Proc. Nat. Acad. Sci. USA
Vol. 73, No. 4, pp. 1293–1296, April 1976
Immunology

Immunoglobulin allotypes of rabbit kappa chains: Polymorphism of a control mechanism regulating closely linked duplicated genes?

(constant region sequences/multiple amino acid substitutions/complex allotypes)

V. FARNSWORTH*, R. GOODFLIESH*, S. RODKEY†, AND L. HOOD*

* Division of Biology, California Institute of Technology, Pasadena, Calif. 91125; and † Division of Biology, Kansas State University, Manhattan, Kans. 66502

Communicated by Henry G. Kunkel, January 9, 1976

ABSTRACT **The amino acid sequence of the constant (C_κ) region from the kappa immunoglobulin chains of a b9 rabbit is compared with the C_κ sequences, taken from the literature, of a b4 rabbit. These C_κ regions differ by 33% of their amino acid sequences and by three sequence insertions or deletions (sequence gaps). These extensive differences together with other published observations suggest that the b9 and b4 C_κ genes may not be simple alleles, but rather they may be encoded by closely linked C_κ genes present in every rabbit whose expression is regulated by a polymorphic control mechanism.**

Immunoglobulins are comprised of light and heavy chains, each of which is divided into a variable (V) and a constant (C) region (1). The variable and constant regions of a single immunoglobulin molecule are coded by separate germ line genes. Three families of genes which are unlinked in the mammalian genome code for immunoglobulins—two code for light chains (λ and κ) and the third codes for heavy chains.

Studies of rabbit immunoglobulins have made important contributions to our contemporary understanding of the genetics, diversity, evolution, and expression of antibody molecules because of the serological polymorphisms found in light as well as heavy chains and in variable as well as constant regions of such molecules (2). The kappa family of the rabbit has four polymorphic forms which can be detected by serological techniques. These are the group b allotypes of the kappa chain and are designated by the symbols b4, b5, b6, and b9. These polymorphic forms segregate in a Mendelian fashion, implying, according to classical genetic dogma, they are coded by alternative forms of a single structural gene. Preliminary amino acid sequence studies on the constant kappa (C_κ) regions of the group b allotypes have suggested that they are distinguished by multiple residue differences (3–5). We report in this paper the nearly complete amino acid sequence of the C_κ region from a b9 rabbit and compare it with the previously published C_κ region of a b4 rabbit (6, 7). The C_κ regions of the b4 and b9 chains differ by 35% of their amino acid sequences. These extensive sequence differences suggest the C_κ genes of the group b allotypes have had an unusual evolutionary history. The genetic and evolutionary implications of these observations are discussed.

MATERIALS AND METHODS

Source, Purification, and Chain Separation of Immunoglobulins. Pooled immunoglobulin was obtained from homozygous b9 rabbits. The purification of immunoglobulin and the separation of light chains was carried out as previously described (8). This procedure included the separation of the two light chain subtypes (κ_A and κ_B).

Preparation and Sequence Analysis of C_κ Regions. The κ_B chains were completely reduced and alkylated, succinylated, and cleaved with dilute acid as previously described (9). This procedure succinylates the NH_2-terminus, making it unavailable for Edman degradation on the sequenator, and cleaves the kappa chains at the labile aspartic acid-proline bond near the beginning of the C_κ region. The succinylated and acid-cleaved chains were then analyzed for 35 cycles on a Beckman model 890A sequencer using methodology and analytic systems previously described (10).

Preparation of Peptide Fragments. After reduction and alkylation, aliquots of intact b9 chains were subjected to proteolytic cleavage by trypsin (1% wt/wt; 2 hr; 37°; about 10 mg/ml of κ chain in 0.2 M NH_4HCO_3), thermolysin (same conditions as trypsin), and subtilisin (same conditions as trypsin). The tryptic peptides were initially fractionated on Sephadex G-100 (1.0 M NH_4OH) and Sephadex G-50 (0.2 M NH_4CHO_3). Final resolution of the smaller tryptic peptides as well as the thermolysin and subtilisin peptides was obtained by the preparative fingerprint technique (11).

