Proc. Nat. Acad. Sci. USA
Vol. 72, No. 12, pp. 5046–5050, December 1975
Genetics

Structure and regulation of immunoglobulins: Kappa allotypes in the rat have multiple amino-acid differences in the constant region

(complex allotypes/amino-acid sequences/multiple amino-acid substitutions/control gene model)

GEORGE A. GUTMAN*‡, ELWYN LOH†, AND LEROY HOOD†

* Department of Pathology, Stanford University, Stanford, California 94305; and † Division of Biology, California Institute of Technology, Pasadena, Calif. 91125

Communicated by E. B. Lewis, September 15, 1975

ABSTRACT **Immunoglobulin kappa chains from various inbred strains of rats have two serologically detectable forms that segregate in a Mendelian fashion (allotypes *a* and *b* of the *RI-1* locus). Partial amino-acid sequences from the constant regions of these two forms have been compared. Of the 81 residues of the constant region studied, 10 amino-acid substitutions as well as one size difference (sequence gap) were found. This large number of sequence differences among alternative forms of the κ allotype raises provocative questions as to the genetic and evolutionary implications of these light chain allotypes. We designate allotypes whose alternative forms differ at multiple residue positions as complex allotypes. There are basically two genetic models that might explain complex allotypes. First, these allotypes are alleles of a single structural gene with an unusual evolutionary history. Second, all rats have genes that code for each of the light chain allotypes and a control mechanism that permits them to be expressed so that they mimic a Mendelian pattern of segregation. We discuss evidence from other immunoglobulin systems that is compatible with this second model.**

The immune system is one of the most complex physiological systems that has been studied at the molecular, genetic, and cellular levels. The general chemical structure of the immunoglobulin molecule is well understood (1, 2). This molecule is composed of two different polypeptides, light and heavy chains. There are two types of light chains, lambda (λ) and kappa (κ). Each immunoglobulin polypeptide is composed of an NH_2-terminal variable (V) and a COOH-terminal constant (C) region. Some general features are known about the organization of antibody genes (1, 2). Three families (clusters) of genes, unlinked in the mammalian genome, code for the λ, κ, and heavy chain polypeptides. It is generally accepted that the variable and constant regions in these families are coded by separate but closely linked genes. However, very little is known about the genes that regulate the immune response. The immune response genes, linked to the major transplantation locus of the mouse, appear to constitute a family of control genes, unlinked to the structural genes for antibodies, that in some unknown manner regulate the ability of an animal to respond to a variety of different antigens (3). This paper suggests that a new system of control genes may regulate the expression of certain immunoglobulin allotypes.

Inbred rats have a genetic locus, *RI-1*, that controls the expression of two serologically detectable forms of κ light chains, a and b, which segregate as Mendelian codominants (4–6). We report here on the amino-acid sequences of kappa constant (C_κ) regions of the a and b allotypes. The a and b allotypes have multiple amino-acid substitutions, as previously suggested on the basis of differences in peptide maps (7).

Abbreviations: C and V, constant and variable region, respectively, of immunoglobulin polypeptides.

‡ Present address: Walter & Eliza Hall Institute of Medical Research, P.O. Royal Melbourne Hospital, Victoria 3050, Australia.

MATERIALS AND METHODS

Details of the techniques used will be published elsewhere. Briefly, pooled light chains [which are 95% kappa type (8)] were prepared from two rat strains differing at the *RI-1* locus, DA(RI-1^a) and LEW(RI-1^b) (6). Peptide maps of trypsin digests were prepared and the peptides eluted and studied for amino-acid composition and for sequence by manual and automated Edman degradation.

