

Genetic heterogeneity in inherited disease

HARRY HARRIS

From the Department of Human Genetics, University College, and the MRC Human Biochemical Genetics Unit

During the last few years a quiet revolution has been taking place in our ideas about the genetical basis of inherited disease. One way in which this has been manifested is by the increasing use in the literature of the term 'genetic heterogeneity', a term which is usually meant to imply that the same or very similar clinical syndromes may be produced by a number of quite different gene mutations. That this could indeed occur has of course been recognized for a long time (Harris, 1956). But whereas in the past it was considered by most workers, at least implicitly, to be an unusual and somewhat infrequent phenomenon, it is now becoming increasingly apparent that it is a common and widespread feature of many inherited disorders (Harris, 1970; McKusick, 1973). In other words it now seems that 'genetic heterogeneity' will prove to be the rule rather than the exception. But it has also become clear that such heterogeneity can arise in several different ways. And in the present paper I would like to consider the basic types of cause which may give rise to the phenomenon, and attempt to arrive at a classification which might be useful in practice.

Multiple Allelism

It is convenient to start from a somewhat theoretical point of view by considering the question of multiple allelism, that is to say, the alternative forms or alleles of a particular gene which may occur among different individuals in a given population.

In theory, a very large number of structurally different alleles may be generated by separate mutational events within the confines of a single gene. For example, from a typical gene containing a DNA sequence of say 900 bases and coding for a protein of 300 amino acids, as many as 2700 different alleles may be generated simply from mutational events which result in no more than a single base change at some point in the sequence. This is so because each base may be changed to one of three others. In addition it is also known that other types of mutational event resulting in small deletions, duplications, frame shifts, etc, may also

occur within the confines of a single gene and each will generate its own structurally distinct mutant allele.

Now the alleles which determine the inherited disease and other variations which we find among living members of our species today must be attributed to specific gene mutations which occurred in single individuals among our ancestors in earlier generations, some living perhaps relatively recently, others in the more remote past. So the question arises as to how many of the different possible mutants, which in theory could have been generated by separate mutational events at any given gene locus, actually occur among living members of human populations at the present time.

One can approach this question in a simple-minded way by thinking about the variants of haemoglobin, since this protein happens to have been investigated from this point of view more extensively than any other. Consider the gene which codes for the β chain of haemoglobin (table Ia and b). There are 146 amino acids which are presumably coded by a

Haemoglobin β Chain

Number of amino acids	146	
Number of bases in gene	438	(146 × 3)
Number of possible alleles generated by single base mutational	1314	(438 × 3)

Table Ia

Effect of Mutant	%	Number of Alleles	Different Variant Proteins	Variants with Charge Change
Synonymous	23.4	307	0	0
Chain termination	3.7	49	?	?
Amino acid substitution	72.9	958	858	287
Total		1314		

Table Ib

Table Ia and Ib *Possible alleles which may be generated by one step single base change mutations in the gene coding for β -polypeptide chain of human haemoglobin*

Known variants with charge change	= 60
... Percentage of possible variants detected	= $\frac{60}{287} \times 100$
	= 20.9%

DNA sequence of 438 bases. So point mutations each resulting in only a single base change could generate 1314 different alleles at this locus. From the known amino acid sequence and the genetic code one can classify their effects approximately as shown in table I, provided we make the simplifying assumption that each of the possible codons for any amino acid is equally likely to occur. Because the code is degenerate, about 23.4% (307) of the mutant alleles generated in this way would not result in any change in structure of the haemoglobin. Such mutants are called 'synonymous' because the base change simply alters a codon specifying one amino acid to another codon specifying the same amino acid. In about 3.7% of cases (49 alleles) the single base change will result in the alteration of a codon specifying a particular amino acid at some point in the polypeptide sequence to a so-called 'nonsense' codon which will specify chain termination. Such mutations would be expected to result in the synthesis of an abbreviated polypeptide chain which lacks the series of amino acids normally coded by the base sequence in the gene beyond the point where the mutation occurred. In most cases such an abbreviated polypeptide is unlikely to give rise to a stable and detectable form of haemoglobin. However, in some 72.9% of cases (958 alleles) a single base change will result in the alteration of a codon specifying one amino acid to a codon specifying another and thus lead to the synthesis of an altered β polypeptide chain which differs from the original by a single amino acid substitution. Because of the degeneracy of the code, the 958 mutant alleles of this sort will give rise to only 858 structurally different haemoglobin variants.

