A Simplified Electrophoretic System for Determining Molecular Weights of Proteins

By CLYDE MANWELL

Department of Zoology, University of Adelaide, Adelaide, South Australia 5001, Australia

(Received 10 February 1977)

Electrophoresis of 31 different proteins in commercially prepared polyacrylamide gradient gels, Gradipore, yields a linear relationship between a hypothetical limiting pore size (the reciprocal of a limiting gel concentration, G_L) and the cube root of the mol.wt., over the range 13 500–900 000. A regression analysis of these data reveals that 98.6% of all variability in $1/G_L$ is explained by the molecular weight, and this degree of accuracy compares favourably with existing methods for the determination of molecular weight by retardation of mobility in polyacrylamide. This new procedure has the additional advantages that molecular-weight standards can be obtained from readily available body fluids or tissue extracts by localizing enzymes and other proteins by standard histochemical methods, and that the same electrophoretic system can be used in determining molecular weights as is used in routine surveys of populations for individual and species variation in protein heterogeneity.

For the determination of molecular weights of proteins and their subunits, methods based on retardation by gel media during electrophoresis have become popular because they are 'experimentally less demanding but still ... reliable.' (Weber & Osborn, 1969, p. 4406).

The retardation of electrophoretic mobility is described by the following equation, originally empirically deduced by Ferguson (1964) and subsequently given a theoretical foundation by Rodbard & Chrambach (1970):

$$\log (M/M_0) = -K_R G \quad (1)$$

where M is the electrophoretic mobility (dx/dt) at a particular total gel concentration, G , M_0 is the mobility at zero gel concentration (free mobility), and K_R is the retardation coefficient (the logarithmic base being 10 in most treatments).

The usual procedures involve plotting a curve for K_R as an appropriate function of molecular weight (or geometrical mean radius; Rodbard & Chrambach, 1970, 1971) for a series of standards and then finding the molecular weights of unknowns by interpolation. When proteins are electrophoresed in the presence of an anionic detergent, retardation depends more on size alone, and it is sufficient to determine the mobility at a single gel concentration for a given set of standards (Weber & Osborn, 1969). These procedures require special gels distinct from those usually used in routine surveys for individual variation of proteins, and of course the presence of an anionic detergent often results in dissociation to subunits and loss of specific activity.

There are two reasons why it would be desirable to have a method for the determination of molecular weights which could utilize the same gels as are used

in studies on allelic variation of proteins or on species differences.

First, it is common in studies on proteins to observe a multiplicity of zones with activity on the same substrate, e.g., electrophoresis of extracts of individual calanoid copepods reveals over 20 different zones with esterase activity on α -naphthyl acetate; possible homology of some zones could be established by using a range of substrates and inhibitors, but the combination of small size of the organisms with great individual and species variability made a precise count of esterase loci impossible (Manwell *et al.*, 1967; Manwell & Baker, 1970). In a frequently cited paper on genetic variability and strategies of adaptation in animals, the measure of heterozygosity for several species had to be 'increased by 70% over observed values to compensate for the fact that esterases, which are highly polymorphic enzymes ..., were not included in the samples of loci assessed.' (Selander & Kaufman, 1973). With the widespread use of electrophoretic methods to study mechanisms of evolution (Manwell & Baker, 1970; Lewontin, 1974; Nei, 1975), a simple method of measuring molecular weight would be highly desirable to assist in determining the number of protein loci and possible interspecific homologies.

Secondly, where specific enzyme-localization methods exist, it is possible to use well-characterized enzymes from extracts of tissues from better-known species as molecular-weight standards, thus saving in the expense of purchasing a number of purified proteins.

The use of commercially prepared gels would have an additional advantage in that it would facilitate the comparison of protein variation by workers in different countries. The prepared polyacrylamide

gradient gels, Gradipore, provide excellent resolution of such different proteins as the 'hidden' 1-F, 1-S polymorphism of human haptoglobin (Margolis & Kenrick, 1968) and the isoallelic variants of the multiple glucose 6-phosphate dehydrogenases of the protochordate *Branchiostoma* ('amphioxus') (Manwell, 1975).

Electrophoretic Theory and Experiment

In the original description of the polyacrylamide-gradient technique, Margolis & Kenrick (1968, p. 360) referred to protein molecules of different size coming after many hours of electrophoresis to a 'dead stop'. Slater (1969), using a linear gradient in gel concentration, and Anderson *et al.* (1972), using 8M-urea, reported that gradient gels could be used to determine molecular weights.

However, Rodbard *et al.* (1971, p. 138) maintain from theoretical considerations that 'the molecule never comes to a "dead stop"; given enough time, it can reach or exceed any distance.' They plot data from Margolis & Kenrick (1968) and show that $\log(dx/dt)$ does not bear a linear relationship to distance traversed (or to gel concentration, as these data were obtained on a linear gradient).

