

Patterns of Interallelic Divergence at the Rabbit *b*-Locus of the Immunoglobulin Light Chain Constant Region Are in Agreement With Population Genetical Evidence for Overdominant Selection

Wessel van der Loo and Berlinda Verdoodt

Institute for Molecular Biology, Vrije Universiteit Brussel, St-Genesius-Rode, Belgium

Manuscript received April 24, 1992

Accepted for publication September 9, 1992

ABSTRACT

Population studies at the *b*-locus of the "constant" regions of the rabbit immunoglobulin κ 1 light chain ($c_{\kappa 1}$) revealed patterns of gene diversity resembling those that mark the peculiar nature of the major histocompatibility complex, such as large number of alleles, high heterozygosity levels, consistent excess of heterozygous individuals and long allele coalescence times. This paper documents the evolutionary patterns at the *b*-locus as inferred from DNA sequence comparisons. Among alleles, synonymous substitutions outnumbered expectations for neutral alleles by an order of magnitude. They were distributed randomly throughout the $c_{\kappa 1}$ coding region while interallelic amino acid differences did cluster into segments overlapping with the regions exposed to the solvent. Within these regions, acceptance rates of mutation at amino acid replacement sites were even higher than those at synonymous sites ($d_r/d_s = 1.6-3.0$), while in the intervals between these regions the opposite was found ($d_r/d_s \approx 0.3$). Under the assumption that allelic variation is adaptive at the molecular surface, the divergence patterns at the *b*-locus are therefore very similar to those reported for the major histocompatibility complex. An analysis at the quasi silent *bas*-locus ($c_{\kappa 2}$), which is linked to the *b*-locus, and comparisons among genes of the "variable" region of the κ 1 light chains ($v_{\kappa 1}$), revealed patterns of divergence which differed markedly from those observed at the $c_{\kappa 1}$ constant regions. It is suggested that allelic variability at immunoglobulin constant regions can be due to mechanisms similar to those enhancing diversity at histocompatibility loci.

THE mechanisms sustaining the extensive genetic diversity at loci of the major histocompatibility complex (*MHC*) received much interest in recent literature (BODMER 1972; KLEIN 1986; KLEIN and FIGUERA 1986; O'BRIEN and EVERMANN 1988; HUGHES and NEI 1988, 1989; HAMILTON, AXELROD and TANESE 1990; TAKAHATA and NEI 1990; POTTS, MANNING and WAKELAND 1991). The *MHC* complex is characterized by; (1) large number of alleles, (2) extensive amino acid differences between alleles, (3) a rate of amino acid altering substitutions exceeding that of synonymous substitutions (at least in the peptide binding region) and (4) unusually long allele coalescence times (*i.e.*, interspecies polymorphisms; see TAKAHATA and NEI 1990). This situation was considered without precedent and conflicts with predictions of variation at selectively neutral alleles. According to NEI and co-workers it constitutes evidence for balancing selection, either overdominant or frequency dependent (HUGHES and NEI 1988, 1989; TAKAHATA and NEI 1990).

In the European rabbit, population genetical evidence has been presented for overdominant selection at the immunoglobulin (Ig) constant region (VAN DER LOO 1987; VAN DER LOO *et al.* 1987). In natural populations, heterozygosity levels higher than 80%

were recorded at the *b*-locus of the κ light chain (VAN DER LOO, FERRAND and SORIGUER 1991). In areas where only two or three effective alleles were present, a heterozygote excess of 10% was observed which was specific for the *b*-locus, while the analysis of variances revealed a distribution of gene diversity among populations which was consistent with overdominant selection (VAN DER LOO 1992).

In the rabbit genome there are two distinct genes coding for the constant region of Ig light chains (L-chain) of the κ class, defining the K_1 and the K_2 subclasses (reviewed in MAGE 1987). The rabbit K_1 L-chain differs from other known mammalian immunoglobulin L-chains by having a disulfide bridge between the variable ($v_{\kappa 1}$) and the constant ($c_{\kappa 1}$) domain. The L-chains of the vast majority of the antibodies expressed in rabbits are of the K_1 subclass. The K_1 L-chain is polymorphic and defines the *b*-locus mentioned above (OUDIN 1960; MAGE *et al.* 1973; CAZENAVE *et al.* 1987). Although confirmed as a single gene locus by direct genome analysis (AKIMENKO, HEIDMANN and ROUGEON 1984; EMORINE *et al.* 1984; MATTHIJSSSENS *et al.* 1985), differences between K_1 alleles can be extensive (up to 40% amino acid substitutions) [REISFELD, DRAY and NISONOFF (1965), reviewed in MAGE (1987)]. At the *bas*-locus (KELUS and

WEISS 1977) of the rarely expressed K_2 L-chain the situation is different: only two alleles have been distinguished so far, which differ by a single amino acid substitution (BERNSTEIN *et al.* 1984). The $c_{\kappa 1}$ and $c_{\kappa 2}$ genes are located on the same chromosome (BENAMMAR and CAZENAVE 1982).

Population genetical and evolutionary studies on the allele diversity at the rabbit b -locus might contribute to a more general test of the theoretical models on interallelic divergence under balancing selection as proposed by TAKAHATA (1990) for *MHC* loci. In an attempt to evaluate the possible degree of similarity in evolutionary pathways, we have compared the patterns of allelic divergence at the rabbit b -locus to those documented by HUGHES and NEI (1988, 1989) for *MHC* loci. In light of the study on the variable region of the Ig heavy chain (v_H) by TANAKA and NEI (1989), the mode of variation at the constant region genes ($c_{\kappa 1}$) was also compared to that at the gene family defining the variable regions of the $\kappa 1$ L-chain ($v_{\kappa 1}$).

MATERIALS AND METHODS

Sequence data: All available sequence data on rabbit κ L-chain were considered. The accession numbers and PC-gene identifications of the sequences analyzed in this study are displayed in Table 1. For $c_{\kappa 1}$ (b -locus), all (and only those) sequences were used representing different alleles that can be distinguished by serological means: *i.e.*, the $b4$ (or $b41$), $b4var$ (or $b42$), $b5$, $b9$ and $b95$ allotypes (K_1 sequences). These are from domestic breeds of *Oryctolagus cuniculus cuniculus*, except for the $b95$ gene which was derived from a wild specimen of the subspecies *Oryctolagus cuniculus algiris* (CAZENAVE *et al.* 1987). For $c_{\kappa 2}$ (bas -locus), the sequences of three distinct genes of the $bas2$ allotype were compared to the one $bas1$ sequence available (K_2 sequences). These $bas2$ genes were on chromosomes differing at the b -locus (*i.e.*, $b6$, $b9$ and $b95$). The B41H gene is not part of this database. It was cloned from a mouse-rabbit hybridoma line expressing a complete rabbit L-chain of the $b41$ allotype (SUOMALAINEN *et al.* 1983; MATTHYSSENS *et al.* 1985) and was sequenced by W. VAN DER LOO. The B41H sequence is displayed in the APPENDIX.

Sequence comparisons: Estimates of substitution rates at synonymous sites (d_s) and at amino acid replacement sites (d_r) were done following NEI and GOJOBORI (1986). In short d_s , the estimated number of synonymous substitution per synonymous site is derived from the fraction p_s of observed number of substitutions per synonymous site as suggested by JUKES and CANTOR (1969): $d_s = -3/4 \ln(1 - 4/3 p_s)$. For $p > 0.6$ see TAJIMA and NEI (1984). The significance of the differences of d_r *vs.* d_s can be evaluated by a Z-test. The variances of estimates of mean of d were derived as in HUGHES and NEI (1989) or according to NEI and JIN (1989). For the variable region genes, where divergence can be extensive, it would be more rigorous to present p_r and p_s values rather than the transformations (see TANAKA and NEI 1989): p_r and p_s are easily obtained from the d_r and d_s values presented. The distribution of conserved and variable sites among K_1 alleles was tested for randomness following the procedure proposed by STEPHENS (1985). The definition of regions of high antigenicity (AG blocks) was based upon measures of average structural dissimilarity following PADLAN (1977, 1979), hexapeptide hydrophilicity (HOPP and

TABLE 1
Accession numbers and PC GENE-identifications of
Immunoglobulin κ L-chain sequences analyzed

