HIDDEN ALLELES AT THE α-GLYCEROPHOSPHATE DEHYDROGENASE LOCUS IN COLIAS BUTTERFLIES[1]

GEORGE B. JOHNSON[2]

Department of Biology, Washington University, St. Louis, Missouri 63130

Manuscript received August 18, 1975
Revised copy received November 14, 1975

ABSTRACT

By varying polyacrylamide gel pore size, the α-glycerophosphate dehydrogenase locus of Colias butterflies is shown to contain at least five alleles, rather than the two which had been reported previously. Two of the alleles have the same apparent net charge, and presumably are detected electrophoretically because of conformational differences. Additional variation occurs in the isoelectric points of the proteins. It is suggested that electrophoresis employing a single gel of intermediate pore size will fail to discriminate between many alleles, and that the concept of electrophoretic alleles as differing simply in charge may not always be appropriate.

"How does one know what it is one believes
When it's so difficult to know what it is one knows?"
—*Jumpers* (STOPPARD, 1972)

OVER the last decade, an increasing body of data has accumulated indicating that genetic variability is common within natural populations. Large numbers of variants have been detected by systematic electrophoretic screening of samples, followed by genetic analysis of electrophoretically distinct types. In this fashion it is possible to describe alleles at an enzyme locus in terms of their electrophoretic mobilities, and to enumerate their relative frequencies by suitable sampling procedures. It has been clear from the outset, however, that this approach reveals only a fraction of the total variability—specifically, that fraction involving electrophoretically detectable differencies. The magnitude of the "hidden" fraction undetected by electrophoresis has not been clear. It is possible to make a preliminary estimate by assuming that electrophoretic screening detects only those differences which involve a change in net charge: if alleles at a locus represent proteins which differ by single amino acid substitutions to or from one of the six "charged" amino acids (GLU, ASP, TYR, HIS, ARG, LYS), one can calculate from the genetic code that the fraction of detectable amino acid substitutions ought to be approximately one-third of all possible substitutions (LEWONTIN 1974).

[1] CIW-DPB Publication Number 561.

[2] Present address: Carnegie Institution of Washington, Department of Plant Biology, 290 Panama Street, Stanford, California 94305.

In the absence of direct empirical data addressing the issue, most workers have accepted this simple model. This conceptualization of unit charge differences has led to a redefinition of the effective number of alleles (n_e) to compensate for undetected variation resulting from the even cancelling of off-setting unit positive and unit negative charges (OHTA and KIMURA 1973). The new theory is based upon the conceptualization of electrophoretic bands as representing one or more alleles with the same net charge. Under such a model, many variants would not be detected because unit positive and negative charges cancel, producing families of alternative alleles with the same net charge which appear as one elecrophoretic band. The lumping of different alleles into single net charge categories visualized by OHTA and KIMURA implies that significantly less than one-third of all variants are detected by electrophoresis (KING and OHTA 1975), and suggests that the relatively even spacing often observed for the bands of alternative alleles results from a progressive series of such unit difference in net charge (RICHARDSON, RICHARDSON and SMOUSE 1975). This "ladder" model has gained rapid acceptance (EWENS and FELDMAN 1975), despite some objections (JOHNSON 1974).

Over the last several years there have been indications that heterogeneity within electrophoretic allelic classes does indeed exist. Variation in electrophoretic mobility unexplainable by experimental error was reported four years ago (JOHNSON 1971). Examination of the heat stabilities of allozymes has yielded direct evidence of heterogeneity (BERNSTEIN, THROCKMORTON and HUBBY 1973, SINGH, HUBBY and LEWONTIN 1974). This paper describes a high resolution approach to electrophoresis which permits characterization of previously "hidden" allelic variation, and suggests that the concept of electrophoretic alleles as differing simply in charge may not always be appropriate.

An Analytical Approach to Electrophoresis. In the last ten years a great deal has been learned about the nature of electrophoretic migration (FERGUSON 1964; CHRAMBACH and RODBARD 1971). Theoretical descriptions of the process have been advanced which in a relatively simple way account for the observed behavior of proteins in polyacrylamide gel electrophoresis (FAWCETT and MORRIS 1966; RODBARD and CHRAMBACH 1970, 1971, 1974; GONENNE and LEBOWITZ 1975).

The theory stems from the observation that a protein's mobility in gel electrophoresis is a logarithmic function of gel pore size (FERGUSON 1964). This implies a frictional interaction of the protein with the gel pores, or perhaps a hydrodynamic interaction with the fibers. The theory may be stated quite simply:

$$(1) \qquad R_f = \frac{M_o}{u_f} \cdot e^{(K_r T)}$$

where R_f = mobility of protein relative to the front

u_f = apparent mobility of a moving boundary in front of the resolving phase (a constant known for most common buffer systems)

M_o = free electrophoretic mobility of protein

K_r = the retardation (frictional) coefficient

T = % acrylamide (which determines pore size and is inversely proportional to it)

In this model, the rate of migration is seen to vary not only as a function of net charge (here expressed as free electrophoretic mobility, M_o, corrected by a constant, u_f, for the buffer employed), but also as a *logarithmic* function of the pore size (expressed as % acrylamide) and of the gel-protein interaction as the protein passes through these pores (expressed as the retardation coefficient, K_r, which may also be defined in terms of log to the base 10, K_R).

In this conceptualization of RODBARD and CHRAMBACH, the retardation of the protein by the gel is essentially a geometric steric problem relating to the available fraction of the pores which the molecule can enter. Thus the contribution of the protein to the gel fiber interaction is expressed in terms of the protein's mean geometric radius, $\bar{R}$. It is not clear whether or not the dimensions of polyacrylamide gel pores would produce interactions such as RODBARD and CHRAMBACH's initial model describes. A hydrodynamic or viscus interaction between the protein molecules and the pores of the gel offers an alternative conceptualization (ORNSTEIN 1964). Both types of interaction probably exert important influences upon migration during electrophoresis. In either case, the retardation coefficient K_R would be expected to be a *polynomial* function of the protein's effective radius, and thus quite sensitive to small changes in that radius. In principle, then, K_R offers a sensitive means of characterizing the non-charge contribution of a protein to its electrophoretic migration rate.