When it was noticed that the commercially available Gradipore gels were largely of an exponential gradient in concentration (see Fig. 1), it was deduced

Fig. 1. *Logarithm of the gel concentration (%) as a function of distance along the gel (x, in cm)*

The Figure demonstrates that the Gradipore gels are made up of two exponential gradients described by eqns. (5) and (6) in the text. Manufacturer's data for gels made in 1973(O) and in 1975(Δ) are shown.

that this should balance the logarithmic relationship between mobility and gel concentration, the Ferguson (1964) equation (1). In a gel with such a concave gradient, the electrophoretic mobility, dx/dt , is then simply related to a hypothetical limiting distance, x_L , and the actual distance traversed, x , by the equation:

$$dx/dt = k(x_L - x) \quad (2)$$

with initial conditions $x = 0$ when $t = 0$. The solution of eqn. (2) is:

$$\frac{x_L}{x_L - x} = e^{kt} \quad (3)$$

which approximately equals $1 + kt$ for t sufficiently small ($kt < 1$). Then the linearizing approximation is the same as the Lineweaver-Burk transformation of the Michaelis-Menten equation (Dixon & Webb, 1964) and the approximation becomes:

$$\frac{1}{x} - \frac{1}{x_L} = \frac{1}{kx_L} \cdot \frac{1}{t} \quad (4)$$

The questions then become:

1. Will eqn. (4) reasonably describe data on distance migrated as a function of time?
2. Can such a linear approximation be used to estimate a hypothetical limiting pore size, $1/G_L$?
3. Will a plot of $1/G_L$ as a function of molecular weight to the one-third power be linear, as Tombs (1965, 1968) has shown in studies using acrylamide gels of varying porosity?

Materials and Methods

Erythrocytes and minced organs were extracted with a 1% digitonin/0.02M-glycylglycine solution adjusted to pH 8.0 with LiOH, an extracting solution found to be useful in studies on genetic variation of 52 enzymes and other proteins (Baker & Manwell, 1977). For most purposes, 1 ml of centrifuged erythrocytes or 1 g of finely minced organ was extracted with 2 ml of the above solution. The extracts were centrifuged for 14 min at 4000 rev./min; if the supernatant was turbid, the extracts were either filtered through Whatman 3MM paper or diluted with an equal volume of the above extracting solution and re-centrifuged. Serum or haemolymph was diluted with an equal volume of the electrophoresis buffer described below and centrifuged or filtered. Egg white was best diluted with 5 vol. of electrophoresis buffer, to which 5% (w/v) sucrose was added together with enough of an iron solution to saturate the ovotransferrin; one drop of 0.1% ferrous ammonium sulphate solution to 5 ml of diluted egg white was sufficient. Extracts and haemolysates were made within 1 day of the electrophoretic run and kept on ice. The extraction procedure was varied by using frozen organs, freshly thawed, and diluted electrophoretic

EXPLANATION OF PLATE I

Electrophoresis of sheep serum caeruloplasmin in a Gradipore gel, localized by the oxidation of o-dianisidine (Owen & Smith, 1961)

Samples were inserted every hour for 7 h, starting at the left-hand side of the Plate. For the samples on the right, which have been electrophoresed for only 1 and 2 h respectively, the caeruloplasmin zone has weaker *o*-dianisidine ferroxidase activity, and these two weak zones are marked with arrows. The anode is at the top of the Plate. The Plate provides a demonstration of the progressive retardation as the zones, going from right to left, appear to approach a limiting distance. Distances, x , are measured from the bottom of the Plate (sample-insertion region) to the middle of the zone. The reciprocals are plotted as shown for L-lactate dehydrogenase in Fig. 2; x_L is converted into G_L by using Fig. 1 or eqns. (5) and (6). The reciprocal of the limiting gel concentration, $1/G_L$, is converted into an estimate of molecular weight by using Fig. 3, or the regression equation (eqn. 8).

buffer to which 5% sucrose was added without significance changes.

The Gradipore gels were obtained from Gradient Pty. Ltd., P.O. Box 342, Lane Cove, N.S.W. 2066, Australia, through a local distributor. The general electrophoretic procedure was described by Margolis & Kenrick (1968). For many purposes we found that a modified Aronsson-Grönwall Tris/EDTA/borate buffer gave best results (Baker & Manwell, 1977). A stock solution is prepared by dissolving 121 g of Tris (Sigma Chemical Co. 'Trizma' grade), 15.6 g of EDTA (disodium salt, from BDH Chemicals Ltd., Poole, Dorset, U.K.) and 9.2 g of boric acid, all in 1 litre. For electrophoresis the stock solution is diluted 1:20 with water. To improve the resolution of transferrin and caeruloplasmin, 4 mg of ferrous ammonium sulphate was added per litre of diluted electrophoretic buffer. For certain NAD- and NADP-dependent dehydrogenases, electrophoretic resolution is improved by the presence of the appropriate coenzyme in the diluted electrophoretic buffer (40 μ M-NAD⁺ or 10 μ M-NADP⁺ respectively). Haemoglobin and myoglobin were first converted into the carbon monoxide derivative.