Species		Reference ^a	
<i>Oryctolagus cuniculus</i>			
K ₂ constant region ($c_{\kappa 2}$): <i>bas</i> allele (and associated <i>b</i> allele)			
<i>K_{2bas2}(b9)</i>	OCIGK2B9	X05801	(1)
<i>K_{2bas1}(b9)</i>	OCIGKLCR	K01280	(2)
<i>K_{2bas2}(b6)</i>	OCIGK2B6	X05800	(2)
<i>K_{2bas2}(b95)</i>	OCIGK2B95	M22543	(3)
K ₁ constant region ($c_{\kappa 1}$): <i>b</i> allele			
<i>K_{1b9}</i>	OCIGK	X00674	(4)
<i>K_{1b95}</i>	OCIGKCI	M22542	(3)
<i>K_{1b5}</i>	OCIGKB5C	K01363	(4)
<i>K_{1b4}</i>	OCIGKLC	K01358/J00667	(5)
<i>K_{1b4var}</i>	OCIGKB4C	K01362	(4)
K variable region (v_{κ})			
V18A	OCIG06	X00997	(6)
V18B	OCIG07	X02336	(6)
V19A	OCIG08	X02337	(6)
V19B	OCIG09	X02338	(6)
V20	OCIGKVA	K02131	(7)
V3C8	OCIGKLCR	K01280	(1)
V17D9	OCIGKB9	K01359	(8)
B4D5	OCIGKLC	K01358/J00667	(5)
B5F2	OCIG03	X00032	(9)
B41H			(10)
Other species:			
Kappa constant region (c_{κ})			
<i>Homo sapiens</i>	HSIGKCA	M11736	(11)
<i>Rattus norvegicus</i>	RNIGKX	V01241	(12)
<i>Mus musculus</i>	MMIGKC	V01569	(13)

^a (1) BERNSTEIN *et al.* (1984), (2) MARIAME, AKIMENKO and ROUGEON (1987), (3) H. AYADI, N. MARCHE and P. A. CAZENAVE (personal communication), (4) AKIMENKO, HEIDMANN and ROUGEON (1984), (5) DREHER *et al.* (1983), (6) HEIDMANN and ROUGEON (1984), (7) LIEBERMANN, EMORINE and MAX (1984), (8) MAGE *et al.* (1987), (9) BERNSTEIN, SKURLA and MAGE (1983), (10) VAN DER LOO, this paper (APPENDIX, Figure 4), (11) HIETER *et al.* (1980), (12) SHEPPARD and GUTMAN (1981), (13) STAVENEZER-NORDGREN, KEKISCH and ZEGERS (1985).

WOODS 1981) and immunogenic potential (PADLAN 1985) as they were proposed in MCCARTNEY-FRANCIS *et al.* (1986). In view of ambiguities in the definition of regions of high *vs.* low antigenicity an algorithm was searched which would assign the highest fraction of interallelic amino acid replacements to a minimal number of regions totaling a minimal fraction of the K_1 sequence. These regions of increased amino acid variation are designated as AV blocks. Unrooted trees of rabbit c_{κ} sequences were constructed using the program FITCH of the software package PHILIP which was generously donated by JOE FELSENSTEIN (University of California, Berkeley).

RESULTS

Clustering of interallelic amino acid replacements in diversity blocks: In Figure 1 the nucleotide sequences of different alleles of the rabbit $c_{\kappa 1}$ are displayed together with those of other mammalian c_{κ} genes. According to the test outlined in STEPHENS (1985), the regions conserved between K_1 alleles are not distributed randomly ($\chi^2 = 22$, d.f. = 8, $P < 0.01$).

TABLE 2

Mean numbers of nucleotide substitutions per synonymous site (d_s) and per amino acid replacement site (d_r) expressed as percentages, with their standard errors, between genes of the constant region of the Ig κ L-chain

Comparisons (nb)	Entire sequence ($n = 309-318$)			AV blocks ($n = 156-165$) AG blocks ($n = 186-195$)			AV intervals ($n = 153$) AG intervals ($n = 123$)			
	d_s	d_r	d_r/d_s	d_s	d_r	d_r/d_s^a	d_s	d_r	d_r/d_s	
Rabbit										
Intralocus:										
K_1 vs. K_1 (10)	AV	10.9 ± 3.9	14.3 ± 2.6	1.3	11.9 ± 5.7	26.3 ± 5.1	2.2*	9.9 ± 5.3	2.9 ± 1.6	0.2
	AG				9.3 ± 4.6	25.9 ± 4.7	2.8**	15.4 ± 7.5	2.4 ± 1.6	0.16
K_2 vs. K_2 (3)	AV	0.9 ± 1.1	0.4 ± 0.4	0.5	0.9 ± 1.5	0.8 ± 0.8	0.5	0.9 ± 1.6	0.0 ± 0.0	0.0
Interlocus:										
K_1 vs. K_2 (5)	AV	18.3 ± 5.1	16.9 ± 2.8	0.9	20.7 ± 7.7	28.0 ± 5.2	1.3	15.7 ± 6.8	5.7 ± 2.2	0.27
	AG				19.2 ± 6.8	23.8 ± 4.4	1.2	19.3 ± 8.5	8.2 ± 3.0	0.42
All comparisons (15)	AV	13.4 ± 4.4	15.2 ± 2.7	1.1	14.9 ± 6.4	26.9 ± 5.2	1.8	11.8 ± 5.8	3.8 ± 1.8	0.32
	AG				13.3 ± 5.6	25.3 ± 4.6	1.5	16.9 ± 7.9	4.7 ± 2.3	0.29
Rabbit, rat, mouse and human										
All comparisons (6)	AV	76.1 ± 14.8	28.6 ± 3.9	0.4	75.9 ± 22.4	36.6 ± 6.5	0.5	80.1 ± 22.1	20.4 ± 4.5	0.26

The K_1 alleles compared represent respectively the $b4$, $b4var$, $b5$, $b95$ and $b9$ allotypes. For the K_2 alleles, three distinct sequences of the $bas2$ allotype were compared to a sequence of the $bas1$ allotype. AV blocks refer to eight regions showing increased interallelic amino acid variation, AG blocks are regions of increased antigenicity as defined in McCARTNEY-FRANCIS *et al.* (1986). Standard deviations displayed were estimated from the sampling variances as in HUGHES and NEI (1989). Standard deviations are lower and significance levels higher if variances are estimated according to NEI and JIN (1989).

^a d_r is significantly larger than d_s by a Z-test at 5% (*) or 1% level (**).

Ninety-two percent of variable sites could be allocated to eight regions of increased allelic variation (AV blocks), representing 50% of the $c_{\kappa 1}$ coding region. This was achieved by limiting to one the number of consecutive conserved amino acid positions allowed within blocks while requiring at least three variable sites per block. These AV blocks are indicated on Figure 1, together with the regions of increased antigenicity (AG blocks). Variability regions were also defined by "sliding windows" as in IRWIN, KOCHER and WILSON (1991): they were virtually identical to the AV blocks defined above, and are not shown.

Estimation of d_r and d_s at regions of high and of low diversity: In Table 2 the mean values for the estimates of rates of synonymous and nonsynonymous mutant acceptance between Kappa constant regions (respectively designated as d_s and d_r) are displayed for the AV and AG blocks and for the corresponding intervals, in analogy with the data presented on MHC loci in HUGHES and NEI (1989). The d_s values in AG or AV blocks were similar to d_s values observed in the corresponding intervals. At variability blocks, d_r among K_1 alleles was always larger than d_s (Tables 2 and 3). At the intervals, d_r was always smaller than d_s . The d_r and d_s values are displayed for pairs of c_{κ} genes in Table 3. The standard deviations shown in the Tables were estimates as in HUGHES and NEI (1989). It would be more accurate to derive these estimates from the variances of the total nucleotide diversity according to the method outlined in NEI and JIN (1989), where covariances due to phylogeny between

paired comparisons are taken into account. This procedure reduces the estimates of the standard deviations compared to those displayed in the tables by a factor which is smaller than 0.76 for comparisons among K_1 alleles and among the rabbit K sequences (*i.e.*, K_1 and K_2). The differences between the means of d_r and d_s at AV and AG regions are therefore more significant than indicated in Tables 2 and 5 (the corresponding Z values calculated by the method of NEI and JIN (1989) were 1.9 and 2.5 for, respectively, the K and the K_1 sequences).

It appears that among K_1 alleles, the ratio of the means of d_r and d_s (d_r/d_s) would be higher if the $b9$ allele was excluded from the comparisons. The $b9$ allele shows more homology with $c_{\kappa 2}$ than do the other $c_{\kappa 1}$ alleles and might not be monophyletic with the $b4$, $b5$ and $b95$ alleles (see APPENDIX). In Figure 2 unrooted trees for synonymous and replacement substitutions are displayed while Figure 3 visualizes the correlations of d_s and d_r among pairs of rabbit c_{κ} genes. Variation was much more limited among K_2 alleles. Comparisons between $c_{\kappa 1}$ and $c_{\kappa 2}$ are also shown.