RESULTS AND DISCUSSION

Rat κ Allotypes Exhibit Multiple Amino-Acid Differences. The amino-acid sequences of C_κ regions of normal serum light chains from two inbred strains of rat, DA and LEW, are compared in Fig. 1 with the previously published C_κ sequences of myeloma (Bence-Jones) κ chains from the LOU strain of rat (S211) (9) and the BALB/c strain of mouse (M321) (10). The LEW and LOU strains are of the b serotype, whereas DA belongs to the a serotype. The LEW and DA C_κ regions differ by 10 residues plus one sequence gap, whereas the LEW and S211 C_κ regions differ by only two residues. This is a minimum estimate of the total number of differences for several reasons: (*i*) only 81 of the 108 residues of the C region were compared; (*ii*) the acid and amide forms of aspartic and glutamic acid were not distinguished; and (*iii*) the V regions were not examined.

These allotype-associated differences are distributed in a nonrandom manner. Ten of the 11 differences occur at positions where the LEW sequence differs from that of the mouse, indicating that certain positions are more likely to accumulate changes than others. The fact that the DA sequence is identical to that of the mouse at five of these positions is a puzzling point which may indicate a more rapid accumulation of changes in the LEW gene. Further, the distribution of the substitutions in the tertiary structure of the light chain is not random. Most of the substitutions lie on the external portion of the polypeptide chain (only position 136 is internal) (ref. 11; R. Poljak, personal communication). Two clusters of differences (one including 153 and 155, the other 184, 185, and 188) are external, and both lie very close to one another in a region already known to encompass the serological markers Oz and Inv, and the sequence marker Kern. Since it is known the RI-1 determinants lie exclusively in the C-region (12), it seems likely that one or more of these external substitutions will determine the a and b serological specificities.

The LOU Strain of Rat Appears to Have Two C_κ Genes. Since the LOU and LEW strains of rat are identical at the *RI-1* locus by serological analysis (13), the two sequence differences found between the C_κ regions of the pooled LEW and S211 light chains were surprising. However, the S211 protein seems to be a relatively unusual variant among LOU

