AN EXPERIMENTAL INVESTIGATION OF THE UNIT CHARGE MODEL OF PROTEIN POLYMORPHISM AND ITS RELATION TO THE ESTERASE-5 LOCUS OF *DROSOPHILA PSEUDOOBSCURA*, *DROSOPHILA PERSIMILIS*, AND *DROSOPHILA MIRANDA*

GARY COBBS* AND SATYA PRAKASH

Department of Biology, University of Rochester, Rochester, New York 14627

Manuscript received January 27, 1977
Revised copy received August 11, 1977

ABSTRACT

The relationship between charge changes and electrophoretic mobility changes is investigated experimentally. The charge of several proteins is altered by reaction with small molecules of known structure and the change in electrophoretic mobility is measured. The method of Ferguson plots is used to separate charge and shape components of mobility differences. The average effect of an amino acid charge change on the mobility of the esterase-$5^{1.00}$ allele of *Drosophila pseudoobscura* is estimated to be 0.046. This estimate is then used to apply the step model of OHTA and KIMURA (1973) to electrophoretic mobility data for the esterase-5 locus of *D. pseudoobscura* and *D. miranda*. The variation in electrophoretic mobility at this locus was found to be in agreement with the predictions of the step model.

MANY workers in the area of protein polymorphism have assumed that the method of gel electrophoresis is generally insensitive to most amino acid changes in protein structure and that only 1/4 to 1/3 of all new mutants can be detected in a single set of electrophoretic conditions. It is generally held that only substitutions interchanging members of the groups of acidic, basic and neutral amino acids will be detectable by electrophoresis (LEWONTIN 1974). This argument relies on the assumption that the methods used in population surveys do not have sufficient resolution to distinguish between the more subtle noncharge amino acid changes. Very little experimental information has been used to test this assumption.

The relationship between amino acid changes in proteins and their effect on electrophoretic mobility is of fundamental importance in the interpretation of enzyme polymorphism data. OHTA and KIMURA (1973) have proposed a model of the stepwise production of neutral alleles in which mutations change electrophoretic mobility in discrete jumps of unit value. Such a model is a direct consequence of the above assumptions about the limited resolution of electrophoretic methods. They were able to show that under this model (step model) the effec-

* Present address: Department of Biology, University of Louisville, Louisville, Kentucky 40208.

tive number of neutral mobility classes (n_e) that will be maintained in a finite population of effective size N_e is

$$n_e = \sqrt{1 + 8 N_e v} \; . \tag{1}$$

Here v is the mutation rate from a mobility class to an adjacent class. The step model was applied to data on protein polymorphism by OHTA (1975). The fit was moderately good with a large excess of rare mobility classes. This deviation was attributed to either nonequilibrium population structure or to the presence of some directional selection.

JOHNSON (1974) has criticized the step model as being biochemically unrealistic. He argues that partial rather than unit charge changes are more likely. His argument derives from the fact that the dissociation constants of side groups of amino acids are often different when in a protein. He also argues that subtle conformational changes caused by conservative amino acid changes may also be detected by electrophoresis. JOHNSON's criticisms imply that electrophoresis may be much more sensitive to amino acid changes than the step model presumes. He feels the true situation in nature is somewhere between the step model and the infinite alleles model of KIMURA and CROW (1964).

MARSHALL and BROWN (1975) support the step model by arguing that it is approximately correct. They discuss some data on human hemoglobin alleles that show that known charge changes cause large changes in mobility while noncharge changes cause small changes in mobility. They then argue that most routine methods of electrophoresis would not detect these smaller changes.

In this paper we will be concerned with the application of the step model to electrophoretic mobility information. If for a given protein the effect of an amino acid charge change on mobility were known, then it may be possible to partition the observed mobility variation into classes which, on the average, differ by one charge change. Alleles within such classes may or may not be distinguishable by electrophoresis but would be considered indistinguishable for the purpose of applying the step model. A problem that naturally arises is that the effect of an amino acid charge change on mobility is most likely not constant but has some distribution. The distribution is probably different for every allele in the population. Here we must assume that the distributions are roughly the same and that a partition using the average effect of a charge change on mobility will be appropriate. If the average effect of charge change is large with respect to the average difference between adjacent mobility alleles, then the above method may be considered better than the method used by OHTA (1975). However, if the average effect of a charge change is roughly equal to or smaller than the average difference between adjacent alleles, then there would be no improvement over OHTA's method.

There are convincing data in the literature that suggest the effect of a charge change on relative mobility may be large. RICHARDSON, RICHARDSON and SMOUSE (1975) report relative mobility differences of 22%, which they interpret as simple charge changes at the malic dehydrogenase locus. HENNING and YANOFSKY (1963) have published photographs of acrylamide gels of mutant A46 of

tryptophan synthetase A. This mutant is known to result from the substitution of a glutamic acid for a glycine in the wild-type allele. This represents a unit charge change and the resulting change in mobility is 30%. These changes in mobility are surprisingly large when one considers that many protein loci have alleles that differ in mobility by only a few percent or less. For example, at the esterase-5 (*est-5*) locus in *D. pseudoobscura* PRAKASH, LEWONTIN and HUBBY (1969) report alleles which differ in mobility by only 1%. Also, COBBS and PRAKASH (1977) report alleles at this locus in *D. persimilis* and *D. miranda* that differ in mobility by 1% or less.

These considerations suggest that for some loci, many of the electrophoretic alleles differ in mobility by substantially less than what might be expected for an amino acid charge change. If this is true, the application of the step model to these loci used by OHTA (1975) would be inappropriate. In this paper we will attempt to estimate the average effect of an amino acid charge change on the electrophoretic mobility of the $est\text{-}5^{1.00}$ allele in *D. pseudoobscura*. This information will then be used to analyze goodness-of-fit to the step model of OHTA and KIMURA (1973).

THE METHOD OF INVESTIGATION

For a protein of a given shape and size the electrophoretic mobility is proportional to the net charge on the molecule. If the net charge can be changed by a known amount, then the mobility of the protein will change by an amount depending on the net charge of the unaltered protein. If Q is the net charge on a protein and δ is the amount of charge change caused by an alteration, then the fraction of change in absolute mobility, ΔU, may be related to δ and Q as

$$\Delta U = \frac{\delta}{Q} \; . \qquad (2)$$

A charge change on a protein may also cause a change in shape. Consequently, an observed change in mobility may be due to both charge and shape changes in the protein. Fortunately, there are methods by which the charge and shape differences between proteins can be at least partially separated. The method of Ferguson plots is now being recognized as a method that gives information on both charge and shape or size differences between proteins (RODBARD and CHRAMBACH 1971). The method relies on a linear relation between the log of the relative mobility of two proteins and the concentration of gel. The slope of this relationship is a function of both size and shape differences between the proteins and is unaffected by charge differences. Evaluation of the linear relation at zero concentration of gel yields a quantity whose antilog is at least approximately equal to the ratio of the free electrophoretic mobilities of the two proteins. In general we will represent this linear relation by the equation $y_t = y_0 + K_r t$, where y_t is the log of the relative mobility of the two proteins in a gel of concentration t, and K_r will be called the retardation coefficient after RODBARD and CHAMBACH (1971). We wish to emphasize that in this paper the K_r values are

defined using relative mobility of two proteins. We therefore use a lower case subscript to distinguish this coefficient from the one of RODBARD and CHRAMBACH (1971), who used an upper case subscript. It may be shown that the K_r value defined in this way is equal to the difference of K_R values estimated by the methods of RODBARD and CHRAMBACH (1971). Let R_m be the mobility of protein 1 relative to protein 2, and let K_r be the retardation coefficient obtained as the slope of the linear regression of log R_m on concentration of gel. Let $K_R(1)$ and $K_R(2)$ be the retardation coefficients of proteins 1 and 2, respectively, determined by the methods of RODBARD and CHRAMBACH (1971). Then it may be shown that $K_r = K_R(1) - K_R(2)$.

The parameters K_r *and* y_0 *may be regarded as physical constants characterizing structural and charge differences between the proteins, respectively.* The antilog of y_0 will be denoted by M_0 and is the ratio of the electrophoretic mobilities at zero gel concentration. More generally we shall define $M_t = \log^{-1} y_t$, which is the ratio of mobilities in a gel of concentration t.

The value of M_0 is considered to be dependent primarily on charge and size differences between proteins and independent of shape differences. The ratio of the charges of two proteins may be calculated from the value of M_0. The method of calculation results from the equations of GORIN (1941) and are given in the paper by COBBS and PRAKASH (1977). The derivation of GORIN's equations involves the assumption that the molecules are spherical and rigid, and that the charge is uniformly distributed on the surface of the protein. Consequently the above calculations are subject to error due to deviation from these assumptions. We wish to emphasize that we are always working with relative mobilities of two proteins and their relative charges. Nowhere is it necessary to estimate absolute mobility or valence of any molecule. This minimizes errors stemming from assumptions about the physical structure of proteins. This is especially true if the proteins being compared are of roughly the same size and shape.

The alteration of charge on proteins was accomplished by several methods. In all the methods we have used, a small molecule of known charge is bound to the protein in some way. The first method to be described is the covalent binding of various molecules to free sulfhydryl (SH) groups of the protein. This method was used with ovalbumin, human hemoglobin A and bovine hemoglobin. The second method involves ovalbumins which differ in the number of covalently bound phosphate groups. The last method is the binding of a cupric ion by bovine serum albumin.

Numerous compounds are known which readily react with free SH groups in proteins. The thiol compounds used in this study are *p*-chloromercuriphenylsulfonic acid (*p*CMSA), and 2,2′ dithiobis (ethylamine) diHCl which is commonly called cystamine. For each reactive SH in a protein, *p*CMSA adds a single sulfonic acid group and cystamine adds a single amino group. Under the step model, a single molecule of these compounds should alter the electrophoretic mobility by one unit in the increasing and decreasing direction respectively. The electrophoretic buffer system used here has a pH of 9.0 and all proteins studied have a net negative charge. We will always be dealing with net negative charge in electron units.

