

An Accurate Method for Estimating Sizes of Small and Large Plasmids and DNA Fragments by Gel Electrophoresis

By P. A. ROCHELLE, J. C. FRY,* M. J. DAY AND M. J. BALE

Department of Applied Biology, University of Wales Institute of Science and Technology,
PO Box 13, Cardiff CF1 3XF, UK

(Received 3 June 1985; revised 5 September 1985)

Several regression methods were tested for estimating the sizes of a wide range of plasmids (1.37–312 MDa) and restriction fragments (2.2–14.2 MDa) by agarose gel electrophoresis. The most accurate and least variable method was the multiple regression of \log_{10} molecular size against \log_{10} relative mobility and the reciprocal square root of the relative mobility. This method gave a good fit to all the data with low percentage errors of the molecular size estimates ($\leq 3.0 \pm 1.5\%$). It is suggested that with this method the molecular size of unknown plasmids can be accurately estimated using the plasmids from *Escherichia coli* V517 and *E. coli* IR713 as standards.

INTRODUCTION

For many years molecular sizes of plasmids and other pieces of DNA have been estimated by their mobility in electrophoresis gels. Aaij & Borst (1972) showed a linear relationship between the logarithm of molecular size and mobility that could be used to estimate unknown sizes of bacteriophage and mitochondrial DNA up to 10 MDa. This approach was extended for plasmids of up to 95 MDa by using a log/log plot of molecular size and mobility (Meyers *et al.*, 1976). Other workers have consistently been unable to accurately estimate molecular sizes of large plasmids with this method (Willshaw *et al.*, 1979) because the log/log plot is intrinsically curved. More complex approaches have been developed using mobility data obtained from restriction fragments of DNA to obtain linear relationships (Southern, 1979; Schaffer & Sederoff, 1981), but these have rarely been used subsequently. Most currently accepted methods for measuring the molecular sizes of plasmids, such as by electron microscopy (Grindley *et al.*, 1973) and from restriction endonuclease digests (Grinter & Barth, 1976), are much more difficult and time consuming than direct gel electrophoresis. Consequently there is a need for a better method of estimating molecular sizes of DNA from their electrophoretic mobilities, particularly for the larger plasmids. The aim of this work was to develop such a method that would be easy to use routinely.

METHODS

Bacterial plasmids and cultural conditions. Molecular sizes of the plasmids used have all been estimated independently, by methods other than straightforward electrophoresis. In the following list of the plasmids used the reference quoted refers to the molecular size estimation. The plasmids, which were mainly in *Escherichia coli* and *Pseudomonas* spp., were pVA517A-H (Macrina *et al.*, 1978), R300B (Grinter & Barth, 1976), pGSS15, pGSS6 (Barth *et al.*, 1981), pRP1 (Lanka & Barth, 1981), RP4 (Currier & Morgan, 1981), R1 *drd*19 (Jacoby *et al.*, 1983), pWW0 (Lehrbach *et al.*, 1983), TP116 (Grindley *et al.*, 1973), R91, CAM (Jacob *et al.*, 1977), RIP64 (Korfhagen *et al.*, 1978) and pMG1 (Hansen & Olsen, 1978). All cultures were started from single colonies and streaked onto nutrient agar (Oxoid; CM3) plates which were incubated at 30 °C for 18 h. The larger single colonies were then used for the plasmid preparations. All strains were maintained on drug-free nutrient agar without plasmid loss.

DNA procedures. Plasmid DNA was isolated as described by Kado & Liu (1981) for single colonies; following phenol/chloroform extraction the aqueous layer was removed using a wide bore (2.5 mm inside diameter)

disposable 200 μl tip. Using such a tip avoids shearing large plasmids and minimizes disturbance of the interface between the organic and aqueous phases. The restriction endonuclease *EcoRI* and bacteriophage λ DNA were obtained from Sigma. λ DNA was digested as recommended by the supplier.