Estimation of d_r and d_s for hypervariable (CDR) and framework (FR) regions of the $v_{\kappa 1}$: The divergence between 10 variable region genes (v_{κ}) was analyzed in analogy with what is described for the constant region genes. Regions of variability were either defined by the procedures used to define blocks of allelic variability (V blocks), or by adopting the definition of hypervariable regions as proposed in KABAT

TABLE 3

Estimates of d_r and d_s , expressed as percentages, their sampling variances and Z-statistics at AV regions for pairs of C_k Genes

	d_r/d_s	d_s	s SD	d_r	s SD	Z d_r-d_s	P $d_s > d_r$ AV (%)	P $d_s > d_r$ INT (%)
Interspecies C_k								
Hum:Mus	0.47	91.0	26.3	43.0	7.3	-1.75	96.0	97.6
Hum:Mus	0.47	91.0	26.3	43.0	7.3	-1.75	96.0	97.6
Hum:Rat	0.50	72.2	20.5	35.8	6.4	-1.70	95.5	99.9
RK ₂ :Mus	0.38	113.0	34.9	43.1	7.3	-1.96	97.5	99.2
RK ₂ :Rat	0.41	89.3	25.8	36.4	6.4	-1.99	97.7	99.6
RK ₂ :Hum	0.70	66.7	19.0	46.4	7.7	-0.99	83.9	78.8
Rat:Mus	0.63	23.2	8.3	14.6	3.5	-0.95	82.9	91.6
Mean	0.48	75.9	22.5	36.5	6.2	-1.69	95.1	96.2
Rabbit C_k interlocus (C_{k1}/C_{k2})								
K_2/K_{11}								
K ₂ :b5	1.34	26.6	9.0	35.8	6.3	0.84	20.0	89.4
K ₂ :b95	1.31	23.5	8.3	30.7	5.7	0.72	20.6	94.8
K ₂ :b4v	1.23	22.0	8.0	27.1	5.2	0.53	30.1	87.1
K ₂ :b4	1.27	20.9	7.8	26.4	5.1	0.60	27.4	87.1
K_2/b_9								
K ₂ :b9	1.87	10.7	5.4	20.0	4.3	1.35	8.8	93.3
Rabbit C_{k1} intralocus								
$b9/K_{11}$								
b9:b5	1.78	23.6	8.3	42.1	7.0	1.70	4.5	80.5
b9:b95	2.10	19.4	7.4	40.8	6.9	2.11	1.7	94.0
b9:b4v	1.82	17.6	7.0	32.1	5.8	1.58	5.7	80.5
b9:b4	1.90	16.5	6.8	31.4	5.7	1.68	4.6	80.5
K_{11}/K_{11}								
b5:b4v	2.66	11.5	5.6	30.7	5.7	2.41	0.8	96.6
b5:b4	2.90	10.3	5.3	30.0	5.6	2.57	0.5	96.6
b95:b4v	2.66	8.9	4.8	23.6	4.8	2.17	1.5	91.6
b95:b4	3.05	7.6	4.4	23.1	4.7	2.39	0.8	91.6
b95:b5	2.84	2.6	2.6	7.3	2.5	1.33	9.3	88.1
b4:b4v	1.61	1.3	1.8	2.1	1.3	0.35	36.7	50.0

s SD is the standard deviation due to sampling variance estimated as $p(1-p)/m(1-4/3 p)^2$, p the frequency of substitutions and m the number of sites. Comparisons were done between the regions of increased interallelic variability (AV) as defined among rabbit b -locus alleles. Among species, the regions compared were homologous to the AV regions. $Z = (d_r - d_s) / \sqrt{(s SD_r)^2 + (s SD_s)^2}$. The probabilities P that $d_s > d_r$ are shown for AV regions and for the intervals (INT).

(1987) (CDR for "complementarity defining regions") in analogy with the study of TANAKA and NEI (1989) for heavy chain (H-chain) variable regions. The positions of CDR and V segments are indicated on Figure 4, together with the cDNA sequence of the B41H L-chain. CDR regions correspond more or less to segments of high amino acid variation (*i.e.*, CDR3 is identical with V3), except for the N terminal region (V_{NT}), where no CDR region is defined but amino acid variation is high ($d_r = 0.3$). Comparisons between v_k germline genes of the same individual (V18A, V18B, V19A, V19B), between v_k genes expressed in different individuals (17D9, B52, B41H, B4D5) or between v_k genes grouped according to the corresponding b -locus allotype, gave similar results. A detailed analysis can be found in VERDOODT (1991). Estimates of d_r and d_s at hypervariable regions (CDR or V) and at the corresponding intervals (Framework) are shown in Table 4. d_r was four times larger at CDR regions than at Framework ($d_r^{CDR} = 0.402$ *vs.* $d_r^{FR} = 0.090$). However, in contrast with findings on C_{k1} alleles and on

murine v_{11} regions (TANAKA and NEI 1989), where d_s values were similar among variable and conserved segments, d_s was, like d_r , much larger at CDR than at the Framework regions ($d_s^{CDR} = 0.314$ *vs.* $d_s^{FR} = 0.105$). This was more pronounced in cases where amino acid variability was the criterion for partitioning ($d_s^V = 0.432$ *vs.* $d_s^{Int} = 0.077$, Table 4). Within V blocks, d_s was even slightly larger than d_r ($d_s^V = 0.432$ *vs.* $d_r^V = 0.415$). The d_r/d_s ratios were by consequence similar among hypervariable and conserved regions.

Estimate of the relevant mutation rate ν of b -locus genes: The relevant mutation rate is the rate at which mutations that can possibly be favored by selection arise at a given locus in the population. ν can be estimated as the product $\nu = \mu r \mathcal{P}$, where μ is the mutation rate, r is the number of sites where substitutions are potentially adaptive, and \mathcal{P} is the probability that a mutation is favorable (TAKAHATA 1990). By adopting divergence times among mammalian orders as in LI *et al.* (1990), the mean d_s between C_k genes of primates, lagomorphs and rodents indicate a

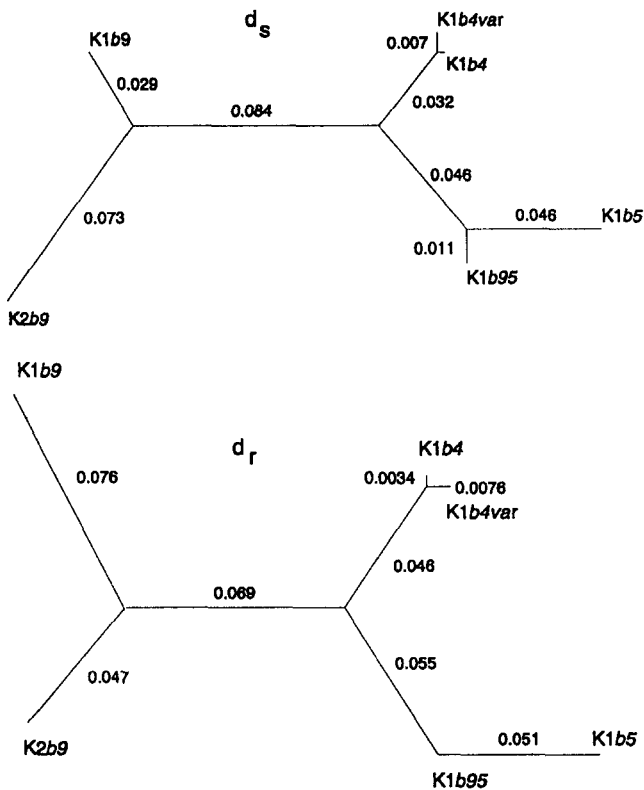


FIGURE 2.—Unrooted tree of the constant region of rabbit Ig κ L-chain, based upon synonymous substitutions d_s (top) and on upon amino acid replacement substitutions d_r (bottom).

rate of synonymous substitutions very similar to $\mu = 5.3 \times 10^{-9}$ per site per year, the overall mutation rate estimated for proteins in KIMURA (1985). A maximal estimate of ν at the b -locus is then obtained by setting $r = 195$ (the largest number of nucleotides within the AG regions), and $\rho = 0.8$ (the probability that a mutation is not synonymous). One might also consider that part of the amino acid changes in AV or AG regions are not selected but might be a corollary of balancing selection which increases also the rate of neutral mutations (mainly, but not only, because of increased allele persistence times). If, as indicated by the data, d_r/d_s equals 2 to 3 in regions exposed to selection, 33–50% of the observed replacement mutations might actually be neutral. The areas where mutations are potentially favorable could then be more restricted and/or the probability of adaptive mutations smaller than 0.8. A reasonable low estimate of ν might be derived from $r = 156/2$ (half of the smallest number of nucleotides within AV blocks). Therefore, with $\mu = 5.3 \times 10^{-9}$ per year per site, the yearly rate of relevant evolution at c_κ genes could lie between the values:

$$\nu_{\max} = 8.4 \times 10^{-7} \quad \nu_{\text{low}} = 3.4 \times 10^{-7}.$$

DISCUSSION

This study proposes estimates of d_r and d_s for sections of the $c_{\kappa 1}$ chain that might be exposed to differ-

ent selective constraints. For the *MHC class I* alleles, the regions exposed to positive selection were postulated to coincide with the 57 amino acids of antigen binding regions (ARS), a hypothesis which was supported by the observation that the ratios of d_r/d_s were much higher within the ARS than outside. The relevant mutation rate ν was therefore estimated as $\nu = 2.9 \times 10^{-9} \times 57 \times 3 \times 0.8 = 3.9 \times 10^{-7}$ amino acid replacements per ARS per year, given that the mutation rate μ at this locus was estimated at 2.9×10^{-9} per nucleotide per year (HAYASHIDA and MIYATA 1983) and that 80% of substitutions alter the amino acid composition (TAKAHATA 1990). The region of putative adaptive divergence at *MHC class II* genes was defined by analogy with *MHC class I*.