As the Gradipore sample inserts hold 14 samples, the usual procedure was to run two different sets of samples simultaneously. Every hour, electrophoresis was stopped briefly and a pair of samples inserted by a disposable micropipette (5 μ l). This procedure provided seven points for each standard on the graph of $1/x$ as a function of $1/t$.

The electrophoretic conditions were a constant 150 V at 12–14°C, the temperature being maintained by addition of ice to the bath which cools the circulating buffer (Margolis & Kenrick, 1968). The current drawn varied between 25 and 35 mA. Gels were 'pre-electrophoresed' for 1 h before insertion of the first pair of samples.

The electrophoretic run terminated 7 h after the introduction of the first pair of samples, and gels were then sliced and stained, although certain brightly coloured proteins such as haemoglobin, iron-ovotransferrin, and haemerythrin could have their distances of migration measured *in situ*. Enzymes were localized on the sliced gels by the use of conventional histochemical procedures (Shaw & Prasad, 1970). Carbonic anhydrase was localized by adapting the usual colorimetric procedure involving *p*-nitrophenyl acetate with the additional check of sensitivity to acetazolamide (Tanis & Tashian, 1971) and serum cholinesterase was localized by adapting the technique of Ellman *et al.* (1961), using acetylthiocholine and 5,5'-dithiobis-(2-nitrobenzoic acid). Haptoglobin and caeruloplasmin were detected by peroxidation and oxidation of *o*-dianisidine respectively (Owen & Smith, 1961). Haemocyanin was detected by methods described elsewhere (Manwell & Baker, 1963). Serum transferrin, ovotransferrin (conalbumin) and

haemerythrin were localized by staining for iron with Nitroso R Salt (Mueller *et al.*, 1962; Manwell, 1964).

Both Amido Black 10B and Coomassie Brilliant Blue G 250 were used in staining for total protein (Maurer, 1971). The usual 2:9:9 (by vol.) mixture of acetic acid, methanol or ethanol, and water was often found to distort the gels sufficiently to make measurements inaccurate with reference to the original gel distances. Coomassie Brilliant Blue gave best results, and the following procedure was used: 100 mg of the dye was dissolved in 100 ml of acetic acid/ethanol/water (1:2:7, by vol.). The solution was heated to 60°C and the sliced gel placed in this for 2–4 h. If inadequate fixing occurred, the acetic acid was replaced by formic acid (to a final concentration of 5%, v/v). After staining, the gels were rinsed repeatedly in acetic acid/ethanol/water (1:5:4, by vol.). A few hours before measurement of the position of total protein-stained zones, the gels were placed in water, 10% (v/v) ethanol, or 10% (v/v) glycerol to restore the size as close as possible to the original.

The distance migrated, x , was measured from the sample insert region. Individual gels varied slightly in length, especially after staining. By using water or 10% glycerol, gels were expanded or shrunk to the range 7.0–7.5 cm in length and the measured distances were corrected to a standard distance of 7.1 cm.

Results

The relative retardation of seven samples of sheep caeruloplasmin after specific staining for this serum protein is shown in Plate 1. The protein is progressively retarded in samples that have from 1 to 7 h for electrophoresis.

That the exponential gradient in gel concentration forms a suitable linear transformation for $1/x$ as a function of $1/t$ is shown for the five isoenzymes of L-lactate dehydrogenase from sheep (Fig. 2). The first point on the curve, at $t = 1$, receives very little weight because this point is sometimes either high or low as a consequence of two opposing factors; the different gradient in the first 2 cm of the gel and the slight variation in the surface of the gels at the point of sample entry. However, if the next six points are taken and a linear-regression analysis (Bailey, 1959) is made for $1/x$ as a function of $1/t$ for each of the five isoenzymes of L-lactate dehydrogenase, the usual measure of linearity, the coefficient of determination, r^2 , is 0.9952, 0.9985, 0.9974, 0.9981 and 0.9976 respectively, for the data on isoenzymes M₄, HM₃, H₂M₂, H₃M and H₄ in Fig. 2. Regression analyses made on 50 different runs of other proteins yielded coefficients of determination in excess of 0.99. In other words, the experimental data confirm the predicted linearity of $1/x$ as a function of $1/t$ to well in excess of 99%.

Fig. 2. Plot of $1/x$ (cm^{-1}) as a function of $1/t$ (h^{-1}) as a test of approximation 4 in the text of the paper for the five isoenzymes of L-lactate dehydrogenase from skeletal muscle of the lamb

The Figure demonstrates that a linear transformation holds and that the lines can be extrapolated to a common point, $1/x_L$, which is not at $1/t = 0$, but at $1/t = 0.032$. Hence the concept of a hypothetical pore limit is valid, but not as a 'dead stop' extrapolation.

