Available online at www.sciencedirect.com
SCIENCE DIRECT
Molecular Immunology 43 (2006) 731–744
Molecular Immunology
www.elsevier.com/locate/molimm

Substitution patterns in alleles of immunoglobulin V genes in humans and mice

Tania Romo-González, Enrique Vargas-Madrazo *

Instituto de Investigaciones Biológicas, Universidad Veracruzana, 2a Schubert No. 4 Indeco Animas, Xalapa, Ver., C.P. 91190, Mexico

Received 1 March 2005
Available online 1 June 2005

Abstract

Immunoglobulins (Igs) constitute a subfamily of rapidly evolving proteins. It is postulated that this characteristic is due mainly to the participation of these proteins in highly diverse functions of recognition and defense. Although this vision of rapid evolution in Igs is widely accepted, various studies have demonstrated that diverse and contradictory forces not yet completely understood converge in the evolution of these receptors.

In a recent study of the substitution patterns in the alleles that form the human IGHV locus, we found that the variation in genetic and structural information does not occur homogeneously among the different genes, nor among the regions and positions conforming said locus. In view of these results and of the importance of a better understanding of the basic evolutionary process in specific receptors (such as Igs) for both immunology and molecular evolution, it is important to explore the nature of the diversification process in these proteins in detail. In this work, therefore, we analyzed the substitution patterns in all the alleles reported for loci IGKV and IGLV in humans and mice, and we compared the results with those previously observed in the human IGHV locus. We found that the process of evolutionary variation of the Igs reflect the diversity of selective pressures operating on the different loci, genes, sub-regions and positions; for example, diversification through substitution is generally centered on CDRs, but only few positions inside the CDRs were frequently substituted. In spite of this general tendency, it is possible to observe differences in the degree of diversification among loci, families and genes. These tendencies to modify only certain attributes of IGV genes seem to be in agreement with differential strategies associated with the restrictions of the molecular immune recognition mechanism. The complexity of the evolutionary patterns observed in this study leads us to think that the predispositions observed herein may also be due in part to processes of DNA dynamics.

Keywords: Patterns of nucleotide substitutions; Protein evolution; Immunoglobulins; V genes; Immune recognition

1. Introduction

One of the basic processes in the evolution of DNA sequences is the change of nucleotides over long periods of time; therefore, the study of these variations has been one of the fundamental topics of molecular evolution (Li, 1997). This type of analysis has made it possible not only to understand the mechanisms by which DNA evolves, but also to estimate the evolutionary patterns for reconstructing the history of organisms (Harvey et al., 1994; Tourasse and Gouy, 1997; Li, 1997; Takano, 1998). In addition, the analysis of this evolutionary process has permitted the observation of phylogenetic differences among several types of genes. As an example, it is well known that certain structural proteins such as histones, actins and ribosomal proteins are extremely conserved; whereas others have evolved intermediately (hemoglobin, myoglobin, carbonic anhydrases, etc.) or rapidly (Igs, MHC, interferons, interleukines, etc.) (Dayhoff, 1972; Ferguson, 1980). This rate of evolution and its substitution patterns have been correlated with the function or functions developed by the proteins (Zuckerkandl, 1976).

* Corresponding author. Tel.: +52 228 8418900x13401; fax: +52 228 8418911x15911.
E-mail addresses: romisnaider@yahoo.com.mx, envargas@yahoo.com (E. Vargas-Madrazo).

0161-5890/$ – see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.molimm.2005.03.018

As we mentioned, MHCs and Igs are the proteins with the highest evolutionary rates (Wilson et al., 1977; Kataoka et al., 1982; Perlmutter et al., 1985; Klein et al., 1993; Klein, 1986). Some estimates indicate that their rate of evolution is 500 times faster than that of histones, which are the slowest-evolving proteins (Ferguson, 1980). This tendency toward rapid evolution seems to be associated with the need for generating sufficient diversity to contend with an enormous spectrum of antigens (Ag).

Although this vision of rapid evolution of MHCs and Igs is widely accepted, some studies show results that contradict this logic; for example, Booker and Haughton (1994) found that some Ig genes present an extremely low rate of evolution. Likewise, phylogenetic studies of various species have shown that Ig families have remained quite stable in time, a moderate degree of similarity among the Ig sequences in different species being observable (Andersson and Matsunaga, 1995; Ota et al., 2000). Similar behavior has been noted in other multigenetic antigen-binding families (Litman et al., 1999). The evidence indicating the convergence of diverse and contradictory forces in the evolution of Igs demonstrates the importance of delving more deeply into the nature of the diversification processes occurring in these proteins.