Differences in K_R between proteins may reflect molecular weight differences resulting from small differences in size, or from differences in the degree of subunit aggregation. A regular relationship is seen between K_R and molecular weight (Figure 1). In comparing allelic proteins the first alternative is not often expected (but see SCHLESSINGER, BLOCK and KELLEY 1975), as the alleles are commonly thought to represent single amino acid substitutions. Binding of ions or other low molecular weight compounds to protein may alter K_R. Suitable controls may be run to investigate the possibility of binding artifacts. Differences in K_R may also

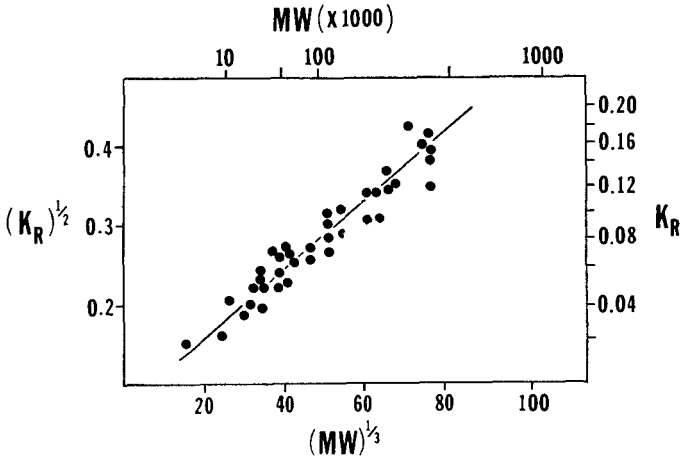

FIGURE 1.—The retardation coefficient as a function of molecular weight. The data are those of HEDRICK and SMITH 1968; RODBARD and CHRAMBACH 1971; JOHNSON 1975a.

reflect conformational differences, or, in principle, localized fractional charge interactions between the protein and the gel fibers. It is not possible from electrophoretic analysis alone to clearly distinguish between the third and fourth alternative, although the difference is conceptually important.

The power of the above theory is that it permits indpendent characterization of the contributions of charge and of interactive effects such as size or conformational difference to the electrophoretic behavior of a protein. These parameters may be independently estimated by the simple expedient of taking the log of both sides of equation (1), yielding an equation of linear form:

$$\underbrace{\log R_f}_{y} = \underbrace{\log \frac{M_o}{u_f}}_{\text{intercept}} + \underbrace{K_R}_{\text{slope}} \cdot \underbrace{T}_{x} \tag{2}$$

Regressing log R_f on T, one obtains a linear plot with an intercept whose antilog is M_o divided by a constant, u_f, and whose slope is K_R (Figure 2). One is thus able to estimate the charge and interactive contributions to mobility independently, *with associated error statements on each estimate.* It is important to note that the slope K_R is usually negative.

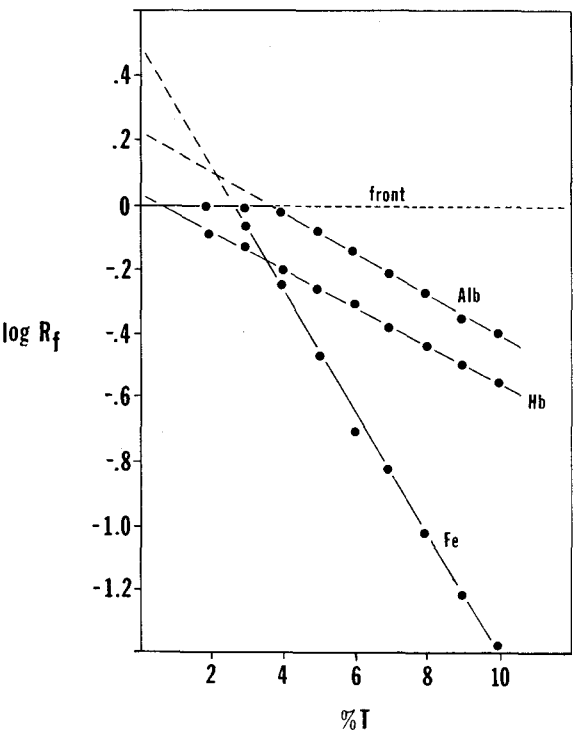

FIGURE 2.—Relative mobility as a function of acrylamide concentration. The proteins are horse spleen ferritin (Fe), human hemoglobin (Hb), and bovine serum albumin. %T is the % acrylamide of the running gel in the Davis-Ornstein polyacrylamide disc gel system (DAVIS 1964), run in TRIS-glycine buffer, running pH 9.5.

Experimental Error. In order to insure the comparability of results, it is particularly important that experiments be adequately standardized. I employ two internal standards throughout (JOHNSON 1971, 1975a).

In all of the work, uniform experimental conditions were employed and u_f is taken as 1.18. Determination of the value of the constant u_f is subject to large errors (RODBARD and CHRAMBACH 1974). It may be preferable to simply characterize mobility as the intercept Y_o (where $Y_o = M_o/u_f$) (RODBARD and CHRAMBACH 1970); values of Y_o for these experiments may be rederived simply by dividing M_o by 1.18. Because u_f is a constant, scaling error introduced in estimating u_f will not alter estimates of mobilities relative to one another.

I have used human hemoglobin extensively as an internal standard when varying gel pore size (JOHNSON 1975a). Data on the electrophoretic behavior of this protein thus provide a means of assessing the errors involved in the approach. In 409 separate experiments, each involving a minimum of six gels of differing acrylamide concentration (a total of greater than 2500 polyacrylamide gels), the mean estimates for the retardation coefficient and for the free electrophoretic mobility of hemoglobin $\pm$ σ(one standard deviation) were: $K_R = -0.057 \pm 0.006$, $M_o = 1.255 \pm 0.114$. The units of K_R are (% acrylamide)$^{-1}$; the units of M_o are $cm^2V^{-1}sec.^{-1}$ Because of the large sample, the standard errors ($\sigma/\sqrt{N}$) of these two estimates are very small, less than 0.5%of the estimates in each case, indicating 95% confidence intervals which are less than 2% of estimates.

Of course, one rarely estimates K_R from 2500 gels. The mean standard errors within each six gel experiment, averaged for the 409 experiments considered above, were 0.003 (K_R) and 0.013 (M_o). Thus the within-experiment coefficient of variation ($\sigma/\bar{x} \cdot 100$) averaged 12.8 for estimations of K_R and 17.8 for estimations of M_o. The larger coefficient of variation of M_o reflects the fact that the coefficient of variation of an intercept is usually greater than that of a slope, as the intercept includes errors of slope as well as errors of mean. The additional error largely reflects the fact that the intercept is an extrapolated parameter.

These results for hemoglobin indicate that in general for a single six-gel determination, the standard error in estimating K_R will be about 5% of the estimate. Similarly, the standard error in estimating M_o will be about 8% of the estimate.

It should be noted that the use of an unweighted linear regression to estimate K_R and M_o is tenable only if the R_f values cover a fairly restricted range. In this work, R_f typically varied from 0.3 to 0.9 When R_f becomes less than 0.1, estimation of experimental error requires a "weighted" regression. (As R_f approaches zero, the logarithm of R_f becomes infinite, and small changes in R_f will result in astronomical errors in the logarithm of R_f.) Such a weighted regression procedure is described in RODBARD and CHRAMBACH (1971).

To statistically test the identity of two proteins on the basis of their $K_R - M_o$ values, it is necessary to compare the slope and intercept simultaneously (in other words, evaluate the entire line). This may be done by carrying out an F test, which has two degrees of freedom in the numerator. A suitable statistical procedure is described by RODBARD and CHRAMBACH (1974).