The lines for each of the five lactate dehydrogenase isoenzymes in Fig. 2 do not intersect at $1/t = 0$, but at $1/t = 0.032$. The statistical proof is as follows: Regression analysis shows a systematic deviation in that α (i.e. $1/x$ evaluated when $1/t = 0$) tends to increase progressively. For the five isoenzymes, in the order M_4 , HM_3 , H_2M_2 , H_3M and H_4 , the values of α (together with, in parentheses, the 95%-confidence interval, based on the t distribution value 2.776 multiplied by the standard error of α) are: 0.105 (0.193–0.017); 0.135 (0.158–0.112); 0.136 (0.155–0.116); 0.157 (0.168–0.145); and 0.166 (0.175–0.156). For 11 out of 20 possible cases the mean value of α for any one particular isoenzyme lies outside the 95%-confidence interval for α for some other isoenzyme. The probability that a deviation as large as this, or larger, could occur solely by chance, under the null

hypothesis that the lines intersect at $1/t = 0$, is less than 10^{-10} , and thus the null hypothesis is rejected.

However, if a similar test is performed on the values of $1/x$ evaluated not at zero but at $1/t = 0.032$, the systematic deviation disappears and most of the estimates of $1/x$ fall within the 95%-confidence intervals for the other curves. For the five isoenzymes M_4 , HM_3 , H_2M_2 , H_3M and H_4 , the values of $1/x$ evaluated at $1/t = 0.032$ from the regression equations (together with the 95%-confidence intervals based on the t distribution and the standard error of $1/x$ on $1/t$) are: 0.204 (0.295–0.115); 0.180 (0.204–0.157); 0.165 (0.185–0.145); 0.177 (0.188–0.165); and 0.181 (0.191–0.172). Now the estimate of $1/x$ for any particular isoenzyme falls within the 95%-confidence interval for any other curve 16 out of 20 possible times, and three of the exceptions are for isoenzyme M_4 , whose line is subject to the most variability, as the isoenzyme does not move far from the sample insert. For the other four isoenzymes, only one estimate of $1/x$ falls outside the 95%-confidence interval (out of 12 possible); the probability of such an occurrence on chance alone is $1 - 0.54 = 0.46$. In other words, just on chance alone we would expect that of 12 opportunities one or more deviations at the 5% level would occur approx. 50% of the time. We thus accept the null hypothesis that the regression lines intersect near $1/t = 0.032$. A similar intersect point was observed for different haemoglobins of similar molecular weights.

Distances migrated, x , are converted into the equivalent gel concentrations, or x_L is converted into G_L , by using the exponential relationships shown in the linear plots of $\log G$ as a function of x in Fig. 1. The gel concentrations are based on manufacturer's data. Although one might reasonably expect some degree of inter-batch variation, this has been negligibly small in my experience over a 2-year period. The two sets of points in Fig. 1 are for batches of gels received in 1973 and 1975 respectively. The variation of gel concentration with distance is described closely by the following pair of equations:

$$\log G = 0.091x + 0.785 \quad \text{for } x > 2.6 \text{ cm} \quad (5)$$

$$\log G = 0.133x + 0.670 \quad \text{for } x < 2.6 \text{ cm} \quad (6)$$

Regression analysis (Bailey, 1959) reveals that this logarithmic transformation of the exponential relationship between G and x fits the two regions in Fig. 1 with coefficients of determination (r^2) of 0.998 for eqn. (5) and of 0.986 for eqn. (6). That the two regions of the plot of $\log G$ as a function of x are distinct can be tested by testing whether or not the two regression coefficients $\beta_1 = 0.091$ and $\beta_2 = 0.133$ are statistically significantly different (Bailey, 1959, formula 52 on p. 99). The calculated t -statistic from comparing the two regression coefficients is 8.739, whereas with 12 degrees of freedom any deviation of

$t = 4.318$ or more is statistically significant at the $P = 0.001$ level. Thus the manufacturer's gel-concentration data are described by the two distinct curves represented by eqns. (5) and (6); obviously one curve blends into another at the region near 2.6 cm, but this is not important in the present study. What is important is that most of the gel distance is described by a logarithmic relation with an accuracy better than 99%.

The final test of the technique is to plot the estimate of the hypothetical limiting pore size, proportional to $1/G_L$, as a function of the cube root of the molecular weight. These data are organized in Table 1, together with the literature sources for the molecular-weight values used. (It was not always possible to obtain data for a given enzyme from the species used as a convenient reference, the sheep *Ovis aries*; however, data were then taken for other mammals and usually were available for related artiodactyls such as the cow or the pig.)

In agreement with Tombs (1965, 1968), there is a linear relationship between the measure of a limiting pore size ($1/G_L$) and the effective molecular radius, directly proportional to the cube root of the molecular weight (Fig. 3).

The data in Table 1 and Fig. 3 can be analysed by standard linear-regression techniques (e.g. Bailey, 1959), the general equation being:

$$Y = \alpha + \beta X \quad (7)$$

and the specific equation being:

$$1/G_L = -3.034 + 0.1732 (\text{mol.wt.})^{1/3} \quad (8)$$

The coefficient of determination, r^2 , is 0.986, i.e. 98.6% of the total variation in $1/G_L$ can be explained by differences in molecular weight. The standard error of Y on X , in this case $1/G_L$ on $(\text{mol.wt.})^{1/3}$, is 0.3955. The standard error of α is 0.148 and the standard error of β is 0.00280.