The number of reactive sulfhydryl groups per protein molecule is established from a variety of methods. One method is to react equimolar amounts of protein and thiol reagent. If the mobility of all molecules is altered by the same amount, the change must be due to a single reacting SH group per protein molecule. For some proteins the number of reactive SH groups has been titrated and reported in the literature.

Another method of observing known charge changes involves ovalbumin. Commercial preparations of purified chicken ovalbumin contain major components of three different electrophoretic mobilities. PERLMAN (1955) has shown that these three forms result from different amounts of phosphorylation. Furthermore, she has shown that the phosphates are involved in monoester bonds with ovalbumin. Under the step model these proteins should differ by two charge units. The fast, medium, and slow mobilities of ovalbumin (OVA) will be referred to here as OVA-1, OVA-2, and OVA-3, respectively.

The last method of modification involves the binding of cupric ions to bovine serum albumin (BSA). BRADSHAW, SHEARER, and GURD (1968) have shown that BSA monomer will bind cupric ions on a mole to mole basis. They also have suggested that BSA dimer will bind a total of 4 cupric ions per dimer molecule. The BSA monomer-cupric ion complex should have a charge reduced by two units and the BSA dimer-cupric ion complex should have a charge lowered by eight units.

All these methods of altering the charge of proteins by predicted amounts are subject to the same criticisms as the step model itself, namely that shape as well as charge may be changed by the alteration. Also the dissociation constants of the molecules may be significantly altered when bound to the protein. By employing a number of different modifications on different proteins we hope to get some idea of the average effect of a charge change on electrophoretic mobility. We will also obtain information on how variable this effect is.

The final stage of the method is the comparison of the mobility changes of different proteins by the various methods. We do this by calculating the mobility changes that would be expected for the *est-5*$^{1.00}$ allele of *D. pseudoobscura*. This serves two purposes. First it is a means of comparing all the mobility changes. Second, we may then interpret the mobility differences of the esterase-5 alleles in terms of what the effect of a charged amino acid substitution is expected to be.

As mentioned earlier, if we ignore shape changes, the change in electrophoretic mobility resulting from a given charge change depends on the net charge of the protein. If a protein has net charge Q_i and we alter the charge amount by δ, the resulting change in absolute mobility is given by equation (2). We then wish to know what the effect of δ would be on a protein of net charge Q_j. Letting ΔU_i and ΔU_j be the mobility changes for proteins i and j and using equation (2), we obtain

$$\Delta U_j = \left(\frac{Q_i}{Q_j}\right) \Delta U_i \ . \qquad (3)$$

Equation (3) indicates that all we need to know in order to calculate the expected ΔU of a protein from an observed ΔU of another protein is the ratio of the

charges of the two proteins. As discussed here and elsewhere (COBBS and PRAKASH 1977), we may estimate the charge ratio of two proteins by the method of Ferguson plots and by using the equation of GORIN (1941), which is also given in TANFORD (1961, eq. 24–8). The equation used to estimate the charge ratio, CR, is

$$CR = \hat{M}_0 \left(\frac{R_i}{R_j}\right) \left\{\frac{1 + \kappa(R_i + r)}{1 + \kappa(R_j + r)}\right\} \left\{\frac{f(\kappa R_j)}{f(\kappa R_i)}\right\}. \quad (4)$$

Here R_i is the radius of protein i, r is the radius of the counterion, κ is the Debye-Huckle constant of the buffer, and $f(\kappa R_i)$ is Henry's function of κR_i, and $\hat{M}_0$ is the antilog of the least squares regression estimate of y_0. The estimate of y_0 results from linear regression of y_t on t where y_t is the log of the relative mobility of the two proteins in a gel of concentration t. The value of R_i may be estimated from knowledge of molecular weight (see RODBARD and CHRAMBACH 1971).

MATERIALS AND METHODS

Proteins: Proteins used in electrophoresis were the following:
(1) crystallized and lyophilized egg ovalbumin from Sigma Chemical Company;
(2) crystallized and lyophilized bovine serum albumin from Sigma Chemical Company;
(3) bovine hemoglobin for protease assay from Nutritional Biochemical Corporation;
(4) purified human hemoglobin A obtained from DR. RICHARD DOHERTY; (5) esterase-5 proteins were obtained by electrophoresing crude extracts of homozygous strains of *D. pseudoobscura*, *D. persimilis* and *D. miranda* according to the methods of HUBBY and LEWONTIN (1966).

Electrophoresis: Vertical slab acrylamide gel electrophoresis was performed in EC 470 gel boxes from Electrochemical Corporation. Gels were made from cyanogum-41, a commercial mixture of acrylamide and bis-acrylamide, obtained from American Cyanamid Corporation. The cyanogum concentration varied from 3% to 13% and the electrophoresis was carried out at a field strength of 11.6 volts/cm throughout. Cooling solution maintained at $-4 \pm 1°$ was circulated through the gel apparatus to maintain a constant temperature in all gels during electrophoresis. Gels were run until the sample path was roughly 2 to 10 cm and in most cases was greater than 5 cm. A continuous system of trisborate-EDTA, pH = 9.0 buffer as described by HUBBY and LEWONTIN (1966) was used. The buffer ionic strength is 0.06 and the Debye-Huckle constant is 8.035×10^6. The radius of the tris cation is assumed to be 2.3×10^{-8} cm.

Staining and marking gels: Protein bands were stained with 0.005% Coomassie blue in a solution of H_2O, methanol, and acetic acid in a volume ration of 5:5:1, respectively. Esterase activity was visualized in gels using a solution of 0.3 mg/ml α-naphthyl acetate and 0.5 mg/ml fast red TRN in 0.1 M phosphate buffer at pH 6.5. The α-naphthyl acetate was dissolved in a small volume of 75% acetone before mixing with phosphate buffer. The position of bromophenol blue bands was marked by injecting a small amount of charcoal powder suspension into the gel. Distance from origin to bands was measured with a ruler.

Determination of percent polymerization: The percent polymerization of cyanogum-41 was determined by gravimetric analysis. Equal volumes of each of 3% to 13% cyanogum were polymerized using 0.05 g ammonium persulfate and 0.2 ml tetramethylethylenediamine per 100 ml of cyanogum-41 solution. This concentration of catalysts was used for all gels in this study. One of the resulting gels for each concentration was then leached using 8 successive changes of distilled-deionized water at 12-hr intervals. The volume of each leaching solution was five times that of the gel. The control gels were not leached and were stored covered to prevent desiccation. Both the control and leached gels were desiccated at 145° for 7 days and then weighed. The

weight of the buffer constituents were subtracted from the control gels' desiccated weight to yield the weight of the control gel. In all cases this weight was within a few percent of the predicted amount. The fraction of polymerization was then calculated as the ratio of the weight of the leached gel to that of the unleached gel.

Thiol treatments: Equal volumes of the thiol reagent solution and a 10^{-4} M solution of protein were mixed and incubated for one hour at room temperature before layering onto the gels. The human and bovine hemoglobins were dissolved in a 0.1 M tris-borate-EDTA buffer with 6% sucrose. The OVA was dissolved in H_2O with 6% sucrose.

Cupric ion treatments of BSA: Equal volumes of 0.1 M $CuSO_4$ and 10^{-4} M BSA in 6% sucrose were mixed and incubated for one hour at room temperature before layering onto the gels. A solution of K_2SO_4 was used as a control.

Experimental design and statistical analysis: The procedure described here was used in all cases where Ferguson plot analysis was used to compare two proteins. Each comparison of two proteins involved 4 adjacent pockets, here numbered 1 through 4, on each of several different concentrations of gel. Let the two proteins be denoted by A and B. Pockets 1 and 3 contain protein A and pockets 2 and 4 contain protein B. After electrophoresis and staining, the migration distances are measured and the relative mobility of A to B or B to A is calculated as the ratio of the averages of the two measurements for each protein. The replicate measurements of migration distances were in all cases very close and in nearly all cases differed by less than one percent.

Ferguson plot experiments, in most cases, involved gels with percent cyanogum-41 set at integer values from 3 to 13. The odd concentration gels were all run together on one day and the even concentration gels were also run together on a different day. When the experiment involved an esterase-5 strain, a large number of flies, usually 10 to 20, were homogenized together in a glass tissue grinder in 50 μl/fly of tris-borate-EDTA buffer containing 5% sucrose. This homogenate was then centrifuged to remove insoluble material and 15 μl was pipetted into appropriate pockets on all gels run on that day. A new homogenate was made each day immediately before electrophoresing. All the strains used were known to be homozygous for electrophoretic alleles and were most likely isogenic for *est-5* alleles. The strains were highly inbred by repeated sib mating and have been maintained in the laboratory for roughly 10 years.

Estimates of the slope, K_r, and intercept, y_0, of Ferguson plots were obtained using least squares regression of log of relative mobility on percent acrylamide. The regression analysis involved a single value of log R_m for each gel concentration and the standard errors of the regression parameters, obtained using well-known methods, contain components due to both experimental error and to deviation from a linear relation between these variables. Percent acrylamide was obtained as percent cyanogum-41 multiplied by the fraction of polymerization determined for gels of the given concentration of cyanogum-41.

RESULTS

Modified proteins: Both *p*CMSA and cystamine were found to alter the mobility of proteins in the expected direction. Human hemoglobin A, bovine hemoglobin and OVA all changed mobility after reaction with *p*CMSA. It was found, using Ferguson plots, that bovine Hb was dissociated into dimers by reaction with *p*CMSA. This did not occur with human Hb. Both human and bovine Hb reacted with cystamine and neither tetramer was dissociated by this reaction. The change in mobility in these proteins in an 8% gel after reacting with 10^{-2} M *p*CMSA or 10^{-2} M cystamine is shown in Figure 1. Here one can see that *p*CMSA and cystamine change the mobility of Hb by roughly equal amounts but in opposite directions, as expected.