The screening technique which has been most widely used and systematically applied to searching for haemoglobin variants is the method of electrophoresis. This will in general be expected to detect amino acid substitutions which involve a change in charge, ie, substitution of a neutral for a basic or acidic amino acid or vice versa, and substitutions of a basic acid for an acidic amino acid or vice versa. One can estimate that 287 of these possible 858 variants would involve substitutions of this sort, ie, almost exactly one third. Thus on purely theoretical grounds we can expect that about 287 different electrophoretic variants might be generated by single one-step base change mutations in the β Hb gene. In fact among all the haemoglobin variants so far identified at least 60 appear to have involved this type of amino acid substitution in the β chain. So we can say that of all the possible mutants of this sort which might exist and be detectable by electrophoresis, more than 20% are known to occur.

Although electrophoretic screening of Hb and the necessary amino acid sequencing of variants has now been carried out very extensively both in patients with haemolytic disease of various sorts and in random population surveys, it is clear that only a minute fraction of the world's population can have been examined in this way. So the fact that as many as 20% of the possible examples of this type of mutant have already been found is quite remarkable. It suggests indeed that a very high proportion of all possible mutants of this sort actually exist somewhere in the species, though the majority are likely to be very rare.

The haemoglobin variants produced by these different mutant alleles vary widely in their properties, and only a proportion of them are associated with overt clinical abnormality. But, for example, at least 20 structurally different β -chain variants have been identified which are so unstable as to give rise in the heterozygous state to a chronic haemolytic disease (Stamatoyannopoulos, 1972). These different haemolytic states, although they vary somewhat in degree, appear for the most part to be difficult to distinguish from each other on clinical or pathological grounds alone. And before their characterization by structural studies on the haemoglobin present, they would no doubt have been regarded as a single genetic entity with perhaps some superimposed variability, or at most two or three separate syndromes each determined by a single mutation. Thus this represents a very striking example of genetic heterogeneity.

There is no reason to think that the haemoglobin β -gene locus is in any way peculiar in this respect (Harris, 1974). More than 80 different variants of the enzyme glucose-6-phosphate dehydrogenase (Yoshida, Beutler, and Motulsky, 1971; Beutler and Yoshida, 1973), for example, have been reported, all apparently the consequence of different mutations at a single locus. And at least 20 of these variants have been shown to give rise to chronic haemolytic syndromes which resemble each other very closely at the clinical level, though they differ one from another in the molecular basis of the enzyme deficiency (Yoshida, 1973).

The deficiency of the enzyme hypoxanthine-guanine phosphoribosyl transferase provides another example. There is now evidence for the existence of quite a number of distinct mutant alleles, some of which result in such a severe degree of enzyme deficiency as to give rise to the full clinical manifestations of the Lesch-Nyhan syndrome in infancy (Lesch and Nyhan, 1964; Seegmiller, Rosenbloom, and Kelley, 1967; Kelley and Meade, 1971), while others result in less severe enzyme defects which are expressed clinically as

gout in adolescence or early adult life (Kelley, 1968). Numerous examples of multiple allelic variants of other proteins, such as albumin, transferrin, and serum α_1 -antitrypsin, have been reported, and surveys on a wide variety of different enzymes carried out in our own laboratory have revealed a widespread occurrence of variants strictly analogous to those identified in the cases of haemoglobin and glucose-6-phosphate dehydrogenase (Harris, Hopkinson, and Robson, 1974). Indeed it now appears probable that, at virtually all gene loci coding for enzyme and protein structure, a large number of the different possible mutant alleles which may be generated by separate mutations actually exist among living members of the species.