For the constant regions of the Ig L-chain no particular region is known to benefit from allelic variability. However, microorganisms can manipulate the immune response by interacting with constant regions of the immunoglobulin molecules (FUJIYAMA *et al.* 1985; HAMERS, VAN DER LOO and DE BAETSELIER 1986). Allelic diversity at Ig constant regions might therefore be part of a host strategy (VAN DER LOO 1987). For obvious reasons such adaptive variation should in the first place concern regions exposed to the environment, *i.e.*, the molecular surface. The amino acid compositions of the $c_{\kappa 1}$ chains indicates regions of high antigenicity, separated by mainly hydrophobic regions. The latter represent in essence the β pleated sheet core of the immunoglobulin fold while the antigenic regions (AG blocks) occupy the connecting loops and are exposed to the solvent (see MCCARTNEY-FRANCIS *et al.* 1986). The degree of replacement substitutions d_r was clearly higher in the AG regions than in the intervals between them (Table 2) confirming differences in selective constraints. The distinction between high *vs.* low antigenicity is necessarily somewhat arbitrary. The mean ratio of d_r/d_s was larger for AG blocks than for AV blocks, indicating that the latter does not lead to overestimation of this parameter (Table 2). In view of the extensive overlap between antigenic regions and those of allelic variation we present estimates on d_r and d_s for AV regions and intervals.

The observation that some regions of protein molecules are more conserved than others, is not surprising but it is interesting that in the AV and AG blocks, replacement substitutions among alleles were consistently more frequent than synonymous substitutions ($d_r > d_s$). This makes it unlikely that the differences in variability are merely due to selection that prevents structural change in the conserved regions. SHEPPARD and GUTMAN (1981), who reported increased d_r/d_s ratios for the alleles of the rat κ L-chain, regarded this increase as evidence for selection preventing synonymous mutations ("strong selection at the level of the

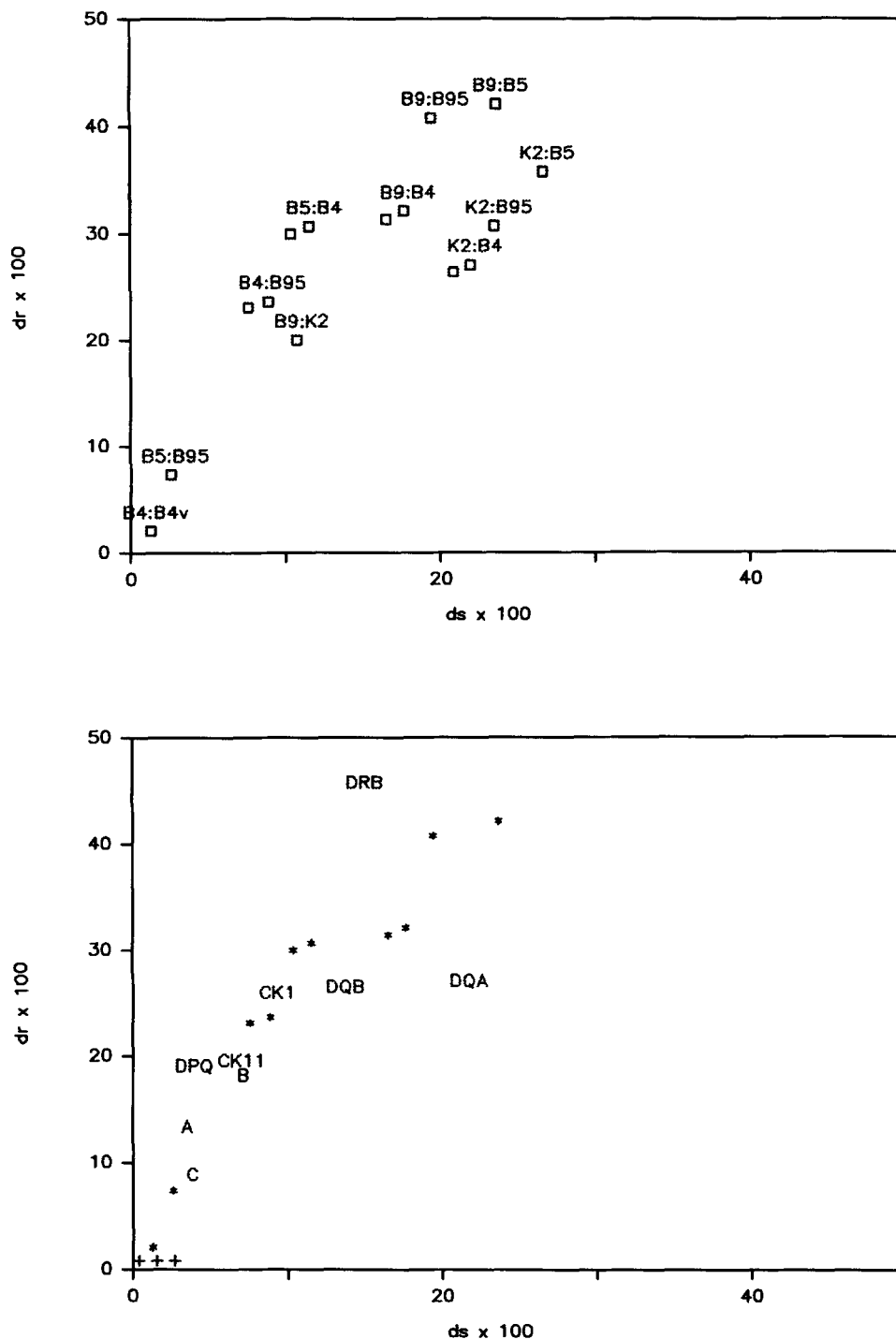


FIGURE 3.—Correlation of d_s and d_r values among AV regions of rabbit κ L-chains (top) and among ARS regions of human MHC class I and class II molecules (bottom). The (d_s, d_r) coordinates are displayed for two by two comparisons between AV regions of rabbit κ genes and compared to (d_s, d_r) coordinates for the mean d values at antigen recognition site regions of human MHC loci as they were reported in HUGHES and NEI (1988, 1989). These are *class I* loci: *A*, *B*, *C* and *class II* loci *DPB*, *DQB*, *DRB*. Datalabels are centered on datapoints. *CK1* indicates mean (d_s, d_r) among all *b*-locus alleles, *CK11* that among the *b*-locus alleles excluding the *b9* allele. *Correspond to pairwise comparisons among *b*-locus (*CK1*) genes. (d_s, d_r) points for comparisons between *bas*-locus (*CK2*) alleles fall within the area expected for neutral alleles ($d_s > d_r$ and $d_s < 2\%$, indicated by +++).

nucleotide sequence"). Also the number of silent replacements between *b*-locus genes was repeatedly described as "low" (BERNSTEIN, SKURLA and MAGE 1983; MAGE *et al.*, 1987; VAN DER LOO, FERRAND and SO-RIGUER 1991). It is important to emphasize that the

incidence of synonymous substitutions among rabbit *K₁* alleles is, on the contrary, unusually *high* for alleles. According to KIMURA (1968, 1985), for neutral alleles, the turnover times *T* are, on theoretical grounds, expected to be of the order of $4N_e$ genera-