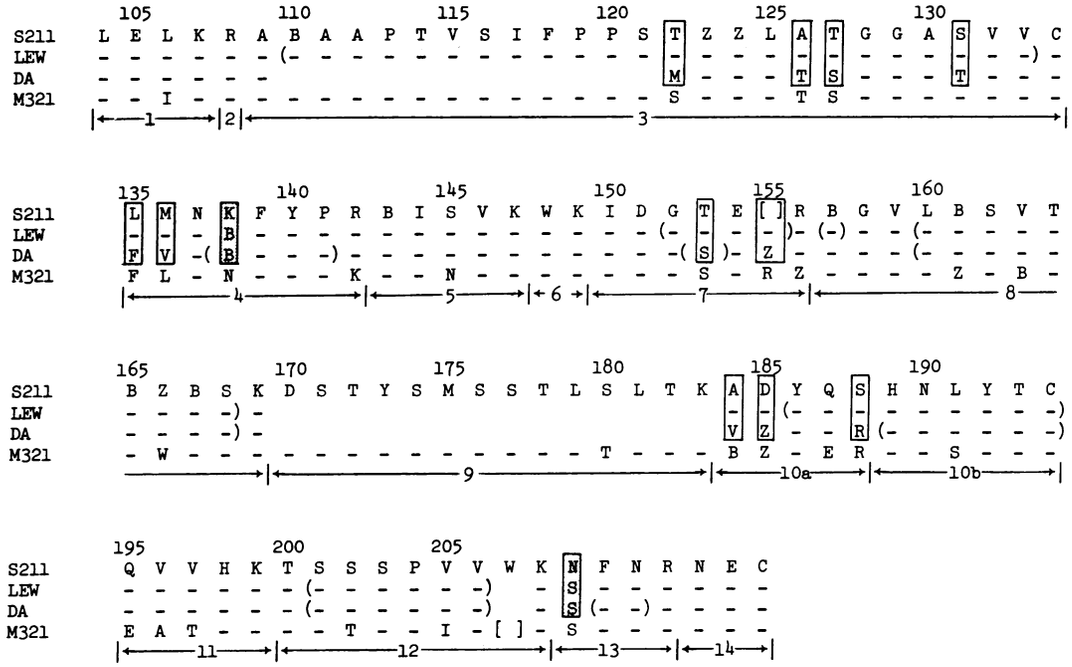

FIG. 1. Amino-acid sequences of rodent κ chain constant regions. S211 is a myeloma κ chain from the LOU rat (11), LEW and DA are sequences from pooled light chains, and M321 is a mouse myeloma κ chain (17). The numbering of peptides is given below the sequences. Differences between DA and LEW are boxed. The two differences between pooled LEW and S211 C_κ regions are boxed and shaded. Dashes indicate an amino acid identical to the S211 sequence. Parentheses indicate that the sequence of the corresponding residues has not been determined. The one letter code of Dayhoff (44) is used for the amino acids.

κ chains, as three other κ myelomas show sequences that are identical at both these positions to the pooled LEW C_κ sequence (P. Querinjean, personal communication). Accordingly, at least two C_κ genes appear to be present in the germ line of the LOU strain of rat. Presumably the same is true of the LEW strain.

Immunoglobulin Allotypes Fall into One of Two Categories, Simple and Complex. It is useful to define two categories of immunoglobulin allotypes, each with very different genetic and evolutionary implications. Alternative forms of *simple allotypes* segregate in a Mendelian fashion in mating studies and differ by one or a few amino-acid substitutions. The InV marker of the human κ chain (14) and the Gm(3) and Gm(17) markers of the human γ1 chain (see ref. 15) are examples of simple allotypes. Simple allotypes are probably coded by alternative alleles at a single structural locus. In contrast, alternative forms of complex allotypes differ by multiple amino-acid residues and *generally* segregate in a Mendelian fashion. The group a allotypes (a1, a2, a3) of the rabbit are V_H markers that differ by multiple amino-acid residues (16, 17). Likewise, the group b allotypes (b4, b5, b6, b9) of the rabbit are C_κ markers that also differ by multiple amino-acid residues (18, 19). Multiple serological specificities have been defined in alternative forms of human γ_3 chains (Gm markers) as well as in certain mouse γ chains. These are designated serologically complex allotypes. If these serological specificities correlate with multiple amino-acid differences, certain human and mouse γ allotypes may represent additional examples of complex allotypes (see ref. 15). Finally, the C_κ regions of the inbred rats described in this paper differ by multiple amino-acid residues. The importance of making a distinction between complex and simple allotypes lies in the very different types of evolutionary or genetic mechanisms they imply. Complex allotypes may be coded by alternative alleles at a single structural locus with an unusual evolutionary history (Fig. 2a and b) or they may result from duplicated genes and the operation of an unusual control mechanism (Fig. 2c). Similar proposals have been made by others (see ref. 15). These three models for complex allotypes will be discussed in subsequent sections using the rat C_κ allotypes as an example.

Complex C_κ Allotypes in the Rat May Evolve by the Divergence of Two Alleles at a Single Genetic Locus. This is the simplest model accounting for the genetics of the RI-1 specificities and it assumes that the two forms have diverged from one another by 10 substitutions and one sequence gap (Fig. 2a). The S211 C_κ gene may represent a very recent duplication in the b strains (LOU and LEW) that may be expressed at low levels in the serum. Two objections can be raised against this model. First, there are a large number of amino-acid differences between the alternative forms. This model assumes that a variant C_κ gene arises and is fixed in the population by natural selection. This new C_κ gene must then incur a second mutation that is once again fixed, and the entire process must be repeated 11 times. Indeed, each new variant gene must be improved in function over its predecessor in order for natural selection to fix (or partially fix) it in the rat population. The question arises as to whether rats as a species have had sufficient evolutionary time for alleles of structural genes to evolve to be so different.