An interesting point emerges from all this. In the past it was usually assumed that if a particular disorder segregated in families in the manner of a typical autosomal recessive characteristic, then the affected individuals were ipso facto homozygous for a single mutant gene. It is, however, now apparent from the degree of multiple allelism which evidently occurs that this conclusion will often be mistaken. Although a condition segregates as an autosomal recessive, it will in many cases be due to heterozygosity for two different mutant alleles each producing a functional abnormality in the particular enzyme or protein but in a different way. Such heterozygous individuals would have received one of the abnormal alleles from one parent and one from the other, each of whom would usually also be carrying the normal allele. So typical segregation patterns with quarter ratios would be obtained. A well known example is sickle-cell-haemoglobin C disease. But other examples involving enzyme defects in inborn errors of metabolism are now beginning to be reported, for instance in such conditions as pyruvate kinase deficiency (Paglia, Valentine, and Rucknagel, 1972) and phosphohexose isomerase deficiency (Detter, Ways, Giblett, Baughan, Hopkinson, Povey, and Harris, 1972).

Compound heterozygosity of this sort is probably by no means uncommon in recessively inherited diseases due to enzyme defects. But it may often be peculiarly difficult to recognize in conditions determined by genes on autosomal chromosomes because it involves in the patients the characterization of the products of two different alleles in a mixture where the total enzyme activity is by the very nature of the situation markedly reduced.

Before leaving the question of multiple allelism it is perhaps worth pointing out that while this is an important cause of genetic heterogeneity in the sense that the same or very similar clinical syndromes may result from different mutant alleles at a single gene locus, multiple allelism can also give rise to a

different and indeed opposite type of phenomenon that is, the occurrence of a series of disorders which may differ quite markedly from each other and yet be due to different mutants at the same gene locus. The point is well illustrated by the disparate clinical appearances of conditions such as sickle cell anaemia, haemoglobin C disease, methaemoglobinemia, and chronic anaemia due to an unstable haemoglobin, all of which can be determined by different mutant alleles at the haemoglobin β gene locus. The very different clinical manifestations of the Lesch-Nyhan syndrome in infants, and of gout due to overproduction of uric acid in adults, both of which states can arise from allelic mutants affecting the enzyme hypoxanthine-guanine phosphoribosyl transferase, illustrate the same point. In the past such disorders, if considered on clinical grounds alone, would probably have been regarded as due to mutations of quite different genes. This possibility clearly has to be thought about in considering the genetics of inherited syndromes whose primary molecular basis is still obscure.

Multiple Loci

So far we have only been considering the genetic heterogeneity of disease which arises in consequence of multiple allelism at a single gene locus. But the phenomenon can also occur because of mutations at a number of different gene loci.

In some cases this is because a particular protein may contain two non-identical polypeptide chains coded by separate gene loci. Mutants at either locus may result in structural abnormalities of the protein and cause overt clinical disease. And, since the same protein is involved, the clinical consequences of particular mutants at the two different loci may be very similar. Haemoglobin again provides an apt example. Hb A ($\alpha_2\beta_2$) contains α and β polypeptides coded by separate genes which are not closely linked and probably occur on different chromosomes. And mutant alleles at either the α - or the β - gene loci, can, for example, result in unstable forms of the protein which give rise to chronic haemolytic disease. The α and β thalassaemias illustrate the same general phenomenon.

More often, however, genetic heterogeneity arises from mutants at gene loci which determine different and quite distinct enzymes or proteins. That this should be so is not very surprising. Loss of function of one or another of a series of enzymes involved in a sequential series of reactions in a metabolic pathway, or which are associated together in a complex of metabolic relationships, can well

result in the same or very similar consequences at the clinical level.