The Principle of Perversity. Several studies have now been reported in which

K_R and M_o values are estimated for a wide range of proteins (KAPADIA, CHRAMBACH and RODBARD 1974; JOHNSON 1975a). In general, as M_o increases, the absolute value of K_R increases. This means that for plots such as Figure 2, lines with a steeper slope will have a higher intercept—at intermediate values, the lines cross (Figure 3). This implies that *at intermediate pore sizes* (in the case of Figure 3, roughly 3%–5% acrylamide) *discrimination between different proteins will be minimal.* As much of the screening for electrophoretic variants has been conducted as just such pore sizes, many potentially distinguishable alleles may have gone undetected, "hidden" because the electrophoretic comparisons were carried out at minimal resolution.

This paper reports a survey of electrophoretically detected variability at the α-glycerophosphate dehydrogenase locus in Colias butterflies. Enzymes from individual butterflies were run on six gels of different acrylamide concentration, and K_R and M_o values determined ± stated confidence intervals. This maximizes the resolving power of the electrophoretic approach, revealing hitherto unsuspected allelic variability.

METHODS

Five species of Colias butterflies were examined (*C. meadii, C. scudderi, C. alexandra, C. philodice,* and *C. eurytheme*); the populations were sampled respectively from Colorado popula-

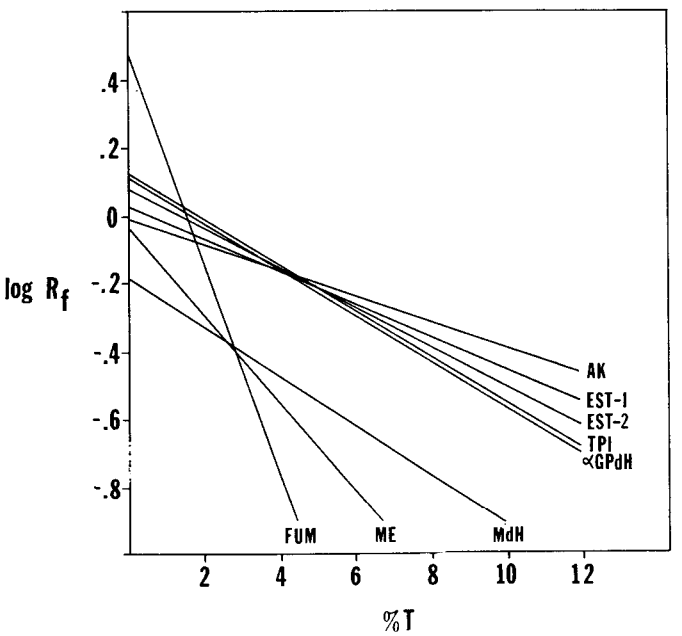

FIGURE 3.—A variety of enzymes of *C. meadii.* Data are presented for the most common allele (JOHNSON 1976). The enzymes are adenylate kinase (AK), α-glycerophosphate dehydrogenase (α-GPdH), triose phosphate isomerase (TPI), fumarase (FUM), malic enzyme (ME), and esterase (EST).

tions at Cumberland Pass, Mesa Seco, Taylor Park, East River, Hotchkiss, and Cement Creek, as described in JOHNSON (1975b). Sixty-two individuals of *C. meadii* were examined and 18 to 25 individuals of each of the other four species.

Butterflies were individually fed on a 1:3 honey:water solution and mailed from Colorado to St. Louis by air. Freshly killed butterflies, minus wings, were ground in 0.6 ml of a 0.1 *M* TRIS-HCl buffer, pH 7.8, containing 10^{-4} M dithiothreotol, 1 mM EDTA, and 10^{-4} M NaN_3, in 10% sucrose. Samples were then drawn up into five capillary tubes, centrifuged in a hemocrit centrifuge for five minutes to remove cellular debris and particulate matter, and frozen at —35°.

Electrophoresis was performed on polyacrylamide gels as described in JOHNSON (1975a). 0.02 ml of sample was loaded on each gel, together with an equal volume of internal standard solution [bromphenol blue, hemoglobin (Hb), and ferritin (Fe) in 10% sucrose]. Each individual was examined at six concentrations of acrylamide (usually 3%, 4%, 5%, 6%, 7%, and 8%) polymerized with a constant concentration of bis-acrylamide. Electrophoresis was carried out in the Ornstein-Davis TRIS-Glycine buffer system (DAVIS 1964) at 100 volts for one to two hours (JOHNSON 1975a). At the end of a run each gel was removed from its tube and the position of the bromphenol blue front marked.

Gels were individually assayed at 25° in 5 ml of a standard assay solution [containing 25 mg nicotine adenine dinucleotide (NAD), 80 mg α-glycerophosphate (α-GP), 380 mg EDTA, 2 mg phenazine methosulfate (PMS), 20 mg nitro blue tetrazolium (NBT), in 100 ml of 0.1 M TRIS-HCl buffer, pH 7.1]. Assays were generally complete within 20 minutes. Each gel was immediately scanned on a Gilford 240 spectrophotometer with linear transport, and the position of αGPdH, Hb, Fe, and dye peaks noted.

Each sampled individual was thus examined at six different acrylamide concentrations; six corresponding R_f values for αGPdH were obtained, each value relatable to two internal standards run in the same gel. An unweighted linear regression was performed on each such data set to obtain values of M_o and K_R ± one standard deviation. The use of a weighted regression did not substantially affect the results (JOHNSON 1975a).

Verification of the Mendelian nature of variants was carried out by crosses as described in JOHNSON (1975b), using procedures developed by WATT and made available to me by him.

Isoelectric focusing was carried out using procedures modified from O'BRIEN and MACINTYRE (1972). Focusing was performed at 300 V for 3 hours, using 0.16% ampholine, pH 3–10, in 6.6% acrylamide gels containing 0.44% glycerol. For each gel the pH gradient was first established by applying the electrical field at 1 mA per gel for 30 minutes, prior to application of the sample. Hemoglobin was run as an internal standard in each gel. The pH of bands is determined in this work by staining a longitudinal slice of the disc gel to locate the appropriate band positions, excising with a razor blade the corresponding region of the unstained gel (usually within 20 to 30 minutes of the completion of the run), homogenizing gel slices in 5 ml of distilled water and determination of pH carried out on a digital Orion meter freshly calibrated at pH 7 and pH 10 with Corning pH standards. Much of the experimental error is associated with determining the pH of the bands.

