GERALDINE A. MCDOWELL* AND MIRIAM G. BLITZER†
Section on Human Biochemical Genetics, Human Genetics Branch, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda; and †Division of Human Genetics, Departments of Pediatrics and Obstetrics and Gynecology, University of Maryland School of Medicine, Baltimore

Reference

Deiler JH (1909) The settlement of the German coast of Louisiana and the Creoles of German descent. Americana Germanica, Philadelphia

© 1993 by The American Society of Human Genetics. All rights reserved.
0002-9297/93/5303-0027$02.00

Am. J. Hum. Genet. 53:783-785, 1993

The Designation of Mutations

To the Editor:

Many different conventions are used for the primary designation of mutations. Commonly, the amino acid change that has been deduced from the nucleotide substitution is employed, but often the cDNA number, the genomic number, or even a "nickname" based on a restriction site or a patient's name has been used. These notations are not, of course, mutually exclusive, and several of them are used in first describing a mutation. However, when the existence of a mutation becomes well established, it usually acquires a designation that is used exclusively, and that common name is the one to which I refer here.

An ideal nomenclature would be one that is entirely unambiguous. One might hope that a geneticist of the year 2493 could pick up a 1993 copy of *The American Journal of Human Genetics* and quickly understand, from the designation of a mutation and without extensive study of other sources, the location of a nucleotide change. However, the complexity of the genome and its functions is such that a perfect nomenclature is unachievable.

Amino acid–based mutation designation.—Surely a convention based on the amino acid change, embraced by so many geneticists, must have something to commend it. And so it does. One reason for the use of amino acid–based designations is the historical fact that a number of proteins, most notably hemoglobin, were sequenced at the protein level even before the DNA code was known. This, quite understandably, established a tradition from which it has sometimes been difficult to break. Another major reason for the use of this nomenclature seems to be the wish to divine the change in the gene product brought about by the mutation. Commendable as it may be, the idea that this can be achieved is often an illusion. Certainly a change to a stop codon near the amino terminus of the protein tells us much about the effect of the mutation. Knowing that a change in the nucleotide sequence does not change an amino acid is also useful, although the usual conjecture that such a mutation is "neutral" may in the future sometimes prove to be incorrect. It is entirely possible that some such mutations may exert an effect either because of their effect on the stability or translatability of the message or because of the abundance of the needed tRNA. Between these extremes, the data often do not reveal much about the effect of the mutation on the protein, although this may change with advances in understanding of protein structure.

There are, however, are a number of compelling disadvantages intrinsic to the use of the amino acid mutation as the primary nomenclature for the designation of mutations:

1. It is more logical to report what we actually find, rather that what we deduce. Genes are not composed of amino acids but of purine and pyrimidine bases. In the vast majority of cases it is the base sequence that is determined in the laboratory. Moreover, the deduction is occasionally wrong. Notable is the fact that in glutathione peroxidase the UGA codes for selenium cysteine, not for a stop codon (Chambers and Harrison 1987), and that the putative β141-deleted leucine in hemoglobin Atlanta-Conventry has actually been changed posttranslationally to hydroxyleucine (Brennan et al. 1992). There is actually no mutation at this location at all. The mutation in the β-chain of hemoglobin E produces an amino acid substitution, but it also causes aberrant splicing.

2. Any good notation should be not only logical but also unambiguous. The amino acid notation for the description of mutations has a number of serious, glaring ambiguities.

a. At least three different starting points for the numbering of amino acids are employed. Is one to use the sequence of the primary translated product, or is one to use the processed proteins? Does the start methionine

count? No consistent rule has been applied. Some investigators have counted the first amino acid after the putative leader sequence cleavage site as number 1. In the case of many other proteins, the first methionine is counted as amino acid 1. This is not true, however, in the case of hemoglobin, where the amino acid sequence has actually been determined. Here amino acid 1 of α-, β-, and δ-globin are valine, γ is glycine, etc. Sickle cell hemoglobin is designated as $\beta^{6\ Glu \rightarrow Val}$, but, had the sequence been determined on DNA, the mutation would involve amino acid 7. How is our 2493 reader to know what the 1993 scientist had in mind when numbering amino acids to designate the site of a mutation? Even with an agreed-on starting point, an amino acid base-numbering system cannot deal with proteins in different tissues that are the product of the same gene but with a different splicing motif in different tissues. An example is the pyruvate kinase PKL gene. Here the red-cell form of the enzyme contains amino acids from an exon that is not translated in the liver form of the enzyme.

b. The amino acid change can be deduced from the change in the nucleotide sequence, but the reverse is not true. Thus the amino acid notation of many mutations is ambiguous in that the same change is produced by different mutations. For example, when the normal codon is defined as CAU for His, a change to Glu can be a change of U to A or G, to give CAA or CAG. If Trp is AAA, a mutation to Asn could be AAC or AAU. There are about 15 other such examples. The situation is compounded, of course, if one does not know the sequence of the original triplet, and it is worse still when there is no amino acid change. For example, the L210L mutation can represent 25 different nucleotide substitutions.

c. There are some mutations that cannot be described at all by using an amino acid–based nomenclature. For example, the second most common Jewish mutation that causes Gaucher disease is due to insertion of a single extra nucleotide into the coding region. Nor can deletions that produce frameshifts or splicing mutations be described in an amino acid–based nomenclature.

cDNA-based mutation designation.—The advantage of a cDNA-based system is that it is generally unambiguous and that it accommodates most but not all mutations. It shares, however, some of the deficiencies of the amino acid–based system. It cannot describe splicing mutations, nor does it take into account different splicing patterns in different tissues.