A rather obvious example is provided by the complex network of metabolic reactions which is present in the red cell, and is necessary for maintaining its functional and structural integrity. Disruption of the normal metabolism of the red cell, with its consequent premature destruction and hence chronic haemolytic disease, can clearly occur as a result of defects in a number of quite different enzymes. And indeed, a series of clinically not very dissimilar forms of genetically determined haemolytic anaemias have already been differentiated and shown to be due to deficiencies of quite different enzymes determined by separate gene loci. These various haemolytic diseases vary considerably in severity. But in practice it seems that they are difficult to differentiate one from another in the absence of information provided by enzyme studies. This difficulty is further enhanced by the fact that at any one of these loci, multiple alleles, which differ in the degree of enzyme deficit they produce and hence in the clinical severity, may also occur. So superimposed on the variation due to mutations at separate loci affecting different enzymes is the added variation due to multiple allelism at a single locus.

Storage diseases, such as the sphingolipidoses, the mucopolysaccharidoses, and the glycogen storage diseases, provide further examples of the same general phenomenon. The normal degradation of each of these classes of complex macromolecules is effected by the sequential action of a series of different enzymes, and a specific deficiency of one or another enzyme in this sequence results in progressive accumulation of the partially degraded macromolecules. It is not surprising, therefore, that very similar consequences may often result from mutations affecting quite different enzymes involved with the degradation of the same macromolecule. For example, in the case of the mucopolysaccharidoses, the syndromes now known as Sanfilippo A and B are apparently indistinguishable clinically (McKusick, 1972). Yet they are due to mutations at separate loci determining different enzymes, namely, heparan sulphate sulphatase in Sanfilippo A (Kresse and Neufeld, 1972) and N-acetyl- α -glucosaminidase in Sanfilippo B (O'Brien, 1972; Figura and Kresse, 1972). Both enzymes are concerned in the degradation of heparan sulphate and their separate deficiencies both result in the progressive lysosomal accumulation of partially degraded products of this substance.

The classical Hurler and Hunter syndromes furnish another example. They are very similar clinically and both are due to the progressive

accumulation of partially degraded forms of dermatan sulphate and heparan sulphate. In this case it had long been known from pedigree studies that two different gene loci were involved, because while Hunter's syndrome was evidently determined by a gene on the X chromosome, the gene determining Hurler's syndrome was found to be autosomal. More recently the defective enzymes have been identified. In Hurler's disease the faulty enzyme is α -iduronidase (Matalon and Dorfman, 1972; Bach, Friedman, Weissmann, and Neufeld, 1972), while in Hunter's disease it is sulphoiduronate sulphatase (Bach, Eisenberg, Cantz, and Neufeld, 1973). Both enzymes are normally concerned in degrading the mucopolysaccharides which accumulate, and it is not surprising therefore that the two syndromes resemble each other so closely.

Other Causes of Heterogeneity

It is important to remember that variation in the expression of a particular inherited disease in different individuals may derive not only from differences in the particular mutant genes causing the specific enzyme or protein defect which is the basic cause of the disease in question, but also from differences in the rest of the genetical constitution of the individuals in which the particular abnormal gene occurs. There is now an extensive body of data which indicates that probably no two individuals, with the exception of monozygotic twins, are exactly alike in their enzyme and protein constitutions, because of the common allelic differences (so-called polymorphisms) which occur at very many gene loci (Harris, 1970; Harris and Hopkinson, 1972). Such so-called 'normal variations' in genetic background, because they must define many of the details of the biochemical milieu against which the effects of particular abnormal genes are expressed, may well result in differences in the manifestations of a particular disorder in different affected patients. Some combinations of genes at other loci might minimize the pathological consequences and others accentuate them. Since such different combinations of genes will often not result in obvious differences between normal individuals not carrying the particular abnormal gene, the detailed manner in which they may contribute to variation in manifestation of particular disease is generally very difficult to define.