Generally alleles at a single genetic locus are assumed to differ by only one or two residues. Indeed, more than 200 human hemoglobin variants (alleles) have been examined (20). Most differ by one residue, a few differ by two residues, and only one differs by as many as three residues. The allelic forms of the somewhat larger bovine carboxypeptidase A differ by three out of 307 residues (21). Likewise, alternative forms of human κ chains (14), human haptoglobins (22), and a variety of other serum proteins show one or a few amino-acid differences. On the other hand, the "allelic" A

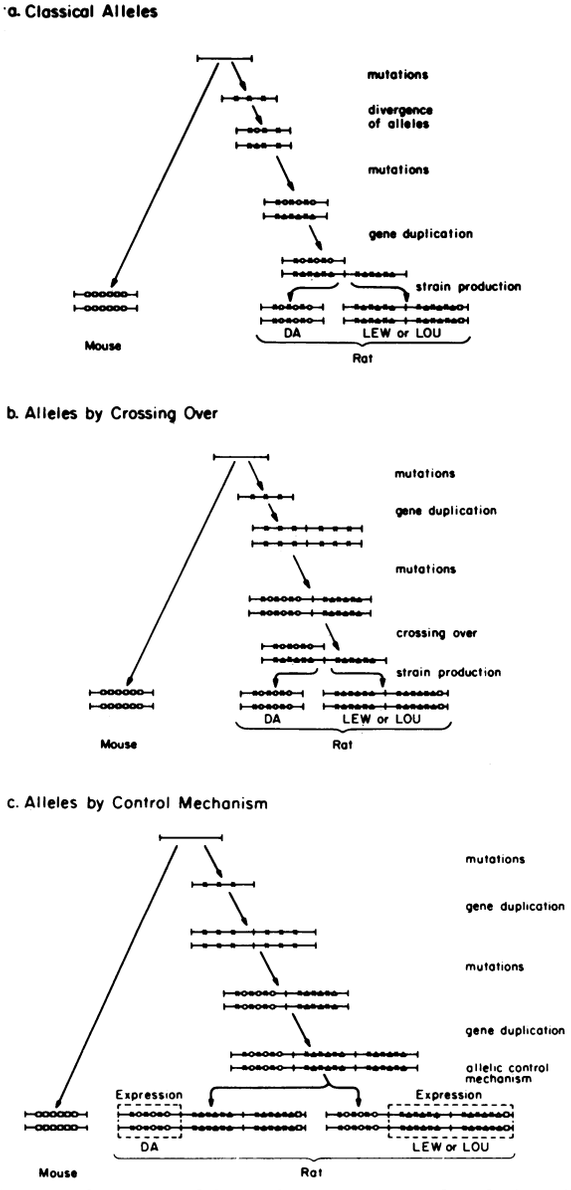

FIG. 2. Genetic models for the expression and evolution of the rat C_κ allotypes. (a) Classical alleles. (b) Alleles by crossing-over. (c) Alleles by gene duplication and a control mechanism. (See *text.*)

and B forms of sheep hemoglobins differ by seven residues out of 145 (23). If the A and B forms are true alleles, which as subsequent discussion will show is often difficult to determine, then alleles with multiple substitutions are possible.

A second objection to this model is the apparent absence (serologically) of any intermediate forms of the rat C_κ region. For example, the white tailed deer has at least seven allelic forms of the β hemoglobin chain that differ by as much as 10% of their amino-acid sequence in the portion of these molecules examined (24). These allelic forms generally express one of two alternative residues at positions that differ. These intermediate forms could arise as (*i*) true intermediates that are maintained in the population between the most extreme alleles and (*ii*) the products of intragenic crossing-over which could scramble the multiply substituted forms. The lack of intermediate forms in κ allotypes of the rat could be explained by hypothesizing that the a and b C_κ gene products have a selective advantage over any of their intermediates.