Miscellaneous. Peptides were hydrolyzed *in vacuo* for 20 hr in 5.7 M HCl and analyzed on a Durrum D-500 Amino Acid Analyzer. Large peptides were sequenced directly by the dansyl chloride procedure of Gray (12). Smaller peptides were degraded by the multishot method of Gray (13) and analyzed by subtractive amino acid analysis. The state of the acidic group, for peptides with a single amide or acid group, was examined by thin-layer electrophoresis at pH 6.5.

RESULTS

The amino acid compositions of the subtilisin and thermolysin peptides are given in Tables 1 and 2. The amino acid sequence data and the amino acid compositional data are summarized in Fig. 1. In addition to our data, we have included two pepsin peptides previously published by Goodfliesh (5). The amino acid sequence of the C_κ region of the b9 chains are complete except for the presence of an unidentified residue at position 172, compositional but not sequence data at positions 199 and 200, and the uncertainty of an amide or acid assignment at positions 127 and 203. Details of sequence proof are discussed in the legend to Fig. 1.

DISCUSSION

The C_κ Region of b9 Rabbits Differs in Amino Acid Sequence from that of b4 Rabbits by 35%. The C_κ regions of the b4 and b9 chains are given in Fig. 1. The b9 and b4 chains differ by 33 of 103 amino acid residues and three sequence gaps (positions 109, 141, and 189). Nine out of 33 of these substitutions are separated by two base changes in the

Abbreviations: C_κ, kappa light chain constant region; V_H, heavy chain variable region; V_κ, kappa light chain variable region.

Table 1. Amino acid composition of subtilisin peptides*

	S-2	S-7	P3S-6	S-21	S-24
Asx		1.0(1)	0.8(1)	1.0(1)	1.2(1)
Ser		1.1(1)			1.4(1)
Thr			1.0(1)	1.0(1)	
Glx	1.0(1)	1.0(1)		1.0(1)	1.0(1)
Gly		0.9(1)			
Ala	0.9(1)				
Val				1.1(1)	
Ile		1.0(1)			
Tyr			1.2(1)		0.7(1)
His				0.8(1)	0.8(1)
Lys	1.0(1)				
Cys			+†		

* Values reported are amino acid residues. Amino acids present at a level of less than 0.2 residue are omitted. Values in parentheses represent the nearest integral number of residues.
† Carboxymethyl cysteine was present but not quantitated.

genetic code dictionary. These amino acid substitutions are scattered throughout the entire C_κ region, although there is a particularly heavy concentration of differences in the region from 150 to 161.

A single residue could not be identified in the b9 chains at position 172. In homogeneous b4 chains at least three different residues have been reported at this position (6, 7, 14), indicating either that b4 rabbits have duplicated C_κ genes [similar to the C_γ genes of man (15)] or a genetic polymorphism [similar to the C_κ genes of man (16)]. Our inability to identify this residue in the pooled chains from a single individual suggests that multiple alternatives may also exist at this position in b9 rabbits or that a deletion of this particular residue may have occurred in the C_κ gene of b9 rabbits.

To our knowledge these differences constitute the largest amino acid differences ever reported for two "alleles." This difference is emphasized by the nine two-base substitutions and the three sequence gaps. Indeed, the C_κ regions of mouse and man, two species which diverged about 75 million years ago, differ by only 40% of their amino acid sequence and show no size differences (17). It will be interesting to determine whether the C_κ regions of b5 and b6 chains also show such extensive sequence differences. Let us consider the genetic and evolutionary implications of these extensive differences between two allotypes.