Gel electrophoresis. Plasmid DNA (50 μl) was mixed with 5 μl of loading dye (50% w/v, glycerol; 0.05 M-Tris buffer; 0.25% w/v, bromocresol purple; pH 7.9), and 25 μl of this mixture was loaded into the wells of a 0.8% (w/v) agarose gel (BDH; product no. 44302); in one experiment a 1.4% (w/v) agarose gel was used. The gel was made up and run in Tris/borate buffer (pH 8.0; Tris, 10.8 g; EDTA, 0.93 g; boric acid, 5.5 g; distilled water, 1 litre). Gels were run at 6 V cm^{-1} for 10 min followed by 12 V cm^{-1} for a further 2 h. In one experiment the gel was run at twice this voltage gradient and in another experiment at 8 V cm^{-1} . DNA was stained for 1 h in ethidium bromide (0.4 $\mu\text{g ml}^{-1}$) and washed in running tap water for 20 min.

Gel examination. Gels were illuminated with an ultraviolet transilluminator (Ultraviolet Products, San Gabriel, California, USA) and photographed (Burton *et al.*, 1982). Negatives were scanned with a Chromoscan 200 (Joyce Loebel, Gateshead, UK) to determine the position of the peak absorbance of each covalently closed circular plasmid or DNA restriction fragment band. Migration distance was measured from the front of each well to the absorbance peaks. Relative mobilities for plasmids and DNA fragments were calculated from the migration distance, as a percentage of the distance moved by the smallest plasmid in *E. coli* V517.

Statistical analysis. Regression analysis was done with Minitab (Ryan *et al.*, 1976) using appropriate transformations (x , $x^{1/2}$, $\log_{10} x$, $1/x$, and $1/x^{1/2}$) of the molecular sizes and relative mobilities of the plasmids. These transformations were chosen because they have been recommended for use when curved regression lines need to be straightened (Velleman & Hoaglin, 1981). A separate regression analysis was done for all the plasmids on each gel and the squared correlation coefficient (R^2) and Durbin-Watson statistic noted (Chatterjee & Price, 1977). The percentage errors of the predicted values for the plasmid molecular sizes were calculated from the residuals (predicted value - actual value) as a percentage of the actual molecular sizes. A multiple regression analysis was also done; the regression equation was

$$\log_{10} M_r = a + b_1(\log_{10} r) + b_2(1/r^{1/2})$$

where M_r is the estimated molecular size, r is the relative mobility and a , b_1 and b_2 are regression coefficients. The method of Southern (1979) used a modified relative mobility (r'):

$$r' = M - M_0$$

where M is the distance moved (mm) by the DNA band and M_0 is an empirical correction factor which helps fit a linear regression line (Southern, 1979).

RESULTS AND DISCUSSION

Eighteen independent gel electrophoresis experiments were done to determine the relative mobility of a wide size range of plasmids (1.37–312 MDa). Regression analysis of these data was done to find a method which could be used to estimate sizes of unknown plasmids within this size range. As anticipated the regression of \log_{10} molecular size against \log_{10} relative mobility (LL) gave a reasonable fit for the smaller plasmids but grossly underestimated the sizes of the largest plasmids (Fig. 1*a, b*; Table 1).

To find a better pair of transformations, regressions were done on all the data with twelve combinations of different transformations. The efficiency of transformations in providing a good fit to a set of data can be assessed by the squared correlation coefficient (R^2) and the Durbin-Watson statistic, both of which were calculated during the regression analysis (Chatterjee & Price, 1977). The R^2 value gives the proportion of the variance in the data explained by the linear relationship between the regression variables. Thus for an R^2 value of 100% all the points would lie exactly on the straight line calculated by the regression, so the higher the value of R^2 the closer the data points are to a straight line. The Durbin-Watson statistic tests for autocorrelation in the residuals; this is only a meaningful test when the data points are ordered, and in this work the data were always ordered by molecular size. Durbin-Watson tables are unusual in that two table values are given. A Durbin-Watson statistic lower than the lowest table value indicates significant autocorrelation; a value higher than the highest table value shows that no significant autocorrelation exists and the result for intermediate values is uncertain. For a bivariate regression analysis with eighteen data points, which was the maximum used here, the table values of the Durbin-Watson statistic were 1.151 and 1.391. Significant autocorrelation between residuals indicates either curvature or a poorly fitting