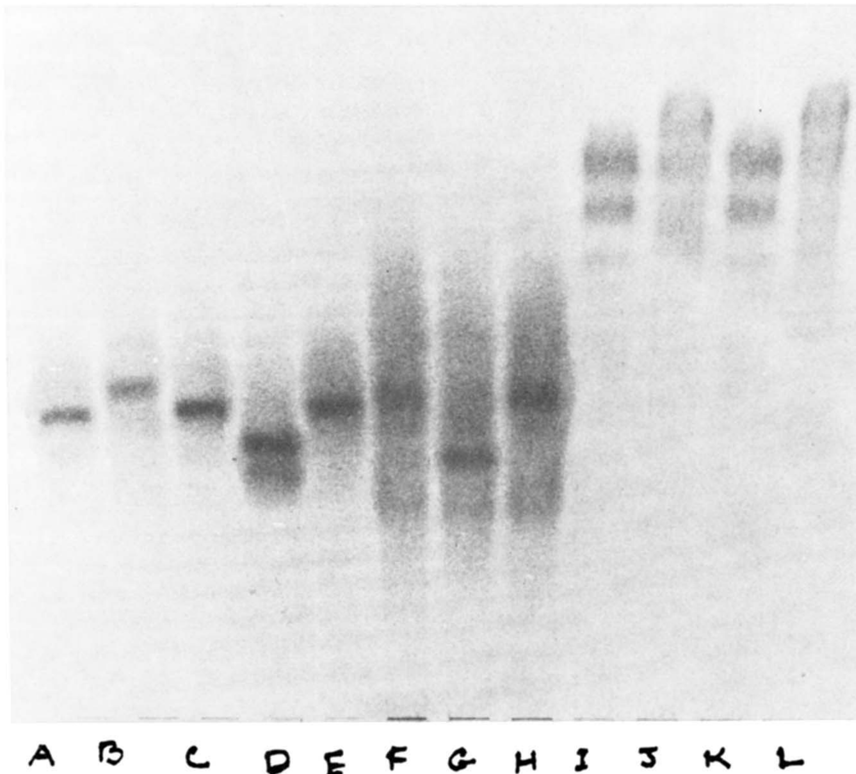

FIGURE 1.—Effect of thiol reagents on hemoglobins and ovalbumin. Pockets A, C, E are untreated human hemoglobin-A. Pockets B and D are Hb-A treated with 10^{-2} M *p*CMSA and cystamine respectively. Pockets F and H are untreated bovine Hb and pocket G is bovine Hb treated with 10^{-2} M cystamine. Pockets I and K are untreated OVA and pockets J and L are OVA treated with 10^{-2} M *p*CMSA. Origin is toward bottom of photograph.

In order to determine how many SH groups are reacting per molecule, a series of roughly equimolar concentrations of protein and thiol reagent were used. An experiment with human Hb-A in which the *p*CMSA concentration ranged from 0.5×10^{-4} to 4×10^{-4} M is shown in Figure 2. Here we see that a molar ratio of 2 of *p*CMSA to Hb-A was required to move the entire band to a new mobility. This change in mobility is most likely due to the reaction of two SH groups per Hb-A tetramer. Higher concentrations of *p*CMSA did not cause any further change in mobility. The change in mobility of Hb-A caused by 2×10^{-4} M *p*CMSA was compared to that caused by 10^{-2} M *p*CMSA in adjacent pockets of the same gels and were found to be identical. The same type of experiments were done using cystamine with human Hb-A and bovine Hb and the same results were observed. Thus the change in mobility of human Hb-A and bovine Hb with both *p*CMSA and cystamine is due to the reaction of two SH groups per molecule.

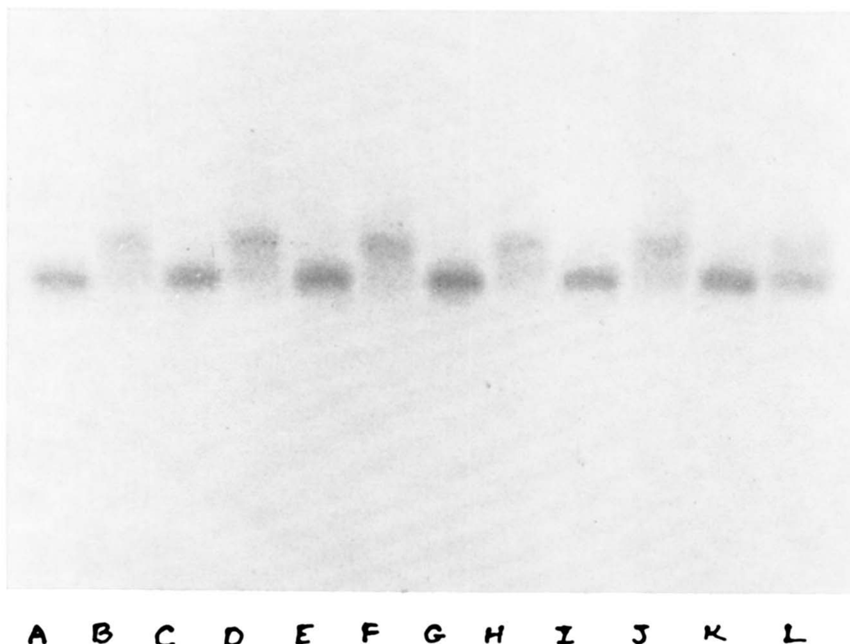

FIGURE 2.—Effect of varying concentrations of pCMSA on human Hb-A. Pockets A, C, E, G, I and K are all untreated Hb-A. Pockets B, D, F, H, J and L are treated with 5, 4, 3, 2, 1 and 0.5×10^{-4} M pCMSA, respectively. Origin is toward bottom of photograph.

Experiments similar to the above were done using pCMSA and OVA. Here it was found that increasing the concentration of pCMSA above 10^{-4} M did cause further increase in mobility. A comparison of the effect of 10^{-2} M, 10^{-4} M, and 2×10^{-4} M pCMSA is shown in Figure 3. Here we find that 10^{-2} M pCMSA causes a mobility change of roughly three times that of 10^{-4} M pCMSA. The ΔM_0 estimates given in Table 3 give values of 2.9, 2.8 and 3.6 for OVA-1, -2 and -3, respectively. This is in good agreement with the findings of FOTHERGILL and FOTHERGILL (1970) who found a total of three reactive SH groups per OVA molecule. The three forms of OVA which differ in the number of phosphate groups may also be seen in Figure 1 pockets I, J, K, and L.

Electrophoresis of the BSA preparation yielded two predominant bands as shown in Figure 4 (pockets A, C and E). The faster, more intensely staining band was suspected to be monomer BSA and the slower band, dimer BSA. This conjecture was tested by measuring molecular weight of the bands by the method of Ferguson plots. The molecular weight of BSA monomer and dimer are known to be 67,000 and 134,000, respectively. The Ferguson plots were constructed using *est-5*$^{1.00}$ dimer of *D. pseudoobscura* as the mobility standard. The least squares regression estimates and their standard errors are given in Table 2 and the relation between K_r and molecular weight is shown in Figure 5. This relation is roughly linear suggesting that our identification of BSA monomer and dimer is correct.

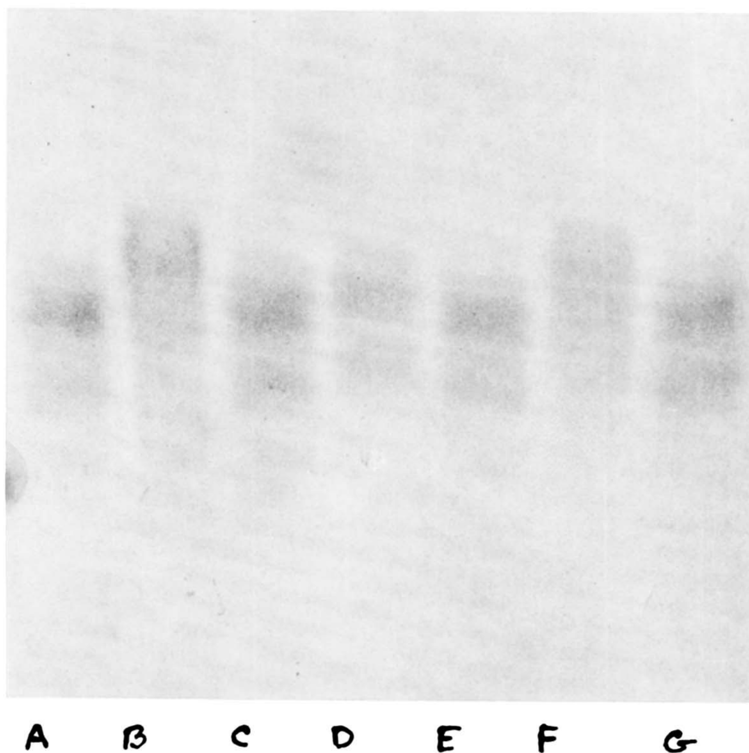

FIGURE 3.—Effect of varying concentrations of *pCMSA* on ovalbumin. Pockets A, C, E and G are all untreated. Pockets B, D, and F are treated with 10^{-2} M, 10^{-4} M and 2×10^{-4} M pCMSA respectively. Origin is toward bottom of photograph.

Incubation of BSA with 0.1 M $CuSO_4$ was found to alter its electrophoretic mobility. An experiment using 0.1 M K_2SO_4 as a control is shown in Figure 4. Here we see that the presence of cupric ions lowers the mobility of both BSA monomers and dimers. In other experiments it was found that K_2SO_4 treatment was identical to treatment with distilled H_2O or 0.1 M tris-borate-EDTA buffer.