The b4 and b9 Allotypes May Be Explained by One of Three Evolutionary Models. (*i*) The b4 and b9 allotypes may have evolved by the divergence of alleles at a single genetic locus. If so, intense selective pressures are required to fix many substitutions in a relatively short period of evolutionary time. This model appears unlikely for a number of reasons. First, by serological analysis a single heterozygous (b4/b5) rabbit has expressed three group b allotypes (b4, b5, and b6) (18). If amino acid sequence analysis of the group b gene products confirms that this rabbit did express three allotypes, then the group b allotypes cannot be alleles at a single genetic locus (a heterozygous rabbit should only be able to express two alleles). Second, the two-base substitutions at nine of the 33 positions indicate these allotypes are separated by 42 nucleotide substitutions (and three sequence gaps). It is not obvious what intense selective forces might generate these significant sequence differences in allotypes during the evolution of rabbits as a species. (*ii*) The b4 and b9 allotypes may have evolved by gene duplication, mutation divergence, and subsequent crossing-over events to delete the C_κ gene number. In different populations, different combinations of the C_κ genes could remain. Thus, some rabbits may have a single C_κ gene (e.g., b4) and others may have two, three, or even four C_κ genes (e.g., combinations of b4, b5, b6, and b9). Presumably each C_κ gene carried on the two homologous chromosomes would be expressed in individual rabbits. The amino acid polymorphism at position 172 in the b4 chains previously mentioned is consistent with the supposition that b4 (and possibly b9) rabbits have multiple closely linked C_κ genes. If so, the C_κ gene family could readily undergo the gene expansion and gene deletion events postulated by this model. This model is particularly interesting in view of the observations made on two serological markers in rabbit λ chains (c7 and c21). In some populations of rabbits these markers behave as alleles; in others they behave as closely linked duplicated genes (2). This is precisely the pattern the gene duplication and deletion model would predict. In the vast majority of rabbits examined, the group b serotypes behave as classical Mendelian alleles. Accordingly, one

Table 2. Amino acid composition of thermolysin peptides*

	R1-2	R1-5	R1-6	R1-7	T4-2†	T4-5	T4-6	T4-11‡	T5-1	T3-7	T3-9	T2-4	T2-8	T2-11
Asx				1.0(1)	(1)	1.0(1)	1.2(1)		1.0(1)				1.1(1)	
Thr	0.9(1)	1.1(1)	3.0(3)								1.1(1)	1.1(1)		
Ser		1.9(2)		0.9(1)		2.2(2)	2.1(2)	1.4(1)		1.0(1)				
Glx	1.0(1)		1.1(1)	1.1(1)		1.3(1)		1.3(1)						
Pro				2.1(2)										
Gly							0.9(1)		1.0(1)					
Ala				0.9(1)		0.8(1)	0.8(1)						1.1(1)	
Val							0.9(1)	0.6(1)				0.9(1)	0.9(1)	1.0(1)
Ile								0.7(1)						1.0(1)
Leu		0.9(1)	0.9(1)	0.9(1)						1.0(1)	1.0(1)			
Tyr	1.0(1)					0.8(1)								
Phe				1.0(1)	(1)									
His						0.9(1)	1.1(1)							
Lys											0.9(1)		1.0(1)	
Cys	+								+					+
Arg					(1)									

* See legend to Table 1.
† The composition was determined qualitatively by high voltage paper electrophoresis.
‡ The sequence Iso-Val is hydrolyzed very slowly.

