Proc. Natl. Acad. Sci. USA
Vol. 74, No. 2, pp. 698–701, February 1977
Genetics

Polymorphism and the subunit structure of enzymes: A contribution to the neutralist-selectionist controversy

(subunit number/subunit size)

HARRY HARRIS*, D. A. HOPKINSON†, AND Y. H. EDWARDS†

* Department of Human Genetics, University of Pennsylvania, Philadelphia, Pa. 19174; and † MRC Human Biochemical Genetics Unit, Department of Human Genetics, University College, Wolfson House, 4 Stephenson Way, London NW1, England

Contributed by Harry Harris, November 19, 1976

ABSTRACT **The occurrence of polymorphism in a series of 87 different loci coding for enzyme structure in human populations has been related to the size and the number of subunits in the corresponding enzymes. Polymorphic and nonpolymorphic enzymes did not differ on average in subunit size. However, multimeric enzymes showed a significantly lower incidence of polymorphism than did monomeric enzymes. A particularly low incidence of polymorphism was noted among multimeric enzymes in which interlocus molecular hybrids occur. The findings are discussed in terms of the "neutralist" and "selectionist" hypotheses of polymorphism.**

The widespread occurrence of enzyme polymorphism in natural populations is now well established, but the biological causes of the phenomenon remain a matter of considerable controversy. There are two extreme and contrasting views. One view attributes the occurrence of most polymorphisms to neutral or near-neutral mutations and random genetic drift; the other attributes it to balancing selection. Much data, much mathematics, and much argument have been deployed on both sides of the controversy, but the basic issues can hardly be said to have been resolved (for example see refs. 1–3).

One aspect of the matter which does not so far appear to have been considered in any detail is the possible relationship of enzyme polymorphism to the structural characteristics of enzymes. In this paper we focus on two features of enzyme structure: the number of polypeptide subunits present in the enzyme, and the sizes of the subunit molecules. We will first outline certain predictions that appear to follow naturally from the neutral mutation-random drift hypothesis but do not necessarily seem to follow from the selectionist hypothesis. We then examine these predictions in a series of 87 loci coding for enzyme structure in man, about which information on polymorphism, subunit number, and subunit size is available.

The selectionist and neutralist hypotheses

In principle a large number of point mutations may occur within the confines of the coding region of any given gene, each resulting in a specific alteration in the amino acid sequence of the corresponding enzyme protein. According to the selectionist hypothesis of polymorphism, the great majority of these possible mutational changes are in some degree deleterious to the function of the enzyme and hence to the biological fitness of the organism, so that most mutant alleles are eliminated or kept at a low frequency by the pressure of natural selection. Occasionally, however, mutants will occur which confer some selective advantage on the organism in the prevailing environmental circumstances and consequently tend to spread, and according to this hypothesis it is such mutants that give rise to the polymorphisms. The origin and maintenance of enzyme polymorphisms is held to derive from differential fitness of the various enzyme phenotypes consequent on differences in enzyme function in particular environments.

The neutralist hypothesis of polymorphism accepts the idea that, of the different possible mutants, many are deleterious and are therefore eliminated or kept at a low frequency by natural selection. However, it is argued that there are also many mutants which, because they make little or no difference to the function of the enzyme protein, are effectively neutral or near-neutral from the standpoint of natural selection, so that their incidence in a population depends entirely on random genetic drift. Some by chance are eliminated, others persist at low frequencies, and occasional ones spread fortuitously and achieve relatively high frequencies in the population and thus give rise to polymorphism. The occurrence of genuinely advantageous mutants is not excluded in this view, but such mutants are regarded as extremely uncommon compared with the so-called neutral mutants, which consequently, it is argued, account for the great majority of polymorphisms (4).

Theoretical analysis of the hypothesis of neutral mutations and random drift shows that the levels of heterozygosity achieved, and hence the occurrence of polymorphism, depends on N_e, the effective population size, and u, the mutation rate to neutral or near-neutral alleles (5, 6). This mutation rate to neutral alleles depends on the number of amino acid sites in the enzyme at which amino acid substitutions may occur without significantly altering enzyme function. In general the number of such sites present in the molecule will vary from enzyme to enzyme and hence the mutation rates to neutral alleles will vary from locus to locus. This provides a basis for variations in the level of heterozygosity and in the occurrence of polymorphism from locus to locus in any given population because the greater the mutation rate, the higher the level of heterozygosity.