Discussion

Although Rodbard *et al.*, (1971) are correct in criticizing the concept of an absolute pore limit, the concept is sufficiently close to reality that it can still be used in determination of mol.wts. of proteins over the range 13 500–900 000 at least. Most of the deviation from linearity in the analysis by Rodbard *et al.* (1971) of the data from Margolis & Kenrick (1968) corresponds to times of migration of 20–30 h. It is not unexpected that the mode of protein migration will change as protein molecules encounter successively smaller pores in the gel matrix. Also, I have observed that, in slicing acrylamide gels that have run for different periods of time, the mechanical resistance changes over time; presumably this reflects changes in the pore structure.

It is, however, necessary to find the best place to estimate $1/x_L$ by experiment. A simple 'dead stop' pore-limit theory would predict that $1/x_L$ should be evaluated at $1/t = 0$, i.e. when time had approached infinity. In fact, as shown in Fig. 2 and the statistical analysis, for families of proteins differing in isoelectric point but not in molecular weight, the point of convergence is not at $1/t = 0$ but at $1/t = 0.032$.

Nevertheless, the concept of a hypothetical pore limit is useful; as the regression analysis of the data in Table 1 and Fig. 3 has shown, this concept explains 98.6% of all the observed variation.

How does the accuracy of the present method compare with other techniques for determining molecular weight by electrophoresis?

Weber & Osborn (1969, p. 4411), using mobility in 0.2% sodium dodecyl sulphate, say 'one is inclined to assume that the technique ... yields molecular weights with an accuracy of better than $\pm 10\%$.' Hedrick & Smith (1968), using disc electrophoresis in polyacrylamide gels with non-denaturing buffers, and using 17 different standards, observed a linear relationship between the relative retardation in gels of different concentration and the molecular weight and claim (p. 162) that 'the average deviation of proteins ... from the line ... is $\pm 4\%$.' Unfortunately, no details of any regression analysis are supplied in these or most other papers using these techniques.

However, Rodbard & Chrambach (1971), using 37 different proteins, provide a regression analysis for retardation coefficient versus molecular weight. They obtained a correlation, $r = 0.97259$ (Table 6 in their paper). That value yields a coefficient of determination, r^2 , of 0.94593, i.e., that 94.6% of the variation in retardation coefficient is explained by the molecular weight.

In the present study, 31 different proteins were used, although the total number of standards was 55, as polymorphic systems with the same molecular weight were counted singly (except that for the L-lactate dehydrogenase isoenzymes which were used to calibrate the system in finding the point of convergence, $1/t = 0.032$). The coefficient of determination for the data in Table 1 and Fig. 3 is 0.986.

Rodbard & Chrambach (1971) stress the importance of expressing the accuracy of a method by stating the 95%-confidence interval for the molecular weight applicable to any single estimate of the retardation coefficient. In Fig. 4 of their paper they show that the 95%-confidence interval for a single determination for a hypothetical protein with a mol.wt. of 100 000 is contained between 20 000 and 180 000. This relatively large 95%-confidence interval arises from the fact that such an estimate includes three sources of error: experimental procedure itself, uncertainty in both α and β of the regression equation and uncertainty arising from variability about the regression line.

Table 1. Estimate of hypothetical limiting pore size, $1/G_L$ (from experiment), versus molecular weight (from biochemical literature)