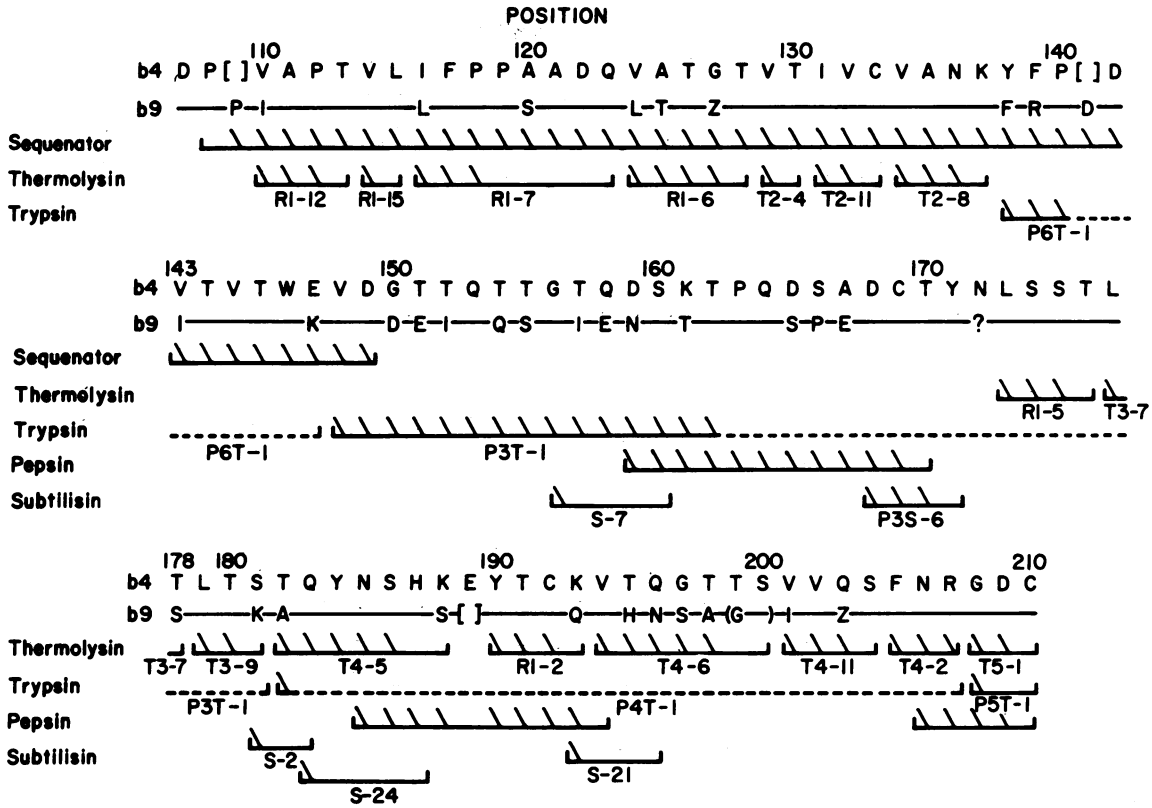

FIG. 1. The amino acid sequence of the C_κ regions isolated, respectively, from rabbits of b4 and b9 allotype. The b9 sequence is indicated as a straight line where it is identical to the b4 sequence. Substitutions are indicated by letters. The b4 sequence is taken from ref. 7. The pepsin peptides were taken from ref. 5. The one letter code of Dayhoff is used (25). Brackets indicate residue deletions. Parentheses indicate that the sequence of the enclosed residues is unknown. Sequenator indicates sequence determined by automatic analysis on the half chain fragment. A preliminary report of the NH_2-terminal portion of this fragment, differing at a number of residues, has been published by others (26). Peptide compositions are indicated by straight lines terminated by vertical bars. Slanted bars on the peptide fragments indicate that the sequences of the corresponding residues were determined by the manual Edman-dansyl technique. Dotted lines on the tryptic peptides indicate that quantitative amino acid compositional data were not obtained.

must postulate that rabbits with two or more of the group b C_κ genes on a single chromosome are rare. (*iii*) The b4 and b9 allotypes may have evolved by gene duplication with a control mechanism that permits them to be expressed so they mimic a Mendelian pattern of genetic segregation. This model suggests that the rabbit chromosome which encodes the κ family has four closely linked C_κ genes (i.e., b4, b5, b6, and b9). The extensive amino acid C_κ differences reported in this paper are consistent with this hypothesis (e.g., consider the 19% difference between the closely linked δ and γ chains of human hemoglobins). The genetic polymorphism may reflect a control mechanism which operates at the chromosomal level to determine which of the four C_κ genes is to be expressed. This theory postulates that ordinarily this commitment (to the expression of one C_κ gene) is transmitted through germ line and is very stable. Rarely, this control mechanism may fail and permit a "wrong" allotype to be expressed. Accordingly, this model is also consistent with the observation that a single rabbit can produce three group b allotypes (18).

The allelic model would be ruled out as an explanation for the b4 and b9 allotypes if it is unequivocally demonstrated (e.g., by structural analysis) that a rabbit can produce allotypes that it could not have inherited from its parents (i.e., "wrong" allotypes). The control model could be verified if it could be demonstrated (e.g., by nucleic acid hybridization) that every rabbit has C_κ genes for all four group b allotypes. The gene duplication and deletion model would predict that different rabbit chromosomes could have different numbers of C_κ genes. In this regard, it will be particularly interesting to examine progeny from the single rabbit that expressed three group b allotypes to determine whether they inherit the ability to express "wrong" allotypes.

