Volume 12 Number 11 1984 Nucleic Acids Research

Complex allotypes of the rabbit immunoglobulin kappa light chains are encoded by structural alleles

Marie-Andrée Akimenko, Odile Heidmann and François Rougeon

Unité de Génétique et Biochimie du Développement, Institut Pasteur, ERA CNRS 851, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France

Received 2 March 1984; Revised and Accepted 8 May 1984

ABSTRACT

We have isolated the rabbit immunoglobulin b9 Ck light chain gene and compared its nucleotide sequence with the b4, b4var, b5 and bas Ck sequences. In spite of the high number of substitutions found between the different rabbit Ck coding regions, only very few changes are silent. Furthermore, the nucleotide changes are clustered in segments which correlate with the bends and helical regions found in the tertiary structure of the Ck domain of the protein. The flanking regions present a higher degree of conservation than the coding regions. The two genomic EcoRI fragments hybridizing to a b4cDNA probe have been correlated with the two distinct loci, Ck1 and Ck2 : one encodes for the nominal b9 Ck allotype and the other contains the information for the bas Ck region. The b allotypes are true alleles which could have evolved by intergenic conversion.

INTRODUCTION

Genetic and molecular studies have shown that in the human (1,2) and mouse genome (3), the constant region of the immunoglobulin k light chain is encoded by only one Ck gene. Allelic variants of the Ck gene, differing by few substitutions have been described in the human species. Such variants, leading to antigenic differences are called allotypes (4,5). In contrast to this situation, the rabbit shows an interesting complexity. Although the genetic markers of the b series characterizing the rabbit Ck light chain are transmitted in a Mendelian fashion, some findings have called the allelism of these allotypes in question and led to postulate the presence of multiple Ck genes in the rabbit genome.

First, the four major allotypic forms b4, b5, b6 and b9 encountered in the domestic rabbit, exhibit an abnormaly high degree of divergence at the protein level (b4 and b9 constant regions diverge by 33% in their amino acid sequence).

Secondly, Kelus and Weiss (6) have established a new strain, called Basilea, which in contrast to the other rabbit strains, mainly expressed λ-type light chain. The k light chains named k bas, synthetized by this strain, are characterized by the lack of b series determinants. Chemical and serological studies made on several wild rabbits families (7) have shown that individuals homozygous or heterozygous for allotypes of the b series could also express a kbas allotype. This result suggests that the gene encoding the kbas chain is not an allele at the b locus.

Finally, genomic blot analysis have shown that a b4cDNA probe could detect multiple related sequences in homozygous rabbits for b4, b5, b6 or b9 allotypes (8). Sequences coding for b4var, a b4 allelic variant (9) and bas Ck genes have been isolated from a b4b4var rabbit genomic library (10,11). The presence of a restriction fragment bearing the bas sequence in the genome of b4, b5, b6 and b9 homozygous rabbits added to the serological and genetic data supports the hypothesis that bas protein is a second isotypic form of the Ck locus. The expression of the rabbit k light chains is therefore governed by two loci : the Ck1 locus encoding light chains characterized by the b series allotypes and the Ck2 locus bearing the information for the bas protein.

In order to compare the genetic content of b4b4var rabbits with that of rabbits homozygous for another b allotype, we have undertaken to characterize the Ck related sequences detected in a b9b9 rabbit by genomic blot analysis.

b9b9 genomic DNA digested with EcoRI was analysed by the Southern blot technique (12). When the blot is hybridized with a b4cDNA probe and washed under conditions of low stringency, two intense bands and some weaker bands are revealed (data not shown). The 16,5 kb major EcoRI band corresponds to the fragment containing the bas Ck gene previously characterized in the b4b4 var rabbit (11). To identify the second major band, 11,5 kb, we have cloned EcoRI DNA fragments of the appropriate size of a rabbit in lambda phage vector and screened the obtained clones with the b4cDNA probe. This procedure has allowed us to isolate a recombinant phage, λ 104, containing the nominal b9 Ck gene.

In this paper, we report the nucleotide sequence of the co-

ding and flanking regions of the b9 Ck gene. We compare this sequence with the published sequences of b4, b4var, b5 or bas Ck genes.