Protein (name of species), Enzyme Commission no. (if any)	Mol. wt.	Citation (other species)	$1/G_L$	(Mol.wt.) [†]
1. Myohaemerythrin (<i>Themiste cymodoceae</i> , stout sibling species)	13 500	Klippenstein <i>et al.</i> (1972) (<i>Themiste pyroides</i> or <i>T. zostericolum</i> ?)	1.03	23.81
2. α -Lactalbumin (ox)	16 000	Gordon (1971)	1.92	25.20
3. Myoglobin (sheep)	17 000	Han <i>et al.</i> (1972)	1.51	25.71
			1.98	25.71
4. (a) Carbonic anhydrase I (human), EC 4.2.1.1	30 000	Liljas <i>et al.</i> (1972)	2.10	31.07
(b) Carbonic anhydrase II (human)	30 000		1.89	31.07
(c) Carbonic anhydrase II (sheep)	30 000		1.93	31.07
5. Riboflavin-binding protein (chicken egg white)	35 000	Baker (1968)	2.10	32.71
			2.42	32.71
6. β -Lactoglobulin (ox)	36 000	McKenzie (1971)	2.69	33.02
7. Ovalbumin (chicken egg white; average of three closely moving zones)	45 000	Baker (1968)	2.86	35.57
			2.88	35.57
8. Coelomic haemerythrin (<i>Phascolosoma arcuatum</i> , major component)	56 000	C. Appleby, personal communication (see also data on <i>P. agassizii</i> by Liberatore <i>et al.</i> , 1974)	3.68	38.26
			3.72	38.26
9. (a) L-Malate dehydrogenase, NAD-linked, supernatant locus (human), EC 1.1.1.37	62 000	Murphey <i>et al.</i> (1967) (various mammals)	3.40	39.58
(b) L-Malate dehydrogenase, NAD-linked, supernatant locus (sheep)	62 000		4.37	39.58
10. Haemoglobin A (human)	66 000	Rossi-Fanelli <i>et al.</i> (1961)	4.19	40.41
(b) Haemoglobin A ₂ (human)	66 000		4.42	40.41
(c) Haemoglobin A (sheep)	66 000	Huisman <i>et al.</i> (1965)	4.12	40.41
			4.45	40.41
(d) Haemoglobin B (sheep)	66 000		4.07	40.41
			4.48	40.41
11. (a) Serum albumin (human)	69 000	Schultze & Heremans (1966)	4.45	41.02
(b) Serum albumin (ox)	69 000		4.25	41.02
(c) Serum albumin (sheep)	69 000		4.07	41.02
			4.40	41.02
12. (a) Transferrin A (sheep)	77 000	Spooner <i>et al.</i> (1975)	4.15	42.54
(b) Transferrin B (sheep)	77 000		4.66	42.54
			4.70	42.54
(c) Transferrin D (sheep)	77 000		4.37	42.54
13. Ovotransferrin (chicken conalbumin)	80 000	Baker (1968)	4.81	43.09
14. Hexokinase (sheep heart), EC 2.7.1.1	97 000	Easterby (1971) (pig)	4.63	45.95
15. NADH diaphorase (sheep heart, major component), EC 1.6.4.3	100 000	Dixon & Webb (1964) (pig)	4.88	46.42
16. Coelomic haemerythrin (<i>Themiste cymodoceae</i> , stout sibling species)	107 000	Klotz & Keresztes-Nagy (1963) (<i>Phascolopsis gouldii</i>)	5.34	47.48
			5.49	47.48
17. Caeruloplasmin (sheep) = ferro-oxidase, EC 1.16.3.1	130 000	Fee (1975) (several mammals)	5.28	50.66
			5.34	50.66
			5.35	50.66
			5.78	50.66
			5.44	52.17
18. L-Lactate dehydrogenases H ₄ , H ₃ M, H ₂ M ₂ , HM ₃ and M ₄ (sheep), EC 1.1.1.27	142 000	Kaplan (1964) (bovine)		
19. D-Glyceraldehyde 3-phosphate dehydrogenase (sheep liver, major component), EC 1.2.1.12	143 000	Harris & Perham (1968) (pig)	5.72	52.29
20. Haptoglobin 2-1, haemoglobin complex = Hp 2-1, Hb (human)	150 000	Margolis & Kenrick (1968)	5.75	53.13
21. 'Malic enzyme' = L-Malate dehydrogenase (decarboxylating) (major isoenzyme of sheep heart, F variant), EC 1.1.1.40	200 000	Frenkel (1972) (ox, both supernatant and mitochondrial forms)	6.89	58.48

Table 1. — *continued*

Protein (name of species), Enzyme Commission no. (if any)	Mol. wt.	Citation (other species)	$1/G_L$	(Mol.wt.) [†]
22. Haemoglobin (<i>Artemia salina</i> , parthenogenetic race, major component)	230000	Bowen <i>et al.</i> (1976)	7.90	61.27
23. Cholinesterase (sheep serum), EC 3.1.1.8	300000	Schultze & Heremans (1966) (man)	8.96 9.20	66.94 66.94
24. (Hp 2-1, Hb) ₂ (human)	300000	Margolis & Kenrick (1968)	8.17	66.94
25. Haemocyanin (stenoglossan prosobranch <i>Dicathais</i>), dissociated to 1/20's	443000	Wood <i>et al.</i> (1971) (stenoglossan prosobranch <i>Murex trunculus</i>)	8.95	76.23
26. (Hp 2-1, Hb) ₃ (human)	450000	Margolis & Kenrick (1968)	10.00	76.63
27. (Hp 2-1, Hb) ₄ (human)	600000		11.75	84.34
28. (Hp 2-1, Hb) ₅ (human)	750000		13.37	90.86
29. (a) Slow α_2 -macroglobulin (human)	820000	Schultze & Heremans (1966)	13.20	93.60
(b) Slow α_2 -macroglobulin (sheep)	820000		13.87	93.60
30. Haemocyanin (<i>Helix aspersa</i>), dissociated to 1/10's	900000	Siezen & van Bruggen (1974)	13.24	96.55
31. Ovomacroglobulin (chicken egg white)	900000	Baker (1968)	13.70	96.55

* The systematics of the sipunculids used as sources of haemerythrin has been confusing. Specimens sold by a supplier to biochemists as '*Dendrostomum pyroides*' or '*Themiste pyroides*' which I examined were, in fact, *Themiste zosteriolum*. Electrophoretic studies reveal that '*Dendrostomum cymodoceae*' is two distinct sibling species (C. Manwell & C. M. A. Baker, unpublished work); these are now placed in the genus *Themiste*.