Complex C_κ Allotypes in the Rat May Evolve by Gene Duplication and Gene Loss Through Crossing-Over. This model suggests that C_κ gene in rat underwent an early gene duplication and that many differences were fixed in these duplicated genes (Fig. 2b). Later in the evolution of the rat line, two unequal but homologous crossing-over events occurred to generate two populations of rats—one (e.g., DA) with a chromosome coding for the *a* form of the C_κ gene and a second (e.g., LEW) with a chromosome coding for the *b* form of the C_κ gene. Once again, the two C_κ forms would be coded by alleles at a single structural locus, but with an evolutionary history that avoids, in part, the need for the highly stringent selective pressures described in the preceding model. For example, one of these C_κ genes may be freed to accept many substitutions while the second is temporarily the functional C_κ gene. Again, the S211 C_κ gene is a recent gene duplication. There are at least two precedents for the evolution via a crossing-over event of alleles that differ by multiple substitutions [β hemoglobin alleles of inbred mice (25) and of Barbary sheep (26)]. This model appears to be a reasonable mechanism for evolving alleles that differ extensively without the necessity of selective pressures to eliminate many intermediate forms. However, intragenic crossing-over might still be expected to scramble these alleles and generate intermediate forms.

Complex C_κ Allotypes in the Rat May Evolve by Gene Duplication and Be Differentially Expressed Via a Control Mechanism. A third model postulates that all rats have genes that code for each of the C_κ allotypes and a control mechanism that permits them to be expressed so that they mimic a Mendelian pattern of genetic segregation (Fig. 2c). Under this model, C_κ gene duplication may have occurred even prior to speciation. Accordingly, complex allotypes could evolve significant differences in their alternative forms. As noted in the previous model, gene duplication can free one gene to accept many substitutions. An important implication of the control gene model is that the serologically detected allotypes follow the inheritance of a control gene(s) that may or may not be closely linked to the corresponding structural (C_κ) genes.

The hypothetical control gene(s) could be operating in a qualitative or quantitative manner. In the latter case small amounts of the "wrong" allotype may be synthesized in homozygous rats. There are precedents for both types of expression in closely linked structural genes. The β-like hemoglobin genes of man (β, δ, γ, and probably ϵ) are closely linked and expressed in a qualitative fashion at different times of development. In the early embryo, the ϵ gene alone is expressed, while later in embryonic life only the γ gene is expressed (27). In the normal adult only the β and δ genes are expressed, at a ratio of about 50 to 1. In addition, an unusual α hemoglobin gene, termed Wazoo, probably present in many primates (e.g., chimpanzee, gorilla), is expressed only in certain individuals of these species (28). Thus, closely linked genes may be expressed in a qualitative manner that varies during development (ϵ, γ), or among individuals (Wazoo), or in a quantitative fashion (adult β and γ). The corresponding control genes may be inherited as Mendelian alleles, as suggested by the inheritance patterns of the ability to express differing ratios of two nonallelic α chains of the stump-tailed macaque (29).

The major objection to the control gene model is its relative complexity. It must mimic with an unknown control

mechanism precisely the same form of expression as the simpler model of allelic structural genes. This is difficult to justify evolutionarily, unless a regulatory mechanism with "allelic" expression has some innate biological advantage. Perhaps it might reflect a strategy to maintain a level of population diversity that may be selectively advantageous in certain environments (30). Two duplicated genes with no control mechanism would result in identical individuals, whereas a control mechanism of the type described in Fig. 2c yields individuals of three kinds (two "homozygotes" and one "heterozygote").