The diversification of the recognition capacity of Igs is mainly the result of changes in the sequence that modify the characteristics of the antigen-binding site, that is, of variation in the general form of this site and in the composition of the amino acids of the hypervariable loops (Wilson and Stanfield, 1993). From this perspective, the number of possible antigen-binding sites would be almost unlimited if the Igs had a high rate of mutations with respect to the postulated time. Nevertheless, diverse studies have shown that the hypervariability presented by the Igs is not as random and elevated as had been postulated (Cocho et al., 1993; Vargas-Madrazo et al., 1994; Lara-Ochoa et al., 1996), due to restrictions operating at both functional and structural levels.

The antigen-binding site of Igs consists of six hypervariable loops: three of VH and three of VL, named H1, H2, H3 and L1, L2, L3, respectively (Poljak et al., 1973). Although these regions are hypervariable, it has been demonstrated that, with the exception of H3, the remaining five hypervariable loops adopt only a small number of canonical conformations or canonical structures (CS) (Chothia and Lesk, 1987; Wilson and Stanfield, 1993). Besides, not all the combinations of CS types exist in the sequences of known antibodies, only a few having been found thus far (Vargas-Madrazo et al., 1995). These possible combinations or canonical structure classes (CSC) have been denominated structural repertoire (Chothia et al., 1992; Tomlinson et al., 1995; Vargas-Madrazo et al., 1995). In addition, it has been suggested that the forms of the antigen-binding site present in the structural repertoire correlate with the type of antigen recognized (Martin et al., 1991; Vargas-Madrazo et al., 1995; Lara-Ochoa et al., 1996; MacCallum et al., 1996), that is, a correlation between the form of the antigen-binding site and the type of ligand recognized would point to the existence of additional restrictions on the evolution of these proteins.

The germline genes of Igs are subject to evolving processes (presumably rapid) that introduce changes in the sequence to make the immune system more apt or adaptable. Within these processes we find gene duplication, allelic polymorphism and germline conversion (Pascual and Capra, 1991; Sitnikova and Su, 1998). Although much has been said about the high degree of polymorphism occurring in Ig genes, few studies evaluate the content of genetic and structural information of the sequences at the level of allelic polymorphism in detail. In most studies on the evolution of Ig genes, the distance between sequences is considered as a measure of temporal separation, it being tacitly supposed that the probabilities of substitution are uniform and invariant in all positions (Eigen et al., 1988). The studies realized under these premises have generally been applied to genes of invariant or slightly variable proteins, since it is more plausible to assume sequence homogeneity in such a situation. Nevertheless, it is well known that diversification in the variable regions of Igs is not distributed homogeneously in all positions (Wu and Kabat, 1970; Vargas-Madrazo et al., 1994). From this point of view, variability can range from total conservation in some positions to hypervariable in others (Eigen et al., 1988). In addition, a closer view of biological complexity at the macromolecular level tells us that not all the amino acid substitutions in a given protein have the same impact on evolution (Zuckerkandl, 1976). As we have already mentioned, Igs are subject to a vast and complex menu of different selective forces. From this perspective, those studies on proteins with a high mutational rate (Igs, for example), in which the content of the genetic and structural information of the sequences is considered (vertical and horizontal analyses), can contribute to a more profound understanding of the nature of evolutionary processes.

Recently, we studied the substitution patterns in alleles composing the IGHV locus of humans (Romo-González and Vargas-Madrazo, 2005), which is one of the most studied of loci, since all of its members are mapped, sequenced and classified in phylogenetic trees (Krawinkel et al., 1989; Tomlinson et al., 1992; Milner et al., 1995).

Through the structural analysis of the substitution patterns in alleles of this locus, we found that the genetic and structural information is not homogeneously distributed among the genes, nor among the distinct regions that codify for them. Explicitly, we observed that: (i) not all the genes are equally substituted; (ii) among the different families and genes of the IGHV, there are preferences for mutating only certain regions (FRs or CDRs) and positions and (iii) some of these preferences are common to the majority of these genes (Romo-González and Vargas-Madrazo, 2005). These results suggest that the diversification by substitution in alleles of the human IGHV locus is an evolutionary phenomenon that creates diversity in genes and organisms through complex forces

and selective pressures of diverse origin (Tutter et al., 1991; Nagylaki and Petes, 1982; Reynaud et al., 1989; McCormack and Thompson, 1990; Rothenfluh et al., 1995). The combination of evolutionary factors molding genetic diversity operates under a certain preferential strategy of change associated with diverse aspects of the functions of the locus (Zuckerkandl, 1976).

