100,000 RECRUITS GIVEN PRIMAQUINE:
650 DIE OF HEMOLYTIC ANEMIA

50,000 PATIENTS UNDERGO GENERAL ANESTHESIA:
25 SUFFER PROLONGED RESPIRATORY PARALYSIS

GENETIC VARIATION AND HUMAN DISEASE

Medical Grand Rounds
Parkland Memorial Hospital
September 23, 1976
Michael S. Brown, M.D.

10,000 LEGIONNAIRES ATTEND PHILADELPHIA CONVENTION:
27 DIE OF RESPIRATORY FAILURE

150,000,000 PEOPLE GIVEN SWINE FLU VACCINE:
? ?

GENETIC PRINCIPLES

More than one eighth of the proteins (and hence genes) in each human being exist in a form that differs from the one present in the majority of the population. This remarkable degree of genetic variability or polymorphism among "normal" people accounts for much of the naturally occurring variation in body traits such as height, intelligence, and blood pressure. Moreover, these genetic differences produce marked variations in the ability of individuals to handle every environmental challenge, including those that produce disease. Thus, every human disease can be considered to occur as a result of an interaction between a given individual's genetic makeup and his environment. In certain diseases, however, the genetic component is so overwhelming that it expresses itself in a predictable manner without a requirement for extraordinary environmental challenges. We call such diseases genetic disorders.

Molecular Basis of Gene Expression

All hereditary information is transmitted from parent to offspring through the inheritance of specific molecules of deoxyribonucleic acid (DNA). DNA is a linear polymer composed of purine and pyrimidine bases whose sequence ultimately determines the sequence of amino acids in every protein molecule made by the body. The four types of bases in DNA are arranged in groups of three, each group forming a code word or codon that signifies a particular amino acid. A gene represents the total sequence of bases in DNA that specifies the amino acid sequence of a single polypetide chain of a protein molecule.

In order to be translated into a polypeptide each DNA region corresponding to a gene must first be transcribed within the cell nucleus into a molecule called messenger RNA (mRNA). The mRNA represents a sequence of purine and pyrimidine bases that is "complementary" to that of the DNA. Thus, each adenine of DNA becomes a uridine of RNA, each cytosine of DNA becomes a guanine of RNA, each thymine of DNA becomes an adenine of RNA, and each guanine of DNA becomes a cytosine of RNA. Figure 1 shows the DNA and mRNA code words for each of the 20 amino acids that are utilized to form proteins.

The mRNA leaves the cell nucleus and enters the cytoplasm where it becomes associated with ribosomes and thereby serves as a template for the ribosomal synthesis of proteins. Each of the 20 precursor amino acids for protein synthesis is attached in the cell cytoplasm to specific molecules called transfer RNA (tRNA). Each tRNA contains a sequence of purine and pyrimidine bases that is "complementary" to a specific codon in the mRNA. These tRNA molecules with their attached amino acids line up along the mRNA molecule in the precise order dictated by the mRNA code. Under the action of a variety of cytoplasmic enzymes (initiation factors, elongation factors, and termination factors), peptide bonds are formed between the various amino acids, and the completed protein is released from the ribosome. A schematic diagram of the genetic control of protein synthesis is shown in Figure 2.