The complex allotypes in rabbits (groups a and b) appear to be coded for by duplicated genes rather than alleles of a single structural locus. Two lines of evidence support this supposition. First, under certain conditions an individual rabbit may express three group a or three group b allotypes (31). If the group a (or group b) allotypes were true alleles, an animal should express at most two alleles. Second, certain rabbits may express low levels of group a allotypes that they should not have according to the genotypes of their parents (32). Two additional examples of serologically complex allotypes may be coded by duplicated genes. A particular congenic strain of mouse, ICR CB-17, has been developed which carries a heavy chain locus homozygous for the C57 Bl/Ka allotype superimposed on a BALB/c background. This strain can express the supposedly absent BALB/c heavy chain allotype under certain conditions (33). This observation implies that the C57 *Bl/Ka* heavy chain locus includes a gene coding for the BALB/c C_H allotype that is ordinarily not expressed. Finally, certain human Gm allotypes not present in donor serum can be found in the supernatant of mixed leukocyte cultures of donor and foreign lymphocytes. Once again this observation implies that humans have C_H genes that are not ordinarily expressed (34). The experiments described above, however, were all carried out using serological methods and will need to be confirmed by structural studies of the "wrong" allotype gene products.

Alternative forms of the complex allotypes of the rabbit, and possibly those of man and mouse, appear to be coded by multiple germ line genes. Complex allotypes may appear on V genes (group *a*, rabbit) or C genes (group *b*, rabbit). In these cases their expression appears to be regulated by a control mechanism that generally causes them to mimic a Mendelian pattern of segregation (Fig. 2c). The obvious parallel between the complex allotypes of the rabbit and the allotypes described in this paper renders the control gene model attractive for the complex κ allotype of rats.

Allelic and Control Gene Models May Be Distinguished by Demonstrating in a Particular Strain "Wrong" Genes or Gene Products. Direct support of the control gene model could be adduced by finding the "wrong" allotype being produced by an inbred rat. A search could be made for the production of a "wrong" allotype in immunologically manipulated situations, as has already been described in rabbits (31), mice (33), and humans (34). Since rat light chains are now fairly well characterized chemically, it will be an easy matter to support serological data with structural studies, even on very small amounts of material. Ultimately, DNA·RNA or DNA·DNA hybridizations of C_κ messenger RNA to somatic or germ cell DNA under very stringent conditions may allow one to determine directly whether the DA and LEW C_κ genes are present in the DNA of all rats.

Rodent C_κ Regions Appear to Be Evolving Rapidly. Mouse and human C_κ regions, which diverged about 75 million years ago, differ by 40 amino-acid residues (35). The rat S211 κ chain differs from its mouse counterpart by 26 substitutions. If mutations were fixed in the rat, human, and mouse evolutionary lines at similar and constant rates, these differences would suggest that the rat and mouse lines diverged more than 40 million years ago [using the naive calculation $(75 \times 26)/40$] rather than 10 million years ago, as is suggested by paleontologic evidence (36). Furthermore, the LEW and DA allotypes are separated by 11 substitutions, suggesting that the divergence of these genes occurred more than 15 million years ago, prior even to the presumed speciation of rat and mouse. There are two possible explanations for these paradoxes. First, the rodent C_κ genes must be diverging considerably more rapidly than their primate counterparts. Recent DNA reassociation studies on primate and rodent single-copy DNA suggest that evolutionary divergence is related to generation time rather than absolute time and, accordingly, rodent genes would be expected to diverge more rapidly than primate genes (37). Second, there is a controversy as to the divergence times of many mammals (38). For example, the mouse and rat evolutionary lines may have diverged from one another much earlier than 10 million years ago. If so, perhaps there is adequate time for the evolution of complex allotypes.

Two types of studies would be valuable in discerning the evolution history of the rat C_κ genes. As examination of the light chain allotypes of wild populations of *Rattus norvegicus* would yield information about the number and range of variation among variants for C_κ chain allotypes. Large sample numbers will be necessary to obtain useful results, however, since it is known that human populations regularly contain rare alleles (with frequencies less than 1%) for many loci (39). The only rat population study reported to date has not demonstrated any new alleles (40). Likewise, the analysis of C_κ regions from other rodents, particularly of the family Cricetidae (lemmings and voles) which is closely related to the Muridae (mice and rats), would answer more general questions relating to the evolution of these light chain genes.