Given that a better understanding of the basic evolutionary processes in the repertoires of specific receptors such as Igs is important for both immunology and molecular evolution, and in view of the foregoing results, it becomes necessary to study various loci of the Ig protein family and to compare the results obtained in different species. As we mentioned previously, the antigen-binding site is composed of six hypervariable loops, three of VH and three of VL; therefore, not only the heavy chain participates in the conformation of the antigen-binding site and in the recognition of Ags (Wilson and Stanfield, 1993; Vargas-Madrazo et al., 1995; Saul and Poljak, 1993; Vargas-Madrazo and Paz-Garcia, 2003). Since the functional properties of the VL domain vary in respect to the VH, it would be interesting to contrast the properties observed in the human IGHV with the results reported here in VL loci IGKV and IGLV. In addition to being the two species most studied in immunology, humans and mice present many similarities in their mechanism of Ig diversification (Weill and Reynaud, 1996) and in the characteristics of their structural repertoire (Almagro et al., 1997, 1998); consequently, it is also interesting to compare the results for the Ig loci in these two species. We also conducted studies of substitution patterns in the alleles of the IGHV locus in mice, but since we obtained results contrasting strongly with those for other Ig loci in humans and mice, we decided to report them in a separate article (now in preparation).

2. Methods

2.1. Construction of the database

Starting from the IMGT database (http://imgt.cines.fr), all the alleles reported (up to November 2003) for the IGKV and IGLV loci in both humans and mice were compiled and compared in detail with the predominant allele in each gene segment. We aligned each of the alleles with the allele representing each gene and, following the criterion of maximum homology, we assigned the sequences to the corresponding allele. An analysis of the sequences made it possible to reassign some of the alleles that presented errors in their assignment in the original database due to problems of alignment. In order to have qualitative elements for this study, each of the original articles was reviewed to obtain the information relevant to each sequence. The original sequences were checked; the experimental conditions, the source of the DNA, and other data were collected in order to evaluate the quality of the database.

2.2. Classification of the type of replacement

For each of the substitutions reported in the alleles, the type of amino acid substitution (alterations of physico-chemical properties) presented in each residue was analyzed; this was done in accordance with the system of the grouping and analysis by Grantham, 1974, Go and Miyazawa (1980) and Romo-González and Vargas-Madrazo (2005). In these systems, the alteration of physico-chemical properties is mainly determined by the composition, polarity, molecular volume, exteriority and interiority of the lateral chains. Considering all these characteristics, we classified the amino acid substitutions into three groups: (i) conservative, (ii) non-conservative and (iii) radical.

2.3. Calculating the R/S ratio

Due to redundancy in the genetic code, base pairs changes in a codon may yield either a replacement (R) of one amino acid by another or preservation of the same residue (a silent (S) mutation). Because of this, it is possible to characterize the evolutionary forces shaping the diversification of the different sub-regions of a gene by studying the replacement and silent substitutions ratio (*R*/*S* ratio). The *R*/*S* ratio was calculated by dividing the replacement substitutions by the total number of silent substitutions (*R*/*S*) (Jukes and King, 1979; Shlomchik et al., 1987). The substitutions are counted as nucleotide changes found in an allele with respect to the predominant allele in a gene. Codons undergoing random mutation are predicted to yield an *R*/*S* ratio of 2.9. Values below 2.9 indicate conservation and those above 2.9 diversification (Jukes and King, 1979; Shlomchik et al., 1987).