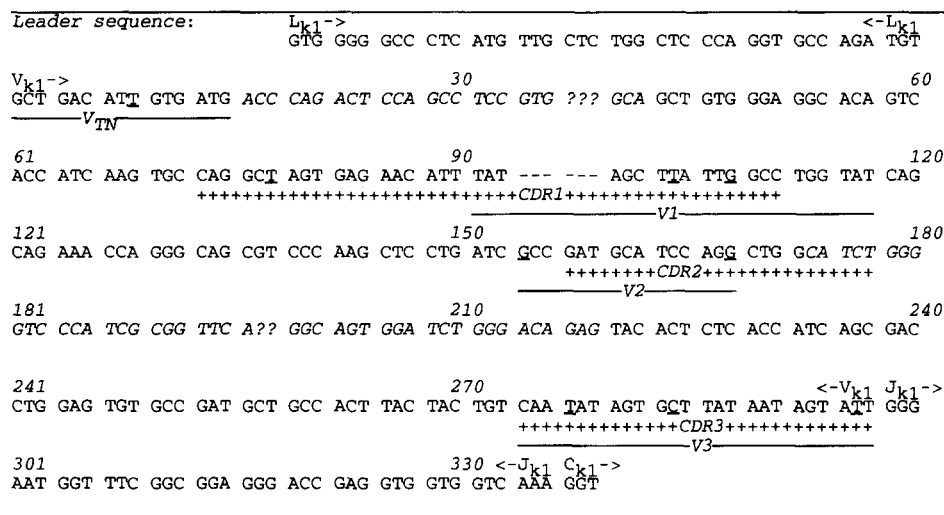


FIGURE 4.—DNA sequence of the B41H variable region of a κ L-chain of the *b41* allotype as expressed in a mouse-rabbit hybridoma line and position of CDR regions and variability (V) blocks: cDNA was derived from the cell line RX-54.3, which was kindly donated by GEORGE KÖHLER (Basel Institute for Immunology, Switzerland) and was described in SUOMALAINEN *et al.* (1983) and MATTHYSSENS *et al.* (1985). The sequence starts at position 9 of the leader (L_{K1}), followed by, respectively, the variable region (V_{K1}) and the joining region (J_{K1}). The constant region (C_{K1}) which was found identical to other published *b4* sequences and noncoding regions are not shown. The nucleotides underlined are unique at that position among the 10 rabbit v_{κ} sequences here analyzed. *Italics* indicate likely nucleotides at positions where data were unclear. The CDR regions (++++) and V blocks (—) are indicated.

TABLE 4

Mean numbers of nucleotide substitutions per synonymous site (d_s) and per amino acid replacement site (d_r) expressed as percentages between genes of the Ig variable region

Comparisons (nb)	V blocks (n = 81) CDR-regions			Intervals (n = 216) framework		
	d_s	d_r	d_r/d_s	d_s	d_r	d_r/d_s
Rabbit Kappa 1 light chain v_{κ}						
All (45)						
V	43.2 ± 16.9	41.5 ± 9.6	0.9	7.7 ± 3.9	8.5 ± 2.3	1.1
CDR	31.4 ± 13.7	40.2 ± 9.5	1.3	10.5 ± 4.6	9.0 ± 2.4	0.9
V18A&B; V19A&B; V20 (10)						
CDR	30.1 ± 13.4	38.3 ± 9.1	1.3	14.5 ± 5.5	7.1 ± 2.1	0.5
17D9; B5F2; B41H; B4D5 (6)						
CDR	31.4 ± 13.8	40.2 ± 9.4	1.3	9.5 ± 4.4	7.4 ± 2.2	0.7
Mouse heavy chain v_{H1} subgroup II ^a						
All (210)						
CDR	22.2 ± 5.8	30.3 ± 3.0	1.3	25.1 ± 2.8	7.0 ± 0.8	0.3

CDR are complementarity defining regions of the antigen binding site, as defined in KABAT (1987). V blocks are segments where amino acid replacements among the compared v_{κ} sequences is high and are operationally analogous to the AV blocks of the c_{K1} regions. The overlap between V and CDR is shown in Figure 4. The d_s and d_r values on variable regions of Mouse Ig heavy chain were derived from p_s and p_r values presented by TANAKA and NEI (1989) as $d = -3/4 \ln(1 - 4/3 p)$ (JUKES and CANTOR 1969).

^a After TANAKA and NEI (1989).

tions (N_e is the effective phylogenetical population size) and this was found to be in agreement with observations on eukaryotic genes where interallelic differences vary between 0.0001 to 0.02 per nucleotide site (NEI 1987). N_e of the rabbit in its natural habitat (Mediterranean coastal chaparral) is probably smaller than N_e of mouse populations, which can be estimated at 10^5 (NEI 1987). N_e smaller than 10^5 was also suggested by studies on mitochondrial DNA diversity in natural rabbit populations (BIJU-DUVAL 1992). As average generation times in wild rabbit

should be smaller than 2 years, persistence times of neutral alleles should be shorter than 10^6 years. Consequently, because $d_s = 2 \times T \times \mu$ with $\mu = 5.3 \times 10^{-9}$, d_s should be smaller than 0.01. The d_s values between pairs of alleles displayed in Tables 2 and 3 are clearly larger by an order of magnitude, except for the *b4-b4var* pair. Considering that d_s between rat and mouse c_{κ} genes is 0.20, the d_s values among K_1 alleles appear indeed larger than can be reasonably expected for neutral alleles.

The allelic divergence of the $c_{\kappa 2}$ alleles ($d_s \approx 0.01$)

is, on the contrary, in agreement with the hypothesis of selective neutrality and indirectly confirms the validity of the theoretical expectations outlined above. The *bas1* and *bas2* alleles associated with a *b9 K₁* gene did not show a single synonymous nucleotide substitution. It could be argued that such strong homology between the *bas1* and *bas2* alleles reflected a very recent divergence. For this reason, the K2*bas2(b6)* and K2*bas2(b95)* sequences were included in the analysis. The comparison of sequences of *c_{κ2}* genes of the *bas2* allotype from chromosomes harboring different *b*-locus alleles (*b9*, *b6* and *b95*) confirmed that the comparatively limited divergence between *c_{κ2}* genes was not due to sampling and suggests allelic turnover rates and substitution rates as indeed expected for neutral alleles. Also at the 5' and 3' flanking regions of the *c_{κ1}* genes, intralocus substitution rates (respectively 0.04 and 0.02, not shown) are compatible with neutral drift (MAGE 1987). The differences in divergence at *b*-locus, *bas*-locus and flanking regions are reminiscent of the differences in substitution rates among exons of *MHC* genes. As documented by HUGHES and NEI (1988, 1989), exons *not* containing the ARS region often showed substitution rates compatible with neutral evolution. For example, at the *K*, *L* and *D* loci of the murine *MHC class I* complex, *d_s* was 0.132 in the ARS, while *d_s* equaled 0.012 in Exon 4. Other examples are the *DQA* locus of the human *MHC class II* and the *E_β* locus of the murine *MHC class II*, where the D2 exons showed much stronger synonymous intralocus homology than observed at the ARS containing exons.

Because balancing selection is bound to increase allelic turnover times, the increase in *d_s* constitutes on its own an indication of balancing selection (TAKAHATA 1990) (*cf.* KREITMANN and HUDSON 1991). The fact that *d_r* is nevertheless consistently larger than *d_s* in those regions where allelic variation is potentially adaptive, suggest the existence of deterministic processes favoring population diversity at this Ig constant region.

Comparison between *b*-locus and *MHC* evolution:

In summary, the observed situation meets the predictions on allelic divergence under overdominant selection. The rates of mutant acceptance at replacement sites are increased and the prolonged persistence times of alleles in populations favors the accumulation of selected as well as of neutral mutations among allelic lineages. This evolutionary mode has been discussed and documented in detail for the *MHC* alleles ((TAKAHATA and NEI 1990; HUGHES and NEI 1988, 1989). Table 5 summarizes similarities in evolutionary patterns at the *b*-locus and *MHC* loci, similarities which are also sustained by the estimates of the relevant mutation rates (see above) and by the distribution of (*d_s*, *d_r*) coordinates as shown in Figure 3. It seems that both polymorphisms could conceivably be the out-

come of similar evolutionary pathways most likely related to immune defense strategies. It is interesting that estimates of the rate of synonymous substitutions (*d_s*) at the *b*-locus did not differ markedly from those at the *MHC* loci which is in agreement with phylogenetical observations suggesting that allelic coalescence times could be similar (VAN DER LOO, FERRAND and SORIGUER 1991). The somewhat smaller *d_s* values at the *MHC class I* loci could be accounted for by the difference in mutation rate μ which is lower for *MHC class I* than for *c_κ* (2.9×10^{-9} vs. 5.3×10^{-9}), difference which may be due to differences in codon bias (*cf.* SHARP and LI 1986).