RESULTS

Previous work with α-GPdH in Colias butterflies has suggested that two principle alleles occur among the Colorado species, the same common alleles in all the species as judged by mobilities relative to hemoglobin when run in 7% acrylamide gels (JOHNSON 1972, 1975a). If the only difference between the two alleles involves a single amino acid *charge* difference, one might expect that the two alleles would have similar retardation coefficients, K_R, and would differ primarily in free electrophoretic mobility, M_o. The range of variability in their K_R values would be expected to be limited to about that observed for the hemoglobin standard. What one sees, however, is a distribution very much broader than that

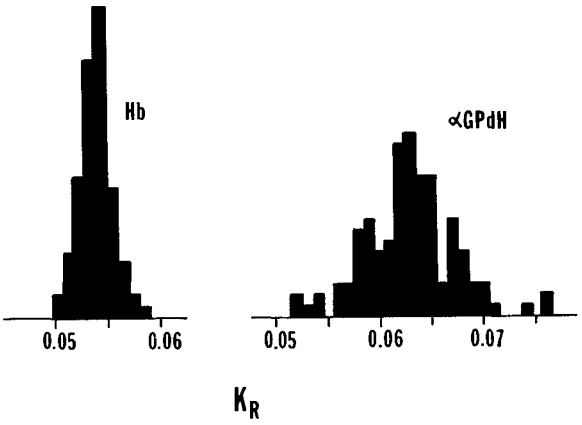

FIGURE 4.—Retardation coefficients of α-GPdH in Colias compared to hemoglobin standards run *in the same gels*. Note that the larger standard deviation in K_r does not reflect differing sample sizes.

of hemoglobin (Figure 4). As judged by a molecular weight calibration curve (Figure 1), the differences in K_R indicated by the data of Figure 4 are very substantial. The unavoidable implication is that Colias butterflies contain non-charge variation in α-GPdH.

The values of free electrophoretic mobility, M_o, estimated from the same data sets also exhibit a complex distribution, even when normalized to the hemoglobin standard (Figure 5). α-GPdH in insects is thought to be a dimer with identical subunits producing a third "heterodimer" band in heterozygotes (O'BRIEN and MACINTYRE 1972). By placing the distal bands of such multi-banded individuals into different classes, four presumptive variant classes emerge. These are indicated on Figure 5. Within class III, however, there occur six individuals which exhibit three bands at high % acrylamide, all three bands in each case having

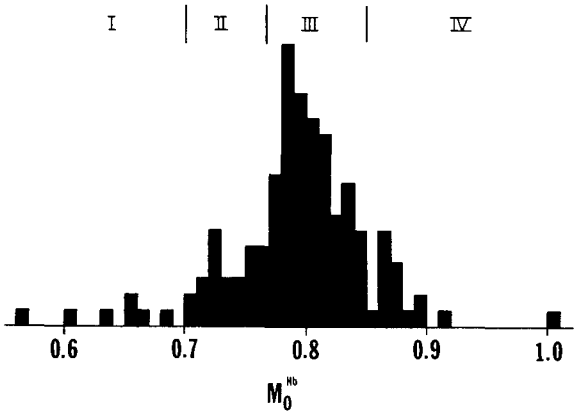

FIGURE 5.—Free electrophoretic mobilities of α-GPdH in Colias, expressed relative to hemoglobin standards run in the same gels.

identical M_o values. There thus appear to be *five* classes of proteins, two of the classes being rare and two of the common classes differing only in K_R while having identical M_o values [being identical in net charge] (Figure 6).

The experimental error associated with K_R and M_o estimates is small (Table 1). In Figure 7, individual values of M_o are plotted *vs.* corresponding K_R values. Note that alleles with higher M_o values have correspondingly higher absolute values of K_R. This is in agreement with the results of other studies (KAPADIA, CHRAMBACH and RODBARD 1974; JOHNSON 1975) and leads to the intersecting plots seen in Figures 3 and 8. Errors in estimating M_o and K_R are highly correlated, as error in intercept is not independent of error in slope. Thus significance is best judged in terms of the joint confidence limits for M_o and K_R. RODBARD and CHRAMBACH (1974) have suggested a variety of approaches to estimating the significance of differences in M_o–K_R pairs.

When individuals from the variant classes are crossed and their progeny analyzed, the resulting ratios are consistent with normal Mendelian segregation, indicating that the variants represent true alleles (Table 2). Genetic analyses was carried out principally in the lowland species *Colias philodice*. In a typical cross, 50 to 100 adults were reared from a single mating. From hatched egg to adult, larval mortality was consistently about 60%. When two individuals presumed homozygous for the same allele were crossed, the progeny exhibited K_R

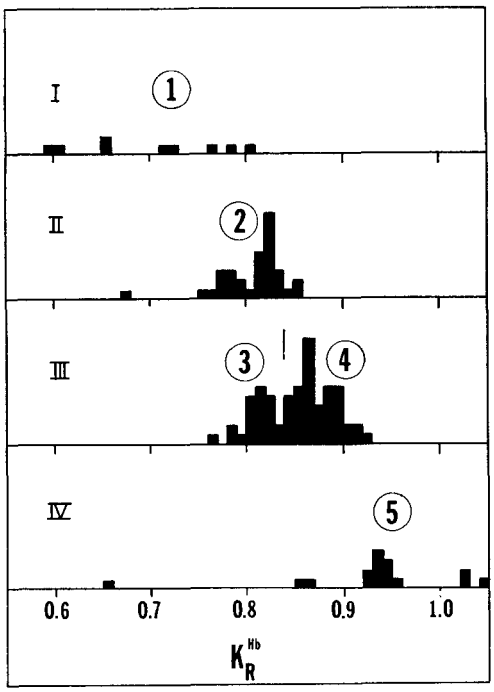

FIGURE 6.—Retardation coefficients of each of the M° classes. Values are expressed relative to hemoglobin standards run in the same gels.

TABLE 1

α-Glycerophosphate dehydrogenase variants in Colias

α-GPdH variant class	Retardation coefficient		Free electrophoretic mobility	
	$\bar{K}_R^{Hb} \pm$ S.E.	$\hat{K}_R$	$\bar{M}_o^{Hb} \pm$ S.E.	$\hat{M}_o$
1	0.65 ± .011	—0.088	0.73 ± .019	1.72
2	0.82 ± .004	—0.070	0.76 ± .005	1.65
3	0.83 ± .003	—0.067	0.81 ± .005	1.55
4	0.89 ± .003	—0.064	0.81 ± .003	1.55
5	0.96 ± .013	—0.059	0.89 ± .015	1.41

The α-GPdH variant classes are those indicated in Figure 6. K_R^{Hb} is the mean K_R value of α-GPdH standardized to the corresponding K_R value of *Hb* determined from the same experiments; the standardized value is expressed as the ratio *Hb*/ α-GPdH, ± one standard deviation. The value listed for each class represents the grand mean of that class for all six surveys of Table 3, and the listed standard error refers to the overall variance, all data considered with respect to that mean. $\hat{K}_R$ is the estimated value of K_R for α-GPdH obtained by dividing the mean K_R of hemoglobin in these experiments (—0.057 ± .006, based upon 409 replicates) by $\bar{K}_R^{Hb}$, the experimental values of α-GPdH standardized to hemoglobin. In an analogous fashion, $\bar{M}_o^{Hb}$ is the mean M_o value standardized to that of hemoglobin, and $\hat{M}_o$ for α-GPdH is obtained by dividing the mean M_o of hemoglobin (1.255 ± 0.114, based upon 409 replicates) by $\bar{M}_o^{Hb}$.

values for α-GPdH with standard errors of approximately 6% of the mean, very similar to the 5% standard error associated with the K_R values of hemoglobin determined from the same gels. This is in marked contrast to the behavior seen in Figure 3. In crosses involving a homozygote and a presumed heterozygote 1:1 segregation was observed. In only one case was I successful in obtaining a heterozygote × heterozygote cross (2/3 × 2/3, Table 2). The results were consistent with a 1:2:1 segregation pattern.