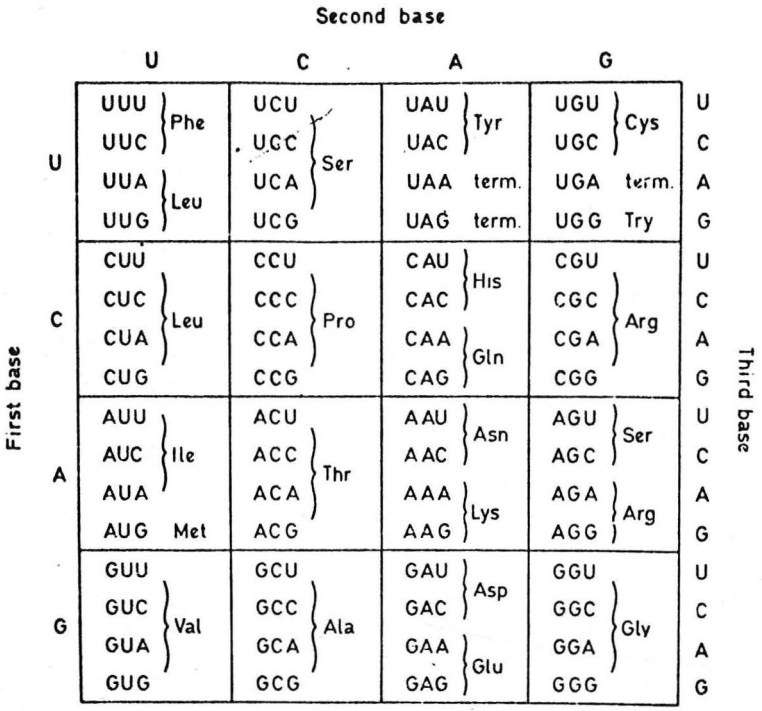

Fig. 1. The genetic code

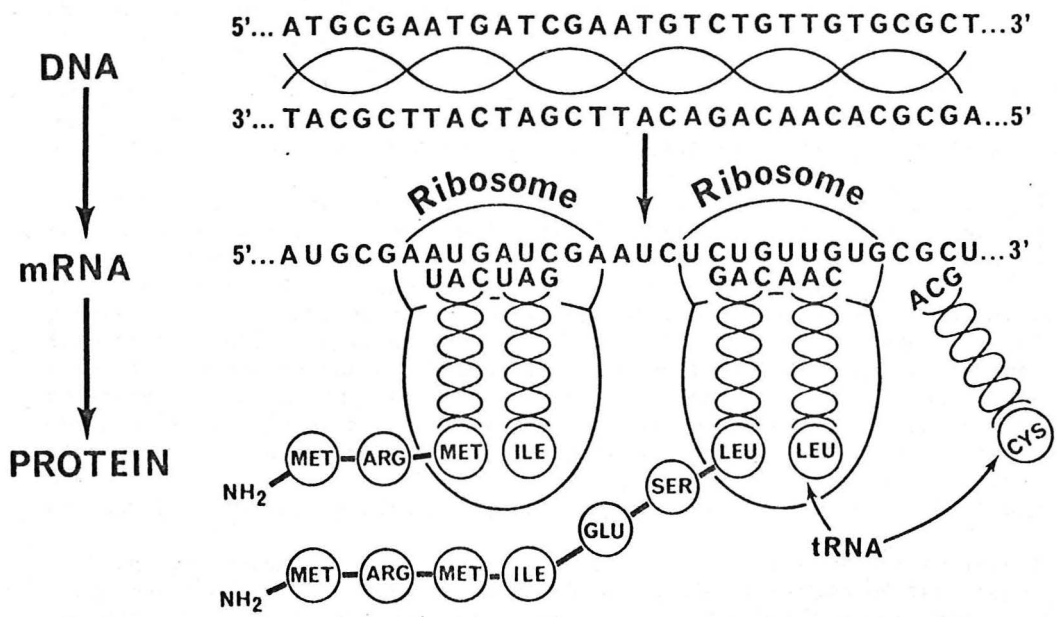

Fig. 2. Machinery for protein synthesis

Maintenance of Genetic Diversity through Transmission and Segregation of Genes

It is estimated that the amount of DNA in the nucleus of each human cell is sufficient to code for more than 100,000 genes and hence to specify more than 100,000 polypetide chains. The genes are arranged in a linear sequence of DNA that together with certain histone proteins form rod-shaped bodies called chromosomes. Each human cell contains 46 chromosomes, arranged in 23 pairs, one of each pair derived from each of the individual's parents. Thus, each individual inherits two copies of each chromosome and hence two copies of each gene. The chromosomal location of the two copies of each gene is termed the genetic locus. When a gene occupying a genetic locus exists in two or more different forms, these alternate forms of the gene are referred to as alleles.

In man, a given gene resides at a specified genetic locus on one particular chromosome. For example, the genetic locus for the Rh blood group in man is on chromosome No. 1; at this chromosomal site there are two Rh genes, one on chromosome No. 1 derived from the mother and the other on chromosome No. 1 derived from the father. When two genes at the same genetic locus are identical, the individual is a homozygote. When the two genes differ (i.e., two alleles are present at the locus), the individual is a heterozygote. Each individual is homozygous at some loci and heterozygous at others. Figure 3 shows a map of human chromosome No. 1, illustrating the location of those genes that have been assigned loci on this chromosome.

The genetic information carried on chromosomes is transmitted to daughter cells under two different sets of circumstances. One of these occurs whenever a somatic cell (i.e., a non-germ cell) divides. This process, called mitosis, functions to transmit identical copies of each gene to each daughter cell, thus maintaining a uniform genetic makeup in all cells of a single organism. The other set of circumstances prevails when genetic information is to be transmitted from one individual to his offspring. This process, called meiosis, functions to produce germ cells (i.e., eggs or sperms) that possess only one copy of each parental chromosome, thus allowing for new combinations of chromosomes to occur when egg and sperm cells fuse during fertilization.

During the process of meiosis, the 46 chromosomes of an immature germ cell arrange themselves in 23 pairs at the center of the nucleus, each pair being composed of one chromosome derived from the mother and its homologous chromosome derived from the father. At a specified point in the meiotic process, the two partner chromosomes separate, only one of each pair going into each daughter cell, or gamete. Thus, meiosis produces gametes with a reduction in the number of chromosomes from 46 to 23, each gamete having received one chromosome from each of the 23 pairs. The assortment of the chromosomes within each pair is random so that each germ cell receives a different combination of maternal and paternal chromosomes. During the process of fertilization, the fusion of egg and sperm cells, each of which has 23 chromosomes, results ultimately in an individual with 46 chromosomes.

The independent assortment of chromosomes into gametes during meiosis produces an enormous diversity among the possible genotypes of the progeny. For each 23 pairs of chromosomes, there are 2^{23} different combinations of chromosomes that could occur in a gamete. Thus, the likelihood that

one set of parents will produce two offspring with the identical complement of chromosomes is one in $2^{23} \times 2^{23}$ or one in 7×10^{13} (assuming no monozygotic or identical twins).

Recombination

Adding even further to the enormous genetic diversity in man is the phenomenon of genetic recombination. During meiosis, when homologous chromosomes are paired, bridges frequently form between corresponding regions of the chromosome pair. These bridges, or chiasmata, are regions in which the two chromosomes break at identical points along their length and subsequently rejoin, the distal segments having been switched from one homologous chromosome to another. This process is designated crossing-over. Although no net change in the amount of genetic material occurs during crossing over, a recombination of genes does occur. For example, consider a chromosome with two loci, A and B, located at opposite ends of the same chromosome. On this particular chromosome, the A locus has a rare allele x and the B locus also has a rare allele y. Without the phenomenon of recombination every offspring that inherited the x allele at the A locus would also inherit the y allele at the B locus. However, if recombination occurs the A locus with the x allele would now be on the opposite chromosome from the B locus with the y allele. In this case any offspring that inherited the x allele at the A locus could not inherit the y allele at the B locus.

Crossing-over in man occurs with great frequency in every meiosis, and the resultant recombination of genes may occur at any point on a chromosome. The further apart are two genes on the same chromosome, the greater is the likelihood that a crossing-over event will occur in the space between them. When two genes are on the opposite ends of a long chromosome, the probability of recombination is so great that their respective alleles are transmitted to offspring almost independently of one another, just as if the two gene loci were on different chromosomes. On the other hand, gene loci that are close together on the same chromosome are said to be "linked" so that there is a great likelihood that offspring will inherit the same combination of alleles that are present on the parental chromosome.

Several examples of gene linkage can be seen from the map of human chromosome No. 1 (Figure 3). For example, the locus for the gene specifying the Rh blood group factor and the locus for the gene producing one form of the dominant trait, hereditary eliptocytosis, occur in close proximity on this chromosome. Thus, if a subject with hereditary eliptocytosis transmits the disease to his offspring, the offspring will usually inherit the allele that is present at the Rh locus on this chromosome. If the Rh allele happens to be a rare one in the population (such as r'), then one can assume that whichever offspring inherits the r' allele at the Rh locus will also inherit the abnormal allele at the eliptocytosis locus. On the other hand, if an offspring does not exhibit the r' allele, he will not usually have eliptocytosis. It is important to note that the concept of linkage does not imply an association between any particular set of Rh alleles and the disease state eliptocytosis, but rather between the two genetic loci. Thus, in different families the abnormal eliptocytosis allele may be linked to the R^1,R^0, r_2 or any other allele at the Rh locus, depending on the allele that happened to be at that locus when the eliptocytosis mutation occurred. Stated another