The expected relationship between the degree of heterozygosity (and hence polymorphism) from locus to locus and the subunit number in the enzyme products of the different loci arises from the neutralist hypothesis because the amino acid sites at which neutral substitutions may be expected to occur are in general likely to be located on the outer surface of an enzyme molecule rather than in its inner regions. In multimeric enzymes a significant fraction of the amino acid sites located on the surface of each of the subunits will be concerned with the intersubunit contacts and will be involved in maintaining molecular stability and enzyme function. They are therefore much less likely to be susceptible to neutral substitutions than are those sites located on the outer surface of the subunit, which also lie on the outer surface of the enzyme molecule as a whole. Consequently, one would expect that on average the mutation rates to neutral alleles would be greater in monomeric enzymes, in which such subunit interactions do not occur, than in multimeric enzymes. Hence, the neutralist hypothesis of polymorphism would appear to predict increased average heterozygosity per locus and a higher incidence of polymorphism at loci coding for monomeric enzymes than at loci coding for multimeric enzymes.

Table 1. Incidence of polymorphism at 87 loci coding for monomeric, dimeric, trimeric, and tetrameric enzymes

Subunit structure	No. of loci		Percent polymorphic loci
	Total	Polymorphic	
Monomeric	27	15	56
Dimeric	37	13	35
Trimeric	4	1	25
Tetrameric	19	4	21

A relationship between variation in heterozygosity from locus to locus and variation in subunit size of the enzyme products of the various loci is suggested by the following considerations. The size of a polypeptide subunit of a particular enzyme is determined by the number of amino acids it contains, and this in turn is determined by the number of bases in the coding region of the corresponding gene. Furthermore, one can argue that the rate at which spontaneous mutations occur in a gene is likely to be roughly proportional to the number of bases it contains. So the variation in subunit sizes of enzymes gives an indication of the expected variation from locus to locus in the rates of occurrence of mutations that produce structural changes in the enzymes. Consequently, from the neutral mutation-random drift hypothesis one would expect that on the average the subunit sizes of enzymes coded by polymorphic loci should be appreciably larger than those coded by nonpolymorphic loci.

According to the selectionist hypothesis, variation in level of heterozygosity (and in the occurrence of polymorphism between loci) will in general depend on the differential effects of selective factors on the enzyme phenotypes. Although they would be expected to differ from locus to locus, these effects would not necessarily be directly related to the subunit structures of the enzymes coded at the different loci. However, because selection must depend on functional differences between the enzyme phenotypes, it is possible that such differences may be more likely to occur with some types of enzyme structures than with others.

Enzyme polymorphism in man

In what follows, the term "polymorphic locus" will be reserved for those loci at which, in a specific human population, the frequency of the commonest allele is no greater than 0.99. This implies that the fraction of the population heterozygous for alleles at that locus must be at least 2%. In addition, consideration is limited to enzyme polymorphisms that have been demonstrated electrophoretically. We have previously reported on the occurrence of electrophoretic polymorphism and the levels of heterozygosity at 71 different loci coding for enzymes in White populations of European origin (7). The aim was to estimate the incidence of polymorphism and the average heterozygosity per locus over a wide range of structural loci. Since then, more information has accumulated and we now consider that acceptable data are available on at least 104 loci. This new material has been combined with the material used in the earlier study, some of which now requires reassessment in the light of further investigations of specific enzymes. In addition, data are available on many, though by no means all, of these loci in other human populations, and we have recently reexamined this material and tabulated the enzyme loci that have been found to be polymorphic in one or another of the major human ethnic groups: Whites of European origin, Blacks, Asiatic Indians, and Orientals. Details of the data on the individual enzymes are given in ref 8.

Table 2. Average heterozygosity at 87 loci coding for monomeric, dimeric, trimeric, and tetrameric enzymes in Whites of European origin

Subunit structure	Number of loci	Average heterozygosity per locus
Monomeric	27	0.096
Dimeric	37	0.071
Trimeric	4	0.045
Tetrameric	19	0.050

In all, 33 of the 104 loci appeared to exhibit electrophoretic polymorphism in at least one of the major ethnic groups, but there are differences between the groups. For example, only 24 of the loci showed polymorphism in Europeans, and among these at least two were not apparently polymorphic in Blacks. And there were several nonpolymorphic loci in Europeans that were polymorphic in Blacks. The data on the other ethnic groups are much less extensive than those on Europeans because many of the loci so far examined in Europeans have not yet been studied in the other populations. Also, the sample of loci actually studied in these non-European populations appear to be somewhat biased against loci not found to be polymorphic in Europeans. Consequently, acceptable estimates of average heterozygosity per locus in these non-European populations cannot at present be derived. In Europeans, our present estimate is 0.063 (8).