The V Genes of the Rabbit May Be Coded by Multiple, Closely Linked Genes whose Expression is Regulated by a Control Mechanism. Three groups of variable (V_H) regions of rabbit heavy chains can be distinguished by serological techniques and are found in every rabbit (groups a, x, and y, see ref. 2). The V_H genes coding for these variable regions are closely linked. Three allotypes (a1, a2, a3) are found in the group a chains which generally segregate in a Mendelian fashion, but they differ in amino acid sequence from one another by multiple amino acid residues (19). Two recent experiments suggest that individual rabbits may express "wrong" allotypes. First, the same heterozygous rabbit that expressed three C_κ allotypes also expressed three V_H allotypes (18). Second, as many as 50% of normal rabbits express low levels of a group a allotype that should not be present according to breeding data (20). Once again it is extremely important that these "wrong" allotypes be verified by structural analysis because of the possibility of serological cross-reactions, particularly in the latter case where very low levels of "wrong" allotypes are detected. If these observations are verified, the possibility that the group a allotypes

are coded by alleles of a single structural gene would be eliminated. Furthermore, the control gene model would be favored over the duplication and deletion model for two reasons. First, if chromosomes with two (or three) allotypes are rare as is suggested for the duplication and deletion model by the breeding data, then it appears very unlikely that the first rabbit to express a "wrong" allotype would express "wrong" allotypes for *both* the C_κ and V_H regions. Second, the expression of low levels of the "wrong" allotype in a high percentage of rabbits suggests (*1*) that there is a control mechanism for quantitatively expressing high or low levels of V_H genes, and (2) that the a1, a2, and a3 V_H genes are present in all rabbits as closely linked genes. Accordingly, the V_H gene family is multigenic and the expression of subsets of V_H genes may be regulated by a control mechanism.

Let us now consider the variable kappa (V_κ) genes of the rabbit. If the C_κ allotypes are indeed duplicated genes, it is interesting to note that each group b allotype may have a unique set of V_κ genes that are expressed with it. Careful quantitative analysis of the NH_2-termini of pooled κ chains isolated from b4, b5, b6, and b9 individuals suggests that there are small but reproducible sequence differences that correlate with the rabbit's allotype (8, 21). More recently, major residue differences have been found at certain positions in the V_κ regions of b4 and b9 rabbits (22). These observations suggest that each C_κ gene selectively is expressed with at least a subset of distinct V_κ genes. The implications of the control gene model presented above are that every rabbit has all the V_κ genes of the species and that these V_κ genes will be expressed in accordance with the type of C_κ genes being expressed. Whether the special association of V_κ and C_κ genes is due to unknown selective pressures on the intact molecule, the mechanistic constraints imposed by the variable-constant joining mechanism, or a special regulatory mechanism of some other type is unknown.

The Groups a and b Allotypes Are Examples of Complex Allotypes. We have recently suggested that alleles (or allotypes) can be divided into two categories (23). Alternative forms of *simple allotypes* segregate in a Mendelian fashion in mating studies and differ by one or a few amino acid substitutions [e.g., the Inv marker of the human κ chain (16)]. In contrast, alternative forms of *complex allotypes* differ by multiple amino acid residues and *generally* segregate in a Mendelian fashion. Allotypes that differ by multiple serological determinants can also be tentatively designated as complex. By these definitions, the groups a and b allotypes fall into the complex category. The importance of the distinction between simple and complex allotypes lies in the very different types of genetic or evolutionary models implied by each (18). Simple allotypes are probably coded by alleles at a single structural locus. In contrast, complex allotypes may be explained by a number of different genetic models, as described above. Complex allotypes are found in a variety of immunoglobulin families and, indeed, in a variety of other complex eukaryotic systems (see ref. 23 for a review). Thus, the presence of complex allotypes in any system raises the possibility that the corresponding alleles may be closely linked duplicated genes regulated by a polymorphic control mechanism.

In summary, the C_κ as well as the V_H genes of the rabbit differ by multiple amino acid residues and sequence gaps. Accordingly, they constitute examples of a complex allotype system. Complex allotypes may have arisen by (*i*) the rapid evolution of alleles of a single gene, (*ii*) gene duplication, divergence, and deletion, or (*iii*) by gene duplication with regulation by a polymorphic control mechanism. Current data are perhaps most consistent with model *iii*. Many multigenic families (immunoglobulins and other eukaryotic systems) exhibit complex allotypes. It will be fascinating to determine whether most complex allotypes are coded by duplicated genes with an unusual control mechanism and, if so, what the evolutionary advantages of such a system might be (see refs. 1, 23, and 24 for a general discussion of this issue).