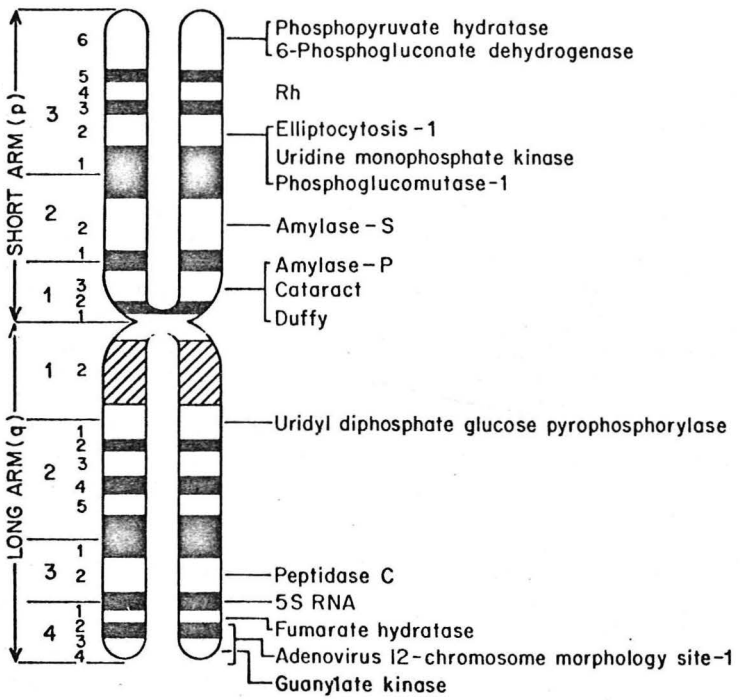

Fig. 3. Genetic map of human chromosome number 1.

way, the eliptocytosis locus is linked to the Rh locus in every family, but the particular Rh allele with which it is associated will differ from family to family.

Mutation

Broadly defined, a mutation is a stable, heritable alteration in DNA. Although the causes of mutation in man are largely unknown, a variety of environmental agents, such as radiation, viruses, and chemicals, are among the factors that are implicated.

Mutations can involve a visible alteration in the structure of a chromosome, such as a deletion or translocation of a portion of a chromosome or they can involve a minute change in one of the purine or pyrimidine bases of a single gene. Most commonly, such "point" mutations consist of the substitution of one base for another, thus changing the meaning of the codon containing that base; hence, their designation as missense mutations. For example, in the gene coding for the β chain of hemoglobin the sixth position normally contains either the nucleotide triplet CTT or CTC, both of which code for the amino acid glutamic acid (Figure 1). The mutation that gives rise to hemoglobin C produces a change of the first base of this triplet from cytosine to thymine, changing the triplet to TTT or TTC, either of which codes for lysine. On the other hand, the mutation that gives rise to hemoglobin S produces a change in the second base of the same triplet (from thymine to adenine) producing either CAT or CAC, which codes for valine. Thus, in the sixth position of the β chain of hemoglobin, the normally occurring glutamic acid may be replaced with either lysine (producing hemoglobin C) or valine (producing hemoglobin S). More than 84 such single-base mutations in the hemoglobin β chain have so far been identified in different population groups, many of which produce a different clinical syndrome. Of all of the mutations so far elucidated in man, the vast majority involve such single base changes.

Besides producing an amino acid substitution, a single base substitution can also cause another abnormality in protein synthesis - premature chain termination. Three mRNA code words (UAA, UAG, and UGA) normally do not specify an amino acid, but rather they constitute the signal that the message has ended and that the protein chain should be released from the ribosome (Figure 1). If a change occurs in DNA that produces one of these mRNA code words (for example, a switch in an mRNA triplet from UAU [tyrosine] to UAA [termination]), the polypeptide chain would be terminated prematurely when translation had reached that point. Such mutations, called nonsense mutations, produce short fragments of proteins that have reduced function.

Cellular Mechanism By Which Mutant Genes Produce Diseases

Critical to the modern understanding of heredity is the concept that the only information transmitted from generation to generation is the sequence of bases in DNA, and that these sequences in turn specify only the primary structure of RNA and protein molecules. All other chemical reactions within a cell - such as the synthesis of complex lipids and carbohydrates, the formation of membranes and other cellular organelles, the accumulation and partitioning of inorganic ions etc. - all occur as a secondary consequence of the action of specific proteins. Many of these proteins are enzymes that catalyze the biochemical conversion of one molecule into another. Others are structural proteins, such as collagen and elastin, and still others are regulatory proteins that dictate how much of each enzyme and each structural protein is to be made.

Since proteins are the cellular molecules whose structures are encoded by

genes, mutations in genes exert their deleterious effects by altering the structure of enzymes, structural proteins, or regulatory proteins. Thus, in a disease such as glycogen storage disease, type I (Von Gierke's disease), massive accumulation of glycogen in the liver is due not to a primary structural abnormality in the polysaccharide glycogen, but to a structural abnormality in a protein, glucose-6-phosphatase, an enzyme that is required to liberate glucose from glycogen.

Genetic Heterogeneity

Genetic heterogeneity is said to exist when two or more mutations can produce a similar clinical syndrome. Hemophilia is one example of such a genetically heterogeneous syndrome. A clinically similar bleeding disorder can be caused by mutations at either of two different loci on the X-chromosome, one leading to a deficiency of factor VIII (classic hemophilia) and the other causing a deficiency of factor IX (Christmas disease). It is now generally believed that most, if not all, hereditary diseases, when carefully analyzed, will be shown to be genetically heterogeneous.

Genetic heterogeneity may result from the existance of a series of different mutations at a single genetic locus (allelic mutations) or from mutations at different genetic loci (nonallelic mutations). The hemoglobinopathies (e.g., sickle cell anemia and SC hemoglobinopathy) are examples of allelic mutations in the gene encoding the β-chain of hemoglobin that can produce a similar clinical phenotype. Hemophilia is an example of a syndrome in which nonallelic mutations can produce a similar clinical picture (see above).

In some cases of heterogeneity, not merely does the genetic locus differ but the mode of inheritance will also differ depending on the mutation. For example, spastic paraplegia, Charcot-Marie-Tooth peroneal muscular atropy, and retinitis pigmentosa are each inherited as autosomal dominant traits in some families, as autosomal recessives in others, and as X-linked recessives in still others. The identification of such genetic heterogeneity in these disorders is of obvious importance for correct genetic counseling.

Genetic Variation Among "Normal" Individuals

The classic view of genetics, which was derived from a synthesis of the ideas of Mendel and those of Darwin, did not allow for extensive genetic variation among members of a single species. It was thought that during the evolution of each species the process of natural selection had operated to select the single genotype that showed the highest fitness. The members of each species were thought to have survived because they were adapted perfectly to their environments. Thus there was thought to be a "perfect" genotype that was "normal" for each species. Since the environment had already selected a perfect genotype, it was thought that any new mutations that occurred could only reduce the fitness of an individual. Hence, in these fully evolved species natural selection operated primarily to "purify" the species so as to eliminate mutant genes with lowered survival value. The idea of an optimal genotype within a species was taught to each medical student educated before 1966 - and it may still be taught in some medical centers today.