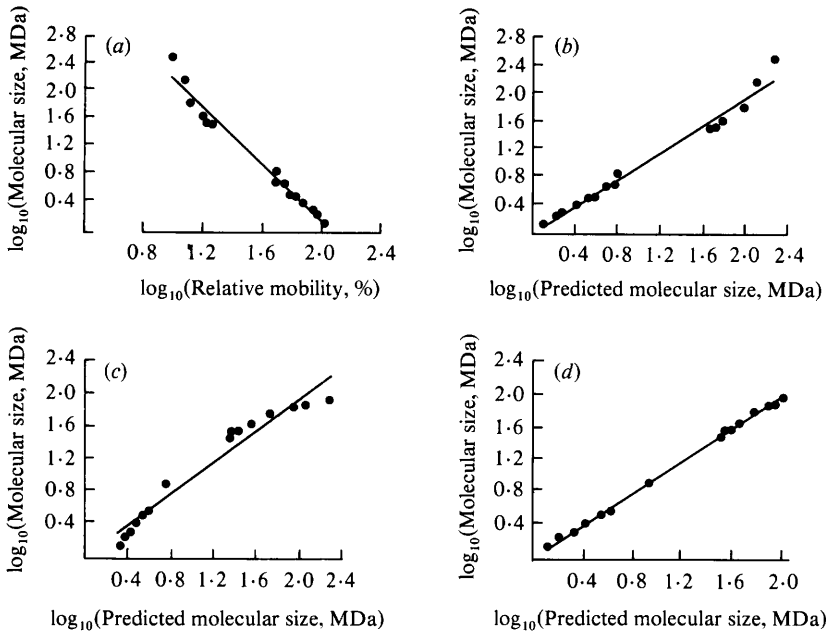


Fig. 1. Typical plots of \log_{10} plasmid molecular sizes against \log_{10} relative plasmid mobility (a) and \log_{10} predicted molecular sizes (b-d). \log_{10} molecular size was predicted by regression of \log_{10} molecular size with \log_{10} relative plasmid mobility (b), reciprocal square root of relative plasmid mobility (c) and both of these variables using multiple regression (d). Regression lines are drawn which, for plots (b)-(d), are also the lines of equality.

straight line; thus the higher the Durbin-Watson statistic the better the regression model fits the data. Using these criteria the most satisfactory transformations were the pair used in LL and \log_{10} molecular size against the reciprocal square root ($1/x^{1/2}$) of the relative mobility regression (LR). For LR the R^2 value and the Durbin-Watson statistic were almost as high as for LL (Table 1).

In the plots of molecular size and predicted molecular size for both LL and LR methods (Fig. 1 b, c) the points did not lie on a straight line. As the curvatures of the points were in opposite directions multiple regression of \log_{10} molecular size against \log_{10} relative mobility and reciprocal square root of relative mobility (MR) should give a better fit to the data. This was clearly true for the typical results presented (Fig. 1 d). Under the electrophoresis conditions tested there was a very good linear relationship between the predicted and real values for the molecular sizes of the plasmids used (1.37-312 MDa). Also MR consistently showed a higher R^2 value and Durbin-Watson statistic than LL or LR. MR on all eighteen sets of data showed much lower percentage errors for the plasmid molecular sizes (Table 1). The mean percentage error for all the molecular sizes of the plasmids examined was very much lower and less variable for MR than for the LL or LR methods. The mean percentage error for the smallest plasmid used (1.37 MDa) was much larger than the rest (11.6%; Table 1); consequently the mean percentage error of plasmids between 1.76 and 312 MDa was only $2.53 \pm 1.06\%$.