***d_r* and *d_s* at hypervariable and framework regions among *v_κ* genes:** A detailed discussion on the comparisons between rabbit *v_κ* genes lies beyond the scope of this paper. The data on the *v_{κ1}* genes in Table 4 illustrate that mechanisms which increases diversity among genes do not necessarily lead to increases in *d_r*/*d_s* ratios, as was observed at AV and AG regions of *c_{κ1}* genes. The mechanisms underlying the diversity at antibody variable regions genes (reviewed in GEARHART 1982) can clearly be of a different nature compared to those sustaining allelic diversity. The former generate diversity among genes within the individual genome and seem, at least in part, to rely upon an increase in mutation rate at the "hypervariable" (CDR) segments: they are expected to enhance both synonymous and replacement substitutions rates (EISEN and REILLY 1985). The latter, which has to do with population diversity, increases only mutant acceptance rates (and allelic persistence times): That is, this process should preferentially affect substitutions that alter amino acid composition. The observations for *v_{κ1}* genes reported here are somewhat different from those reported by TANAKA and NEI (1989) on variable region genes of the mouse Ig H-chain (*v_H*) genes. In mouse, CDR and FR regions showed similar *d_s*, while *d_r* was higher at CDR (Table 4) indicating Darwinian selection. At CDR1 (as well as at *V_{NT}* and *V1*) of the Rabbit *v_κ* chains, synonymous substitutions were, on the contrary, about two times more frequent than replacement substitutions, notwithstanding the fact that these regions were defined as regions of increased amino acid variation (*d_r* is about 0.3). This strongly indicates a nondarwinian component in the enhancing mechanism. However, at CDR2 and at CDR3 (=V3), our data (*not shown*) support the hypothesis of positive selection as it was proposed for the *v_H* variability by TANAKA and NEI (1989).

It is interesting that the overall degree of amino acid divergence between *v_{κ1}* genes (*d_r* = 0.159, not shown) did not differ markedly from that between *c_{κ1}* alleles (*d_r* = 0.143, Table 2). In the framework regions of *v_{κ1}* genes, *d_s* values were similar to those between *c_{κ1}* genotypes. This indicates that the common ancestor of the actual *v_{κ1}* gene family could be contempo-

TABLE 5

Divergence patterns of *MHC* genes compared to those of the *b*-locus of the constant regions of the rabbit *Ig* κ 1 L-chain: d_r and d_s (%) at regions with either high or low variability

Locus (<i>nb</i> of comparisons)	d_s	d_r	d_r/d_s	d_s	d_r	d_r/d_s
<i>MHCI</i>		ARS (<i>n</i> = 171)		Remaining D1 (<i>n</i> = 372–375)		
Human <i>A,B,C</i> (19)	4.7 ± 2.6	14.1 ± 2.4	3.0***	5.1 ± 2.1	2.4 ± 0.8	0.5
Mouse <i>K,L</i> (12)	13.2 ± 4.3	21.2 ± 3.2	1.6*	8.8 ± 2.3	6.3 ± 1.1	0.7
<i>MHCII</i>		ARS (<i>n</i> = 45–60)		Remaining D1 (<i>n</i> = 192–234)		
β chain:						
Human <i>DPB</i> (19)	3.9 ± 5.5	19.0 ± 6.4	4.9	2.4 ± 1.7	2.8 ± 1.1	1.2
Human <i>DQB</i> (28)	13.7 ± 7.7	26.5 ± 5.4	1.9	8.5 ± 2.3	6.7 ± 1.3	0.8
Human <i>DRB</i> (91)	15.0 ± 8.5	45.7 ± 6.2	3.0**	8.0 ± 1.9	4.5 ± 0.9	0.6
Human mean (122)	14.4 ± 8.2	40.6 ± 6.0	2.8	8.0 ± 2.0	5.3 ± 1.2	0.7
Mouse <i>Aβ</i> (21)	0.0 ± 0.0	30.0 ± 6.7	***	4.0 ± 1.6	6.7 ± 1.2	1.7
α chain:						
Human <i>DQA</i> (6)	21.7 ± 11.8	27.0 ± 6.7	1.2	8.0 ± 3.2	4.3 ± 1.3	0.5
Mouse <i>Aα</i> (15)	3.2 ± 3.0	23.7 ± 4.9	7.4***	2.8 ± 1.7	2.6 ± 0.8	0.9
Rabbit <i>Ig cκ1</i> (<i>b</i> -locus)		AV blocks (<i>n</i> = 156–165)		Remaining <i>cκ1</i> (<i>n</i> = 153)		
<i>K$_I$(b9)-K$_{II}$</i> (4)	19.3 ± 7.6	36.6 ± 6.4	1.9	12.0 ± 5.9	5.3 ± 2.2	0.2
<i>K$_I$-K$_I$</i> (10)	11.9 ± 5.7	26.3 ± 5.1	2.2*	9.9 ± 5.3	2.9 ± 1.6	0.3
<i>K$_{II}$-K$_{II}$</i> (6)	7.0 ± 4.2	19.5 ± 4.3	2.8*	8.4 ± 4.9	1.3 ± 1.1	0.2

Data on *MHC* are from HUGHES and NEI (1988, 1989). For the *b*-locus of the rabbit *c κ 1* gene, estimates based on comparisons among *b4*, *b4var*, *b5* and *b95* (*K $_{II}$* alleles) and between the *b9* sequence and these four *K $_{II}$* alleles are displayed together with those on comparisons among the five *K $_I$* alleles (Table 2). Significance levels shown are estimated as in HUGHES and NEI (1988, 1989). If variances of d_r and d_s are estimated according to NEI and JIN (1989) the differences between d_r and d_s are significant at the 1% level among AV blocks of the rabbit *K $_I$* and *K $_{II}$* alleles.

rary to the common ancestor of *c κ 1* alleles (*K $_I$*). While this should be relevant to the evolutionary processes underlying the emergence of the disulfide bond between *v κ 1* and *c κ 1* chains (see appendix), a more recent history of the *v κ 1* gene family, compared to that of the *v H* genes studied in TANAKA and NEI (1989), might also account for differences in divergence patterns among both gene families.

Concluding remarks: Within the context of previous population genetical data, the present study strongly suggests that population diversity at the rabbit *Ig* constant region can be the outcome of evolutionary pathways similar to those underlying the variation at *MHC* loci. While the distribution of *b*-locus alleles in populations of wild rabbit was in agreement with strong overdominant selection, recent studies on *MHC* alleles in mouse populations revealed a *MHC*-allele-specific mating choice which resulted in heterozygote excess at this locus (POTTS, MANNING and WAKELAND 1991). This is not in contradiction with the hypothesis of NEI and coworkers on the origin of *MHC* variation if overdominance is understood in a nonrestrictive way (*i.e.*, any mechanism increasing the heterozygote frequency at a specific locus). Such mechanisms have a cost and cannot be evolutionary stable without selective advantage of the situation they help to create (*cf.* HAMILTON, AXELROD and TANESE 1990; HOWARD 1991). The possibility that the underlying determinism might reside within the organism

itself is interesting as it suggests a feedback of benefits related to gene diversity which overdominance is bound to sustain. The fact that patterns of allele divergence due to direct overdominance cannot be distinguished from patterns caused by such selection-anticipating mechanisms, should therefore not diminish the evidence that the extensive allele diversity at the *MHC* and at *Ig* loci is the outcome of deterministic processes related to adaptiveness of population diversity and evolutionary change. On the other hand, the striking similarity between the divergence patterns at *MHC* and *Ig* loci here reported might either reflect similarities in actual selection forces or a similar efficiency of the anticipating mechanisms.

We are indebted to RAYMOND HAMERS and DOMIEN ROGGEN for interesting discussions and to WILLY VERHEULPEN and ROBERT HERZOG for valuable help in sequence data accession. We are grateful to GASTON MATTHIJSSSENS for his contribution in the cloning of the B41H gene. W. VAN DER LOO is research associate at the Belgian National Science Foundation (NFWO-FNRS).