However, this classic view of biology is no longer tenable. In 1966 a revolution occurred that destroyed forever the concept of a "perfect" or even a "normal" genotype for each species. This revolution stemmed from the startling observation that no two individuals within any species were genetically alike. This observation was made simultaneously on two continents and from studies of two very different species - man and the fruit fly. The studies in man were carried out by Professor Harry Harris at the Galton Laboratories in London; and the studies on the fruit fly (drosophila) were made by Richard Lewontin and co-workers at the University of Chicago. Both observations were made possible by the emergence of a powerful technique - namely, the use of electrophoresis to identify differences in a single amino acid between two homologous proteins. Using this technique Harris and Lewontin both found an enormous diversity in the amino acid composition of enzymes obtained from individuals within each species. These amino acid variations indicated that the genes dictating each enzyme exist in the population in a wide variety of forms. In the sections below I first some specific illustrative examples of this type of polymorphism among enzymes and then discuss the implications of this variation for the physician.

Electrophoresis As A Tool To Explore Enzyme Polymorphisms

The great power of electrophoresis in demonstrating genetic alterations in enzymes stems from the fact that enzymes can be studied in crude extracts without purification. For example, if one is looking for electrophoretic variants in a human enzyme (let's say, acid phosphatase), one can simply take whole blood, isolate the red cells by centrifugation and then hemolyze the red cells in hypotonic solutions. An aliquot of the crude hemolyzate is then applied to a starch gel and exposed to an electric current. If one performs the electrophoresis at alkaline pH most proteins in the hemolyzate will have a net negative charge so they will migrate toward the positive pole (anode). The rate of their migration will depend on the number of negative charges that they possess - i.e., the more negatively charged proteins migrate fastest. After the electrophoresis the starch block is transferred to a solution containing a substrate for the enzyme to be assayed. In the case of acid phosphatase one uses a phosphorylated dye such as naphthyl phosphate that is colorless in its native state but that becomes yellow when the phosphate is cleaved off by the enzyme. When the starch gel is incubated with the substrate in the presence of the proper buffer and co-factors the dye becomes yellow in the area over the enzyme spot, so a particular enzyme can be visualized even though thousands of other proteins are also present.

This method lends itself particularly to three classes of enzymes: phosphatases, whose reaction products are visualized directly; esterases, whose products are acids that react with appropriate dyes; and dehydrogenases that reduce NAD or NADP, whose presence can be detected fluorometrically or in the presence of certain dyes.

The ability of the electrophoresis technique to detect mutant enzymes is exceedingly potent. Thus, even though a given enzyme may contain 150 or more amino acids, the substitution of a single negative for a positively charged amino acid (for example, glutamate instead of arginine) or vice versa can easily be detected. Even more impressive, the technique can frequently detect alterations of the type in which a neutral amino acid is replaced by either a positively or negatively charged amino acid, or when one of the charged amino acids is replaced by a neutral amino acid (for example, leucine).

In evaluating the results using this technique, two points must be kept in mind. First, the enzymes to be studied are chosen at random. There is no reason to suspect that there is any more polymorphism among phosphatases, esterases and dehydrogenases than there is among any other proteins. The enzymes are simply selected on the basis of the ease with which a staining reaction can be performed on the gel. They thus represent essentially a random sample of enzymes.

Second, in order to be detected by this technique the mutant enzymes must have activity so that they can be visualized by the stain. This is in direct contrast to the types of mutations that create disease states - in

which case the enzyme is generally reduced in activity. Since the electrophoretic method depends on the enzyme having measurable activity, this method is perfectly adapted to demonstrating mutations that change the charge and structure of an enzyme but that do not grossly alter its activity.

Serum Cholinesterase: A Polymorphism With Clinical Implications

One of the earliest examples of a protein polymorphism in man came to light following the introduction and widespread use of the drug suxamethonium (succinyl dicholine) as a muscle relaxant in surgery. Normally the effects of this drug are quite short because it is rapidly hydrolyzed into inactive products by serum cholinesterase. However, occasional individuals (about one in 2000 in the European population) are unusually sensitive to its effect. Following a normal dose of the drug such people develop an extremely prolonged muscular paralysis and respiratory arrests often lasting 2 hours or more instead of just a few minutes. In the early 1950's it was found that in these people the level of serum cholinesterase was consistently low and it seemed reasonable to suppose that this was cause of the sensitivity. It was also found that family members of the patients with succinyl choline apnea also seemed to have cholinesterase levels that were on the low side. However, when large populations were screened, there was marked overlap between the various phenotypes and some people with low cholinesterase levels turned out not be sensitive to succinyl choline (Fig. 4).

An important advance was made in the late 1950's when it was discovered that in addition to its low activity the cholinesterase of sensitive individuals differed in certain important properties from the normal enzyme. For example, its affinity for a variety of substrates was much lower than the affinity of the normal enzyme. Even more importantly, when the susceptibility of the enzyme to certain competitive inhibitors was tested it was found that the enzyme from the suxamethonium sensitive individuals required abnormally high concentrations of inhibitors in order to inhibit its activity. The most informative of these inhibitors is dibucaine (Fig. 5). When the sensitivity of the normal enzyme and the mutant enzyme were compared with regard to inhibition by dibucaine a marked difference between the two enzymes was demonstrated. The percent of inhibition obtained with dibucaine under a certain set of standard conditions is called the "dibucaine number". The higher this value the more normal the enzyme. If one now does this kind of screening in a population using the dibucaine number one finds that the population segregates neatly into three groups as shown in the next figure. The atypical group has low dibucaine numbers, the intermediate group has intermediate numbers and the normal group has high dibucaine numbers (Fig. 6).

The simplest explanation for the above distribution is that patients with the "atypical" enzyme have inherited two alleles for a mutant enzyme that has a low affinity both for the substrate and for dibucaine. The members of the "intermediate" group have one "usual" allele and one mutant allele. The normal group has two alleles that are normal. Although subsequent studies have shown that the true situation is somewhat more complicated than this, the data as presented illustrate a very important principle. That is, among a wide variety of normal subjects an enzyme exists in multiple forms that do not apparently influence the behavior of the individuals. However, when the individuals are exposed to a

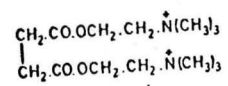

Fig. 4. Succinyldicholine (Suxamethonium)