MATERIALS AND METHODS

Molecular cloning of the 11,5 kb EcoRI fragment

High molecular weight liver DNA of a homozygous b9 rabbit, prepared as previously described (8) was digested to completion with EcoRI restriction endonuclease, then size fractionated by centrifugation on a 5-40% sucrose gradient. DNA fragments ranging from 10 to 20 kb long were selected. Similarly, λL47-1 vector DNA (13) ligated to itself was digested with EcoRI enzyme. Isolation of the ligated arms from the 10 kb stuffer fragment was performed by centrifugation on a 5-40% sucrose gradient. A two fold molar excess of λL47-1 arms to rabbit DNA fragments were ligated. Recombinant DNA was packaged in vitro (14) and propagated in a E.coli host with an efficiency of 3.10^8 pfu/µg of total DNA. Recombinant phages were screened according to the in situ plaque hybridization procedure (15) using the b4cDNA probe (8), ^{32}P labelled by nick translation (16).

Subcloning and DNA sequencing

After Southern blot analysis, a 4,1 kb BamHI restriction fragment was purified by electroelution from a 0,7% agarose gel and inserted into pBR322 vector. Recombinant molecules were used to transform HB101 E.coli. Colonies were screened by hybridization with the b4cDNA probe. Restriction fragments DNA of this subclone were labelled at the 5' end with $\gamma^{32}P$-ATP in an exchange reaction with the T4 polynucleotide kinase and at the 3' end with either the 3'-$\alpha^{32}P$ dATP and the terminal transferase or by filling in protruding sites with E.coli DNA polymerase I large fragment. The nucleotide sequence was determined by the chemical method of Maxam and Gilbert (17).

RESULTS AND DISCUSSION

Construction of a b9b9 rabbit library and isolation of a Ck gene

High molecular weight liver DNA isolated from a homozygous b9b9 rabbit was digested to completion with the restriction

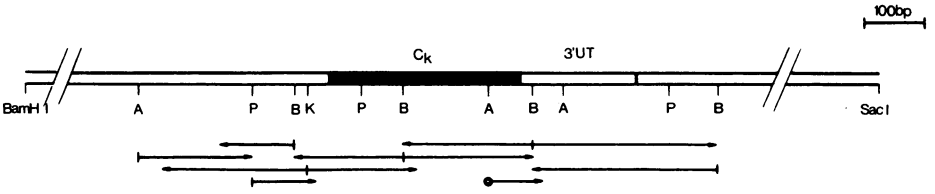


Figure 1 : Strategy used for sequencing the b9 CK gene contained in a 1,4 kb BamHI-SacI restriction fragment. The coding and flanking regions of the gene were sequenced by the technique of Maxam and Gilbert using the restriction sites indicated : A = AvaII ; P = PvuII ; B = BstnI ; K = KpnI. The solid area represents the coding region, the 3' untranslated region is limited by a vertical line. Arrows below the map indicate the direction and the extent of sequencing. All the fragments were labelled at their 3' end except the fragment marked by a circle which was labelled at its 5' end.

endonuclease EcoRI and fractionated on a sucrose gradient. Fractions containing fragments between 10 and 20 kb were ligated into the EcoRI sites of λL47-1 vector arms (13) and after encapsidation propagated into E.coli host. Using the b4cDNA probe to screen the library, we isolated a recombinant phage, λ 104,containing the 11,5 kb EcoRI fragment. To further characterize this fragment, a 4,1 kb BamHI fragment was isolated and subcloned into the vector pBR322. The restriction map of a 1 270 bp BamHI-SacI fragment of this subclone and strategy used to sequence the unique Ck gene detected in this clone are presented in Fig. 1. The 817 bp nucleotide sequence shown in Fig. 2 was determined according to Maxam and Gilbert method (17).