LITERATURE CITED

- AKIMENKO, M. A., O. HEIDMANN and F. ROUGEON, 1984 Complex allotypes of the rabbit immunoglobulin K light chain locus are encoded by structural alleles. *Nucleic Acids Res.* **12**: 4691–4701.
- AKIMENKO, M. A., B. MARIAMÉ and F. ROUGEON, 1986 Evolution of immunoglobulin k light chain locus in the rabbit. Evidence for differential gene conversion events. *Proc. Natl. Acad. Sci. USA* **83**: 5180–5183.
- BENAMMAR, A., and P. A. CAZENAIVE, 1981 *B-96* a 7th allele at

- the rabbit kappa chain *b*-locus. *Eur. J. Immunol.* **11**: 344–346.
- BERNSTEIN, K. E., R. M. SKURLA JR., and R. G. MAGE, 1983 The sequence of rabbit K light chains of the b4 and the b5 allotypes differ more in their constant regions than in their 3' untranslated regions. *Nucleic Acid Res.* **11**: 7205–7214.
- BERNSTEIN, K. E., E. LAMOYI, N. L. MCCARTNEY-FRANCIS and R. G. MAGE, 1984 Sequence of a copy DNA encoding Basilea Kappa light chain suggest a possible relationship of protein structure to limited expression. *J. Exp. Med.* **159**: 635–640.
- BIJU-DUVAL, CH., 1992 Diversité de l'ADN mitochondrial chez les lagomorphes. Thèse de doctorat de Université Paris 6.
- BODMER, W. F., 1972 Evolutionary significance of the *HL-A* system. *Nature* **237**: 139–145.
- CAZENAIVE, P. A., A. BENAMMAR, J. A. SOGN and T. J. KINDT, 1987 Immunoglobulin genes in feral rabbit, pp. 148–160 in *The Rabbit in Contemporary Immunological Research*, edited by S. DUBISKI. John Wiley & Sons, New York.
- DREHER, K. L., L. EMORINE, T. J. KINDT and E. E. MAX, 1983 cDNA clone encoding a complete rabbit immunoglobulin kappa light chain of the b4 allotype. *Proc. Natl. Acad. Sci. USA* **80**: 4489–4493.
- EISEN, H. N., and E. B. REILLY, 1985 Lambda chains and genes in inbred mice. *Annu. Rev. Immunol.* **3**: 337–365.
- EMORINE, L., J. A. SOGN, F. D. TRING, T. J. KINDT and E. E. MAX, 1984 A genomic gene encoding the b5 rabbit immunoglobulin kappa constant region. *Proc. Natl. Acad. Sci. USA* **81**: 1789–1793.
- FUJIYAMA, Y., K. KOBAYASHI, S. SENDA, T. BENNO, T. BAMBA and S. HOSADA, 1985 A novel IgA protease from *Clostridium* sp. capable of cleaving IgA1 and IgA2 A2m(1) allotypes, but not IgA2 A2m(2) allotype paraproteins. *J. Immunol.* **134**: 573–576.
- GEARHART, P. J., 1982 Generation of immunoglobulin variable region gene diversity. *Immunol. Today* **3**, 4: 107–112.
- HAMERS, R., W. VAN DER LOO and P. DE BAETSELIER, 1986 Structure and function of immunoglobulins, the symbiotic interplay between parasites and host antibodies, pp. 324–345 in *The Ruminant Immune System in Health and Disease*, edited by I. MORRISON. Cambridge University Press, Cambridge.
- HAMILTON, W. D., R. AXELROD and R. TANESE, 1990 Sexual reproduction as an adaptation to resist parasites. *Proc. Natl. Acad. Sci. USA* **87**: 3566–3573.
- HAYASHIDA H., and T. MIYATA, 1983 Unusual evolutionary conservation and frequent DNA segment exchange in MHC class I genes. *Proc. Natl. Acad. Sci. USA* **80**: 2671–2675.
- HEIDMANN, O. L., and F. ROUGEON, 1984 Immunoglobulin Kappa light chain diversity in rabbit is based upon 3'-length heterogeneity of germ-line variable genes. *Nature* **311**: 74–76.
- HIETER P. A., E. E. MAX, J. G. SEIDMAN, J. V. MAIZEL and P. LEDER, 1980 Cloned human and mouse kappa-immunoglobulin: constant and J-region conserve homology in functional segments. *Cell*, **22**: 197–207.
- HOPP, T. P., and K. R. WOODS, 1981 Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* **78**: 3824–3828.
- HOWARD, J. C., 1991 Disease and evolution. *Nature* **352**: 565–567.
- HUGHES, A., and M. NEI, 1988 Patterns of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**: 167–170.
- HUGHES, A., and M. NEI, 1989 Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominance selection. *Proc. Natl. Acad. Sci. USA* **86**: 958–962.
- IRVIN, D. M., T. D. KOCHER and A. C. WILSON, 1991 Evolution of the cytochrome *b* gene of mammals. *J. Mol. Evol.* **32**: 128–144.
- JUKES, T. H., and C. R. CANTOR, 1969 Evolution of protein structure. pp 21–132 in *Mammalian Protein Metabolism III*, edited by H. MUNRO. Academic Press, New York.
- KABAT, E., 1987 The structural basis of antibody complementarity. *Adv. Protein Chem.* **32**: 1–75.
- KELUS, A. S., and S. WEISS, 1977 Variant strain of rabbits lacking immunoglobulin kappa polypeptide chain. *Nature* **265**: 156–158.
- KIMURA, M., 1968 Evolution rate at the molecular level. *Nature* **217**: 624–626.
- KIMURA, M., 1985 Diffusion models in population genetics with special reference to fixation time of molecular mutants under mutational pressure, pp. 19–39 in *Population Genetics and Molecular Evolution*, edited by T. OHTA and K. AOKI. Japan Scientific Societies Press, Tokyo.
- KING, J. L., and T. H. JUKES, 1969 Non-Darwinian evolution: Random fixation of selectively neutral mutations. *Science* **164**: 788–798.
- KLEIN, J., 1986 *Natural History of the Major Histocompatibility Complex*. Wiley, New York.
- KLEIN, J., and F. FIGUERA, 1986 Evolution of the major histocompatibility complex. *CRC Crit. Rev. Immunol.* **33**: 70–104.
- KREITMAN, M., and R. R. HUDSON, 1991 Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. *Genetics* **127**: 565–582.
- LI, W-H., M. GOUY, P. M. SHARP, C. O'H UIGIN and Y-W. YANG, 1990 Molecular phylogeny of Rodentia, Primates, Artiodactyla and Carnivora and molecular clocks. *Proc. Natl. Acad. Sci. USA* **87**: 6703–6707.
- LIEBERMANN R., E. EMORINE and E. E. MAX, 1984 Structure of a germline rabbit Ig V-kappa-region gene: implications for V-J recombination. *J. Immunol.* **133**: 2753–2756.
- MAGE, R. G., 1987 Molecular biology of rabbit immunoglobulin and T-cell receptor genes, pp. 106–133 in *The Rabbit in Contemporary Immunological Research*, edited by S. DUBISKI. John Wiley & Sons, New York.
- MAGE, R. G., R. LIEBERMAN, M. POTTER and W. D. TERRY, 1973 Immunoglobulin allotypes, pp. 299–376 in *The Antigens I*, edited by M. SELA. Academic Press, New York.
- MAGE, R. G., N. L. MCCARTNEY-FRANCIS, M. KOMATSU and E. LAMOYI, 1987 Evolution of genes for allelic and isotypic forms of immunoglobulin kappa chains and of the genes for T-cell receptor beta chains in rabbits. *J. Mol. Evol.* **25**: 292–299.
- MARIAME, B., M. A. AKIMENKO and F. ROUGEON, 1987 Interallelic and intergenic conversion events could induce differential evolution of the two rabbit Ig kappa light chain genes. *Nucleic Acids Res.* **15**: 6171–6179.
- MATTHIJSSSENS, G., B. BORREMANS, R. HAMERS and W. VAN DER LOO, 1985 Genomic evidence for the truly allelic nature of the complex *b*-locus allotypes of the rabbit immunoglobulin light chain, pp. 123–126 in *Protides of the Biological Fluids, Proceedings of the 33th Colloquium*, edited by H. PEETERS. Pergamon, New York.
- MCCARTNEY-FRANCIS N. L., E. A. PADLAN, G. H. COHEN and R. G. MAGE, 1986 Localisation of rabbit kappa light chain allootypic determinants. *Mol. Immunol.* **23**: 476–488.
- NEI, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- NEI, M., and T. GOJOBORI, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**: 418–426.
- NEI, M., and L. JIN, 1989 Variance of the average numbers of nucleotide substitutions within and between populations. *Mol. Biol. Evol.* **6**: 290–300.
- O'BRIEN, S. J., and F. EVERMANN, 1988 Interactive influence of infectious disease and genetic diversity in natural populations. *Trends Evol.* **3**: 254–259.
- LOUDIN, J., 1960 Allotopy of rabbit serum proteins. I. Immunol.