It seems particularly important to evaluate the potential genetic basis of the differences in K_R seen between variants 3 and 4. Passage through Sephadex G-50 to remove small molecules does not alter K_R, nor does significantly increasing the concentration of internal standard protein. This argues that the differences seen in K_R do not result from local ionic effects induced by the electrophoresis of impure protein.

To further investigate the genetic basis of the K_R difference between variant classes 3 and 4, I obtained eggs from wild-caught *C. eurytheme* females determined by direct analysis of K_R to be class 3 and class 4, and with several replicates, individually crossed virgin ♀ 3 × ♂ 4 homozygotes; several reciprocal crosses were also made. All the crosses were fertile, and F_1 progeny were multiple-banded at high acrylamide concentrations (Figure 9). The results of these crosses thus clearly establish the genetic basis of the K_R difference between α-GPdH variants 3 and 4.

The nature of the "hidden" alleles is now apparent (Figure 8). A survey conducted at 5% acrylamide will not discriminate between the variants. The "mobility variation" of two segregating alleles previously reported for this locus on

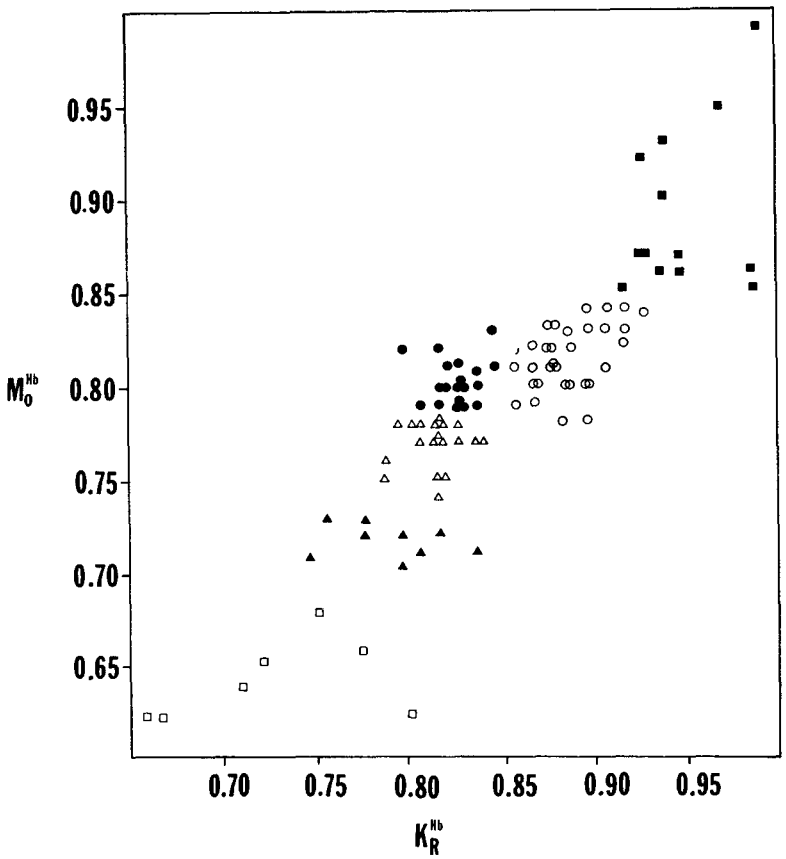

FIGURE 7.—Free electrophoretic mobility as a function of the retardation coefficient: class 5 (■), class 4 (○), class 3 (●), class 2 (▲, △), class 1 (□). Sample of 40 individuals collected at Mesa Seco, Colorado (JOHNSON 1975b). Ellipses represent joint 78% confidence intervals. The distribution of 2 suggests class 2 may be heterogeneous.

7% acrylamide gels undoubtedly reflects the same lack of resolution. Two of the three common alleles have essentially the same R_f value at 7% acrylamide. Only when the full resolving power of the electrophoretic approach is employed can the different alleles be clearly unambiguously distinguished.

It should be noted that the relative frequencies of the alleles vary between species (Table 3), although at intermediate acrylamide concentrations the differences would go largely undetected. The correspondence of K_R and M_o values between species is close, suggesting that the variants of α-GPdH in this genus are often homologous between species.

Characterization of variants in terms of retardation coefficients and free electrophoretic mobilities or Y_o's can only distinguish between variant proteins with significant charge or conformational/interactive differences. Allelic proteins with similar conformations and only fractional charge differences might well go undetected by a gel sieving analysis. Preliminary analysis of variation in

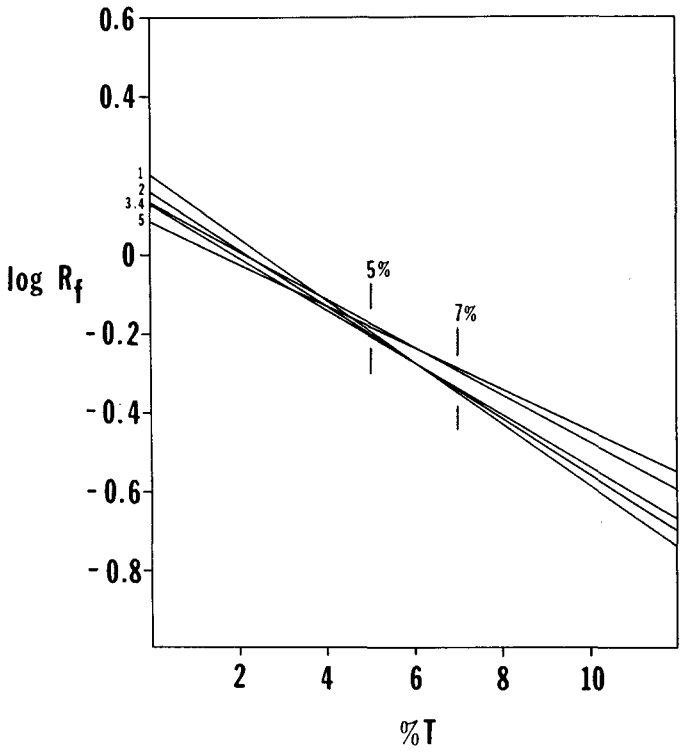

FIGURE 8.—Electrophoretic profiles of the five α-GPdH alleles in Colias. The plots are derived from the values of K_R and M_o indicated in Table I. At 5%T, no distinct alleles are detected, while at 7% two variant classes are detected. Previous work on α-GPdH in Colias reports two alleles, and was done at 7%T (JOHNSON 1971, 1975b).