Subunit number

Subunit number may be derived from a number of different lines of investigation. A discussion of the methods involved and their limitations in deriving subunit numbers is given in ref. 9 with details of the findings in an extensive series of human enzymes.

Among the enzymes coded by the 104 loci on which we had information about the occurrence of polymorphism, there were 87 to which it was possible to assign a subunit number. In all there appeared to be 27 monomers, 37 dimers, 4 trimers, and 19 tetramers. But the distribution differs between the loci that had been found to be polymorphic in at least one major ethnic group and the apparently nonpolymorphic loci (Table 1.) There appears to be a significantly higher incidence of polymorphism among loci coding for monomeric enzymes than among loci coding for multimeric enzymes ($\chi^2 = 5.16$; df = 1; $P < 0.05$). Furthermore, among the multimeric enzymes the incidence of polymorphism appears to decline with increasing subunit number. The same general tendency is illustrated in Table 2 in which average heterozygosity per locus in the White European population group has been estimated for each of the subunit number classes.

Subunit size

We were also able to obtain estimates of the molecular sizes of the polypeptides coded by this series of 87 loci. These estimates with detailed references are given in ref. 9. The distribution of subunit sizes of the enzyme products of these 87 loci was found to be continuous and unimodal. The average subunit size over the whole series was close to 46,000 daltons, which corresponds to about 425 amino acids. The range in subunit size was from about 13,000 to about 116,000 daltons, and this gives a range for the coding regions of the corresponding genes from about 360 bases to about 3200 bases. The subunit sizes showed no correlation with the corresponding subunit numbers (9).

When the average subunit size of the enzymes coded by loci

Table 3. Average subunit sizes of enzymes determined by 33 polymorphic and 54 nonpolymorphic loci

	No.	Mean size (daltons)	SEM
Polymorphic loci	33	45,333	3573
Nonpolymorphic loci	54	45,889	3048
Total	87	45,678	2314

classified as polymorphic was compared with that for loci classified as nonpolymorphic, no significant difference was observed (Table 3). Nor were significant differences found when loci coding for monomeric and multimeric enzymes were considered separately.

Data and the hypotheses

In many respects the data so far assembled are incomplete. For example, information about the occurrence of polymorphism at quite a number of the loci considered is in many instances limited to only one or two of the major ethnic groups, and it is possible that some of the loci classified as nonpolymorphic in one population may turn out to be polymorphic in one of the other populations when these are studied. Also, there are a number of enzymes for which information about the occurrence of polymorphism is available but for which there is at present no satisfactory information about subunit number or subunit size. Because as a rule these represent loci that are considered to be nonpolymorphic, the polymorphic loci tend to be somewhat overrepresented in the subunit number and subunit size tabulations. In addition, many of the subunit number assignments and subunit size estimates require confirmation before they can be accepted without reservation. Clearly, much further work aimed at validating the data so far available and extending it to include other enzymes is required.

Nevertheless, despite these inadequacies, two noteworthy conclusions appear to be emerging from the data, and it is of interest to consider their possible implications. The *first* is that the occurrence of polymorphism over a range of different loci appears to be inversely related to the number of polypeptide subunits in the corresponding enzyme products. The *second* is that the occurrence of polymorphism is not apparently correlated with the size of the polypeptide subunits. Taken at their face value, these conclusions confront us with something of a dilemma. The subunit size data seem to contradict the neutralist hypothesis, while the subunit number data seem to support it.

In considering the subunit size data in terms of the neutralist hypothesis, several further points need to be considered.

1. When one argues that the mutation rate at a structural locus is likely to be proportional to the number of bases in the DNA and hence to the size of the corresponding polypeptide, one is talking about all possible mutations in the coding region of the gene. But the neutralist hypothesis predicts a correlation between heterozygosity and the mutation rate to so-called neutral alleles and this must be smaller than the mutation rate to all alleles. In general one would expect that the actual number of sites in a polypeptide subunit at which amino acid substitutions may occur without significantly altering enzyme function will be greater in larger polypeptides. So one would expect that the mutation rate to neutral alleles would be correlated with the total mutation rates. However, this correlation would presumably be less than 1.0, and the difference between average polypeptide size for polymorphic loci and average polypeptide size for nonpolymorphic loci, as predicted by the neutralist hypothesis, would depend on the magnitude of this correlation. It will therefore be necessary to estimate the expected average difference in subunit size between the products of polymorphic and nonpolymorphic loci with different values assumed for this correlation. Comparing such expected differences with the data obtained from direct observation should then enable one to determine the degree to which the data impose restrictions on the neutralist hypothesis. However, for the neutralist hypothesis to be consistent with the absence of any difference between the average subunit sizes of the polymorphic and nonpolymorphic loci, it would seem necessary to postulate that there is no correlation at all between neutral and total mutation rates, and this seems implausible.