Fig. 3. Graph of the estimate of the hypothetical pore limit, $1/G_L$, as a function of a measure of the effective molecular radius, the cube root of the molecular weight for the 55 proteins (data in Table 1)

The Figure demonstrates that the estimate of a hypothetical pore limit by Gradipore electrophoresis is linearly related to $(\text{mol.wt.})^{1/3}$, described by eqn. (8) in the text.

From the regression analysis of the present data $s_{Y.X} = 0.3955$. Therefore, the 95%-confidence interval is $\pm 1.96 \times s_{Y.X} = \pm 0.7752$. This value is used to calculate the 95%-confidence interval for a single observation for a particular molecular weight. At 100000, the 95%-confidence interval for a single observation is 73 700–131 800.

As Rodbard & Chrambach (1971) point out, the accuracy quickly improves when a number of estimates are taken. If one takes the set of six different estimates of $1/G_L$ for sheep and human haemoglobins (Table 1), the average is 4.288 and the estimate of the standard deviation is 0.182. The 95%-confidence interval for a single value in the haemoglobin series is 4.288 ± 0.182 , corresponding to a mol.wt. interval of 65000–67200. Not surprisingly, accuracy is improved by restricting the consideration to homologous proteins.

As the set of standards includes both proteins and glycoproteins, and these differ significantly in partial specific volume ($\bar{v}$), it has been asked whether or not correction for $\bar{v}$ would improve the accuracy of the method yet further. However, no trend is obvious for glycoproteins, being on either one or the other side of the regression line, as should occur if this was a major cause of residual variance. It is possible that differences in $\bar{v}$ are less important in electrophoretic-retardation methods than in other hydrodynamic methods of molecular-weight determination, and Rodbard & Chrambach (1971, p. 98) used as standards a mixture of proteins and glycoproteins, calculating their molecular radii 'assuming that the molecule is spherical, unhydrated, and has a partial specific volume of 0.74.' However, as a separate study, it would be desirable to compare electrophoretic retardation

for the same protein, with and without its carbohydrate moiety.

In summary, the new modification, using commercially prepared gradient gels has five advantages: 1. Regression analysis reveals that the new modification is more accurate. 2. It is a relatively inexpensive technique in that purified protein standards are not necessary. 3. The very same gels can also be used for detecting many allelic protein variants, and therefore multiple polymorphic protein loci are often more readily distinguished if there are differences in molecular weight. 4. The method can be used with small samples, e.g. 'two-spotted' mites of the *Tetranychus urticae* complex can be studied individually for variation in esterases and dehydrogenases, although specimens must be homogenized under a dissecting microscope. 5. As a new sample is inserted each hour, extracts or body fluid samples can begin electrophoretic separation within a few minutes of their preparation, thus minimizing opportunities for proteolytic modification or other degradative changes.

I thank Dr. C. Appleby, Division of Plant Industry, CSIRO, Canberra, for determination of the molecular weight of the coelomic haemerythrin of *Phascolosoma arcuatum* (= *P. lurco*), Miss C. M. Ann Baker for help in identification of egg-white and serum proteins, and Professor G. M. Badger, Vice-Chancellor of the University of Adelaide, for funds for equipment.