isoelectric point (*pI*) suggests that variants with minor differences in charge may be common. Figure 10 presents *pI* values for α-GPdH from individuals in which the M_o of α-GPdH has also been determined. As expected, the allele of highest mobility in alkaline buffer (class 1) exhibits the lowest *pI* (indicating the most negative net charge), and the allele of lowest mobility the highest *pI*. The variability of *pI* values within a given allelic class is greater than would be predicted from that seen for hemoglobin internal standards run in the same gels: for hemoglobin, the mean *pI* was 6.86, with a standard deviation of 0.18, while the standard deviation in *pI* for the allelic classes of Figure 10 averages 0.37. The source of the added "error" appears to be heterogeneity in *pI* within allelic classes. This is suggested by the frequent observation of multiple-bands upon isoelectric focusing of alleles which had exhibited only one band in electrophoresis at high and low acrylamide concentration (Table 4); fully 32% of the individuals examined fell into this category. No heterogeneity in *pI* is seen among allele 5 individuals. Within each of the other electrophoretic classes, at least two or three variants are indicated. The existence within a natural population of numerous variants differing in isoelectric point tends to confirm recent indications that the

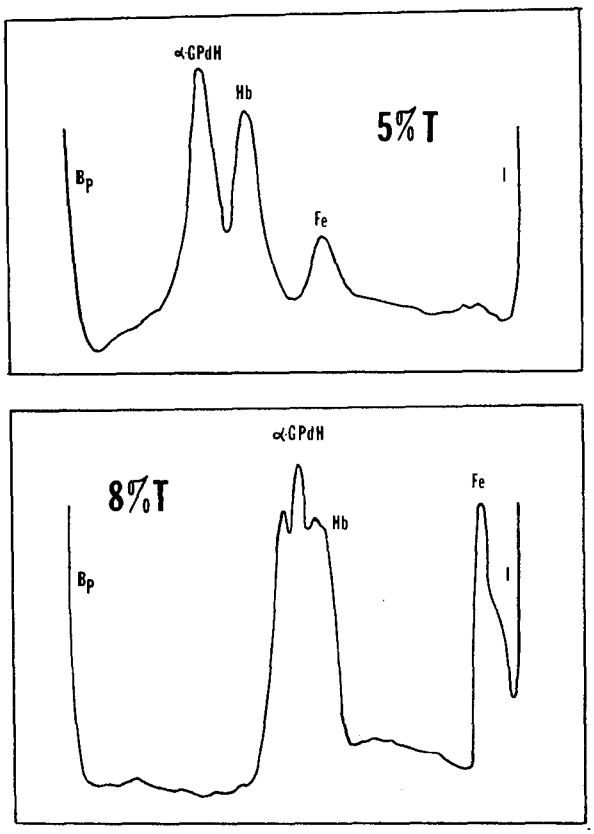

FIGURE 9.—Spectrophotometric scans of α-GPdH of a representative F_1 individual of an allele 3 × allele 4 cross. 5% and 8% acrylamide gels are shown. Multiple bands are seen only at high concentrations of acrylamide: interface (I); ferritin (Fe); α-glycerophosphate dehydrogenase (α-GPdH); hemoglobin (Hb); bromphenol blue dye (Bp).

TABLE 2

Genetic crosses of 2-glycerophosphate dehydrogenase variants of C. philodice

Phenotypic classes of crosses			Numbers (phenotype) of progeny		
♀	×	♂	a	a/b	b
2/3		2/3	18 (2/2)	46 (2/3)	12 (3/3)
4/3		3/3	24 (3/3)	19 (4/3)	0
2/3		3/3	31 (3/3)	27 (2/3)	0
3/3		4/4	0	30 (3/4)	0
4/4		3/3	0	13 (3/4)	0
2/2		2/2	17 (2/2)	0	0
3/3		3/3	62 (3/3)	0	0

Phenotype designations are as in Table 1. All assignments of individuals to variant classes were statistically significant by t-test at $P < 0.05$. (3/4) heterozygotes were multiple-banded only at high acrylamide concentrations.

reservoir of variability detectable at the protein level is large indeed (BERNSTEIN, THROCKMORTON and HUBBY 1973; SINGH, HUBBY and LEWONTIN 1974). There is no evidence as yet to indicate that this heterogeneity in *pI* has a genetic basis.

TABLE 3

The distribution of α-GPdH variants among five species of Colias

Variant class	$\bar{K}_R^{Hb}$ ± S.E.	$\bar{M}_o^{Hb}$ ± S.E.	Frequency
C. eurytheme (N = 25)			
1	—	—	0
2	0.82 ± .01	0.74 ± .01	0.39
3	0.83 ± .01	0.79 ± .01	0.11
4	0.88 ± .01	0.81 ± .01	0.32
5	0.96 ± .02	0.88 ± .02	0.14
C. alexandra (N = 23)			
1	—	—	0
2	0.80 ± .02	0.76 ± .01	0.12
3	0.83 ± .01	0.80 ± .01	0.37
4	0.90 ± .01	0.82 ± .01	0.33
5	0.96 ± .02	0.87 ± .01	0.19
C. scudderi (N = 20)			
1	—	—	0
2	0.81 ± .01	0.75 ± .01	0.39
3	0.83 ± .01	0.81 ± .01	0.14
4	0.89 ± .01	0.81 ± .01	0.36
5	0.92 ± .02	0.85 ± .02	0.05
*C. philodice** (N = 18)			
1	0.65 ± .01	0.73 ± .02	0.16
2	0.81 ± .02	0.77 ± .01	0.46
3	0.84 ± .01	0.81 ± .02	0.19
4	0.90 ± .01	0.81 ± .01	0.19
5	—	—	0
C. meadii (Mesa Seco) (N = 40)			
1	—	—	0
2	0.79 ± .02	0.76 ± .01	0.36
3	0.84 ± .01	0.80 ± .01	0.30
4	0.89 ± .01	0.81 ± .01	0.30
5	0.95 ± .01	0.87 ± .03	0.04
C. meadii (Cumberland Pass) (N = 22)			
1	—	—	0
2	0.78 ± .02	0.74 ± .01	0.11
3	0.83 ± .01	0.81 ± .01	0.58
4	0.90 ± .01	0.82 ± .01	0.24
5	0.98 ± .05	0.91 ± .08	0.08

$\bar{K}_R^{Hb}$ and $\bar{M}_o^{Hb}$ are as described in Table 1, ± one standard error. The listed standard error refers to the variance in K_R observed among a sample of N individuals, each analyzed independently. Variance within individual experiments was similar to that observed for the hemoglobin internal standards run in the same gels, as discussed on page 7.