Table 4. Incidence of polymorphic loci tabulated according to class of loci among loci coding for monomeric enzymes, multimeric enzymes not forming interlocus hybrid molecules, and multimeric enzymes forming interlocus hybrid molecules

Enzyme-coding loci	No. of loci: Total	No. of loci: Polymorphic	Percent polymorphic loci
Monomeric	27	15	56
Multimeric, not forming interlocus hybrid molecules	38	16	42
Multimeric, forming interlocus hybrid molecules	22	2	9

2. Recently it has been shown, in studies on the xanthine dehydrogenase (xanthine:NAD^+ oxidoreductase, EC 1.2.1.37) locus in *Drosophila pseudoobscura*, that a considerable increase in the number of detectable alleles and in observed heterozygosity is obtained by studying the enzyme under a much wider range of electrophoretic and other conditions than had previously been used in routine population studies (10). It is not yet clear whether the dramatic increase in allelic diversity observed in this study is peculiar to the xanthine dehydrogenase locus and perhaps a few other loci or whether it will be found to occur at most other loci coding for enzymes. If it does turn out to be a general phenomenon and if it also turns out that there is a poor correlation between the amount of heterozygosity at different loci as detected by conventional electrophoretic methods and by newer methods that give enhanced resolution of allelic variants, then it is possible that the relationship between polymorphism and subunit size expected from the neutralist hypothesis might have been obscured in the present data. This is because, under these circumstances, the presently available electrophoretic data would provide only a weak reflection of interlocus variation in heterozygosity. However, it should be noted that, if this is indeed the case, it would be difficult to see why the same electrophoretic data when related to the subunit number data show such an apparently good fit with those predicted by the neutralist hypothesis.

3. Recently, Koehn and Eanes (personal communication), in an analysis of published data on the level of heterozygosity as determined electrophoretically at 11 polymorphic loci coding for enzymes in various *Drosophila* species and on estimated subunit sizes of these enzymes, found a significant positive correlation between heterozygosity and subunit size. The reason for this apparent difference between the human and the *Drosophila* data is not clear at present.

Selectionist hypothesis

The inverse relationship between the occurrence of polymorphism and subunit number found in the human data has also been found in data from a number of other species, both animal and plant (11). As has been pointed out, such a relationship is consistent with what is expected from the neutralist hypothesis. It is not, however, an obvious prediction from the selectionist hypothesis. If, then, the relationship turns out to be real, it will be necessary to find some general biological explanation other than neutral mutations and random-drift to account for it if the selectionist position is to be maintained.

In considering this problem, one may note that a critical difference between loci coding for multimeric enzymes and loci coding for monomeric enzymes is that hybrid molecules containing the products of two different genes may occur with multimeric enzymes, whereas with monomeric enzymes this is not possible. Furthermore, there are two distinct situations in which such hybrid enzyme molecules are generated. The first is in heterozygotes in which the hybrid molecules contain polypeptide chains coded by two different alleles at the same locus. This is a general situation that appears to apply to virtually all autosomal loci coding for multimeric enzymes. The second type of situation is more restricted and occurs only in certain multimeric enzymes. In these cases, hybrid molecules containing polypeptides coded by genes at at least two different loci are formed. A critical difference between the two situations is that hybrid enzyme molecules containing polypeptides coded by two alleles only occur in heterozygous individuals, who necessarily represent only a part of the total population, whereas hybrid molecules containing polypeptides derived from genes at different loci will occur in virtually all members of the population.

With these general considerations in mind, we retabulated the data, dividing the loci into three classes as follows: loci coding for monomeric enzymes; loci coding for multimeric enzymes that do not appear to form interlocus hybrid molecules; and loci coding for multimeric enzymes that do form interlocus hybrid molecules (Table 4). The results were unexpected. It now appeared that the incidence of polymorphism in the loci coding for multimeric enzymes that do not form interlocus hybrids is only a little less than that in the loci coding for monomeric enzymes, whereas the group of loci coding for multimeric enzymes that do form interlocus hybrid molecules have a markedly reduced incidence of polymorphism compared with the two other groups of loci. On testing for heterogeneity for the whole table, $\chi^2 = 11.62$, $P < 0.01$; for the monomeric loci versus the multimeric loci not forming interlocus hybrids, $\chi^2 = 1.14$, $P < 0.3 > 0.2$; and for multimeric loci not forming interlocus hybrids versus multimeric loci forming interlocus hybrids, $\chi^2 = 7.23$, $P < 0.01$.