Characterization of the Ck gene

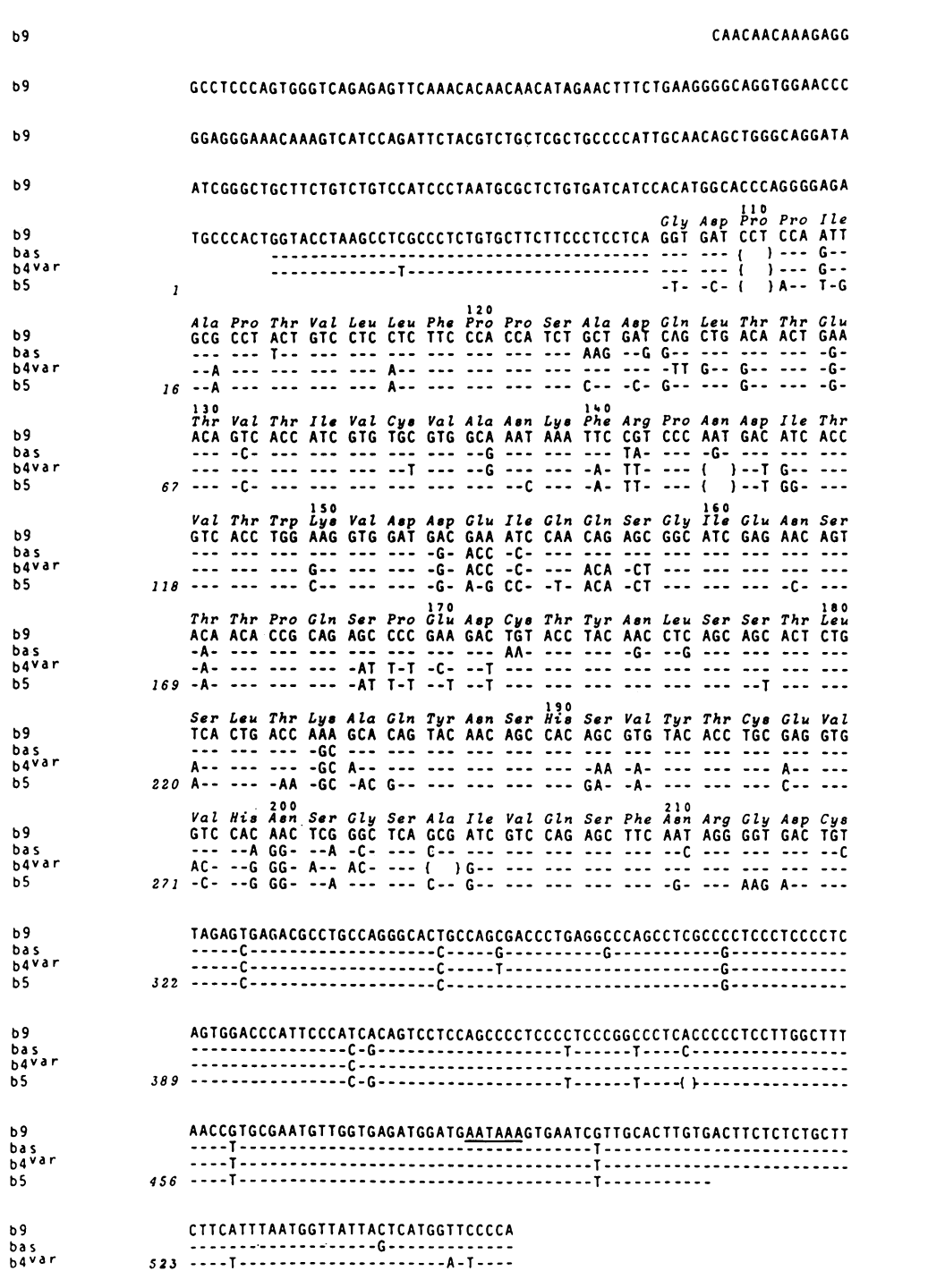
This sequence contains an open reading frame of 321 nucleotides. The 5' end of the sequence contains the AG dinucleotide splice site acceptor needed for the J-C junction event. The end of the coding sequence is marked by the termination codon TAG at nucleotide positions 322-324. As in other eukaryotic genes, the AATAAA polyadenylation signal (18) is present at nucleotides 484-489, followed by the presumed polyadenylation site 17 bp downstream. The amino acid sequence deduced from the nucleotide sequence, allowed us to identify the cloned sequence as the b9Ck

gene. However, at some positions, the deduced amino acid sequence diverges from the b9 amino acid sequence published by Farnsworth et al. (19). Point substitutions are found at positions 143 and 196 where we read, respectively, codons for an asparagine and a glutamic acid instead of an aspartic acid and a glutamine in the published amino acid sequence. Although we cannot exclude the possibility of a polymorphism at these positions, the observed differences are more likely due to artefacts in the amino acid sequence determination. On the other hand, at positions 192 and 197 we have found two codons for Valine residues, which do not appear in the published amino acid sequence, despite the use of different enzymatic cleavages. Finally, we can assign to the unknown positions 202-203-204, the amino acid sequence Gly-Ser-Ala.

Comparison of the Ck sequences

In order to study the relationship between the different allotypic forms of the b series, it was interesting to compare the b9Ck genomic nucleotide sequence with the other Ck nucleotide sequences so far determined.