- chemical analysis leading to the individualization of seven main allotypes. *J. Exp. Med.* **112**: 107–124.
- PADLAN, E. A., 1977 Structural implications of sequence variability of immunoglobulin genes. *Proc. Natl. Acad. Sci. USA* **74**: 2551–2555.
- PADLAN, E. A., 1985 Quantitation of the immunogenic potential of protein antigens. *Mol. Immunol.* **22**: 1284–1254.
- POTTS, W., C. J. MANNING and E. W. WAKELAND, 1991 Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature* **352**: 619–621.
- REISFELD, R. A., S. DRAY and A. NISONOFF, 1965 Differences in amino acid composition of rabbit IgG-immunoglobulin light chains controlled by allelic genes. *Immunochemistry* **2**: 155–167.
- SHARP, P. M., and W.-H. LI, 1986 An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* **24**: 28–39.
- SHEPPARD, H. W., and G. A. GUTMAN, 1981 Allelic forms of rat kappa-chain genes: evidence for strong selection at the level of the nucleotide sequence. *Proc. Natl. Acad. Sci. USA* **78**: 7062–7068.
- STAVENEZER-NORDGREN, P., R. KEKISCH and R. ZEGERS, 1985 Molecular defects in a human immunoglobulin kappa-chain deficiency. *Science* **230**: 458–461.
- STEPHENS, J. C., 1985 Statistical Methods of DNA sequence analysis: detection of intragenic recombination or gene conversion. *Mol. Biol. Evol.* **1**: 593–556.
- SUOMALAINEN, H. A., L. BRYNS, W. VAN DER LOO and J. SCHRÖDER, 1983 Karyotype of a mouse/rabbit hybrid clone secreting rabbit immunoglobulin light chain. *Hereditas* **98**: 83–87.
- TAJIMA, A., and M. NEI, 1984 Estimation of evolutionary distance between nucleotide sequences. *Mol. Biol. Evol.* **1**: 269–285.
- TAKAHATA, N., 1990 A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proc. Natl. Acad. Sci. USA* **87**: 2419–2423.
- TAKAHATA, N., and M. NEI, 1990 Allelic genealogy under overdominance and frequency-dependent selection and Polymorphism of the Major Histocompatibility Complex loci. *Genetics* **124**: 967–978.
- TANAKA, T., and M. NEI, 1989 Positive darwinian selection at the variable region genes of immunoglobulins. *Mol. Biol. Evol.* **6**: 477.
- TANIGUCHI, T., T. MIZUOCHI, M. BEALE, R. A. DWEK, T. W. RADEMAKER and A. KOBATA, 1985 Structure of sugar chains of rabbit immunoglobulin: occurrence of asparagine-linked sugar chains in Fab fragment. *Biochemistry* **24**: 5551–5557.
- VAN DER LOO, W., 1975 Allelic polymorphisms and concomitant evolution of cofunctional genes of the lagomorph immunoglobulins. Ph.D. Thesis, Vrije Universiteit Brussel, Brussels.
- VAN DER LOO, W., 1987 Studies on the adaptive significance of the immunoglobulin alleles (Ig allotypes) in wild rabbit, pp. 164–190 in *The Rabbit in Contemporary Immunological Research*, edited by S. DUBISKI. John Wiley & Sons, New York.
- VAN DER LOO, W., 1992 Variance analysis of Immunoglobulin alleles in natural populations of Rabbit (*Oryctolagus cuniculus*): the extensive interallelic divergence at the *b*-locus of the Immunoglobulin Light chain constant region could be the outcome of balancing selection. *Genetics* (in press).
- VAN DER LOO, W., N. FERRAND and R. SORIGUER, 1991 Estimation of gene diversity at the *b*-locus of the constant region of the immunoglobulin light chain in natural populations of European rabbit (*Oryctolagus cuniculus*) in Portugal, Andalusia and on the Azorean Islands. *Genetics* **127**: 789–799.
- VAN DER LOO, W., C. P. ARTHUR, B. RICHARDSON, M. WALLAGE and R. HAMERS, 1987 Nonrandom allele associations between unlinked protein loci: Are the allotypes of the immunoglobulin constant regions adaptive? *Proc. Natl. Acad. Sci. USA* **84**: 3075–3079.
- VERDOODT, B., 1991 Studie van de evolutiepatronen van de polymorfismen van de immunoglobulines van konijn (*Oryctolagus cuniculus*) gebaseerd op DNA-sequentievergelijkingen. Graduate Thesis, Vrije Universiteit Brussel.

Communicating editor: W.-H. LI

APPENDIX

Note on the phylogeny of *b9* allele of the rabbit *b*-locus of the immunoglobulin κ L-chain: Among K_1 alleles, *b9* shows clearly the strongest homology to the K_2 gene (Figures 1–3). At numerous positions, *b9* is identical to K_2 but differs from the other K_1 alleles *b4*, *b4var*, *b5*, *b95* (below “ K_{11} ” alleles). This will be noted: $K_2 = b9 \diamond K_{11}$. At the same time there is a fair number of sites where $K_2 \diamond b9 = K_{11}$. This observation lead to the hypothesis that the extensive divergence between *b*-locus alleles was due to gene conversion between c_{k2} and c_{k1} genes (AKIMENKO MARIAME and ROUGEON 1986). A closer analysis of the sequence data indicates that the evidence for gene conversion is weak. First, there is no significant clustering of the positions at the *b9* gene where nucleotides are shared either with K_{11} or with K_2 alleles ($P > 0.3$, not shown). Second, as will be outlined here, the *b9* allele might not be monophyletic with the K_{11} alleles but derived from the K_2 gene after the separation of K_{11} and K_2 lineages.

At most positions where $K_2 \diamond b9 = K_{11}$, the *b9* sequence displays what seems to be the ancestral (pleisiomorphic) form (examples are nucleotide positions 19, 45, 49, 191, 204, 321). This indicates mutations in the K_2 lineage rather than a common history of *b9* and K_{11} after the K_1 - K_2 separation. On the other hand, there are 4 sites (48, 190, 200, 310) where $K_2 \diamond b9 = K_{11}$ and where K_2 clearly shows the mammalian consensus nucleotide. These mutations might not be random: The A at nucleotide 190 is at the first position of an ancestral Ser-171(AGC) triplet and was apparently conserved in Asn-171(AAT) of the K_2 gene. All K_1 alleles show the characteristic Cys-171(TGT) triplet at this position. The substitution at nucleotide 200 changes an ancestral Ser-173(AGC) into a K_1 specific Asn173(AAC) triplet. Cys171 forms part of the disulfide bridge of the K_1 L-chain linking the c_{k1} domain to the v_{k1} domain, while Asn-173 is known to play a role as a glycosylation site in this L-chain loop (TANIGUSHI *et al.* 1985). The K_1 specific Asn173 could therefore compensate for the (loss of) Asn-171 found in K_2 . Clearly, these two mutations could reflect convergent evolution imposed by functional constraints related to the emergence of the supplementary interdomain disulfide bridge in stead of common ancestry. At positions 48 and 310, the *b9* and K_{11} genes share a synonymous T. Also this sharing does not imply common history as synonymous substitutions between K_2 and K_1 appear to be directional (GC

→ AT). The observed situation can indeed be explained by the increase in d_s in the $c_{\kappa 1}$ lineages $b9$ and K_{11} if we assume (1) that the K_2 lineage was not affected by such increase and (2) that the K_2 , $b9$ and K_{11} lineages were derived from an ancestor gene showing, like other mammalian c_κ genes, a GC > AT bias in synonymous codon usage.

Of some 20 sites where $K_2 = b9 \leftrightarrow K_{11}$, there are at least five that show the "consensus" nucleotide in K_{11} alleles (positions 15, 28, 150, 179, 248) and at three more sites this is likely (217, 269, 292). This is rather strong evidence for a common history of the $b9$ and the K_2 genes after separation from the K_{11} lineage (Figure 2). Under this assumption, the Cys-171 and Asn-173 shared between $b9$ and K_{11} could be explained by gene conversion or by a recapitulation of the events which created the ancestor of the K_{11} genes, as suggested above. The latter hypothesis was proposed in 1975, prior to the discovery of the K_2 L-chain subclass and before DNA sequence data were

available.¹ In this scenario, $b9$ might have evolved by a more recent duplication of $c_{\kappa 2}$ in a genome which lost, or never harbored, the $c_{\kappa 1}$ type. The $b9$ allele should then be considered pseudoallelic to the K_{11} genes. The proposition of the most likely phylogenetic tree for the K_1 alleles might have to await further sequence analyses of c_κ genes in lagomorphs. At any rate, because the different K_1 allotypes *behave* like alleles, diversity enhancing selection forces should influence the mutant acceptance rates among these genes much in the same way as if they were truly allelic.

¹ It was suggested that a selection pressure existed (possibly due to existing v_κ genes with a free cysteine residue) favoring the emergence in the rabbit genome of a c_κ gene that would allow formation of the supplementary disulfide bridge (*i.e.*, of a c_κ gene with a Cys-171) and it was further argued that this pressure was strong enough to make this happen more than once (in different isolated populations, on different members of a hypothetical c_κ gene family). As soon as a Cys-171 containing c_κ gene existed in the genome, this pressure should vanish, leaving c_κ genes without Cys-171 (*i.e.*, $c_{\kappa 2}$) unexpressed along with expressed genes of the $c_{\kappa 1}$ type behaving like alleles (VAN DER LOO 1975).