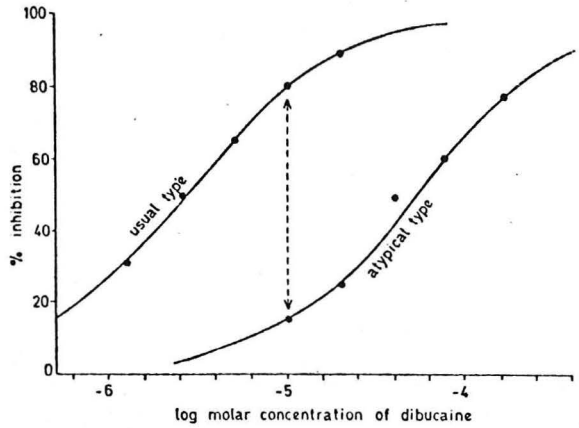

Fig. 5. Inhibition of cholinesterase by dibucaine

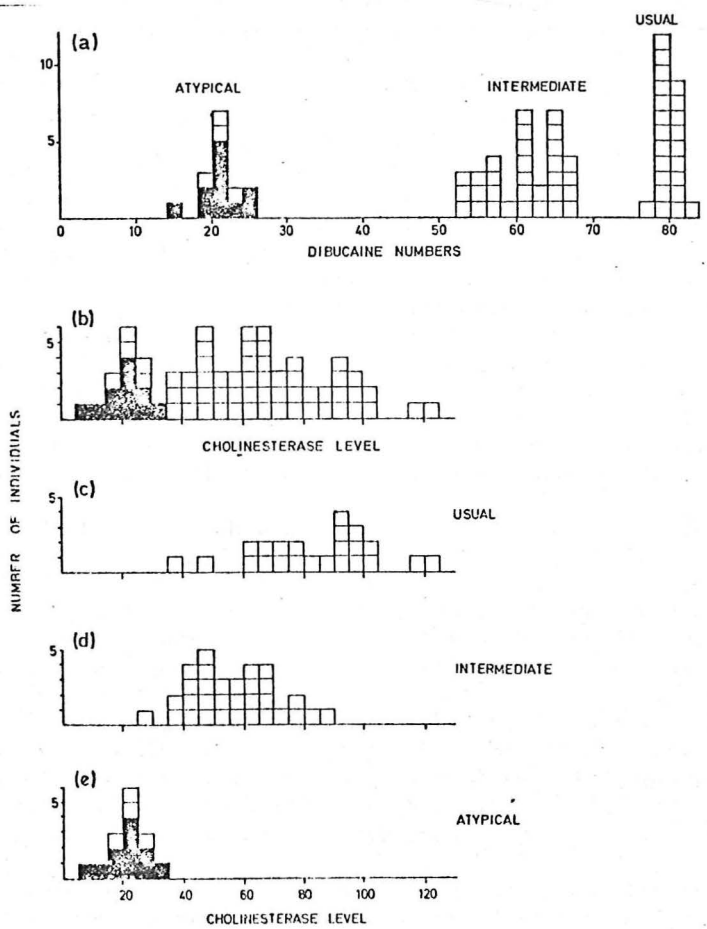

Fig. 6. Cholinesterase level and dibucaine number in blood from patients with "usual", "intermediate", and "atypical" forms of cholinesterase.

particular environmental challenge the enzyme polymorphism suddenly becomes very important and causes disease. It is also important to note that about one in 45 Europeans is a heterozygote for the "atypical" form of cholinesterase, making this an example of a true polymorphism.

Red Cell Acid Phosphatase: A Polymorphism Without Clinical Implications (Yet)

The above example of polymorphism at the cholinesterase locus was detected because the carriers of the atypical enzyme responded to a drug in an idiosyncratic manner. A very similar story was also developed in the 1950's for a polymorphism for a different locus - namely, glucose-6-phosphate dehydrogenase. The third locus at which polymorphism was well described before 1966 was the hemoglobin locus. In fact, sickle cell hemoglobin was actually the first example of a protein polymorphism in man. Each of the aforementioned polymorphisms was discovered because it created a clinical disease. Before 1966 these polymorphisms were thought to represent genetic curiosities. That is, they were thought to represent peculiar situations in which for some reason it was advantageous to the species to maintain variant forms of proteins in the population. However, most enzymes, as well as other proteins, were assumed to be monomorphic in the population.

Early evidence that genetics was not so simple came from the experiments of Harris on red cell acid phosphatase. In contrast to the above examples, no disease state was known in which red cell acid phosphatase activity was abnormal. However, because this enzyme could be studied easily by electrophoresis of red blood cell hemolyzates, Harris selected this enzyme to determine whether any genetic variants exist. The results were striking. When red cell hemolyzates were subjected to electrophoresis and then stained for acid phosphatase activity a variety of bands with different mobilities were observed. By doing family studies Harris was able to cut through the heterogeneity and divide the population into certain specific genotypes. The electrophoretic patterns turned out to be somewhat complicated because each allele actually gives rise to two protein bands on electrophoresis - one fast and one slow. These two bands are both felt to represent the product of the same allele, but they represent some type of conformational change in the protein so that it exists in two forms. Once this is recognized, the situation becomes relatively simple to analyze as show shown in Fig. 7.

Individuals who are homozygous for the type A allele produce one set of bands. Individuals homozygous for the type B and C allele produce different sets of bands. Heterozygotes (that is patients who have a mixed genotype such as B on one chromosome and A on the other chromosome) produce four bands of enzyme activity. In other words each allele behaves independently. The important thing about the red cell acid phosphatase distribution in the European population is that none of the genotypes is sufficently common as to be called normal. Thus, among Europeans 13% of people are type A , 43% type BA, 36% type B, 3% type CA, 5% type CB and about 1 in 600 people are type C. This is a rather remarkable situation that is reminiscent of the heterozygosity at the blood group loci and also at cell surface antigenic loci such as in the HLA system. However, to have such heterogeneity at an enzymatic locus is really surprising.

Even more surprising, was the finding that the activity of acid phosphatase in red cells was different depending on the different allele that was present (Fig. 8).

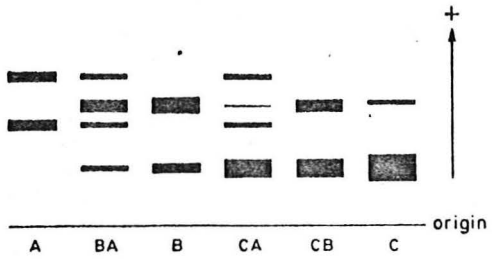

Fig. 7. Electrophoresis of red cell acid phosphatase

Average red cell acid phosphatase activity in individuals of different types. (Spencer et al 1964b.)

Type	Number of individuals tested	Mean activity	Standard deviation
A	33	122.4	16.8
BA	124	153.9	17.3
B	81	188.3	19.5
CA	11	183.8	19.8
CB	26	212.3	23.1

The activity is expressed as μmoles of p-nitrophenol liberated from p-nitrophenyl phosphate in 30 min at 37 °C, per gram of haemoglobin present in haemolysate.

Fig. 8. Average red cell acid phosphatase activity in individuals of different types.