It must also be remembered that the clinical manifestations of an enzyme or protein abnormality determined by a mutant gene may depend in some degree on variations in the environmental circumstances to which individuals with the abnormality are exposed. Sometimes the clinical

expression of a particular abnormality appears to be hardly influenced by such environmental variations, but in other cases the effects may be important and give rise to quite unexpected clinical differences. A very striking illustration of this is provided by the genetically determined deficiency of the serum protein known as α_1 -antitrypsin. Some years ago it was discovered that individuals with this abnormality are exceptionally prone to develop pulmonary emphysema with relatively early onset in adult life (Eriksson, 1964, 1965). However, only a proportion of individuals (perhaps 15-30%) with the defect actually develop emphysema, and there appear to be some as yet undefined environmental factors which precipitate the lung damage in serum α_1 -antitrypsin-deficient individuals. More recently it was found that newborn infants, with the same genetically determined protein defect, are particularly liable to develop neonatal hepatitis which may be rapidly fatal or lead to hepatic cirrhosis (Sharp, Bridges, Krivit, and Freier, 1969; Porter, Mowat, Cook, Haynes, Shilkin, and Williams, 1972). Again only a proportion of the genetically predisposed infants actually develop neonatal hepatitis, presumably because of environmental differences.

It is a remarkable fact that the same genetic peculiarity should in some individuals give rise to neonatal hepatitis or hepatic cirrhosis in childhood, and in others to pulmonary emphysema in adult life. The differences in expression are presumably the consequences of differences in the environments to which different individuals are exposed. But certainly in the past no one would have expected that conditions as different as familial cirrhosis in childhood and familial emphysema in adult life might be due to the same mutant gene.

Conclusion

Table II provides a simple classification of these various cases of heterogeneity in inherited disease.

It is important to note that in any particular clinical syndrome two or more of these different causes may be operating so that the elucidation of the situation will often turn out to be quite complex. A simple model example is provided by the syndrome of congenital methaemoglobinaemia. There

is now evidence that this may be produced by at least 10 different mutant genes which occur at three distinct gene loci: two at the locus coding for the α -chain of haemoglobin, three at the locus coding for the β chain of haemoglobin, and at least five at the locus coding for the structure of the enzyme known as NADH diaphorase or methaemoglobin reductase. One may anticipate that this quite remarkable degree of genetic heterogeneity in what by any standards is an extremely rare condition and what in clinical terms was once regarded as a homogeneous entity, will turn out to be a prototype of the degree of heterogeneity waiting to be uncovered in many other rare inherited syndromes.