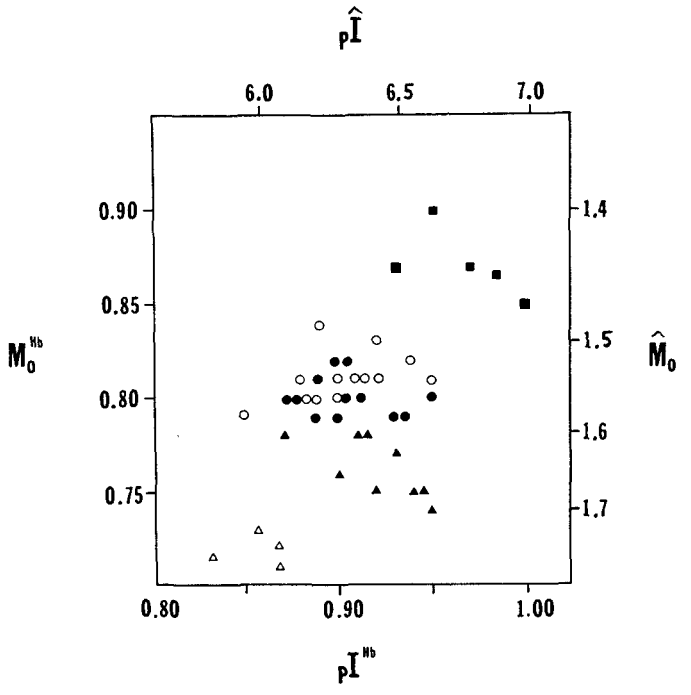

FIGURE 10.—Free electrophoretic mobility as a function of isoelectric point. $M_o{}^{Hb}$ and M_o are as described in Table 1. The isoelectric point (pI) is presented as pI^{Hb} and pI, as described in Table 4: allele 5 (■), allele 4 (○), allele 3 (●), allele 2 (▲, △). Allele 2 appears heterogeneous.

TABLE 4

Isoelectric point heterogeneity among α-glycerophosphate dehydrogenase variants of Colias

Allele	No. individuals analyzed	$\overline{pI^{Hb}}$ ± S.E.	$\hat{pI}$	No. individuals exhibiting multiple α-GPdH bands	$\overline{\Delta pI}$ within individuals ± S.E.
1	4	0.85 ± .02	5.8	1	0.1
2	16	0.90 ± .02	6.2	3	0.4 ± .1
3	14	0.90 ± .01	6.2	8	0.2 ± .1
4	17	0.89 ± .01	6.1	6	0.2 ± .1
5	6	0.93 ± .03	6.4	0	—
	57			18 (32%)	

Each individual considered exhibited a single band in electrophoresis on a range of acrylamide concentrations (3%–8%T). Hemoglobin was run as an internal standard with each sample. $\overline{pI^{Hb}}$ is the measured isoelectric pH of α-GPdH for a given individual, relative to the measured isoelectric point of hemoglobin determined on the same gel; the mean value for each allele is presented ± one standard error. $\hat{pI}$ is the maximum likelihood estimate of pI and is obtained by multiplying $\overline{pI^{Hb}}$ by the known isoelectric point of hemoglobin under these conditions (6.86 ± 0.02, based upon 80 replicates). $\overline{\Delta pI}$ is the mean difference in measured pH between bands of multi-banded individuals.

Alternative possibilities are that the heterogeneity reflects variation in protein folding (CHREIGHTON 1975), binding of ions or other small molecules (CHEN and SUTTON 1967), a buffer interaction (CANN and GOAD 1965), or some other post-translational modification. Whatever their ultimate cause, these data suggest that fractional charge differences among proteins of a locus may be common.

The observation that different *pI* variant proteins often exhibit indistinguishable K_R values argues indirectly that the significant K_R differences seen between alleles are not the result of fractional charge interactions with gel fibers. Also, if fractional charge differences yielded differing K_R values due to electro-osmotic effects, changes in the volume of the hydration shell, or localized gel fiber interactions, then one would not expect variants 2 and 3, which differ significantly in charge, to exhibit such similar K_R values. It seems more likely that these allelic K_R differences reflect conformational differences between the allelic proteins. To demonstrate that conformational change is capable of producing changes in K_R of the magnitude observed in Table 1, the electrophoretic behavior of α-chymotrypsin was examined at several pH's. This dimeric protein is known from single-crystal X-ray crystallography to undergo an asymmetric conformational change (only one of the two subunits changes) at pH 8.0 (MAURIDIS, TULINSKY and LIEBMAN 1974). It is clear from Table 5 that this conformational shift produces a major change in observed K_R. Because a change in conformation *can* alter K_R does not, of course, imply that all differences in K_R need reflect conformational differences. However, while the correspondence of K_R differences and conformational differences will have to be empirically verified by alternative experimental approaches, a conformational interpretation seems the most reasonable working hypothesis.

DISCUSSION

A number of conclusions may be drawn from the results of the analytic survey described above:

1) There are alleles which do not differ in net charge. The amino acid

TABLE 5

The effect of pH upon the K_R of α-chymotrypsin

Buffer	Trial	n	K_R
Imidazole	1	6	.062
(operating	2	6	.068
pH = 7.5)	3	6	.054
			$\bar{x} = .061$
			$\sigma = .007$
Tris-Glycine	1	6	.049
(operating	2	6	.048
pH = 9.5)	3	6	.036
			$\bar{x} = .044$
			$\sigma = .007$

substitutions distinguishing alleles *3* and *4* produce a significant difference in these proteins' retardation coefficients; no significant difference in charge is detectable. The simplest explanation of this variation is that the amino acid difference(s) affect conformation significantly enough to alter electrophoretic mobility, even though not involving a charge difference. Independent evidence of conformational differences will be required to confirm this interpretation. "Ladder" models involving alleles with unit differences in charge are clearly inappropriate for these data, as are any estimates of the fraction of charge-changing alterations permitted by the genetic code. The data of Table 1 indicate that alleles differing only in K_R are detectable by electrophoresis. It follows that only empirically can one determine with confidence the fraction of electrophoretically-detected alleles which actually involve a charge change. This has been done for 14 loci of *C. meadii* (JOHNSON 1976).

2) There are "hidden" alleles which go undetected in a normal 5% acrylamide gel survey. Estimates of polymorphism based upon such surveys are probably too low, and reports of allelic uniformity and homology based upon 5% gel data will require verification.

3) There are clear advantages to characterizing variants in terms of physical state properties such as K_R and M_o, with stated error limits. Workers sampling the same alleles years later should obtain the same values within the stated error limits. This is particularly important where judgments of homology rather than differences are involved: evaluation of homology can be made objectively by t-test with a stated P value. Such an explicit statistical approach to data gathering would greatly improve the value of "genetic distance" measures.