The acid phosphatase polymorphism illustrates another important point about human diversity - namely, that qualitative differences between alleles in a population can contribute to the quantitative variation that constitutes the "bell-shaped curve" of the normal distribution. Fig. 9 shows that the total value for red cell acid phosphatase activity among members of the population shows a continuous variation between about 100 units per gram of hemoglobin and 240 units with a mean of about 170 units. However, on closer inspection this curve is seen to be made up of overlapping frequency distributions from four different curves. Thus, individuals homozygous for the A allele have a mean activity of 122 units, whereas individuals homozygous for the B allele (the most common allele) have a mean activity of 188 units. Heterozygotes who possess one A allele and one B allele have a mean activity exactly halfway between the two homozygotes (154 units). Similar conclusions hold true for AC and BC heterozygotes (the C homozygotes are too rare to contribute to the population data). These data illustrate in a very straightforward way how the understanding of allelic heterogeneity helps us dissect a single quantitatively variable trait into its component parts.

It must be noted that many of the most important human diseases occur in individuals who occupy the extremes of bell-shaped curves such as the one in Fig. 9 . For example, patients with essential hypertension or with "polygenic" hypercholesterolemia are defined by being in the upper 5% of such bell-shaped curves. Such patients are markedly predisposed to the ravages of atherosclerosis. It is likely that such patients owe their diseases to the type of allelic diversity manifested by the acid phosphatase polymorphism. In many of these "polygenic" traits the bell-shaped curve is probably attributable to combinations of unfavorable alleles at several loci.

The Extent of Heterogeneity Among Humans

Studies of the type performed on acid phosphatase have now been performed on 71 human enzymes. The results are rather staggering. Twenty of the 71 loci were polymorphic - that is, the most common allele in the population accounted for less than 99% of the total alleles (i.e., at least 2% of the population were heterozygotes). Thus, at 28.2 per cent of these 71 loci major variant alleles were present in the European population. By summing the observed values for the percentages of heterozygotes at each locus and dividing by the number of loci (71), one obtains an average value for heterozygosity at all loci. This average heterozygosity equals 6.7%. This means that each individual in the population is heterozygous at about 7% of these 71 loci. This figure is even more striking when one considers that these variants were detected electrophoretically and that only one-third of all mutations are detectable by this technique (since 2/3 of mutations result in no change in the charge of the protein). Thus, a conservative estimate is that the true average heterozygosity is 21%. Since the enzymes studied were chosen for their ease of assay rather than because they were suspected of being polymorphic, it is reasonable to suppose that they are representative of all enzymes. Thus, these data indicate that every human being is heterozygous at more than 21% of his genetic loci. Stated another way, more than one-fifth of all enzymes in each "normal" individual exist in a form that is different from the one present in the majority of the population (Fig. 10).

The extreme degree of genetic variability among normal individuals is not restricted to men - it has been found in every diploid species so far tested. Table 1 shows the average heterozygosity per locus among four invertebrate species and six vertebrate species. It can be seen that heterozygosity is

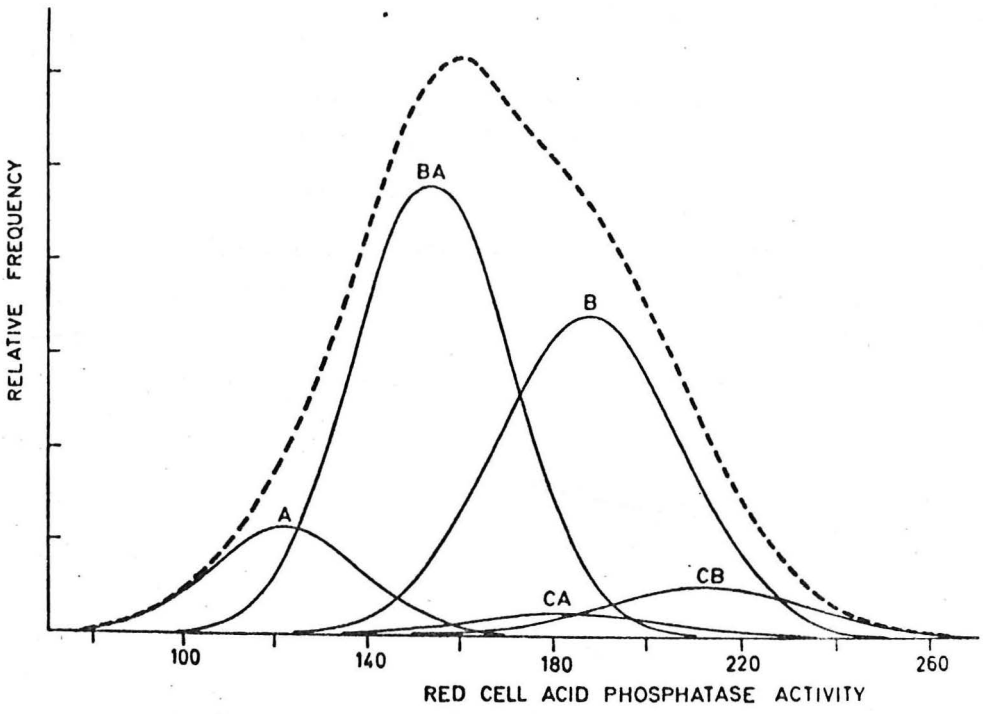

Fig. 9. Breakdown of the acid phosphatase curve for the entire population into separate population curves for each type.

Enzyme polymorphism in Europeans (Harris and Hopkinson 1972).

Number of loci screened	71
Number of loci showing electrophoretic polymorphism (i.e. >0.02 heterozygotes)	20
Percentage of polymorphic loci	28.2
Average heterozygosity per locus (detected electrophoretically)	0.067

Fig. 10. Enzyme polymorphism in Europeans

Table 1. Average heterozygosity per locus in different animal species.

Average heterozygosity per locus in different animal species, for alleles giving rise to electrophoretic enzyme and protein differences. For detailed references see Selander and Kaufman (1973).

Organism	No. of species	No. of loci	Average heterozygosity per locus	
			Mean	Range
Invertebrates				
Drosophila	19	11–33	0.145	0.05–0.22
Field cricket	1	20	0.145	—
Land snails	3	17	0.207	0.14–0.25
Horseshoe crab	1	25	0.097	—
Total	24		0.1507	
Vertebrates				
Fish (Tetra)	1	17	0.112	—
Lizards	4	15–29	0.058	0.05–0.07
Sparrow	1	15	0.059	—
Rodents	14	18–41	0.056	0.01–0.09
Seal	1	19	0.030	—
Man (European)	1	71	0.067	—
Total	22		0.0584	

present in every species and is even more pronounced in the invertebrates than in the vertebrates. These data indicate that the maintenance of such heterozygosity must be a basic biological feature of all diploid organisms.