References

- Bach, G., Eisenberg, F., Jr., Cantz, M., and Neufeld, E. F. (1973). The defect in the Hunter syndrome: deficiency of sulfiduronate sulfatase. *Proc. nat. Acad. Sci. (Wash.)*, **70**, 2134-2138.
- Bach, G., Friedman, R., Weissmann, B., and Neufeld, E. F. (1972). The defect in the Hurler and Scheie syndromes: deficiency of α -L-iduronidase. *Proc. nat. Acad. Sci. (Wash.)*, **69**, 2048-2051.
- Beutler, E., and Yoshida, A. (1973). Human glucose-6-phosphate dehydrogenase variants: a supplementary tabulation. *Ann. hum. Genet.*, **37**, 151-155.
- Blume, K. G., Hryniuk, W., Powars, D., Trinidad, F., West, C., and Beutler, E. (1972). Characterization of two new variants of glucose-phosphate-isomerase deficiency with hereditary nonspherocytic hemolytic anemia. *J. Lab. clin. Med.*, **79**, 942-949.
- Detter, J. C., Ways, P. O., Giblett, E. R., Baughan, M. A., Hopkinson D. A., Povey, S., and Harris, H. (1968). Inherited variations, in human phosphohexose isomerase. *Ann. hum. Genet.*, **31**, 329-338.
- Eriksson, S. (1964). Pulmonary emphysema and alpha₁-antitrypsin deficiency. *Acta med. scand.*, **175**, 197-205.
- Eriksson, S. (1965). Studies in α_1 -antitrypsin deficiency. *Acta med. scand.*, Suppl. 432.
- Figura, K. von, and Kresse, H. (1972). The Sanfilippo B corrective factor; an N-acetyl- α -D-glucosaminidase. *Biochem. biophys. Res. Commun.*, **48**, 262-269.
- Harris, H. (1956). Genes and enzymes in man. In *Novant 'Anni delle Leggi Mendeliane'*, edited by L. Gedda, p. 206. Istituto Gregerio Mendel, Rome.
- Harris, H. (1970). *The Principles of Human Biochemical Genetics*. North Holland, Amsterdam.
- Harris, H. (1974). Common and rare alleles. *Sci. Prog.*, **61**, 495.
- Harris, H., and Hopkinson, D. A. (1972). Average heterozygosity per locus in man: an estimate based on the incidence of enzyme polymorphisms. *Ann. hum. Genet.*, **36**, 9-20.
- Harris, H., Hopkinson, D. A., and Robson, E. B. (1974). The incidence of rare alleles determining electrophoretic variants: data on 43 enzyme loci in man. *Ann. hum. Genet.*, **37**, 237-253.
- Kelley, W. N. (1968). Hypoxanthine-guanine phosphoribosyltransferase deficiency in the Lesch-Nyhan syndrome and gout. *Fed. Proc.*, **27**, 1047-1052.
- Kelley, W. N., and Meade, J. C. (1971). Studies on hypoxanthine-guanine phosphoribosyltransferase in fibroblasts from patients with the Lesch-Nyhan syndrome: evidence for genetic heterogeneity. *J. biol. Chem.*, **246**, 2953-2958.
- Kresse, H., and Neufeld, E. F. (1972). The Sanfilippo A corrective factor. *J. biol. Chem.*, **247**, 2164-2170.
- Lesch, M., and Nyhan, W. L. (1964). A familial disorder of uric acid metabolism and central nervous system function. *Amer. J. Med.*, **36**, 561-570.
- McKusick, V. A. (1972). *Hereditary Disorders of Connective Tissue*, 4th ed. Mosby, Saint Louis.
- McKusick, V. A. (1973). Phenotypic diversity of human diseases resulting from allelic series. *Amer. J. hum. Genet.*, **25**, 446-456.
- Matalon, R., and Dorfman, A. (1972). Hurler's syndrome, an α -L-iduronidase deficiency. *Biochem. biophys. Res. Commun.* **47**, 959-964.

1	Multiple allelism at a single gene locus
2	Multiple gene loci (a) affecting the same enzyme or protein (b) affecting different enzymes or proteins
3	Genetic background
4	Environmental factors

Table II Causes of heterogeneity in inherited disease

- O'Brien, J. S. (1972). Sanfilippo syndrome: profound deficiency of alpha-acetylglucosaminidase activity in organs and skin fibroblasts from type-B patients. *Proc. nat. Acad. Sci. (Wash.)*, **69**, 1720-1722.
- Paglia, D. E., Valentine, W. N., and Rucknagel, D. L. (1972). Defective erythrocyte pyruvate kinase with impaired kinetics and reduced optimal activity. *Brit. J. Haemat.*, **22**, 651-665.
- Porter, C. A., Mowat, A. P., Cook, P. J. L., Haynes, D. W. G., Shilkin, K. B., and Williams, R. (1972). α_1 -antitrypsin deficiency and neonatal hepatitis. *Brit. med. J.*, **3**, 435-439.
- Seegmiller, J. E., Rosenbloom, F. M., and Kelley, W. N. (1967). Enzyme defect associated with a sex-linked human neurological disorder and excessive purine synthesis. *Science*, **155**, 1682-1684.
- Sharp, H. L., Bridges, R. A., Krivit, W., and Freier, E. F. (1969). Cirrhosis associated with alpha-1-antitrypsin deficiency: a previously unrecognized inherited disorder. *J. Lab. clin. Med.*, **73**, 934-939.
- Stamatoyannopoulos, G. (1972). The molecular basis of hemoglobin disease. *Ann. Rev. Genet.*, **6**, 47-70.
- Yoshida, A. (1973). Hemolytic anemia and G6PD deficiency. *Science*, **179**, 532-537.
- Yoshida, A., Beutler, E., and Motulsky, A. G. (1971). Human glucose-6-phosphate dehydrogenase variants. (With table). *Bull. Wild Hlth Org.*, **45**, 243-253.