3. Results

3.1. Analysis of allelism by gene

In Tables 1 and 2 we report those genes in which alleles have been found. This database includes: (i) 13 alleles from human locus IGKV, which come from 12 of the 37 functional genes that make up the locus (Barbie and Lefranc, 1998, 1998); (ii) 27 alleles from mouse locus IGKV, which come from 10 of the 96 functional genes (Martinez-Jean et al., 2001); (iii) 38 alleles from human locus IGLV, which come from 22 of the 33 functional genes (Pallares et al., 1998); (iv) six alleles from mouse locus IGLV, which come from five of the eight functional genes (Scaviner and Guiraudou, 1999). It is important to note that in mouse locus IGKV, the amount of alleles is large if it is compared with the other three loci, and specially because these alleles come from only 10 genes (Table 1). The evaluation of polymorphism by gene and locus, indicates that the number of alleles found in the κ and λ loci of both species is very limited, all the more so in comparison with that observed in human IGHV locus, in which 158 alleles were found in 42 of the

Table 1
Number of alleles per gene reported for locus Igκ V in humans and mice

Clans and families		Number of germline genes	Number of allelic segments	Gene name	Number of alleles
Human					
	IGKV1	18	6	1–5	2
				1-D-12	1
				1–13	1
				1-D-16	1
				1–17	1
Clan I				1–39	1
	IGKV3	7	3	3–11	1
				3-D-15	1
				3–20	1
	IGKV4	1	0	0	0
	IGKV5	1	0	0	0
	IGKV2	10	3	2–29	1
Clan II				2-D-29	1
				2–40	1
Total		37	12		13
Mouse					
Clan I	IGKV8	8	1	8–28	1
	IGKV7	1	0	0	0
	IGKV1	8	2	1–110	1
				1–117	1
Clan II	IGKV2	4	2	2–109	3
				2–112	1
Clan III	IGKV3	9	0	0	0
Clan IV/VI	IGKV4	24	0	0	0
	IGKV5	5	0	0	0
	IGKV6	12	1	6–32	1
	IGKV9	4	0	0	0
	IGKV10	3	2	10–94	9
				10–96	7
Clan V	IGKV11	1	0	0	0
	IGKV12	7	1	12–41	1
	IGKV13	2	1	13–85	2
	IGKV14	3	0	0	0
	IGKV16	1	0	0	0
	IGKV18	1	0	0	0
	IGKV19	1	0	0	0
Clan VII	IGKV17	2	0	0	0
Total	96	10			27

51 functional genes (Romo-González and Vargas-Madrazo, 2005).

In order to acquire greater confidence and control over the sequences under study, we decided to go to the original source of each sequence to analyze diverse data regarding the experimental origin of the information. The following considerations show the quality of the database:

(i) Genes KV1-5, KV2-29, KV2-109, KV10-96, LV1-40, LV2-11, LV1 and LV2 have been studied independently by different research groups, the same allele having always been found for each gene (Lefranc and Lefranc, 2001).
(ii) Nine different alleles have been reported in three independent studies for gene KV10-94 (Lefranc and Lefranc, 2001); for gene KV2-107 on the contrary, no alleles have been found in four independent studies (Lefranc and Lefranc, 2001).
(iii) In the case of those loci with the greatest number of reported alleles (KV10-94, KV10-96, KV2-109, LV2-14 y LV2-18), the different alleles are the result of independent studies (Lefranc and Lefranc, 2001).

The foregoing observations indicate that the allelic diversity among genes is, at least partly, independent of the experimental origin of the sequences.

The fifth and sixth columns of Tables 1 and 2 show the number of alleles reported for each gene. In contrast to the great variability (1–18 alleles) in the number of alleles per gene observed in the human IGHV locus (Romo-González and Vargas-Madrazo, 2005), in loci IGKV and IGLV, both in humans and mice, most of the genes presented only one or two alleles. Only genes KV10-94, KV10-96 and KV2-109 (with 9, 7 and 3 alleles, respectively) of the Vκ locus in mice, and LV2-14 and LV2-18 (with three alleles each) of the Vλ locus in humans were the outstanding exceptions. In any case,

Table 2
Number of alleles per gene reported for locus IgλV in humans and mice

Clans and families		Number of germline genes	Number of allelic segments	Gene name	Number of alleles
Human					
				1–40	2
	IGLV1	5	3	1–47	1
				1–51	1
				2–8	2
				2–11	2
				2–14	3
	IGLV2	5	6	2–18	3
				2–23	1
Clan I				2–33	2
				3–9	2
				3–10	1
	IGLV3	10	5	3–12	1
				3–21	2
				3–25	2
	IGLV6	1	0	0	0
	IGLV10	1	1	10–54	2
				4–60	2
	IGLV4	3	2	4–69	1
				5–39	1
Clan II	IGLV5	3	2	5–45	2
	IGLV9	1	1	9–49	2
	IGLV11	1	0	0	0
	IGLV7	2	1	7–46	1
Clan III	IGLV8	1	1	8–61	2
Total		33	22		38
Mouse					
IGLV1		2	2	1–1	1
				1–2	1
IGLV2		1	0	0	0
IGLV3		5	3	3–4	1
				3–6	2
				3–7	1
				3–8	1
Total		8	5		6

it is generally notable that diversity regarding the number of alleles reported is not equally distributed among the genes.