The striking degree of genetic individuality among humans is illustrated dramatically in Fig. 11 . If one simply examines 15 genetic loci at which polymorphism has been demonstrated one can calculate for each locus the probability that any two individuals will have the same phenotype. If one then combines this data and asks what is the likelihood that any two individuals will have the same phenotype at all 15 loci one concludes that this likelihood is only 0.00017. Thus, when one considers only these 15 loci, there is a tiny chance that any two individuals are identical. When one then applies this reasoning to the other 100,000 loci in man (20,000 of which, at least, are polymorphic) one realizes that there is no chance at all that any individuals will be genetically alike (with the exception of identical twins).

Enzyme individuality in the English population.
Data on 12 enzymes (15 loci).

Enzyme	Number Alleles with frequency	Frequency of Commonest phenotype	Probability of two randomly selected individuals having the same phenotype
Red cell acid phosphatase	33	0.42	0.32
Phosphoglucomutase			
locus PGM_1	22	0.57	0.46
locus PGM_3	22	0.55	0.46
Placental alkaline phosphatase	33	0.39	0.29
Liver acetyl-trans-transferase	22	0.53	0.50
Serum cholinesterase			
locus E_1	22	0.96	0.92
locus E_2	22	0.90	0.82
Adenylate kinase	22	0.92	0.85
Adenosine deaminase	2	0.90	
Phosphogluconate dehydrogenase	2	0.96	0.92
Alcohol dehydrogenase			
locus ADH_2	2	0.94	0.86
locus ADH_3	2	0.48	0.38
Glutamate-pyruvate transaminase	2	0.50	0.38
Esterase-D	2	0.82	0.70
Malic enzyme (mitochondrial form)	2	0.48	0.42
Combined	-	0.0017	0.00017

Fig. 11. Enzyme individuality in the English population

Mechanism for the Maintenance of Genetic Heterogeneity Among Individuals

Two antagonistic theories have developed to explain the existence of such widespread heterogeneity among individuals of a given species. The first is the Balance Theory, which states that heterozygotes for certain traits have a selective advantage over either homozygote, so that natural selection reaches a balance in which a certain optimum gene frequency for each allele is maintained in the population. The second theory is the so-called Random Drift theory which states that the vast bulk of amino acid substitutions in proteins do not affect the function of the enzyme. When such inconsequential mutations occur, most of them are eliminated by chance selection, but because of the randomness of the mating process a few of these mutant genes will spread in the population. Viewed in this manner, the entire genome of a species is in flux, and the polymorphisms we now observe are due to the fact that gene frequencies are fluctuating at random in the population.

The classic example of a balance situation is the case of Sickle Cell hemoglobin. The homozygote for this mutant allele has a genetic fitness of zero, so that homozygotes act to diminish the frequency of the gene in the population. Yet 20% of the individuals in many African countries are heterozygotes for the sickle cell hemoglobin allele. The reason for this high frequency of heterozygotes is that the heterozygotes are resistant to malaria. Thus, a number of studies have demonstrated that the severe forms of falciparum malaria (such as blackwater fever and cerebral malaria) are markedly diminished among sickle cell heterozygotes as compared with homozygous normal individuals in countries of high occurrence of malaria. It has been calculated that even though homozygotes for sickle cell hemoglobin are totally unfit, the gene will maintain a 10% frequency in the population if the heterozygote advantage is only 11% - that is, if heterozygotes have only 11% more children, on the average, than homozygous normals (Figs. 12 and 13).

In addition to sickle cell hemoglobin, the mutant form of glucose-6-phosphate dehydrogenase, and the heterozygous state for beta-thalassemia all seem to protect against malaria, and this explains the high frequencies of each of these mutant traits in countries of high malaria prevalence. A corollary of the above conclusions is that the frequency of these genes will decline when malaria is eradicated.

Unfortunately, the number of instances in which this neat type of balancing selection can be invoked to explain genetic polymorphism in man is very small. It is difficult to think of the reason, for example, why the acid phosphatase phenotypes should give an enhanced fitness to the heterozygote. Thus, a group of population geneticists, led by Kimura, have suggested that most polymorphisms are maintained not by genetic balance, but by random drift. Based on the early theoretical work of R. A. Fisher, these geneticists have calculated that if an amino acid substitution occurs in one individual that does not alter the function of an enzyme, then that mutant allele has a small but finite chance of becoming the predominant allele in the whole population.

Fig. 12.

Incidence of sickle-cell trait among African children whose deaths could be attributed to malaria (Allison 1964).

Locality	Deaths due to malaria	Number with sickle-cell trait	Incidence of sickle-cell trait in the population	Expected number with sickle-cell trait if no selective differential	References
Uganda (Kampala)	16	0	0.16	2.6	Raper (1956)
Congo (Leopoldville)	23	0	0.235	5.4	Lambotte-Legrand, J. and C. (1958)
Congo (Luluaborg)	23	1	0.25	5.7	Vandepitte (1959)
Ghana (Accra)	13	0	0.18	2.3	Edington and Watson-Williams (1964)
Nigeria (Abadan)	29	0	0.24	7.0	Edington and Watson-Williams (1964)
Totals	104	1		23.0	

Fig. 13.

Incidence of sickle-cell trait amongst 818 consecutive admissions to a children's ward at Kampala (thirty-one patients with sickle-cell anaemia admitted during this period are not included). (After Raper 1956.)

Disease group	Total	Number with sickle-cell trait	Incidence of sickle-cell trait
Miscellaneous	186	25	0.13
Pneumonia	118	18	0.15
Upper respiratory infections	59	13	0.22
Diarrhoea and vomiting	106	25	0.24
Poliomyelitis	26	4	0.15
Tuberculosis	37	8	0.22
Meningitis (purulent)	26	5	0.19
Malnutrition	77	11	0.14
Hookworm anaemia	30	2	0.07
Typhoid fever	17	6	0.35
Malaria (a) Uncomplicated	83	13	0.16
(b) Cerebral	47	—	0.00
(c) Blackwater fever	6	—	0.00
Total admissions	818	130	0.16

The strongest evidence in favor of the random drift hypothesis comes from data on the rates of evolution of homologous proteins over a range of different species. For example, if one examines the amino acid sequence in cytochrome C from a variety of species, one finds that 30 of the 100 amino acids are identical in species from candida to man. These 30 amino acids are thought to be critical for the normal functioning of the cytochrome C molecule. On the other hand, the other 70 amino acids vary among the species in a meaningful way. If one arranges the species on the basis of the number of nucleotide differences between their cytochrome C genes, one obtains a pattern that agrees in all respects with the order of evolution of the various species that was deduced from fossil studies (Fig. 14).

Even more striking is the fact that if one looks at all of the currently existing species one can make the remarkable calculation that within each line of descent the total number of nucleotide substitutions is surprisingly constant - that is, within each line of descent approximately 30 nucleotide substitutions have occurred during the course of evolution. This can be calculated to result in an amino acid substitution rate of 3.3×10^{-10} substitutions per amino acid site per year.