References

- Anderson, L.-O., Borg, H. & Mikaelson, H. (1972) *FEBS Lett.* **20**, 199–202
- Bailey, N. T. J. (1959) *Statistical Methods in Biology*, (5th impression, 1969), English Universities Press, London
- Baker, C. M. A. (1968) *Br. Egg Mark. Board Symp.* **4th** 67–108
- Baker, C. M. A. & Manwell, C. (1977) *Aust. J. Biol. Sci.* in the press
- Bowen, S. T., Moise, H. W., Waring, G. & Poon, M.-C. (1976) *Comp. Biochem. Physiol. B* **55**, 99–103
- Dixon, M. & Webb, E. C. (1964) *Enzymes*, 2nd edn., Longmans, London
- Easterby, J. S. (1971) *FEBS Lett.* **18**, 23–26
- Ellman, G. L., Courtney, K. D., Andres, V. & Featherstone, R. M. (1961) *Biochem. Pharmacol.* **7**, 88–95
- Fee, J. A. (1975) *Struct. Bonding (Berlin)* **23**, 1–60
- Ferguson, K. A. (1964) *Metab. Clin. Exp.* **13**, 985–1002
- Frenkel, R. (1972) *Arch. Biochem. Biophys.* **152**, 136–143
- Gordon, W. G. (1971) in *Milk Proteins: Chemistry and Molecular Biology* (McKenzie, H. A., ed.), vol. 2, pp. 331–365, Academic Press, New York
- Han, K., Tetaert, D., Moschetto, Y., Dautrevaux, M. & Kopeyan, C. (1972) *Eur. J. Biochem.* **27**, 585–592
- Harris, J. I. & Perham, R. N. (1968) *Nature (London)* **219**, 1025–1028
- Hedrick, J. L. & Smith, A. J. (1968) *Arch. Biochem. Biophys.* **126**, 155–164
- Huisman, T. H. J., Reynolds, C. A., Dozy, A. M. & Wilson, J. B. (1965) *J. Biol. Chem.* **240**, 2455–2460
- Kaplan, N. O. (1964) *Brookhaven Symp. Biol.* **17**, 131–149
- Klippenstein, G. L., Van Riper, D. Z. & Oosterom, E. A. (1972) *J. Biol. Chem.* **247**, 5959–5963
- Klotz, I. M. & Keresztes-Nagy, S. (1963) *Biochemistry* **2**, 445–452
- Lewontin, R. C. (1974) *The Genetic Basis of Evolutionary Change*, Columbia University Press, New York
- Liberatore, F. A., Truby, M. F. & Klippenstein, G. L. (1974) *Arch. Biochem. Biophys.* **160**, 223–229
- Liljas, A., Kannan, K. K., Bergsten, P.-C., Waara, I., Fridborg, K., Strandberg, B., Carlbom, U., Järäp, L., Lövgren, S. & Petef, M. (1972) *Nature (London) New Biol.* **235**, 131–137
- Manwell, C. (1964) *Oxygen Anim. Org. (I.U.B. Symp. Ser. no. 31)* 49–116
- Manwell, C. (1975) *Nature (London)* **258**, 606–608
- Manwell, C. & Baker, C. M. A. (1963) *Comp. Biochem. Physiol.* **8**, 193–208
- Manwell, C. & Baker, C. M. A. (1970) *Molecular Biology and the Origin of Species*, Sidgwick and Jackson, London
- Manwell, C., Baker, C. M. A., Ashton, P. A. & Corner, E. D. S. (1967) *J. Mar. Biol. Assoc. U.K.* **47**, 145–169
- Margolis, J. & Kenrick, K. G. (1968) *Anal. Biochem.* **25**, 347–361
- Maurer, H. R. (1971) *Disc Electrophoresis and Related Techniques of Polyacrylamide Gel Electrophoresis*, 2nd edn., Walter de Gruyter, Berlin
- McKenzie, H. A. (1971) in *Milk Proteins: Chemistry and Molecular Biology* (McKenzie, H. A., ed.), vol. 2, pp. 257–330, Academic Press, New York
- Mueller, J. O., Smithies, O. & Irwin, M. R. (1962) *Genetics* **47**, 1385–1392
- Murphy, W. H., Kitto, G. B., Everse, J. & Kaplan, N. O. (1967) *Biochemistry* **6**, 606–610
- Nei, M. (1975) *Molecular Population Genetics and Evolution*, North-Holland, Amsterdam
- Owen, J. A. & Smith, H. (1961) *Clin. Chim. Acta* **6**, 441–443
- Rodbard, D. & Chrambach, A. (1970) *Proc. Natl. Acad. Sci. U.S.A.* **65**, 970–977
- Rodbard, D. & Chrambach, A. (1971) *Anal. Biochem.* **40**, 95–134
- Rodbard, D., Kapadia, G. & Chrambach, A. (1971) *Anal. Biochem.* **40**, 135–157
- Rossi-Fanelli, A., Antonini, E. & Caputo, A. (1961) *J. Biol. Chem.* **236**, 391–396
- Schultze, H. E. & Heremans, J. F. (1966) *Molecular Biology of Human Proteins with Special Reference to Plasma Proteins*, vol. 1, pp. 182–204, Elsevier, Amsterdam
- Selander, R. K. & Kaufman, D. W. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 1875–1877
- Shaw, C. R. & Prasad, R. (1970) *Biochem. Genet.* **4**, 297–320
- Siezen, R. J. & van Bruggen, E. F. J. (1974) *J. Mol. Biol.* **90**, 77–89

- Slater, G. G. (1969) *Anal. Chem.* **41**, 1039–1041
- Spooner, R. L., Oliver, R. A., Richardson, N., Buttress, N., Feinstein, A., Maddy, A. H. & Stratil, A. (1975) *Comp. Biochem. Physiol. B* **52**, 515–522
- Tanis, R. J. & Tashian, R. E. (1971) *Biochemistry* **10**, 4852–4859
- Tombs, M. P. (1965) *Anal. Biochem.* **13**, 121–132
- Tombs, M. P. (1968) in *Chromatographic and Electrophoretic Techniques*, 2nd edn., vol. 2 (Smith, I., ed.), pp. 443–458, William Heinemann Medical Books, London
- Weber, K. & Osborn, M. (1969) *J. Biol. Chem.* **244**, 4406–4412
- Wood, E. J., Bannister, W. H., Oliver, C. J., Lontie, R. & Witters, R. (1971) *Comp. Biochem. Physiol. B* **40**, 19–24