GENETICS

Nomenclature of Polymorphic Protein Systems

In this communication some suggestions as to nomenclature of polymorphic protein systems are considered. For these, primarily those of blood, the number of recognized systems increases rapidly in both man and animals. The techniques for detecting and characterizing such systems vary. This bears some influence on the problem of nomenclature. If we, for example, have a system recognized by serological methods only, the common practice has been to name the first-found character *A* (or *a*)[1], the next *B* (or *b*) and so on. The letters therefore merely show the order of detection. After the characters have been classified into genetic systems, letter designations, usually two, have been ascribed to each system. Such a nomenclature is sufficient if characters belonging to the same system do not show a second order of relationship which is contradictory to the alphabetical order. This often happens with polymorphic proteins characterized by their different rates of migration as determined by, for example, the technique of electrophoresis.

Also for characters detected by this last-mentioned technique a common practice has been to name the first protein found *A*, the next *B*, and so on. Smithies[2], however, made use of the letters *B*, *C* and *D*, for human transferrins. Numbers, 1, 2, etc., have also been used[3]. The serum proteins within the same genetic system usually appear as multiple bands after electrophoresis, which have sometimes caused the naming of each band by separate letters[4]. A common practice now, however, is to name the allelic band pattern with one letter only, independent of number of bands. With 'allelic' is understood the product of a single gene. This nomenclature may also be sufficient if new alleles are not detected. If this happens, which is usually the case because the number of alleles within a system seems to depend on the extent of the investigation, difficulties may arise. For example, if in a protein system with two known characters *A* and *B* a third protein with migration speed between that of *A* and *B* is disclosed, it is not consistent to call this new protein *C*.

Nomenclature should be descriptive and simple. For polymorphic proteins we primarily need symbols for the genetic system and for the allelic band patterns, as these appear after electrophoresis. The gene or allele symbol will then be a combination of these two designations.

The genetic system has to be characterized by an abbreviation of the name of the character. An example is *Tf* for transferrin[5]. It might, however, have been even better with *Ti* (transferrin iron). Two letters should in most cases be enough for characterization and distinction between protein systems. Such a nomenclature should cover a high number of characters because of the large number of possible two-letter combinations. There is, of course, also the possibility of using a three-letter symbol if the character has a very long designation and a more descriptive symbol is needed. In my opinion, however, two letters should be enough for such characters, at least for the near future, because familiarity with symbols develops with time and after use.

For the character itself, recognized by its allelic band pattern, one letter only should preferably be used, independently of number of bands or zones. If two alleles are found in the original investigation of a protein system, the fastest allelic phenotype may be called *F* (fast) and the slowest *S* (slow). If there are multiple bands the migration rates for the fastest bands of the respective patterns determine the relationships between these. Such a nomenclature has been used for the serum-albumin system of chicken[6], horse, cattle, sheep[7] and for man[8]. If new alleles are found there are, with this nomenclature, possibilities for logical and consistent symbols for these. Other letters may, of course, also be used, assuming that the whole alphabet is utilized. For example, in horse[9] and reindeer[10] I called the fastest allelic transferrin band pattern *D* and *C*, respectively (alleles Tf^D, Tf^C). The next patterns were named *F* and *E*, respectively, leaving spaces for eventual new allelic phenotypic patterns. This turned out to be an advantage when new alleles were found in reindeer[11]. I must, however, admit that it would have been even better to have left two letter spaces between the first used symbols. The *FS* nomenclature may also be used for multiple allele systems, *F* being the fastest and *S* the slowest allelic band pattern. Such a system of nomenclature is, of course, not perfect. Independent detections of new alleles may thus cause difficulties. These, however, should have far better chances of being solved satisfactorily than by use of the *A*, *B* nomenclature.

If in a genetic system the letters *A*, *B*, etc., have already been adopted and new alleles with products intermediate to the *A* and *B* are detected, I would prefer to call the new ones *A′*, *A″*, *B′* or *B″*, instead of A_1, A_2, B_1 and B_2. In my opinion, numbers should not be used for the major gene products but spared for eventual designations of sub-groups, single bands or zones. The reason behind this is that the use of continuously running numbers for alleles does not leave spaces for eventual new intermediate products. Furthermore, Greek letters should be spared for eventual designations of polypeptide chains. When mentioning polypeptide chains, we must anticipate complete chemical formulae for proteins in the future. However, these formulae will be so extensive and complicated that short designations will still be needed.

In conclusion, I would emphasize the importance of a simple, descriptive designation, preferably two letters, for a polymorphic protein system. For the allelic phenotype I suggest the use of one letter only, large or small, utilizing the whole alphabet but preferably beginning with *F* and ending with *S*.

Mikael Brænd

Department of Medicine,
Veterinary College of Norway,
Oslo.

[1] Landsteiner, K., *Wien klin. Wschr.*, **14**, 1132 (1901).
[2] Smithies, O., *Nature*, **180**, 1482 (1957).
[3] Smithies, O., and Walker, N. F., *Nature*, **178**, 694 (1956).
[4] Smithies, O., and Hickman, C. G., *Genetics*, **43**, 374 (1958).
[5] Giblett, E. R., Hickman, C. G., and Smithies, O., *Nature*, **183**, 1589 (1959).
[6] McIndoe, W. M., *Nature*, **195**, 353 (1962).
[7] Brænd, M., and Efremov, G., *Proc. Fifth Intern. Cong. Anim. Repr. and Art. Ins.* (in the press).
[8] Efremov, G., and Brænd, M., *Science*, **146**, 1679 (1964).
[9] Brænd, M., and Stormont, C., *Nord. Vet.-Med.*, **16**, 31 (1964).
[10] Brænd, M., *Nature*, **203**, 674 (1964).
[11] Brænd, M., *Hereditas*, **52**, 181 (1964).

Estimation of the Frequency of Functioning Gametes in Monosomics

In monosomics, two types of gametes having *n* and (*n*–1) chromosomes are formed, and the segregation of the three types of plant, 2*n* (disomic), 2*n*–1 (monosomic) and 2*n*–2 (nullisomic), in the progeny of a monosomic is dependent on the frequencies of the *n* and (*n*–1) chromosome gametes on the male and female sides. The functional frequencies of *n* and *n*–1 male and female gametes can be determined experimentally by making reciprocal crosses between a normal plant and a monosomic. Estimates of the frequencies of *n* and *n*–1 gametes on either the male or the female side can be obtained from the observed numbers of disomics, monosomics and nullisomics when the frequencies of gametes on one side are known.

If the relative frequencies of *n* and *n*–1 functioning male gametes are *p* and *q* ($p + q = 1$) and those of female gametes are *p′* and *q′* ($p' + q' = 1$), then in a random mating of gametes, the expected frequencies of disomic,