The multiple regression equation

$$\log_{10} M_r = a + b_1(\log_{10} r) + b_2(1/r^{1/2})$$

can easily be used because programs and packages are readily available which can do multiple regression on all types of computer, and present no computational difficulties for calculating the regression coefficients.

Table 1. Comparison of four regression methods for calculating mean errors of predicted molecular sizes of plasmids covering a wide size range

Plasmid	Molecular size (MDa)	<i>n</i> *	Mean† error (%) from <i>n</i> values of predicted molecular sizes obtained with the following regression methods:‡			
			LL	LR	MR	S
pVA517H	1.37	11	-3.8	+23.8	+11.6	+62.9
pVA517G	1.76	14	-5.5	+11.1	+2.2	+87.9
pVA517F	2.02	14	-6.6	+4.2	-2.4	+77.9
pVA517E	2.67	14	-4.3	-3.8	-4.1	+61.8
pVA517D	3.38	14	+1.3	-6.5	-2.3	+54.1
pVA517C	3.64	13	+11.5	-1.4	+0.16	+56.2
pVA517B	4.81	14	+19.8	-0.8	+3.5	+50.2
pGSS15	7.35	6	-11.1	-12.1	-5.1	-0.4
pGSS6	8.19	5	+18.2	+15.1	+5.7	+32.7
pVA517A	35.8	15	+2.3	-3.7	+0.08	-9.7
RP4	36.7	10	+4.0	-6.3	+0.38	-11.0
pRP1	44.5	11	+4.3	-5.1	+0.56	-6.3
R1drd19	61.8	11	-8.6	-12.1	-1.9	-14.3
pWW0	74.9	9	+6.2	+8.2	+0.2	-9.1
RIP64	78	7	+14.1	+21.9	+2.8	+8.6
CAM	91.6	4	-35.6	+71.5	+5.5	+52.1
TP116	143.6	3	-15.6	+5.3	-0.57	+3.0
pMG1	312	5	-45.5	-14.9	-5.6	-8.4
Mean§ percentage error ± 95% confidence interval			12.2 ± 5.6	12.6 ± 8.0	3.0 ± 1.5	33.7 ± 14.4
Mean† <i>R</i> ² (%)			97.5	97.1	98.6	96.0
Mean† Durbin-Watson statistic			1.73	1.41	1.97	1.71

* *n*, Number of times the plasmid mobilities were used in the 18 regression analyses.

† Mean values obtained from 18 separate regression analyses.

‡ LL, $\log_{10} M_r$ against $\log_{10} r$; LR, $\log_{10} M_r$ against $1/r^{1/2}$; MR, $\log_{10} M_r$ against $\log_{10} r$ and $1/r^{1/2}$; S, M_r against $1/r$ (Southern, 1979). (M_r , molecular size; r and r' , relative mobilities.)

§ Absolute mean errors obtained from all the plasmids.

Table 2. Comparison of four regression methods for calculating mean errors of predicted molecular sizes of *EcoRI* restriction fragments of bacteriophage λ DNA

Fragment molecular size (MDa)	<i>n</i> *	Mean† error (%) from <i>n</i> values of predicted molecular sizes obtained with the following regression methods:‡			
		LL	LR	MR	S
2.2	5	+2.18	+9.82	+0.73	+0.18
3.12	5	-1.32	-0.69	-1.05	-1.59
3.6	5	-0.39	-1.89	-0.11	+0.15
3.85	5	-0.42	-2.57	-0.03	+0.18
4.91	5	-0.02	-3.59	+0.60	+0.85
14.17	5	+0.22	+1.37	-0.04	-0.06
Mean§ percentage error ± 95% confidence interval		0.76 ± 0.97	3.32 ± 3.51	0.43 ± 0.45	0.50 ± 0.64
Mean† <i>R</i> ² (%)		99.9	98.4	99.9	100.0
Mean† Durbin-Watson statistic		2.08	1.31	2.46	2.42

* † ‡ §. See Table 1.