Therefore, we have aligned the known genomic nucleotide sequences containing basCk (11) and b4varCk (10) genes with their flanking regions and the b5cDNA sequence (20) with the b9Ck nucleotide sequence. If gaps, which have been introduced to maximize homology are scored as single differences (because they presumably arise from single events), the coding sequence of b9Ck is 79%, 83,5% and 89,7% homologous to, respectively, b5Ck, b4varCk and basCk nucleotide sequences. At the protein level, b9Ck sequence is only 58,5%, 67,9% and 80,2% homologous to, respectively, b5Ck, b4varCk and basCk molecules. The percent of unchanged codons (i.e. 100 x (number of non mutated codons found between two nucleotide sequences)/number of codons compared in the two sequences) calculated between b9 and respectively, b5Ck, b4varCk and basCk sequences are : 52,8%, 63,2% and 75,5%. The comparison of these numbers with the respective percent of homology at the protein level reveals a small difference which reflects the low number of silent substitutions in comparison with the replacement changes. The differences between the b9, b4var, b5 and basCk coding sequences are clustered in 6 regions :



43-55 ; 98-112 ; 137-153 ; 182-194 ; 226-237 ; 271-292. Four of these divergent segments are similar to those already noted in b4var and basCk sequences comparison (11). The beginning of the coding regions of the Ck genes are also highly divergent. The limited number of silent substitutions and the clustering of the differences indicate that the substitutions are not randomly distributed along the Ck coding sequences. Based on X-ray crystallographic studies and amino acid sequences data, tertiary structure models have been established for each immunoglobulin domains (VL, CL ...) (21,22). It appears that all the regions present a similar chain folding composed of seven linear antiparallel β-pleated segments separated by bends or helical segments. The alignment of the different rabbit Ck amino acid sequences with the domains of human Fab' (New) in terms of crystallographic models reveals that the regions conserved at the nucleotide level code for the β-pleated segments of the protein, while the most divergent ones correspond to the bends and the ends of the Ck domain. The conservation of the three-dimensional structure of the immunoglobulin domains probably creates strong selective constraints that reduce the possibilities of variation at the coding sequence level. Some residues in the β-pleated segments are conserved in all the immunoglobulin constant domains, even in species as different as human, mouse, rabbit and Guinea pig (21). These highly conserved residues constitute by their hydrophobic or non polar nature, the domain interior (21,23).

Surprisingly, the analysis of the 3'UT and the 3' and 5' flanking regions of the different rabbit allotypic Ck sequences show that they are more conserved than their coding regions. Table 1 indicates the percent of homology observed between the

Figure 2 : Comparison of the rabbit CK sequences.
The genomic nucleotide sequences of bas (11) and b4 var (10) and the b5 cDNA sequence (20) were aligned to maximize the homology with b9 CK sequence. Dashes indicate nucleotide identity with the b9 CK sequence and deletions are mentioned by brackets. The polyadenylation signal is underlined. The b9 nucleotide sequence is numbered from the first base of the Gly codon 108. The amino acid sequence deduced is shown above the b9 CK nucleotide sequence.

Table 1 : Degree of homology between the nucleotide sequences of the rabbit CK genes.

% homology	b_4	b_5	b_9	bas
b_4		ND 96,8	93,4 95,7	97,4 96,3
b_5	85,7		ND 94,6	ND 98,4
b_9	83,8	79		100 93,5
bas	84,6	76,4	89,7	

5' flanking region

3' UT region

Cκ coding sequence

Length of the 5' sequence determined in bp

b_4	227
b_5	ND
b_9	262
bas	39

The percent homology was separately calculated for the coding and flanking regions, between each allotypic form in this manner : 100 x (number of homologous bases/number of bases compared). Each gap is scored as a single difference.

The lower half of the table indicates the percent homology between the coding sequences and the upper half of the table presents the percent homology obtained between, first, the 5' flanking regions and then between the 3' untranslated regions. The lengths of the 5' flanking sequences compared are mentioned on the right of the table (ND = not determined). The sequences compared were taken from Emorine et al. (25) : b4 ; Bernstein et al. (20) : b5 ; Heidmann and Rougeon (11) : bas and this paper : b9.

coding and flanking regions of the different Ck allotypic forms. The b9 3'UT region is 185 nucleotides long and is 95,7%, 94,6% and 93,5% homologous to, respectively, the b4var, b5 and bas 3'UT regions. This high degree of homology also exists in the

3' flanking region for b9, b4var and bas Ck genes.

The 5' flanking region is the most conserved region of the three genomic sequences analysed. The 39 nucleotides identified in the 5' bas Ck region are identical to those found in the b9 flanking region and the sequence of the 5' b4varCk region differs by only a single transition T-C. The sequence of the 5' region of b4Ck reported by Emorine et al. (25) also shows a strong homology with the b9 sequence : of the 227 nucleotides presented, aside four substitutions, five single deletions and two single insertions, a deletion of eleven nucleotides is observed, with three C nucleotides which do not match at one of its ends with the 5' b9Ck region compared. Two short direct repeats GGGC are observed in b9 sequence near the deletion. One copy of these repeats is present in the b4 sequence. It has been postulated that such short direct repeats could be implicated in the deletion mechanism (26). Despite all of these differences between the 5' b9Ck and the 5' b4Ck regions, the two sequences are 93,4% homologous.