4) Finally, it should be noted that functional studies of differences between allozymes, particularly studies that involve kinetic analysis, are of increasing interest and importance. All such investigations require allelic uniformity within a biochemical analysis if the results are to be interpretable genetically. The results of this paper suggest that such uniformity should be carefully documented.

The author wishes to thank DENISE LUNDRY, CAMILLE LYNCH and GERRY SHEDLOFSKY for able technical assistance, WARD WATT for detailed information on procedures for breeding Colias, LEE SNYDER and CLAY SASSAMAN for stimulating and critical discussion, and the many students of Stanford and Washington University who aided in the collecting of butterflies. Earlier drafts of the manuscript were read by D. RODBARD, WARD WATT and SIMON SILVER. Field facilities were provided by The Rocky Mountain Biological Laboratory. This work was supported by grants from Washington University, The American Philosophical Society, The National Institute of Health, and The National Science Foundation.

LITERATURE CITED

BERNSTEIN, S., L. THROCKMORTON and J. HUBBY, 1973 Still more genetic variability in natural populations. Proc. Nat. Acad. Sci. USA **70**: 3928–3931.

CANN, J. and W. GOAD, 1965 Theory of zone electrophoresis of reversibly interacting systems. Two zones from a single macromolecule. J. Biol. Chem. **240**: 1162–1164.

CHEN, S. and H. SUTTON, 1967 Bovine transferrins: Sialic acid and the complex phenotype. Genetics **58**: 425–430.

CHRAMBACH, A. and D. RODBARD, 1971 Polyacrylamide gel electrophoresis. Science **172**: 440–451.

CHREIGHTON, T., 1975 The two-disulphide intermediates and the folding pathway of reduced pancreatic trypsin inhibitor. J. Mol. Biol. **95**: 167–199.

DAVIS, B., 1964 Disc electrophoresis. Ann. N.Y. Acad. Sci. **121**: 404–427.

EWENS, W. and M. FELDMAN, 1975 The Theoretical Assessment of Selective Neutrality. In: *Proc. Intern. Confer. on Population Genetics and Ecology*. Edited by S. KARLIN and E. NEVO. Academic Press, New York. (In press.)

FAWCETT, J. and C. MORRIS, 1966 Molecular-sieve chromatography of proteins on granulated polyacrylamide gels. Separation Sci. **1**: 9–26.

FERGUSON, K., 1964 Starch gel electrophoresis: Application to the classification of pituitary proteins and polypeptides. Metabolism **13**: 985–1002.

GONNENNE, A. and J. LEBOWITZ, 1975 Estimation of molecular weights of small proteins on polyacrylamide gel electrophoresis. Anal. Biochem. **64**: 414–424.

HEDRICK, J. and A. SMITH, 1968 Size and charge isomer separation and estimation of molecular weights of proteins by disc gel electrophoresis. Arch. Biochem. Biophys. **126**: 154.

JOHNSON, G., 1971 Analysis of enzyme variation in natural populations of the butterfly *Colias eurytheme*. Proc. Nat. Acad. Sci. U.S. **68**: 997–1001. ——, 1972 The selective significance of biochemical polymorphisms in *Colias* butterflies. Doctoral thesis, Stanford University. ——, 1974 On the estimation of effective number of alleles from electrophoretic data. Genetics **78**: 771–776. ——, 1975a The use of internal standards in electrophoretic surveys of enzyme polymorphism. Biochem. Gent. **13**: 833–847. ——, 1975b Polymorphism and predictability at the α-glycerophosphate dehydrogenase locus in *Colias* butterflies: Gradients in allele frequency within single populations. Biochem. Genet. **14**: (In press). ——, 1976 Characterization of electrophoretically cryptic variation in the alpine butterfly *Colias meadii*. (Submitted for publication).

KAPADIA, G., A. CHRAMBACH and D. RODBARD, 1974 Approaches to Macromolecular Mapping by Polyacrylamide Gel Electrophoresis, pp. 115–144. In: *Electrophoresis and Isoelectric Focusing in Polyacrylamide Gel*. Edited by R. Allen and H. Maurer. de Gruyter, New York.

KING, J. and T. OHTA, 1975 Polyallelic mutational equilibria. Genetics **79**: 681–691.

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

MAURIDIS, A., A. TULINSKY and M. LIEBMAN, 1974 Asymmetrical changes in the tertiary structure of α-chymotrypsin with change in pH. Biochemistry **13**: 3661–3666.

O'BRIEN, S. and R. MACINTYRE, 1972 The α-GP cycle in *D. melanogaster*. I. Biochemical and developmental aspects. Biochem. Genet. **7**: 141–161.

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.

ORNSTEIN, L., 1964 Disc electrophoresis. I. Background and theory. Ann. N.Y. Acad. Sci. **121**: 321–349.

RICHARDSON, R., M. RICHARDSON and P. SMOUSE, 1975 Evolution of Electrophoretic Mobility in the *Drosophila mulleri* Complex. In: *Isozymes IV, Genetics and Evolution*. Edited by C. MARKERT. Academic Press, New York.

RODBARD, D. and A. CHRAMBACH, 1970 Unified theory for gel electrophoresis and gel filtration. Proc. Nat. Acad. Sci. USA **65**: 970–977. ——, 1971—Estimation of molecular radius, free mobility, and valence using polyacrylamide gel electrophoresis. Anal. Biochem. **40**: 95–134. ——, 1974 Quantitative Polyacrylamide Gel Electrophoresis: Mathematical and Statistical Analysis of Data, pp. 28–61. In: *Electrophoresis and Isoelectric Focusing in Polyacrylamide Gel*. Edited by R. ALLEN and H. MAURER. de Gruyter, New York.

RODBARD, D., A. CHRAMBACH and G. WEISS, 1974 Optimization of Resolution in Analytical and Preparative Polyacrylamide Gel Electrophoresis, pp. 62–104. In: *Electrophoresis and Isoelectric Focusing in Polyacrylamide Gel.* Edited by R. ALLEN and H. MAURER. de Gruyter, New York.

SCHLESSINGER, M. J., W. BLOCK and P. M. KELLEY, 1975 Differences in the Structure, Function, and Formation of Two Isozymes of *Escherichia coli* Alkaline Phosphatase, pp. 333–343. In: *Isozymes I: Molecular Structure.* Edited by C. L. MARKERT. Academic Press, Inc., New York.

SINGH, R., J. HUBBY and R. LEWONTIN, 1974 Molecular heterosis for heat-sensitive enzyme alleles. Proc. Nat. Acad. Sci. USA **71:** 1808–1810.

STOPPARD, T., 1972 *Jumpers.* Evergreen Press (E626), New York.

Corresponding editor: R. C. LEWONTIN