Southern's method (Southern, 1979), which was designed for determining the size of restriction fragments, was used on our data (S). The results (Table 1) show from the high *R*² and Durbin-Watson statistics that the method gave a good linear fit. However, the errors in the predicted molecular sizes were very large and the mean percentage error for all the plasmids was

Table 3. Comparison of three regression methods for calculating mean errors of predicted molecular sizes of plasmids obtained from a 1.4% agarose gel

Plasmid	Molecular size (MDa)	n*	Mean† error (%) from n values of predicted molecular sizes obtained with the following regression methods:‡		
			LL	LR	MR
pVA517H	1.37	3	+17.9	+30.0	+1.3
pVA517G	1.76	3	+3.6	+7.1	+0.30
pVA517F	2.02	3	-2.5	-1.5	-2.2
pVA517E	2.67	3	-5.3	-8.3	+0.68
pVA517D	3.38	3	-6.1	-9.9	+0.55
pVA517C	3.64	3	-6.4	-10.3	-0.30
pVA517B	4.81	3	+7.3	+9.1	-0.05
R300B	5.7	3	C	C	C
R91	48.1	2	ND	ND	ND
TP116	143.7	2	ND	ND	ND
Mean§ percentage error					
± 95% confidence interval			7.0 ± 4.4	10.9 ± 7.7	0.77 ± 0.64
Mean† R ² (%)			93.1	83.9	99.1
Mean† Durbin-Watson statistic			1.19	1.08	2.40

C, Ran with chromosome.

ND, Not detected because plasmid too large to enter gel at this concentration.

* † ‡ §, See Table 1.

over ten times larger and more variable than our MR method. A modification of Southern's method has also been suggested (Schaffer & Sederoff, 1981). However, both methods are computationally complex and give no advantage over our MR approach. A comparison of the four prediction methods on bacteriophage λ restriction fragments also showed MR to be the best method (Table 2). However, in this case, where the size range of DNA fragments was smaller, the LL and Southern regressions were also much better.

The LL, LR and MR methods were also compared at a higher agarose concentration (Table 3) and at two different voltages (Table 4) for a number of plasmids. In each case MR was the best method, showing the highest R^2 values and Durbin-Watson statistics, but the lowest percentage errors and variabilities. These results suggest that the MR approach will work well under a wide range of electrophoresis conditions.

The degree of supercoiling is known to affect plasmid mobility (Wang, 1980) so the mobilities of pGSS15 and RP4 were tested using six different bacteria as plasmid hosts (three strains of *E. coli* and *Pseudomonas aeruginosa*, *Pseudomonas putida* and *Acinetobacter calcoaceticus*). There was no difference in the mobilities of the plasmids from the different bacteria. Thus MR can be used to accurately estimate the sizes of plasmids isolated from different bacteria because supercoiling appears to be constant in a range of bacteria. However, there is a suggestion that incubation temperature can affect supercoiling (Goldstein & Drlica, 1984). Therefore it would be prudent to use a constant temperature to grow all the hosts in a plasmid sizing experiment.

It would be easier to estimate the size of unknown plasmids on a routine basis if a small number of organisms could supply enough plasmid standards of sufficient size range to calculate a useful regression equation. We tested this possibility by using the pVA517 range of plasmids and TP116 to estimate the sizes of RP4, pRP1, pWW0 and RIP64 which varied in size between 36.7 MDa and 78 MDa. The mean percentage error and 95% confidence limits of the estimated molecular sizes of these five plasmids ($3.63 \pm 1.16\%$) was similar to that observed for MR in Table 1.