In order to separate the charge and shape components of the observed mobility changes, the method of Ferguson plots was used. The percent cyanogum-41 was varied from 3% to 13% and the mobility of the altered protein relative to the unaltered protein was determined at each concentration. The log of this R_m value was then plotted for each treatment as a function of percent acrylamide as shown in Figure 6. Estimates of the least squares regression parameters and their standard error are given in Table 1. As previously mentioned, y_0 or its antilog, M_0, is a measure of the charge difference between proteins. In this case, M_0 is equal to the charge ratio because there is no significant difference in the size of the molecules. A lack of charge change would be indicated by a y_0 value of zero. All modifications, except the reaction of bovine Hb with cystamine, produced highly significant changes in y_0. A K_r value different from zero indicates an

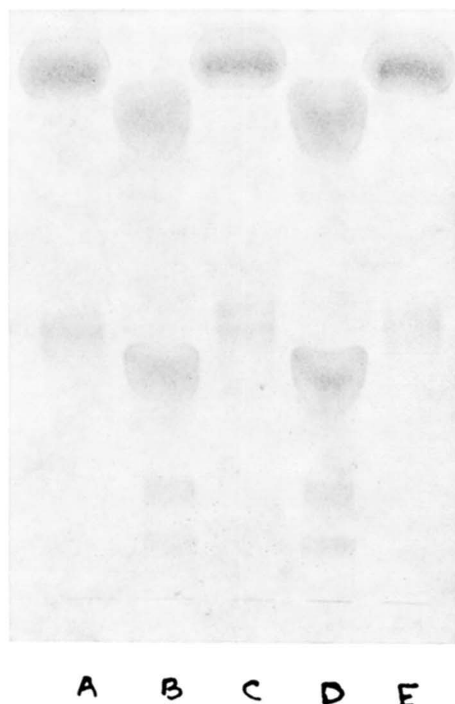

FIGURE 4.—Effect of cupric ions on bovine serum albumin. Pockets A, C and E are treated with 0.1 M K_2SO_4 and pockets B and D are treated with 0.1 M $CuSO_4$. Origin is toward bottom of photograph.

TABLE 1

Estimates of linear regression parameters for Ferguson plots of modified proteins

Protein	Modification	γ_0	Standard error γ_0	M_0	K_r	Standard error K_r
Human Hb-A	10^{-2} M pCMSA	0.0298	0.0070	1.0710	0.000,513	0.000,852
	10^{-2} M cystamine	—0.0261	0.0076	0.9416	—0.003,650**	0.000,920
Bovine Hb	10^{-2} M cystamine	—0.0017	0.0104	0.9962	—0.010,777***	0.001,271
BSA monomer	Cu^{++}	—0.0104	0.0101	0.9764	—0.003,393*	0.001,183
BSA dimer	Cu^{++}	—0.1197	0.0142	0.7591	0.010,391**	0.002,194
OVA-1	10^{-2} M pCMSA	0.0355	0.0029	1.0850	—0.000,368	0.000,356
OVA-2	10^{-2} M pCMSA	0.0488	0.0037	1.1188	—0.001,542**	0.000,446
OVA-3	10^{-2} M pCMSA	0.0760	0.0097	1.1913	—0.007,343***	0.001,188
OVA-1	10^{-4} M pCMSA	0.0125	0.0015	1.0292	—0.000,452	0.000,190
OVA-2	10^{-4} M pCMSA	0.0179	0.0033	1.0420	—0.000,938	0.000,407
OVA-3	10^{-4} M pCMSA	0.0223	0.0036	1.0526	—0.001,654*	0.000,450
OVA-2	$PO_4^{=}$	—0.0430	0.0022	0.9058	0.000,512	0.000,266
OVA-3	$PO_4^{=}$	—0.0790	0.0025	0.8337	—0.000,893	0.000,326

The symbols *, **, and *** denote significant deviation from zero at the 0.05, 0.01 and 0.001 levels, respectively.

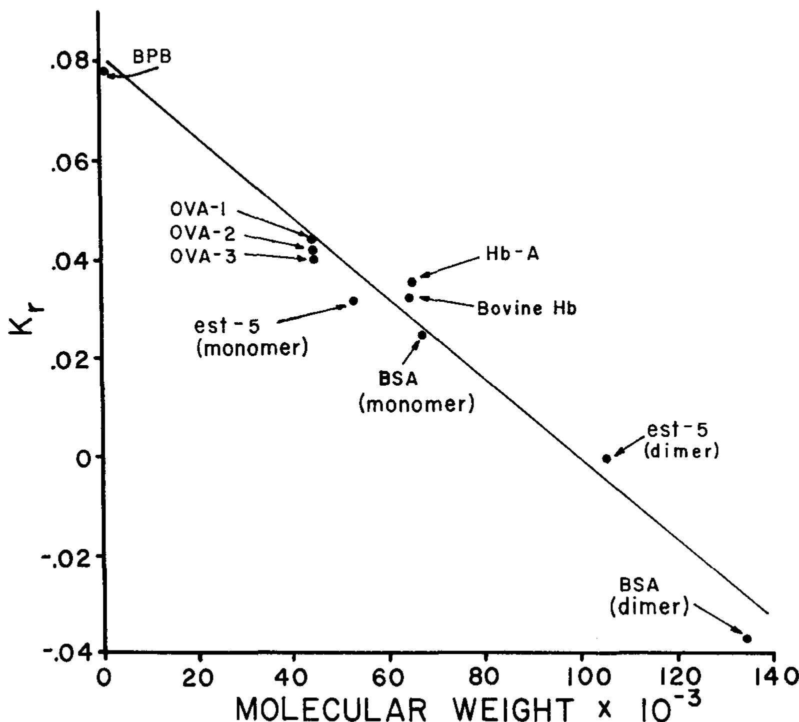

FIGURE 5.—Relationship between K_r and molecular weight of proteins used in this study. The abbreviations are as follows: BPB = bromophenol blue (MW = 670); OVA = ovalbumin (MW = 43,500); Hb = hemoglobin (MW = 68,000); est-5 = esterase-5 (MW monomer = 52,500; MW dimer = 105,000); BSA = bovine serum albumin (MW monomer = 67,000; MW dimer = 134,000).

alteration of the shape or size of the molecule. Eight of the 13 alterations produced K_r values significantly different from zero. The reaction of bovine Hb with cystamine produced a rather large alteration of K_R while not significantly altering y_0. Five of the 13 alterations did not significantly affect K_R and most of the significant changes in K_R were small.

Charge ratios: All the above unmodified proteins, *est-5*$^{1.13M}$ monomer and bromophenol blue were compared to *est-5*$^{1.00}$ dimer by the method of Ferguson plots. The Ferguson plots are shown in Figure 7 and the least squares regression estimates and their standard errors are given in Table 2. The ratio of the charge of these proteins to that of *est-5*$^{1.00}$ dimer was calculated using equation (4) and are also given in Table 2.

We may now compare all the charge alterations given in Table 1. The amount of charge change is estimated as $|M_0 - 1|$ and will be denoted as ΔM_0. The expected ΔM_0 for *est-5*$^{1.00}$ dimer is obtained by multiplying ΔM_0 of protein i by the charge ratio of protein i. The ΔM_0 values for all proteins and treatments are given in Table 3. Also given are the number of elementary charges changed per molecule. The expected ΔM_0 per elementary charge for *est-5*$^{1.00}$ dimer was cal-

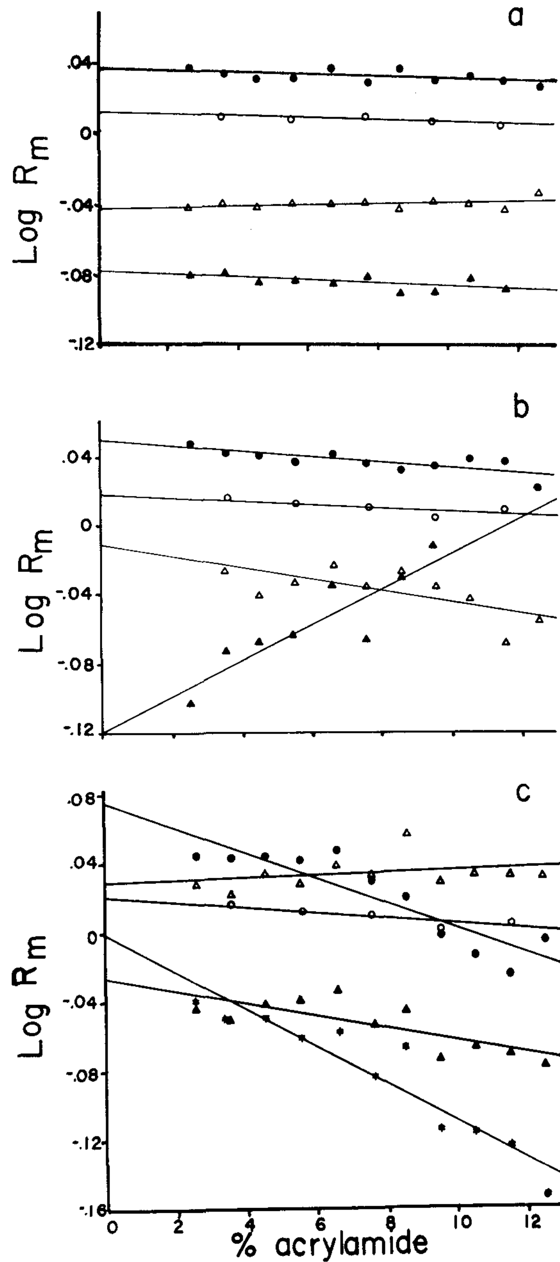

FIGURE 6.—Ferguson plots for modified protein and their linear regression lines. In all cases R_m refers to the mobility of the modified protein relative to the unmodified protein. The parameter M used in the text is synonymous with R_m used here. The symbols represent the following plots: (a), ● is OVA-1 with 10^{-2} M *p*CMSA, ○ is OVA-1 with 10^{-4} M *p*CMSA, △ is OVA-2 relative to OVA-1, ▲ is OVA-3 relative to OVA-1. (b), ● is OVA-2 with 10^{-2} M *p*CMSA, ○ is OVA-2 with 10^{-4} M *p*CMSA, △ is BSA monomer with Cu^{++}, ▲ is BSA dimer with Cu^{++}. (c), ● is OVA-3 with 10^{-2} M *p*CMSA, ○ is OVA-3 with 10^{-4} M *p*CMSA, △ is Hb-A with *p*CMSA ▲ is Hb-A with 10^{-2} M cystamine, * is bovine Hb with 10^{-2} M cystamine.