Similar calculations, on a more limited scale, can be made for other proteins - most notably hemoglobin. In each case when one compares a variety of species for each protein one finds a remarkably similar rate of amino acid substitution. However, between proteins this substitution rate varies somewhat. This variation is attributed to the fact that in different proteins a variable percentage of amino acids will be critical to function and thus invariant.

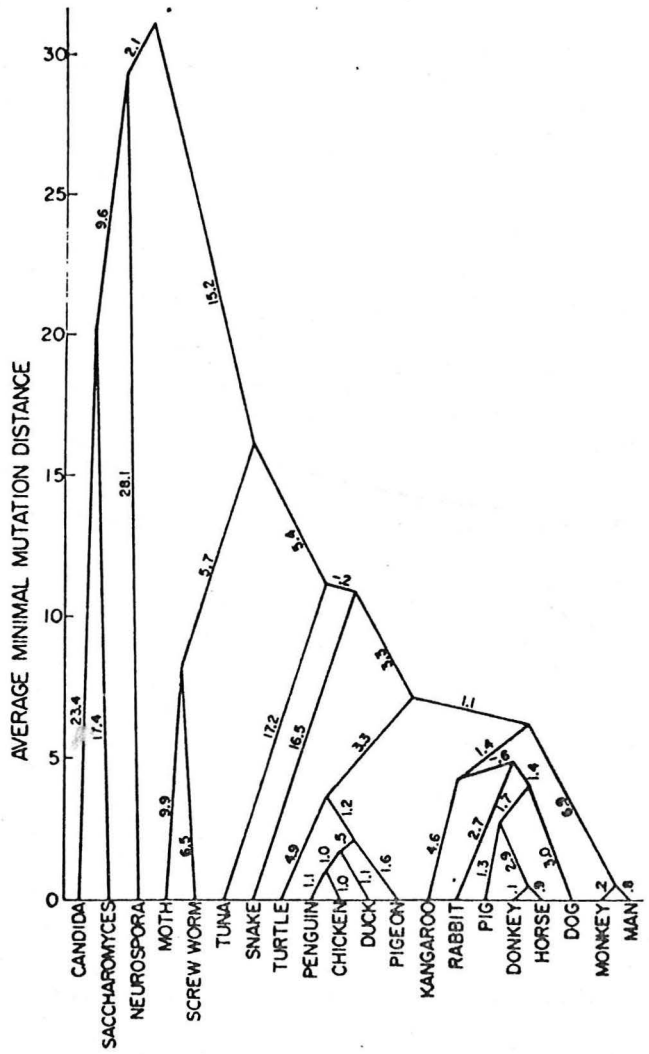

Fig. 14. Number of nucleotide differences in the gene for cytochrome C among various species.

Importance of Genetic Diversity for the Survival of Individuals and Species

The Implications of Being Diploid

When one reviews the literature in a simple-minded way it is difficult to avoid the conclusion that the many polymorphisms are, in fact, due to genetic drift. It seems to me a mistake, however, to conclude that such drift is functionally meaningless. Rather, one must look upon genetic heterogeneity as a phenomenon that provides adaptability in a population - that is, although each species has evolved a set of genes that seems near-optimal for one environment, the species must always be ready for environmental change. Thus, it is a tremendous advantage to the species to contain a reservoir of genes that have been altered in relatively minor ways that don't destroy function, but that provide variant enzymes that may be superior under certain conditions of environmental stress.

An extreme, but illustrative, example of the advantage of genetic diversity can be seen in the immunoglobulin system. Here it is clearly critical that each B cell contain on its surface a genetically distinct immunoglobulin molecule with a specific amino acid sequence that allows it to bind a specific antigen. By having a huge variety of cells with genetically different antibody molecules on their surfaces an individual has the capacity to react to an enormous diversity of structurally different antigens. The important thing is that the antibodies are pre-formed and presumably represent the result of random amino acid substitutions in the variable regions of the immunoglobulins that serve merely to increase the diversity of pre-formed antibodies. Note that the "mutation" that creates the proper antibody has occurred before the subject has ever been exposed to the antigen. If, in order to dispose of a foreign antigen, one had to wait around for a mutation to occur in a B cell, one would certainly be dead before the appropriate antibody was formed. The existence of preformed mutant proteins thus allows an individual to adapt rapidly to new environmental challenges.

I believe that genetic diversity in enzymes accomplishes the same thing within a species as does genetic diversity in immunoglobulins. That is, although the mutations occur at random and drift at random through the population, their very existence gives the species an opportunity to adapt to new environmental challenges.

An example of this is the Duffy null phenotype. It has recently been demonstrated that the plasmodium vivax parasite can only enter human red cells that have the Duffy A or Duffy B antigen. Cells that lack the Duffy A or B antigen are called Duffy null, and patients who are homozygous for this null allele are totally resistant to vivax malaria. It has long been known that nearly all Africans have the Duffy null phenotype and this can now be attributed to the fact that such individuals are resistant to vivax malaria and hence in an

environment that contains malaria organisms these individuals reproduce better than subjects with the Duffy A or B alleles.

The interesting phenomenon, however, is not in the malaria-exposed population but in the malaria-free population. Thus, among European caucasians the frequency of the Duffy null allele is 3% even though malaria is not a selective agent in that population. However, if vivax malaria were to invade northern Europe,the random .023% of people who are homozygous for the Duffy null allele will survive and produce the nucleus of a colony that will re-instate the human species.

As medicine becomes more powerful in altering human metabolism and behavior, I believe that physicians will be encountering this type of allelic polymorphism more and more. If we are to understand these phenomena we must all become more sophisticated in our understanding of the science of genetics itself.

SUGGESTED READING

Harris, Harry (1975). The Principles of Human Biochemical Genetics. 2nd edition. American Elsevier Publishing Company, New York.

A beautifully written, concise, and well-focussed presentation of biochemical genetics from the modern viewpoint. Perfect text for the physician.

Lewontin, Richard C. (1974). The Genetic Basis of Evolutionary Change. Columbia University Press, New York.

A detailed and thoughtful treatise which attempts to solve some of the problems regarding evolution and genetics that have been raised by the finding of widespread genetic polymorphism within species.

Bodmer, W.F. and Cavalli-Sforza, L.L. (1976). Genetics, Evolution, and Man. W.H. Freeman and Company, San Francisco.

A brand new text that presents the topic of population genetics from an intuitive, non-mathmatical viewpoint that is easy to comprehend.

Kimura, Motoo (1976). Population genetics and molecular evolution. Johns Hopkins Medical Journal. 138, 253-261.

A concise review of the contributions of this important proponent of the random drift hypothesis.

Harris, Harry (1966). Enzyme polymorphisms in man. Proc. Roy. Soc. Biol. 164, 298-310.

The first announcement of the finding of widespread genetic heterogeneity in man. A classic paper.