This work was carried out under grants from the National Science Foundation and the National Institutes of Health. R.G. has an NIH Postdoctoral Fellowship. L.H. has an NIH Research Career Development Award. This paper is contribution 1267, Division of Biology, Kansas Agricultural Experiment Station, Kansas State University, Manhattan, Kans. 66506.

1. Hood, L., Campbell, J. H. & Elgin, S. C. R. (1975) *Annu. Rev. Genet.* **9**, 305–353.
2. Mage, R., Lieberman, R., Potter, M. & Terry, W. D. (1973) in *The Antigens*, ed. Sela, M. (Academic Press, New York and London), pp. 300–376.
3. Appella, E., Rejnek, J. & Reisfeld, R. A. (1969) *J. Mol. Biol.* **41**, 473–477.
4. Frangioni, B. (1969) *FEBS Lett.* **3**, 339–341.
5. Goodfliesh, R. (1975) *J. Immunol.* **114**, 910–912.
6. Appella, E., Roholt, O. A., Chersi, A., Radzimski, G. & Pressman, D. (1974) *Biochem. Biophys. Res. Commun.* **53**, 1122–1129.
7. Chen, K. C. S., Kindt, T. J. & Krause, R. M. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 1995–1998.
8. Waterfield, M., Morris, J. E., Hood, L. E. & Todd, C. W. (1973) *J. Immunol.* **110**, 227–232.
9. Paulsen, K., Fraser, K. J. & Haber, E. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 2495–2499.
10. Hood, L., McKean, D., Farnsworth, V. & Potter, M. (1973) *Biochemistry* **12**, 741–749.
11. Bennett, J. C. (1967) in *Methods in Enzymology*, ed. Hirs, C. H. W. (Academic Press, New York and London), Vol. XI, pp. 330–339.
12. Gray, W. R. (1972) in *Methods in Enzymology*, eds. Hirs, C. H. W. & Timasheff, S. N. (Academic Press, New York and London), Vol. XXV, pp. 333–344.
13. Gray, W. R. (1967) in *Methods in Enzymology*, ed. Hirs, C. H. W. (Academic Press, New York and London), Vol. XI, pp. 469–475.
14. Haber, E., Margolies, M. N., Cannon, L. E. & Rosenblatt, M. S. (1975) in *Molecular Approaches to Immunology*, eds. Smith, E. E. & Ribbons, D. W. (Academic Press, New York), pp. 303–333.
15. Natvig, J. B. & Kunkel, H. G. (1973) *Adv. Immunol.* **16**, 1–59.
16. Terry, W. D., Hood, L. & Steinberg, A. (1969) *Proc. Nat. Acad. Sci. USA* **63**, 71–77.
17. Dreyer, W. J., Gray, W. R. & Hood, L. (1967) *Cold Spring Harbor Symp. Quant. Biol.* **32**, 133–146.
18. Strosberg, A. D., Hamer-Casterman, C., van der Loo, W. & Homers, R. (1974) *J. Immunol.* **113**, 1313–1318.
19. Mole, L. E., Jackson, S. A., Porter, P. R. & Wilkinson, J. M. (1971) *Biochem. J.* **124**, 301–318.
20. Mudgett, M., Fraser, B. A. & Kindt, T. J. (1975) *J. Exp. Med.* **141**, 1448–1452.
21. Chersi, A., Mage, R. G. & Appella, E. (1971) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **30**, 1307, abstr. 1485.
22. Thunberg, A. L. & Kindt, T. J. (1975) *Scand. J. Immunol.* **4**, 197–201.
23. Gutman, G., Loh, E. & Hood, L. (1975) *Proc. Nat. Acad. Sci. USA* **72**, 5046–5050.
24. Bodmer, W. F. (1973) *Transpl. Proc.* **5**, 1471–1475.
25. Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Silver Spring, Md.).
26. Zeeuws, R. & Strosberg, A. D. (1975) *Arch. Int. Physiol. Biochim.* **83**, 205–206.