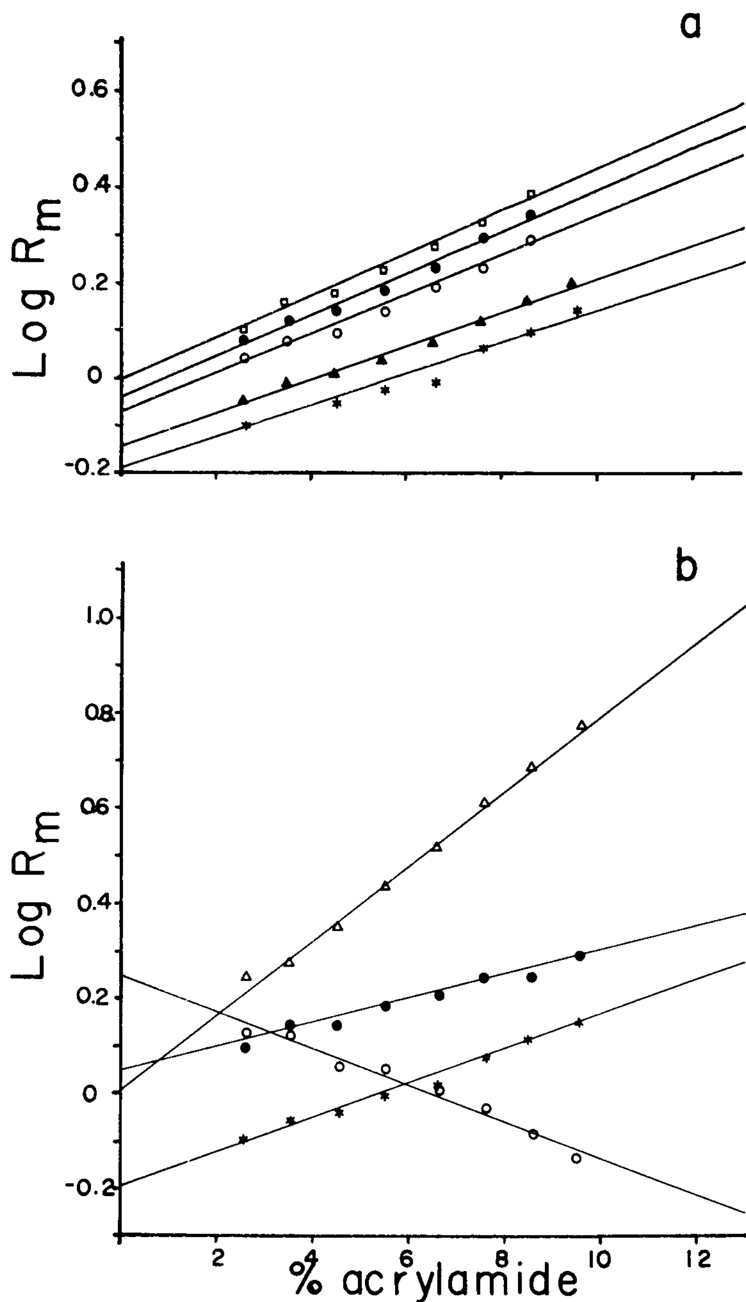

FIGURE 7.—Ferguson plots for unmodified proteins and bromophenol blue. In all cases R_m is mobility relative to *est-5*$^{1.00}$ dimer. The symbols represent the following plots: (a), □ is Ovalbumin-1, ● is ovalbumin-2, ○ is ovalbumin-3, * is bovine hemoglobin and ▲ is esterase-5$^{1.13M}$ monomer; (b), ● is BSA monomer, ○ is BSA dimer, * is hemoglobin A and △ is bromophenol blue.

TABLE 2

Estimates of linear regression parameters in Ferguson plots of unmodified proteins

Protein	Y_0	Standard error Y_0	M_0	K_r	Standard error K_r	CR
Bromophenol Blue	0.0091	0.01648	1.02167	0.07814	0.00254	0.09001
est-$5^{1.18M}$	—0.14011	0.00896	0.72425	0.03465	0.00138	0.50197
OVA-3	—0.07352	0.01290	0.84426	0.04122	0.00218	0.53061
OVA-2	—0.04136	0.01111	0.90916	0.04328	0.00188	0.57140
OVA-1	—0.00609	0.00926	0.98607	0.04414	0.00157	0.61973
Bovine Hb-A	—0.18879	0.02578	0.64746	0.03265	0.00398	0.51399
Human Hb-A	—0.20097	0.01104	0.62955	0.03607	0.00170	0.49977
BSA dimer	0.23481	0.01592	1.71714	—0.03751	0.00246	1.95764
BSA monomer	0.03616	0.01037	1.08682	0.02556	0.00160	0.85607

Charge ratio (CR) to est-$5^{1.00}$ is calculated by equation (4) of text.

culated as described above and are also given in Table 3. The mean and standard deviation of these ΔM_0 per electron values is 0.023 and 0.015, respectively.

Comparison of alleles: A number of different *est-5* alleles of *D. pseudoobscura*, *D. persimilis*, and *D. miranda* were compared using Ferguson plots. Four strains from both *D. persimilis* and *D. miranda* and four strains from *D. pseudoobscura* were examined. Again, the mobilities were all expressed relative to est-$5^{1.00}$ dimer of *D. pseudoobscura*. The Ferguson plots are shown in Figure 8 and the least squares regression estimates and their standard errors are given in Table 4. Of the 12 K_r values, only two deviate significantly from zero. Those that are significantly different from zero are also very small. We may obtain some idea of the relative importance of variation in M_0 *versus* variation in K_R in determining

TABLE 3

Estimates of change in free electrophoretic mobility (ΔM_0) per electron

Protein	Modifier	No. elementary charges changed	ΔM_0	$\Delta M_0/e$	$\Delta M_0/e$ for est-$5^{1.00}$
Human Hb-A	p CMSA	1	0.071,005	0.035,503	0.017,743
	cystamine	1	0.058,390	0.029,195	0.014,591
Bovine Hb	cystamine	1	0.003,854	0.001,927	0.000,990
OVA-3	10^{-2} M pCMSA	3	0.191,263	0.063,754	0.039,510
OVA-2	10^{-2} M pCMSA	3	0.118,837	0.039,612	0.022,634
OVA-1	10^{-2} M pCMSA	3	0.085,042	0.028,347	0.015,041
OVA-2/OVA-1	$PO_4^{=}$	2	0.094,246	0.047,123	0.025,004
OVA-3/OVA-1	2 $PO_4^{=}$	4	0.166,261	0.041,565	0.022,055
BSA monomer	Cu^{++}	2	0.023,634	0.011,817	0.010,116
BSA dimer	4 Cu^{++}	8	0.240,869	0.030,108	0.058,941
OVA-3	10^{-4} M pCMSA	1	0.052,611	0.052,611	0.032,605
OVA-2	10^{-4} M pCMSA	1	0.041,977	0.041,977	0.023,985
OVA-1	10^{-4} M pCMSA	1	0.029,183	0.029,183	0.015,485

The value of $\Delta M_0/e$ for est-$5^{1.00}$ is the product of $\Delta M_0/e$ and CR which is given in Table 2.

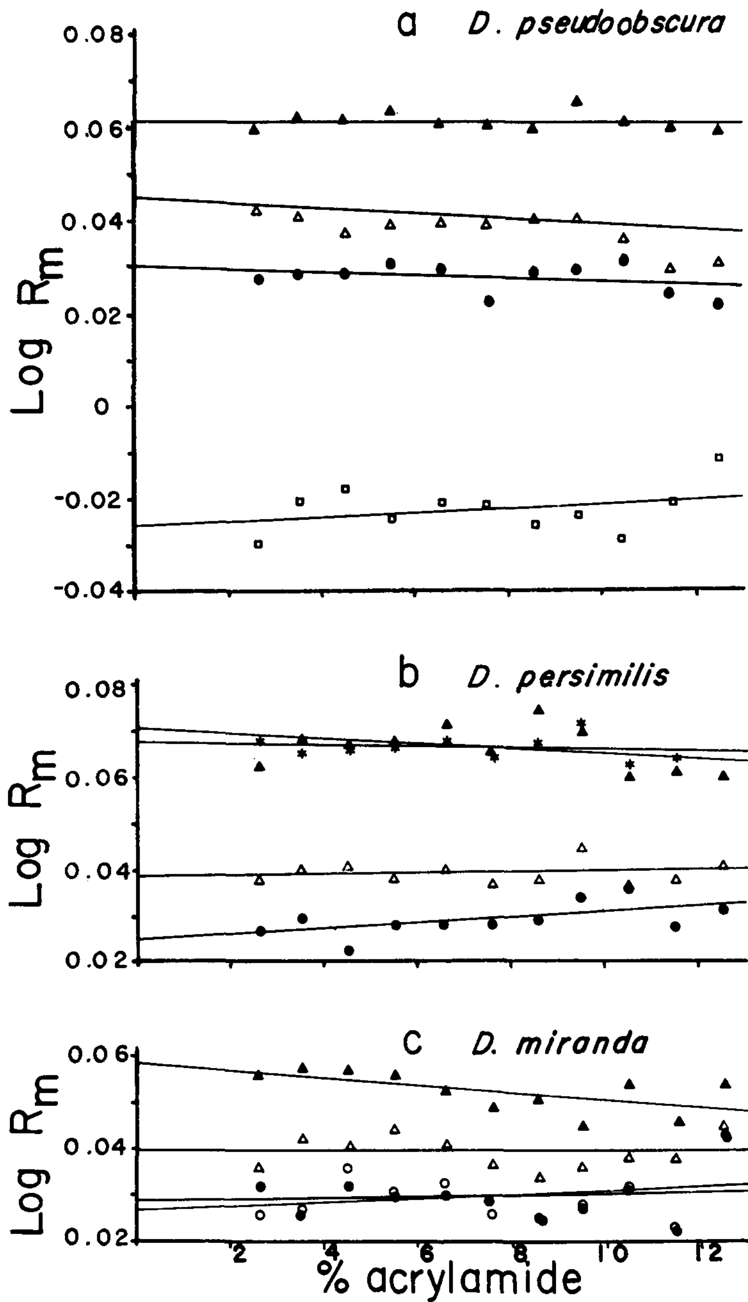

FIGURE 8.—Ferguson plots for esterase-5 alleles. In all cases, R_m is the mobility relative to the *est-5*$^{1.00}$ allele. The symbols represent the following plots: (a), ▲ is *est-5*$^{1.16}$, △ is *est-5*$^{1.12}$, ● is *est-5*$^{1.07}$, □ is *est-5*$^{0.95}$. (b), * is *est-5*$^{1.18}$, ▲ is *est-5*$^{1.17}$, △ is *est-5*$^{1.12}$, ● is *est-5*$^{1.07}$. (c), ▲ is *est-5*$^{1.15}$, △ is *est-5*$^{1.11}$, ○ is *est-5*$^{1.08}$, ● is *est-5*$^{1.07}$.