3.2. Number of substitutions per allele

As we stated in the structural analysis of substitution patterns in the alleles of the human IGHV locus (Romo-González and Vargas-Madrazo, 2005), the number of alleles per locus is not the only essential parameter in the evolutionary diversification of genetic information, since, for example, an allele may present a large number of mutations with respect to the principal allele, whereas another may differ only in a single nucleotide.

In Figs. 1 and 2 we report the number of substitutions (silent and replacement) per allele for loci IGKV and IGLV in humans and mice, respectively. While the number of alleles per gene is quite homogeneous, the great discrepancy existing in the number of substitutions per allele among some genes is very noticeable. When the total substitutions (silent and replacement) are considered, variations ranging from one to nine nucleotide changes among the four loci can be observed. Additionally, it should be noted that the amount of substitutions in the four loci analyzed herein is significantly less than that in the IGHV of humans (Romo-González and Vargas-Madrazo, 2005). The values for the number of maximum substitutions are as follows (the average is reported in parentheses):

	VH human	Vκ		Vλ	
		Human	Mouse	Human	Mouse
Totals	16 (3.4)	4 (1.8)	9 (4.3)	7 (2.0)	6 (3.0)
Replacement	12 (2.3)	2 (1.1)	5 (2.0)	5 (1.3)	5 (1.9)

It can be seen that the alleles in the human IGHV locus are much more mutated (ranging from 1 to 16 substitutions) than the ones studied here (human IGKV 1–4, human IGLV 1–7, mouse IGKV 1–9 and mouse IGLV 1–6 substitutions). Although the differences among the loci are notable, it is interesting to observe that, starting from Figs. 1 and 2, the proportion of alleles with 1–2 total substitutions constitutes a majority in almost all of the loci. The only exception is