Another interesting feature, provided by the comparison of the b9, b4var, bas and b5 Ck coding sequences and their 3'UT regions (a total of 508 nucleotides) is the very small number of positions where three of the four sequences have different bases. Nine of these positions are located in the coding sequences and when they are not deletions, they lead to a substitution in the amino acid sequences. Two other positions are located in the 3'UT regions. On the other positions, if the four sequences are not identical, there are only two alternative nucleotides.

The high degree of conservation of the Ck flanking regions supports the assumption that the b4, b9 and b5 allotypic forms are true alleles. The accelerated evolution observed in their respective coding sequences might have been generated by a non reciprocal intergenic conversion of the allelic forms. The paucity of silent substitutions and the segmental homology observed within the coding regions support this hypothesis. Only the eleven positions previously described could not result from a conversion mechanism. Nevertheless, from these results and those previously reported, it appears that a rabbit phenotypically

homozygous for a given allotype does not contain the genes coding for the other allotypic forms so far characterized.

ACKNOWLEDGEMENTS

We are deeply indebted to B. Mariamé and M.C. Wetzel for helpful discussions and to Drs. G. Langsley and C. Roth for critical reading of the manuscript. The expert assistance of I. Collet for typing this manuscript is gratefully acknowledged.

This work was supported by grants from the C.N.R.S. (A.T.P. n° A1-5054, C.P. n° 960008), the Ministère de la Recherche et de l'Industrie (L.P. n° 82.L.1311) and the Fondation pour la Recherche Médicale Française.

REFERENCES

1- Terry, W.D., Hood, L.E. and Steinberg, A.G. (1969) Proc. Natl. Acad. Sci. USA 63, 71-77.
2- Hieter, P.A., Max, E.E., Seidman, J.G., Maizel, J.V. and Leder, P. (1980) Cell 22, 197-207.
3- Seidman, J.G. and Leder, P. (1978) Nature 276, 790-795.
4- Oudin, J. (1956) C.R. Acad. Sci. (Paris) 242, 2489-2490.
5- Oudin, J. (1956) C.R. Acad. Sci. (Paris) 242, 2606-2608.
6- Kelus, A.S. and Weiss, S. (1977) Nature 265, 156-158.
7- Benammar, A. and Cazenave, P.A. (1982) J. Exp. Med. 156, 585-595.
8- Heidmann, O. and Rougeon, F. (1982) Cell 28, 507-513.
9- Sogn, J.A. and Kindt, T.J. (1976) J. Exp. Med. 143, 1475-1482.
10- Heidmann, O. and Rougeon, F. (1983) Cell 34, 767-777.
11- Heidmann, O. and Rougeon, F. (1983) EMBO J. 2, 437-441.
12- Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
13- Loenen, W.A.M. and Brammar, W.J. (1980) Gene 20, 249-259.
14- Grosveld, F.G., Dahl, H.H.M., de Boer, E. and Flavell, R.A. (1981) Gene 13, 227-237.
15- Benton, W.D. and Davies, R.W. (1977) Science 196, 180-182.
16- Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
17- Maxam, A.M. and Gilbert, W. (1980) Meth. in Enzymology 65, 499-560.
18- Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
19- Farnsworth, V., Goodfliesh, R., Rodkey, S. and Wood, L. (1976) Proc. Natl. Acad. Sci. USA 73, 1293-1296.
20- Bernstein, K.E., Skurla, R.M. and Mage, R.G. (1983) Nucl. Acids Res. 11, 7205-7214.
21- Beale, D. and Feinstein, A. (1976) Quaterly Reviews of Biophysics 9, 135-180.
22- Amzel, L.M. and Poljak, R.J. (1979) Ann. Rev. Biochem. 48, 961-997.

23- Poljak, R.J., Amzel, L.M., Chen, B.L., Phizackerley, R.P. and Saul, F. (1974) Proc. Natl. Acad. Sci. USA 71, 3440-3444.
24- Emundson, A.B., Ely, K.R., Girling, R.L., Abola, E.E., Schiffer, M. and Westholm, F.A. (1974) Prog. Immunol., Ed. L. Brent and J. Holborrow Vol. II 1, 103-113.
25- Emorine, L., Dreher, K., Kindt, T.J. and Max, E.E. (1983) Proc. Natl. Acad. Sci. USA 80, 5709-5713.
26- Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. and Inouye, M. (1966) Cold Spring Harbor Symp. Quant. Biol. 31, 77-84.