TABLE 4

Estimates of linear regression parameters of Ferguson plots for esterase-5 alleles

Species	Allele	Y_0	Standard error intercept	M_0	K_r	Standard error K_r
D. pseudoobscura	*0.95*	—0.02633	0.00410	0.94117	0.000,511	0.000,500
	1.07	0.03035	0.00256	1.07238	—0.000,418	0.000,276
	1.12	0.04461	0.00226	1.10818	—0.000,949**	0.000,276
	1.16	0.06172	0.00164	1.15272	—0.000,156	0.000,200
D. persimilis	*1.07*	0.02489	0.00242	1.05899	0.000,592	0.000,296
	1.12	0.03862	0.00203	1.09300	0.000,094	0.000,248
	1.18	0.07056	0.00367	1.17640	—0.000,508	0.000,448
	1.17	0.06667	0.00231	1.16593	—0.000,142	0.000,303
D. miranda	*1.07*	0.02889	0.00450	1.06878	0.000,205	0.000,549
	1.08	0.02674	0.00453	1.06349	0.000,456	0.000,553
	1.11	0.03902	0.00282	1.09401	0.000,008	0.000,345
	1.15	0.05814	0.00275	1.14324	—0.000,768*	0.000,336

The symbols * and ** denote significant deviation from zero at the 0.05 and 0.01 levels, respectively.

differences between the *est-5* alleles by calculating coefficients of determination (LI, 1955 p. 149). The coefficients of determination of M_0 and K_r on M_5 were calculated according to the method given in the appendix. The coefficients of determination of charge, *i.e.*, M_0, and shape or size, *i.e.*, K_R, on relative mobilities in 5% gels, *i.e.*, M_5, were calculated from the data of Table 4. The values were found to be $d_{M_5 \cdot M_0} = 1.055$, $d_{M_5 \cdot K_r} = 0.010$ and $d_{M_5 \cdot M_0 K_r} = -0.065$. These coefficients imply that variation in M_0 is roughly 100 times more important in determining mobility variation in 5% gels than is variation in K_R. A similar result is found for any gel concentration less than 25%.

DISCUSSION

Average effect of a charge change: We have presented evidence that indicates the average effect of an added electron on the electrophoretic mobility of *est-5*$^{1.00}$ dimer is a change of 0.023. The standard deviation of this effect is 0.015. We have also shown that almost all variation in mobility of the *est-5* alleles is due to charge variation. The average effect of a single electron should be half the effect of an amino acid charge change because the est-5 molecule is composed of two identical subunits. Therefore, the average effect of an amino acid charge change is estimated to be 0.046 with a standard deviation of 0.029. This finding suggests that many of the alleles at this locus that differ in mobility by only 0.01 are probably not unit charge changes in the sense of the step model. It is possible that many of these differences are due to conservative amino acid changes, *i.e.*, acidic for acidic, *etc.*, which cause slight changes in charge.

The method of Ferguson plots was used to compare 13 alleles of the *est-5* locus in *D. pseudoobscura, D. persimilis,* and *D. miranda.* It was found that most electrophoretically distinct alleles had negligible differences in K_R. Two of the

12 K_r values were significantly different from zero but were also very small. This finding is substantially different than the results reported by JOHNSON (1976) for the α-glycerophosphate dehydrogenase locus in Colias butterflies. He reports that almost all alleles that differed in M_0 also differed in K_R. The reasons for this difference in results are not known. One possible reason is that JOHNSON (1976) appears to have used an improper experimental procedure to measure K_R values. JOHNSON (1976) states (p. 155) that in his experiments the *concentration of bis-acrylamide* was kept constant in all gels. Results of both HEDRICK and SMITH (1968) and RODBARD and CHRAMBACH (1971) show that the *ratio of concentrations of bis-acrylamide to acrylamide* must be kept constant in order for the Ferguson relation to be obeyed. The effect of this improper procedure on the detection of K_R differences is difficult to assess. It is possible that some loci will show substantial polymorphism of K_R values while others are essentially monomorphic for K_R, even though they may all be polymorphic for electrophoretic mobility. Further work will be necessary to determine the level of polymorphism for K_R values at these and other loci.

Comparison to hemoglobin data: The general result that unit charge changes cause rather large changes in relative mobility may also be observed in the human hemoglobin-A alleles. Here we have an ideal situation for such an investigation, as we know the precise amino acid change for many of the alleles. HUEHNS and SHOOTER (1965) report data on electrophoretic mobility of a number of Hb-A variants. Using the data in their Figure 8 and Table II, we may examine the effect of charge changes on electrophoretic mobility. The alleles used and the change in relative mobility, ΔM, are given in Table 5. Here we see that an amino acid charge change causes an average $|\Delta M|$ of 0.22 with a standard deviation of 0.09.

TABLE 5

Changes in relative mobility (R_m) *of Human Hb-A alleles, calculated from the data of* HUEHNS and SHOOTER *(1965)*

Hb allele	Charge change	R_m to Hb-A	$\Delta R_m/e$
Hb-A_2*	—4	0.536	0.116
Hb-C	—4	0.481	0.130
Hb-E	—4	0.530	0.118
Hb-O	—4	0.608	0.098
Hb-S	—2	0.735	0.133
Hb-L	—2	0.790	0.105
Hb-M	0	1.000	——
Hb-Seattle	2	1.000	0
Hb-J	2	1.254	0.127
Hb-Norfolk	2	1.348	0.174
Hb-I	4	1.478	0.120

Here $\Delta R_m/e$ denotes the change in R_m per electron charge change.
* Not an allele of the Hb-A locus but is closely related to Hb-A locus alleles.

This value is dramatically larger than the corresponding value for *est-5*$^{1.00}$ dimer. This is expected, as we have shown that the net charge of Hb-A is only half that of *est-5*$^{1.00}$ in our buffer system. The data of HUEHNS and SHOOTER (1965) were obtained in a buffer with a slightly lower pH (pH = 8.6). This difference in conditions would tend to reduce the charge of Hb-A. Also, any differences caused by K_R differences between alleles is included in the data of HUEHNS and SHOOTER (1965). All these factors make the comparison of the Hb-A data to our data for *est-5*$^{1.00}$ difficult. We feel that the data from Hb-A alleles are in rough agreement with our findings. In fact the Hb-A data suggest that the average effect is even larger than our method of using altered protein has shown.

Application of step model: The results we have presented here imply that electrophoresis detects alleles that differ in mobility by much less than what unit charge changes are expected to cause. A large proportion of the alleles at the *est-5* locus in *D. pseudoobscura* (PRAKASH, LEWONTIN and HUBBY 1969; McDOWELL and PRACKASH 1976) and in *D. persimilis* and *D. miranda* (COBBS and PRAKASH 1977) probably differ by much less than a unit charge. Many of these alleles may differ by conservative amino acid changes. That such conservative changes are detectable is apparent in Table 5. Here Hb-A_2, Hb-C, Hb-E, and Hb-O should all be identical under the step model; yet they differ substantially in electrophoretic mobility. The same result is true for other alleles in Table 5 and elsewhere in the literature. It appears that substitution of a neutral amino acid for another neutral amino acid may also be detected by electrophoresis. Such substitutions could alter dissociation constants of nearby charged amino acids. Such conjecture could be tested by analyzing amino acid differences between alleles.

If the above conjectures are true, the implications for the application of the step model are profound. Clearly the step model is not an accurate description of the polymorphism at the *est-5* locus. An application of the step model such as that employed by OHTA (1975) is not appropriate. All detected alleles should not be considered to belong to different mobility classes. A partition of the mobility variation using the average effect of an amino acid charge change on mobility may be more appropriate.

We shall induce a partition on the electrophoretic alleles of a sample by assuming the mobility of a central allele is the center of a mobility class of width w. We complete the partition by successively adding mobility classes of width w to the right and left of the original central class. The original central class will be assigned the value 0 and the successive classes to the right of 0 will have values 1, 2, . . . and those to the left of 0 will have values-1, -2, . . . Here w is the average effect of an amino acid charge change on electrophoretic mobility, and for the *est-5* locus we have estimated w to be 0.046. We now have a discrete distribution on the integers which can be compared to theoretical results for the step model. We then estimate the variance of the population from which this sample was taken by standard methods and denote this estimate by $\hat{s}^2$.