If this result is real, and it will certainly need confirmation with more extensive data, it certainly alters the complexion of the matter. In the first place it suggests that the difference in the incidence of polymorphism between loci coding for monomeric and for multimeric enzymes is derived mainly from the subclass of multimeric enzymes that form interlocus molecular hybrids. In the second place it suggests that, in terms of the selectionist hypothesis, one should focus on the question of why polymorphism may be appreciably less frequent among loci coding for enzymes that form interlocus molecular hybrids.

At present one can only speculate about this question in very general terms, but the following type of argument is perhaps worth considering. The occurrence of two or more loci, which may be presumed to have originated by gene duplication and which code for structurally different polypeptides that form molecular hybrids, presumably endows the organism with a greater degree of metabolic flexibility in the face of environmental variations than that achieved by other loci. Heterozygosity in polymorphisms can be considered, from the selectionist point of view, as another way of achieving metabolic flexibility. But heterozygosity can only occur in some of the individuals in a species and its effects are likely to be less profound because the alleles involved will generally have been differentiated by only one, or, at most, a small number of mutational changes compared with genes at different loci which will usually have a much longer evolutionary history. So it might be argued that the selective pressures to achieve heterozygosity in a significant fraction of the population might in general be weaker for loci whose polypeptide products form interlocus hybrids than for other loci. It is of interest that one hypothesis that has been put forward to account for the selection and subsequent fixation of gene duplications in evolution is that the original duplication in each case occurred in a heterozygous individual, and the evolutionary process can be regarded as selection directed at fixing in all members of the species the advantage previously accruing to only those who were heterozygous for the particular alleles (12).

However, whatever theoretical interpretations are placed on these results, they are of interest because they suggest a relationship between two quite different types of biological phenomenon—namely, the occurrence of enzyme polymorphism in natural populations and the subunit structure of enzymes. Such a relationship, if established as a valid generalization, would need to be taken into account in virtually all other lines of work bearing on the biological significance of polymorphism. For example, it has been variously proposed that the occurrence of polymorphisms in some, rather than in other, enzymes may be related to their functional roles such as their involvement in energy-generating processes (13, 14) or in metabolic regulation (15). It would now seem necessary to reexamine these other proposed relationships in terms of the subunit structures of the enzymes involved, since there would appear to be a danger of confounding different effects.

1. Kimura, M. & Ohta, T. (1971) *Nature* **229**, 467–469.
2. Lewontin, R. C. (1974) *The Genetic Basis of Evolutionary Change* (Columbia Univ. Press, New York).
3. Harris, H. (1975) *The Principles of Human Biochemical Genetics* (North-Holland, Amsterdam), 2nd ed.
4. Kimura, M. & Ohta, T. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 2848–2852.
5. Kimura, M. & Crow, J. F. (1964) *Genetics* **49**, 725–738.
6. Ohta, T. & Kimura, M. (1973) *Genet. Res.* **22**, 201–204.
7. Harris, H. & Hopkinson, D. A. (1972) *Ann. Hum. Genet., London*, **36**, 9–20.
8. Harris, H. & Hopkinson, D. A. (1976) *Handbook of Enzyme Electrophoresis in Human Genetics* (North-Holland, Amsterdam).
9. Hopkinson, D. A., Edwards, Y. H. & Harris, H. (1976) *Ann. Hum. Genet., London*, **39**, 383–411.
10. Singh, R. S., Lewontin, R. C. & Felton, A. A. (1976) *Genetics* **84**, 609–629.
11. Zouras, E. (1976) *Nature* **262**, 227–229.
12. Spofford, J. C. (1969) *Am. Nat.* **103**, 407–432.
13. Gillespie, J. H. & Kojima, K-I. (1968) *Proc. Natl. Acad. Sci. USA* **61**, 582–585.
14. Wade Cohen, P. T., Omenn, G. S., Motulsky, A. G., Chen, S-H. & Giblett, E. R. (1973) *Nature New Biol.* **241**, 229–233.
15. Johnson, G. B. (1974) *Science* **184**, 28–37.