Complex Allotypes of Other Complex Eukaryotic Systems May Also Be Encoded by Duplicated Genes and Expressed by an Unusual Control Mechanism. Serologically complex allotypes have been described in a wide variety of complex eukaryotic systems—the T locus of the mouse (41), the major transplantation locus of mammals (3), the antigens of *Paramecium* (42), the sterility alleles of certain plants (43), etc. Other complex eukaryotic systems may use strategies for the organization, expression, and evolution of genetic information similar to those seen in the vertebrate immune system (1). If so, the presence of complex allotypes may serve as a clue to the presence of multigenic systems with complex regulatory mechanisms.

In summary, alternative forms of complex allotypes may be coded by classical alleles (Fig. 2a), by alleles via gene duplication and crossing-over (Fig. 2b), or by duplicated genes with an unusual regulatory mechanism (Fig. 2c). The latter model can be distinguished from the former two by the presence of the "wrong" genes or gene products in appropriate strains of animals. Two complex allotype systems in the rabbit appear to use the control mechanism model. Perhaps other complex eukaryotic systems will use similar mechanisms for the expression of their information. In any case, the phenotypic expression of complex allotypes raises the possibility that the corresponding genetic system is coded by multiple genes with an unusual control mechanism.

Note Added in Proof. It has come to our attention that W. F. Bod-

mer discussed the concept of complex allotypes in 1973 [W. F. Bodmer (1973) *Transplant. Proc.* V, 1471–1475].

The S211 sequence was kindly provided to us before its publication by Dr. Pierre Querinjean, to whom we are grateful. This work was supported by NSF Grant BMS71-0070 and USPHS Grants AI-10781-04 and AI-09072-06. L.H. has a Research Career Development Award from NIH. This work was carried out while G.A.G. was an NIH Postdoctoral Fellow.