An unbiased estimate of the sample variance of $\hat{s}^2$ is

$$\mathrm{Var}_s(s^2) = N \left\{ \frac{M_4 - G}{(N-1)^2} - \frac{2(M_4 - 2G)}{N(N-1)^2} + \frac{M_4 - 3G}{N^2(N-1)^2} \right\} \tag{5}$$

where M_4 is an unbiased estimate of the fourth central moment and is given on page 352 of CRAMÉR (1946). Here G is an unbiased estimate of the square of the second central moment and may be obtained using formulae on page 348 of CRAMÉR (1946) as

$$G = \frac{N^3 s^2 - (N-1)^2 M_4}{N^3 - 3N^2 + 5N - 3}$$

where s^2 is the sample variance and
N is the sample size.

As an indication of goodness of fit to the step model, we shall compare $\hat{s}^2$ to theoretical expectations for the variance of the above discrete distribution. MORAN (1975) has shown that in a population at equilibrium, the expected value of the variance of mobility classes is given by

$$\mathrm{Var}(i) = 2(N_e - 1)v \tag{6}$$

where i is the value of the mobility class and N_e and v are the same as before. Since v is expected to be very small we shall use $2N_e v$ as the expected value of the variance of i. Let x_i be the proportion of the sample in the ith mobility class and define

$$f = \sum_i x_i^2$$

as the sample homozygosity. The sample effective number of alleles is obtained as f^{-1}. Denote by Θ the estimated value of $2N_e v$ which, in view of (1) will be obtained by the formula

$$\Theta = \frac{1}{4}\{f^{-2} - 1\} \ . \tag{7}$$

This value of Θ may be compared to $\hat{s}^2$ as an indication of goodness-of-fit. To this end we define the statistic

$$T = \hat{s}^2 - \Theta \ .$$

Using a Taylor series expansion of (7) about the expected value of f and ignoring terms of degree higher than two, it may be shown that the expected value of Θ is approximately

$$E(\Theta) \simeq 2N_e v + f^{-4} V_s(f) \ .$$

Here $V_s(f)$ denotes the sample variance of f and may be evaluated using equation (12) of NEI and ROYCHOUDHURY (1974). Since $E(T) = E(\hat{s}^2) - E(\Theta)$ and we are assuming $E(\hat{s}^2) = 2N_e v$, we find

$$E(T) \simeq -(3/4) f^{-4} V_s(f) \tag{8}$$

which holds for large sample sizes. The variance of T may be obtained using the relation

$$V_s(T) = V_s(\hat{s}^2) + V_s(\Theta) - 2\text{Cov}_s(\Theta,\hat{s}^2) \ . \tag{9}$$

We will not attempt here a precise test using the distribution of T. We prefer to concentrate on the standardized variable

$$L = \frac{T - E(T)}{\sqrt{\text{Var}(T)}} \tag{10}$$

which has expectation zero and unit variance. This statistic may be used as an index of deviation from the step model. The significance of L may be determined by a two standard deviation rule. In order to calculate L we need to know the sample variance of both $\hat{s}^2$ and Θ, as well as the sample covariance of $\hat{s}^2$ and Θ.

NEI and ROYCHOUDHURY (1974) have given a formula for the sample variance of f. Their formula may be used to obtain an approximate sample variance of Θ. Using a Taylor series expansion of (7) about the expected value of f, using the definition of the variance of Θ, and taking expectations over the density of f, one may find after ignoring central moments of degree higher than two, that

$$V_s(\Theta) \simeq \frac{V_s(f)}{4f^6} \ . \tag{11}$$

Here $V_s(\Theta)$ and $V_s(f)$ denote the sample variance of Θ and f respectively, and $V_s(f)$ is obtained by the formula $V_s(f) = \frac{2(N-1)}{N^3}\{(3-2N)f^2 + 2(N-2)\sum_i x_i^3 + f\}$ from NEI and ROYCHOUDHURY (1974). The relation (11) is only valid for large sample sizes.

The covariance of Θ and $\hat{s}^2$ is found by a method analogous to that used to obtain $V_s(\Theta)$. Define the vector $\mathbf{x} = \{x_i\}$ where x_i is the same as before and i ranges over the negative and positive integers including zero. We will find it convenient to use the function

$$q_{m,n} = \sum_i x_i^m i^n \ .$$

We may then write the following vector functions

$$\Theta(\mathbf{x}) = \frac{1}{4}\{q_{20}^{-2} - 1\} \tag{12}$$

$$\hat{s}^2(\mathbf{x}) = \left\{\frac{N}{N-1}\right\}\{q_{12} - q_{11}^2\} \tag{13}$$

and

$$g(\mathbf{x}) = \Theta(\mathbf{x})\hat{s}^2(\mathbf{x}) \ . \tag{14}$$

The covariance of Θ and $\hat{s}^2$ may then be found as

$$\text{Cov}(\Theta,\hat{s}^2) = E\{g(\mathbf{x})\} - E\{\Theta(\mathbf{x})\}E\{\hat{s}^2(\mathbf{x})\} \tag{15}$$

where expectations are taken over the density of $\mathbf{x}$.

Now denote the expected value of $\mathbf{x}$ by $\hat{\mathbf{x}}$. Express $g(\mathbf{x})$, $\Theta(\mathbf{x})$, and $\hat{s}^2(\mathbf{x})$ as a multivariate Taylor series about the point $\hat{\mathbf{x}}$. Taking expectations over the density

of $\mathbf{x}$ and ignoring central moments of x_i higher than the second, we may express $E\{g(\mathbf{x})\}$, $E\{\Theta(\mathbf{x})\}$, and $E\{\hat{s}^2(\mathbf{x})\}$ as approximate functions of N, $\mathbf{x}$, $\mathrm{Var}(x_i)$ and $\mathrm{Cov}(x_i x_j)$. The only estimate we have of $\hat{\mathbf{x}}$ is $\mathbf{x}$ so we use $\mathbf{x}$ in place of $\hat{\mathbf{x}}$ and use the well-known result for the multinomial density

$$\mathrm{Var}_s(x_i) = \frac{x_i(1-x_i)}{N} \tag{16}$$

$$\mathrm{Cov}_s(x_i,x_j) = \frac{-x_i x_j}{N} \ . \tag{17}$$

After substituting (16) and (17) into the expanded expressions for $E\{g(\mathbf{x})\}$, $E\{\Theta(\mathbf{x})\}$, and $E\{\hat{s}^2(\mathbf{x})\}$, and substituting the resulting equations into (15) some rearrangement yields

$$\mathrm{Cov}_s(\Theta,\hat{s}^2) \simeq \frac{2(q_{11}q_{21} - q^2_{11}q_{20}) + q_{12}q_{20} - q_{22}}{(N-1)q^3_{20}} + \frac{(q_{12} - q^2_{11})\{6(q_{30} - q^2_{20}) - q_{20}(1-q_{20})\}}{2N(N-1)q^4_{20}} \ . \tag{18}$$

Again this relation is only valid for large sample sizes. Equations (5), (8), (9), (10), (15) and (18) now provide the means to estimate the sample variance of T and thus the ability to calculate L.

The above analysis was applied to the data of PRAKASH (1976) for *est-5* alleles of *D. pseudoobscura* and *D. miranda*. This set of data was chosen because of its large sample size. We were unable to apply the analysis to the data for *D. persimilis* because the *est-5*$^{1.ss}$ allele in this species behaves as a monomer under the conditions of electrophoresis used (see COBBS and PRAKASH 1977). In applying the analysis, the value of w was 0.046 and the mobility used as the center of the zero class was 1.02 for *D. pseudoobscura* and 1.03 for *D. miranda*. The results of this analysis are given in Table 6.

TABLE 6

Results of analysis of goodness-of-fit to step model

	D. pseudoobscura	*D. miranda*
Sample size	350	120
$\hat{s}^2$	1.3939	0.6270
Θ	1.6782	0.5439
$\mathrm{Var}_s(\hat{s}^2)$	0.0212	0.0150
$\mathrm{Var}_s(\Theta)$	0.0554	0.0204
$\mathrm{Cov}_s(\Theta,\hat{s}^2)$	0.0233	0.0136
T	−0.2844	0.0831
$E(T)$	−0.0215	−0.0193
$\mathrm{Var}_s(T)$	0.0300	0.0082
L	−1.5182	1.1320

TABLE 7

Values of L *obtained with various* w *values*

w	*D. pseudoobscura*	*D. miranda*
0.030	—2.407	4.800
0.038	—3.165	1.132
0.046	—1.518	1.132
0.054	—3.821	0.743
0.062	—6.464	—0.888

The statistics presented in Table 6 show that the alleles of both species give an acceptable fit to the step model using a two-standard-deviation rule to test the significance of L. Although not significant by our test, the variation of mobility in *D. pseudoobscura* is lower than expected, while that of *D. miranda* is higher than expected. The two species do not differ from the step model in the same direction. Thus the information in Table 6 gives no indication of significant deviation from the step model.

We should point out a source of error in the above analysis. The value of w is a particularly important variable in the whole procedure and we have not included errors involved in its estimation. Our estimate of w of 0.046 is based on only 13 measurements and the standard error of this estimate is roughly 0.008. In order to assess the effect of variation of w on our analysis, we did the entire analysis using w values one and two standard error units in both directions from our estimated value. The results of this sensitivity analysis are given in Table 7. Here we see that the value of L is sensitive to variation in the value of w. If the value is shifted one standard error unit in either direction from our estimated value the L values for *D. pseudoobscura* become significant. This result makes us less confident that there is a nonsignificant deviation from the step model. Clearly more accurate estimates of w will be necessary to support more precise conclusions. Also we must keep in mind that our statistical analysis is only an approximate one. A more precise statistical test with a stronger theoretical basis is desirable.

The data presented here suggest that the step model is a useful first approximation as an explanation of the polymorphism of this locus in both species examined here. More exhaustive studies will be necessary to determine if other forces are also affecting the polymorphism at this locus.

The authors wish to express appreciation to NANCY ISTOCK for her expert technical assistance. We also thank DR. RICHARD DOHERTY for the purified human hemoglobin. This work was supported by grant GM 19217 and by Genetics Training Grant GM-00658-15, both from the Public Health Service.