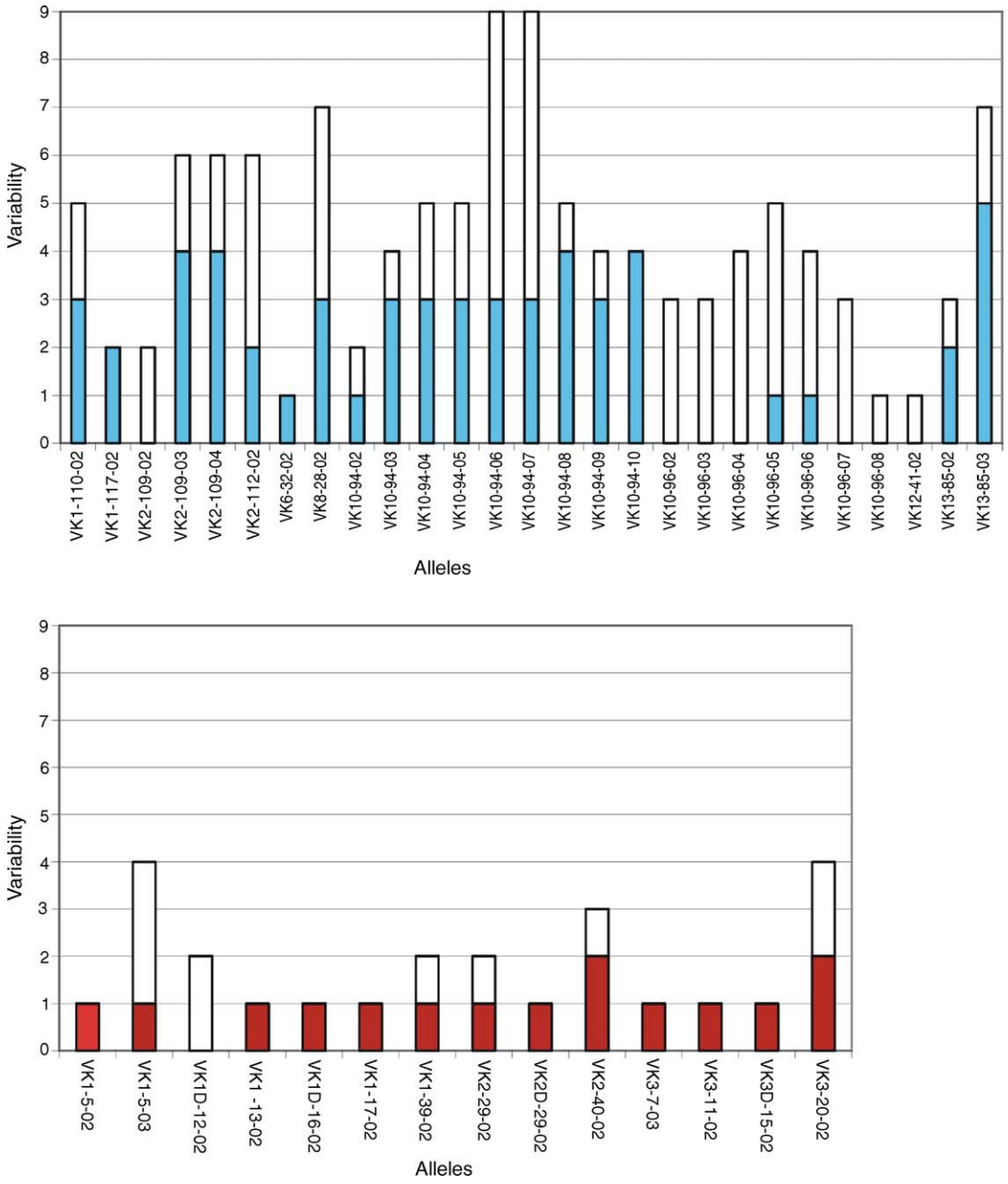

Fig. 1. Number of substitutions occurring by species in each allele in locus IgκV. For both species, all the substitutions are shown in white, while only those changes leading to a replacement at the amino acid level are represented in blue for mice and in red for humans. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

IGLV in the mouse, in which four of the six alleles present more than three substitutions.

These results show that Ig diversification through allelic variation is heterogeneous, for in spite of the marked differences existing among genes, loci and species, they share a common evolutionary strategy and history of mutation.

3.3. Type and localization of substitutions per allele

To create diversity at the protein level, not only is the number of substitutions important, but also the type of amino acid substitution (alteration of the physico-chemical properties) present in each residue. In order to characterize this type of variation in the content of genetic and structural information in the genes, therefore, we classified the type of replacement in three groups: (i) conservative, (ii) non-conservative and (iii) radical (see details in Grantham, 1974; Go and Miyazawa, 1980; Romo-González and Vargas-Madrazo, 2005).

From a global estimate of the type of substitution occurring in the four loci studied, we found that 35% imply a conservative amino acid modification, 55% non-conservative and