1. Hood, L., Campbell, J. H. & Elgin, S. C. R. (1975) *Annu. Rev. Genet.*, in press.
2. Gally, J. A. & Edelman, G. M. (1972) *Annu. Rev. Genet.* **6,** 1–46.
3. Benacerraf, B. & McDevitt, H. O. (1972) *Science* **175,** 273–279.
4. Rokhlin, O. V., Vengerova, T. I. & Nezlin, R. S. (1971) *Immunochemistry* **8,** 525–538.
5. Armerding, D. (1971) *Eur. J. Immunol.* **1,** 9–45.
6. Gutman, G. A. & Weissman, I. L. (1971) *J. Immunol.* **107,** 1390–1393.
7. Vengerova, T. I., Rokhlin, O. V. & Nezlin, R. S. (1972) *Immunochemistry* **9,** 1239–1245.
8. Hood, L., Gray, W. R., Sanders, B. G. & Dreyer, W. J. (1967) *Cold Spring Harbor Symp. Quant. Biol.* **32,** 133–146.
9. Starace, V. & Querinjean, P. (1975) *J. Immunol.* **115,** 59–62.
10. McKean, D., Potter, M. & Hood, L. (1973) *Biochemistry* **12,** 749–759.
11. Poljak, R., Amzel, L., Avey, H., Chen, B., Phizackerley, R. & Saul, F. (1973) *Proc. Nat. Acad. Sci. USA* **70,** 3305–3310.
12. Nezlin, R. S. Vengerova, T. I., Rokhlin, O. V. & Machulla, H. K. G. (1974) *Immunochemistry* **11,** 517–518.
13. Rohklin, O. V. & Nezlin, R. S. (1974) *Scand. J. Immunol.* **3,** 209–214.
14. Terry, W. D., Hood, L. & Steinberg, A. G. (1969) *Proc. Nat. Acad. Sci. USA* **63,** 71–77.
15. Mage, R., Lieberman, R., Potter, M. & Terry, W. D. (1973) in *The Antigens*, ed. Sela, M. (Academic Press, New York and London), Vol. I., pp. 300–376.
16. Wilkinson, J. M. (1969) *Biochem. J.* **112,** 173–185.
17. Mole, L. E., Jackson, S. A., Porter, R. R. & Wilkinson, J. M. (1971) *Biochem. J.* **124,** 301–318.
18. Appella, E., Rejnek, J. & Reisfeld, R. A. (1969) *J. Mol. Biol.* **41,** 473–477.
19. Goodfleisch, R. (1975) *J. Immunol.* **114,** 910–912.
20. Hunt, L. T., Sochard, M. R. & Dayhoff, M. O. (1972) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (National Biomedical Research Foundation, Silver Spring, Md.), pp. 67–88.
21. Petra, P. H., Bradshaw, R. A., Walsh, K. A. & Neurath, H. (1969) *Biochemistry* **8,** 2762–2768.
22. Black, J. & Dixon, G. H. (1968) *Nature* **218,** 736–741.
23. Boyer, S. H., Hathaway, P., Pascasio, F., Bordley, J. & Orton, C. (1967) *J. Biol. Chem.* **242,** 2211–2232.
24. Taylor, W. J. & Easley, C. W. (1975) *Ann. N.Y. Acad. Sci.* **241,** 594–604.
25. Gilman, J. G. (1972) *Science* **178,** 873–874.
26. Huisman, T. H. J. (1975) *Ann. N.Y. Acad. Sci.* **241,** 549–556.
27. Marks, P. A. & Rifkind, R. A. (1972) *Science* **175,** 955–961.
28. Boyer, S. H., Noyes, A. N., Boyer, M. L. & Marr, K. (1973) *J. Biol. Chem.* **248,** 992–1003.
29. Kitchen, H. (1975) *Ann. N.Y. Acad. Sci.* **241,** 12–24.
30. Johnson, G. B. (1973) *Ann. Rev. Ecol. Syst.* **4,** 93–116.
31. Strosberg, A. D., Hamers-Casterman, C., Van der Loo, W. & Hamers, R. (1974) *J. Immunol.* **113,** 1313–1318.
32. Mudgett, M., Fraser, B. A. & Kindt, T. J. (1975) *J. Exp. Med.* **141,** 1448–1452.
33. Bosma, M. J. & Bosma, G. (1974) *J. Exp. Med.* **139,** 512–527.
34. Rivat, L., Gilbert, D. & Ropartz, C. (1973) *Immunology* **24,** 1041–1049.
35. Barker, W. C., McLaughlin, P. J. & Dayhoff, M. O. (1972) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (National Biomedical Research Foundation, Silver Spring, Md.), pp. 31–40.
36. Simpson, G. G. (1959) *Cold Spring Harbor Symp. Quant. Biol.* **24,** 255–271.
37. Kohne, D. E. (1970) *Qt. Rev. Biophys.* **3,** 327–375.
38. Sarich, V. M. & Wilson, A. C. (1973) *Science* **179,** 1144–1147.
39. Harris, H., Hopkinson, D. A. & Robson, E. B. (1974) *Ann. Hum. Genet.* **37,** 237–253.
40. Rokhlin, O. V. & Nezlin, R. S. (1974) *Scand. J. Immunol.* **3,** 209–214.
41. Gluecksohn-Waelsch, S. & Erickson, R. P. (1970) *Curr. Top. Dev. Biol.* **5,** 281–316.
42. Beale, G. H. (1954) *The Genetics of Paramecium aurelia* (Cambridge University Press, London and New York).
43. Burnet, F. M. (1971) *Nature* **232,** 230–236.
44. Dayhoff, M. O., Hunt, L. T., McLaughlin, P. J. & Barker, W. D. (1972) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (National Biomedical Research Foundation, Silver Spring, Md.), pp. 31–39.