LITERATURE CITED

BRADSHAW, R. A., W. T. SHEARER and F. R. N. GURD, 1968 Sites of binding of Copper (II) Ion by peptide (1-24) of Bovine serum albumin. J. Biol. Chem. **243:** 3817-3825.

COBBS, G. and S. PRAKASH, 1977 A comparative study of the esterase-5 locus in *Drosophila pseudoobscura*, *D. persimilis* and *D. miranda*. Genetics **85:** 697-711.

CRAMER, H., 1946 *Mathematical methods of statistics.* Princeton University Press, Princeton, N.J.

FOTHERGILL, L. A. and J. E. FOTHERGILL, 1970 Thiol and dissulphide contents of hen ovalbumin. Biochem. J. **116:** 555–561.

GORIN, M. H., 1941 An equilibrium theory of ionic conductance. J. Chem. Phys. **7:** 405.

HEDRICK, J. L. and A. J. SMITH, 1968 Size and charge isomer separation and estimation of molecular weight of proteins by disc gel electrophoresis. Arch. Biochem. Biophy. **126:** 155–164.

HENNING, U. and C. YANOFSKY, 1963 An electrophoretic study of mutationally altered A proteins of the tryptophan synthetase of *Escherichia coli.* J. Mol. Biol. **6:** 16–21.

HUBBY, J. L. and R. C. LEWONTIN, 1966 A molecular approach to the study of genetic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura.* Genetics **54:** 577–594.

HUEHNS, E. R. and E. M. SHOOTER, 1965 Human hemoglobins. J. Med. Genet. **2:** 48–90.

JOHNSON, G. B., 1974 On the estimation of effective number of alleles from electrophoretic data. Genetics **78:** 771–776. ——, 1976 Hidden alleles at the α-glycerophosphate dehydrogenase locus in the Colias butterflies. Genetics **83:** 149–167.

KIMURA, M. and J. F. CROW, 1964 The number of alleles that can be maintained in a finite population. Genetics **49:** 725–738.

LEWONTIN, R. C., 1974 *The Genetic Basis of Evolutionary Change.* Columbia University Press, New York.

LI, C. C., 1955 *Population Genetics.* University of Chicago Press, Chicago and London.

MARSHALL, D. R. and A. H. D. BROWN, 1975 The charge-state model of protein polymorphism in natural populations. J. Mol. Evol. **6:** 149–163.

MCDOWELL, R. E. and S. PRAKASH, 1976 Allelic heterogeneity within allozymes separated by electrophoresis in *Drosophila pseudoobscura.* Proc. Nat. Acad. U.S. **73:** 4150–4153.

NEI, M. and A. K. ROYCHOUDHURY, 1974 Sampling variances of heterozygosity and genetic distance. Genetics **76:** 379–390.

MORAN, P. A. P., 1975 Wandering distributions and the electrophoretic profile. Theor. Pop. Biol. **8:** 318–330.

OHTA, T., 1975 Statistical analyses of *Drosophila* and human protein polymorphisms. Proc. Nat. Acad. Sci. U.S. **72:** 3194–3196.

OHTA, T. and M. KIMURA, 1973 A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet. Res. **22:** 201–204.

PERLMAN, G. E., 1955 The nature of phosphorus linkage in phosphoproteins. pp. 1–30. In: *Advances in Protein Chemistry*, Vol. X. Academic Press, New York.

PRAKASH, S., 1977 Genetic divergence in closely related sibling species *Drosophila pseudoobscura, Drosophila persimilis*, and *Drosophila miranda.* Evolution: (in press).

PRAKASH, S., R. C. LEWONTIN and J. L. HUBBY, 1969 A molecular approach to the study of genic heterozygosity in natural populations. IV. Patterns of genic variation in central, marginal and isolated populations of *Drosophila pseudoobscura.* Genetics **61:** 841–858.

RICHARDSON, R. H., M. E. RICHARDSON and P. E. SMOUSE, 1975 Evolution of electrophoretic mobility in the *Drosophila mulleri* complex. pp. 533–545. In: *Isozymes IV Genetics and evolution.* Edited by L. MARKERT. Academic Press, New York, San Francisco, London.

RODBARD, D. and A. CHRAMBACH, 1971 Estimation of molecular radius, free mobility and valence using polyacrylamide gel electrophoresis. Anal. Biochem. **40:** 95–134.

TANFORD, C., 1961 *Physical Chemistry of Macromolecules.* Wiley, New York.

Corresponding editor: R. C. LEWONTIN

APPENDIX

Here we consider the problem of determining the relative importance of charge and K_r sensitive shape or size differences in determining the variation in electrophoretic mobility. The coefficient of determination (LI 1955, p. 149) is a natural choice of measure, since all coefficients must sum to one. We will first give a brief description of coefficients of determination and then apply it to electrophoretic mobility variation. The coefficient of determination of variable A on variable X is denoted by $d_{X.A}$ and defined as that fraction of complete determination of X from which the cause A is directly responsible in a given system of related variables (LI 1955, p. 149).

For example, if the random variable X is the sum of the random variables A and B, we may then define

$$d_{X.A} = \frac{\text{Var }(A)}{\text{Var }(X)} \tag{i}$$

$$d_{X.B} = \frac{\text{Var }(B)}{\text{Var }(X)} \tag{ii}$$

$$d_{X.AB} = \frac{2\text{Cov }(A,B)}{\text{Var }(X)} \ . \tag{iii}$$

Here the coefficient $d_{X.AB}$ is a measure of the correlated effect of A and B. The sum of the above 3 coefficients of determination is one.

Relative electrophoretic mobilities of proteins in a given gel system may be described by two parameters. The parameters are the intercept and slope of Ferguson plots (see RODBARD and CHRAMBACH for further discussion) and will be denoted here as y_0 and K_r respectively. The relative electrophoretic mobility of two proteins in a gel of concentration t will be denoted M_t. We then have the following relationship.

$$\log M_t = y_0 + K_r t \ . \tag{iv}$$

The value of $\log^{-1} y_0$ is related to size difference and charge difference of the proteins by equation (4). If the two proteins are of equal size and shape then y_0 is the log of the ratio of their charges. If the proteins are of different sizes, then y_0 is the sum of the log of the ratio of charges and another term depending on the buffer systems and the sizes of the proteins. We let $y_0 = V + S$ where V is the log of the ratio of charges of the two proteins (see eq. (4) of text). Then (iv) becomes

$$\log M_t = V + K_r t + S \tag{v}$$

where S is a constant for any two proteins and buffer system which depends on the size of the proteins. We will limit the present discussion to cases where $S = 0$, *i.e.*, proteins of identical size. In this case equation (iv) is used and y_0 is the log of charge ratio and K_r is the slope of the Ferguson plot.

Equation (iv) is the same form as the example given and equations (i), (ii)

and (iii) could be used to measure the relative importance of y_0 versus K_r differences in proteins. Although useful, such a procedure lacks intuitive appeal and a method dealing with M_t directly is desirable. We rewrite equation (iv) as

$$M_t = M_0\, 10^{K_r t} \qquad \text{(vi)}$$

where $M_0 = 10^{y_0}$. Here equation (vi) may be regarded as a function of variables M_0 and K_r which may be considered measures of charge and shape differences between the proteins respectively. Using a bivariate Taylor series expansion about the mean values of M_0 and K_r and dropping terms of degree higher than two we obtain the approximate equation

$$\begin{aligned} M_t \simeq\ & \overline{M}_0\, 10^{\overline{K}_r t} + 10^{\overline{K}_m t}\,(M_0 - \overline{M}_0) + \overline{M}_0\,(2.303t)\,10^{\overline{K}_r t}\,(K_r - \overline{K}_r) \\ & + (2.303T)\,10^{\overline{K}_r t}\,(M_0 - \overline{M}_0)\,(K_r - \overline{K}_r) \\ & + \overline{M}_0\,(2.303)^2\,10^{\overline{K}_r t}\,(K_r - \overline{K}_r)^2\ . \end{aligned} \qquad \text{(vii)}$$

Using (vii) we find

$$\begin{aligned} \mathrm{Var}(M_t) \simeq\ & 10^{2\overline{K}_r t}\mathrm{Var}(M_0) + \overline{M}_0\,(2.303T)\,10^{2\overline{K}_r t}\mathrm{Cov}(M_0, K_r) \\ & + \overline{M}_0\,(2.303T)^2\,10^{2\overline{K}_r t}\,\mathrm{Var}\,(K_r)\ . \end{aligned} \qquad \text{(viii)}$$

Equation (viii) has partitioned the variance of M_t into 3 components. We then define the coefficients of determination as

$$d_{M_t \cdot M_0} = 10^{2\overline{K}_r t}\left(\frac{\mathrm{Var}\,M_0}{\mathrm{Var}\,M_t}\right) \qquad \text{(ix)}$$

$$d_{M_t \cdot K_r} = \overline{M}_0^2 (2.303T)^2\, 10^{2\overline{K}_r t}\left\{\frac{\mathrm{Var}(K_r)}{\mathrm{Var}(M_t)}\right\} \qquad \text{(x)}$$

$$d_{M_t \cdot M_0 K_r} = \overline{M}_0^2\,(2.303T)\,10^{2\overline{K}_r t}\left\{\frac{\mathrm{Cov}(M_0, K_r)}{\mathrm{Var}(M_t)}\right\} \qquad \text{(xi)}$$

where Var (M_t) is given by (viii). The coefficient $d_{M_t \cdot M_0}$ may be regarded as that proportion of the mobility variation in a gel of concentration t which is due to charge variation alone. The coefficients $d_{M_t \cdot K_r}$ and $d_{M_t \cdot M_0 K_r}$ have analogous interpretations. Equations (ix), (x) and (Xi) should be used only if the standard deviation of both these variables is small. If the standard deviation of these variables is not small, then the first method using equation (i), (ii) and (iii) where $X = \log M_t$, $A = y_0$ and $B = K_r t$ is preferable.