Our results indicate that plasmid sizes can be satisfactorily determined over the entire range we have used from gel electrophoresis of the eight plasmids in *E. coli* V517 and the single larger plasmid, TP116 in *E. coli* IR713. Thus the use of multiple regression with these two bacteria provides a simple and rapid method for determining the size of bacterial plasmids and other

Table 4. Comparison of three regression methods for calculating mean errors of predicted molecular sizes of plasmids from electrophoresis gels at two different voltages

Plasmid	Molecular size (MDa)	n*	Mean† error (%) from n values of predicted molecular sizes obtained from gels run at the following voltages:					
			8 V cm ⁻¹			22 V cm ⁻¹		
			LL	LR	MR	LL	LR	MR
pVA517H	1.37	3	-9.1	+24.7	+9.2	-13.7	+20.6	+0.69
pVA517G	1.76	3	-6.2	+8.9	+1.8	-5.6	+8.9	+0.35
pVA517F	2.02	3	-8.6	+1.3	-3.4	-6.9	+2.3	-3.2
pVA517E	2.57	3	-3.4	-3.3	-2.6	-5.5	-4.4	-5.2
pVA517D	3.38	3	-1.9	-5.5	-2.6	+1.5	-3.2	-0.71
pVA517C	3.64	3	+1.8	-5.2	-2.7	+2.9	-2.9	+0.18
pVA517B	4.81	3	+6.9	-2.7	+1.4	+7.6	-2.2	+3.6
R300B	5.52	3	+10.5	-0.14	+4.4	+10.7	+0.91	+4.4
pVA517A	35.8	3	+2.8	-0.76	+0.65	+2.4	-1.3	+0.90
R91	48.1	2	-0.31	-2.52	-0.39	-3.5	-6.6	-4.7
TP116	143.6	2	-3.4	+2.54	+0.13	-1.6	+4.4	-1.3
Mean§ percentage error ± 95% confidence interval			5.0 ± 2.2	5.2 ± 4.4	2.6 ± 1.6	5.6 ± 2.5	5.3 ± 3.6	2.3 ± 1.2
Mean† R ² (%)			99.4	99.6	99.8	99.3	99.1	99.6
Mean† Durbin-Watson statistic			0.63	0.97	1.23	0.78	1.63	2.01

* † ‡ §. See Table 1.

covalently closed circular DNA. This represents a considerable improvement over existing electrophoretic methods and has similar precision to other methods of accurate plasmid size determination such as electron microscopy contour lengths.

This work was supported by a Natural Environment Research Council grant and research studentship (M. J. B.) for which we are grateful. We also thank the following people for providing us with bacteria: P. Barth, P. Gowland, G. Jacoby, E. Juni, W. Keevil and P. Williams.