Genomic DNA–based designation.—None of the deficiencies inherent in the cDNA-based system exists in a genomic DNA–based system, but new difficulties arise. First of all, the cDNA sequence is often known long before the complete gene sequence has been elucidated. For example, the cDNA sequence of glucose-6-phosphate dehydrogenase became known in 1986, but, because intron 2 was 9,861 nt in length, the genomic sequence remained largely unknown until 5 years later; by this time many mutations had already been described. A second difficulty is that the starting point for

Table 1

Advantages and Disadvantages of Different Methods of Designating Mutations

Designation of Mutation	Advantages	Disadvantages
Amino acid number and substitution	Tradition, based on hemoglobin, etc. Insights into effect on protein product	Deduced, not measured Different starting points Many ambiguities Does not accommodate alternative splicing, insertion, and deletion mutations or intron mutations
cDNA number and substitution	Measured parameter, not deduced Unambiguous	Does not accommodate alternative splicing and intron mutations
Genomic DNA number and substitution	Measured parameter, not deduced Accommodates all types of mutations	Starting point not obvious Full sequence often unknown when mutations are discovered Polymorphisms and sequencing errors likely

numbering is not nearly so obvious in the case of genomic sequences as it is in the case of cDNA sequences. Finally, errors are more likely to occur in determining a genomic sequence, because insertion or deletion of a nucleotide does not produce a telltale frameshift. As errors are discovered, numbering needs to be changed, or flawed sequences need to be retained as standards. Moreover, polymorphisms representing different numbers of repeats occur commonly in introns and not in coding regions.

Conclusions.—The advantages and disadvantages of different notations are summarized in table 1. The development of uniform notation for the designation of mutations would be highly desirable. Such a system should be as broadly applicable to mutations as possible. Amino acid–based designation, although commonly used, has relatively little to commend it. The choice would seem to be between systems based on the nucleotide number in a cDNA- and a genomic DNA–based numbering system. Each of these has advantages and disadvantages, and a cDNA-based system probably represents the most acceptable compromise.

ERNEST BEUTLER
Department of Molecular and Experimental Medicine
The Scripps Research Institute
La Jolla, CA

References

Brennan SO, Shaw J, Allen J, George PM (1992) β141 leu is not deleted in the unstable haemoglobin Atlanta-Coventry but is replaced by a novel amino acid of mass 129 daltons. Br J Haematol 81:99–103

Chambers I, Harrison PR (1987) A new puzzle in selenoprotein biosynthesis: selenocysteine seems to be encoded by the "stop" codon, UGA. TIBS Rev 12:255–256

© 1993 by The American Society of Human Genetics. All rights reserved.
0002-9297/93/5303-0028$02.00

Am. J. Hum. Genet. 53:785–787, 1993

Presymptomatic Testing for Huntington Disease in the United States

To the Editor:

Presymptomatic testing for Huntington disease (HD) by using linkage analysis has been available in the United States on a limited basis since 1986. Guidelines for testing have been published by both the Huntington's Disease Society of America (1989) and the World Federation of Neurology (WFN) Research Group on Huntington's Chorea (Went 1990). Recommended testing protocols include neurological, psychiatric, and psychological screening; pretest counseling; and posttest follow-up.

Unlike Canada and many European countries offering predictive testing, the United States has no central organization to coordinate testing or to gather information on the number of individuals tested and the outcomes of testing. In a meeting held in conjunction with The American Society of Human Genetics annual meeting in 1990, an ad hoc committee was established for the purpose of gathering and disseminating information about testing protocols, results, and outcomes of presymptomatic testing in the United States. The results of two surveys of all the centers offering presymptomatic testing for HD in the United States, conducted in May 1991 and again in May 1992, are presented here.

In May 1991, surveys were mailed to the contact person at each of the 23 sites offering predictive testing, asking their center for information as of December 31, 1990. All 23 centers (100%) replied. A second survey was mailed in May 1992 asking for data as of April 15, 1992. At the time of the second survey, three new centers had begun to offer testing, and two centers that had participated in the original survey were no longer offering this service. Data for this second survey were obtained from all 26 centers.

Results indicate that, after an initial increase, the number of centers offering testing has leveled off. By year, the number of new centers offering testing is as follows: 1986 (2), 1987 (2), 1988 (2), 1989 (8), 1990 (8), 1991 (2), and 1992 (2). Seventeen (65%) of the 26 centers are university based, although in different departments: neurology (3), psychiatry (4), genetics (4), and pediatrics (6). Two are based in health maintenance organizations, three in private genetics clinics, one in a medical center, and two in nonprofit organizations, while one was unspecified. Twelve (46%) of the centers have an HD clinic associated with the program, and 14 do not. Most centers, 19 (73%), have completed fewer than 15 tests.

The testing programs are directed by individuals of various professions and training. These professions include Ph.D. medical geneticist (6), Ph.D. medical geneticist/psychologist (1), Ph.D. psychologist (3), neurologist (2), M.D./Ph.D. medical geneticist (3), genetic counselor (6), M.D. medical geneticist (3), M.D. neuro-