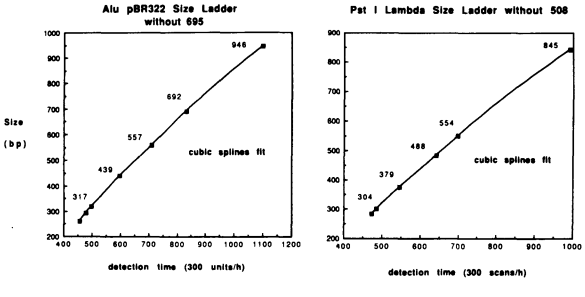

Fig. 2. Removal of bands with irregular mobility from the size ladder can cause different curve fitting methods to yield similar results (shown on the right). See the text for details