REFERENCES

- AAIJ, C. & BORST, P. (1972). The gel electrophoresis of DNA. *Biochimica et biophysica acta* **269**, 192-200.
- BARTH, P. T., TOBIN, L. & SHARPE, G. S. (1981). Development of broad host range plasmid vectors. In *Molecular Biology, Pathogenicity, and Ecology of Bacterial Plasmids*, pp. 439-448. Edited by S. B. Levy, R. C. Clowes & E. L. Koenig. New York: Plenum Press.
- BURTON, N. F., DAY, M. J. & BULL, A. T. (1982). Distribution of bacterial plasmids in clean and polluted sites in a South Wales river. *Applied and Environmental Microbiology* **44**, 1026-1029.
- CHATTERJEE, S. & PRICE, B. (1977). *Regression Analysis by Example*. New York: John Wiley.
- CURRIER, T. C. & MORGAN, M. K. (1981). Restriction endonuclease analyses of the incompatibility group P-1 plasmids RK2, RP1, RP4, R68, and R68.45. *Current Microbiology* **5**, 323-327.
- GOLDSTEIN, E. & DRLICA, K. (1984). Regulation of bacterial DNA supercoiling: plasmid linking numbers vary with growth temperature. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 4046-4050.
- GRINDLEY, N. D. F., HUMPHREYS, G. O. & ANDERSON, E. S. (1973). Molecular studies of R factor compatibility groups. *Journal of Bacteriology* **115**, 387-399.
- GRINTER, N. J. & BARTH, P. T. (1976). Characterisation of SmSu plasmids by restriction endonuclease cleavage and compatibility testing. *Journal of Bacteriology* **128**, 394-400.
- HANSEN, J. B. & OLSEN, R. H. (1978). Isolation of large bacterial plasmids and characterisation of the P-2 incompatibility group plasmids pMG1 and pMG5. *Journal of Bacteriology* **135**, 227-238.
- JACOB, A. E., SHAPIRO, J. A., YAMAMOTO, L., SMITH, D. I., COHEN, S. N. & BERG, D. (1977). Plasmids studied in *Escherichia coli* and other enteric bacteria. In *DNA Insertion Elements, Plasmids and Episomes*, pp. 607-704. Edited by A. I. Bukhari, J. A. Shapiro & S. L. Adhya. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.
- JACOBY, G. A., SUTTON, L., KNOBEL, L. & MAMMEN, P. (1983). Properties of IncP-2 plasmids of *Pseudomonas* spp. *Antimicrobial Agents and Chemotherapy* **24**, 168-175.
- KADO, C. I. & LIU, S. T. (1981). Rapid procedure for

- detection and isolation of large and small plasmids. *Journal of Bacteriology* **145**, 1365–1373.
- KORFHAGEN, T. R., SUTTON, L. & JACOBY, G. A. (1978). Classification and physical properties of *Pseudomonas* plasmids. In *Microbiology*, pp. 221–224. Edited by D. Schlessinger. Washington: American Society for Microbiology.
- LANKA, E. & BARTH, P. T. (1981). Plasmid RP4 specifies a DNA primase involved in its conjugal transfer and maintenance. *Journal of Bacteriology* **148**, 769–781.
- LEHRBACH, P. R., MCGREGOR, I., WARD, J. M. & BRODA, P. (1983). Molecular relationships between *Pseudomonas* IncP-9 degradative plasmids TOL, NAH, and SAL. *Plasmid* **10**, 164–174.
- MACRINA, F. L., KOPECKO, D. J., JONES, K. R., AYERS, D. J. & MCCOWEN, S. M. (1978). A multiple plasmid containing *Escherichia coli* strain: a convenient source of size reference molecules. *Plasmid* **1**, 417–420.
- MEYERS, J. A., SANCHEZ, D., ELWELL, L. P. & FALKOW, S. (1976). Simple agarose gel electrophoretic method for the identification and characterisation of plasmid deoxyribonucleic acid. *Journal of Bacteriology* **127**, 1529–1537.
- RYAN, T. A., JOINTER, B. L. & RYAN, B. F. (1976). *Minitab Student Handbook*. North Scituate, USA: Duxbury Press.
- SCHAFFER, H. E. & SEDEROFF, R. R. (1981). Improved estimation of DNA fragment lengths from agarose gels. *Analytical Biochemistry* **115**, 113–122.
- SOUTHERN, E. M. (1979). Measurement of DNA length by gel electrophoresis. *Analytical Biochemistry* **100**, 319–323.
- VELLEMAN, P. F. & HOAGLIN, D. C. (1981). *Applications, Basics, and Computing of Exploratory Data Analysis*. Boston, USA: Duxbury Press.
- WANG, J. C. (1980). Superhelical DNA. *Trends in Biochemical Sciences* **5**, 219–221.
- WILLSHAW, G. A., SMITH, H. R. & ANDERSON, E. S. (1979). Application of agarose gel electrophoresis to the characterization of plasmid DNA in drug-resistant enterobacteria. *Journal of General Microbiology* **114**, 15–25.