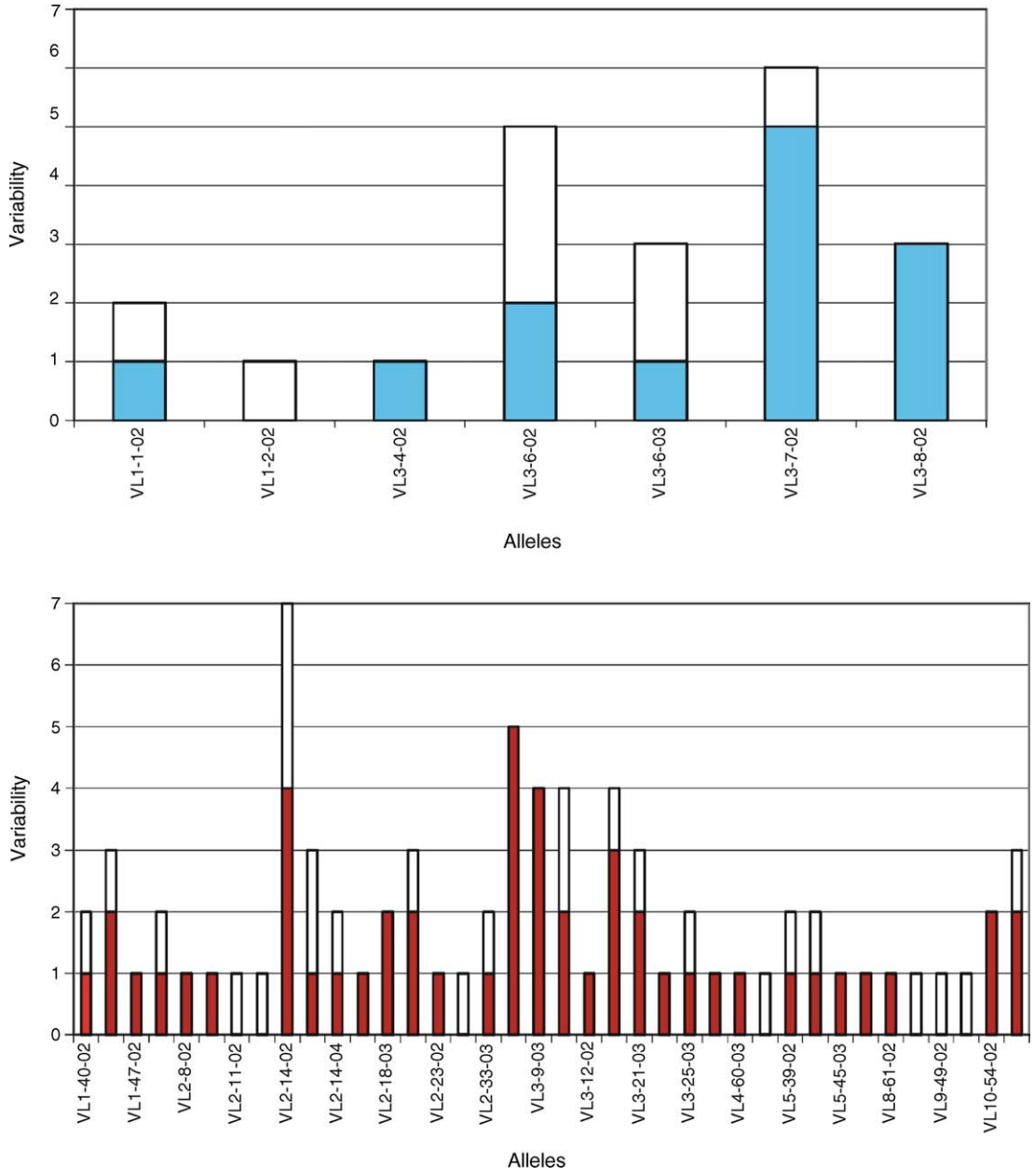

Fig. 2. Number of substitutions occurring by species in each allele in locus IgλV. For both species, all the substitutions are shown in white, while only those changes leading to a replacement at the amino acid level are represented in blue for mice and in red for humans. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

9% radical. An analysis of each one of the loci taken separately produced results (not shown) consistent with the foregoing figures. Thus, more than half of the amino acid substitutions (64%) in the four loci analyzed imply partial or drastic alterations in physico-chemical properties (non-conservative and radical), which suggest a tendency toward diversification. These results are in contrast with those observed in the human IGHV locus, in which a preponderance of conservative substitutions (61%) was found (Romo-González and Vargas-Madrazo, 2005). The previous results show that replacement in the different alleles is not random; on the contrary, they show tendencies toward the partial or drastic alteration of the physico-chemical properties of the amino acids. From the foregoing, it seems interesting to analyze whether the types of substitutions are distributed equally in any sub-region (FR1, FR2, FR3, CDR1, CDR2) or residue of the variable exons studied here.

A tool for differentiating the degree of variability among substitutions per sub-region is the calculation of ratio of the replacement/silent (*R*/*S*) substitutions (see details in Jukes and King, 1979; Shlomchik et al., 1987; Kirkham and Schroeder, 1994; Ota et al., 2000). This analysis is pertinent in view of the results obtained in the human IGHV locus, in which the *R*/*S* ratio indicates se-

Table 3
R/S ratio for sub-segments of VL domain in humans and mice

Clans and families		Number of genes	Number of alleles	FR1	CDR1	FR2	CDR2	FR3
Human								
Clan I	IGKV1	18 (6)[a]	7	2.0	1/0[b]	1/0	0.3	0.5
	IGKV3	6 (3)	3	2/0	1/0	0/3[c]	–[d]	2.0
	IGKV4	1 (0)	0	–	–	0	–	–
	IGKV5	1 (0)	0	–	–	0	–	–
Clan II	IGKV2	10 (3)	3	–	–	2.0	–	2/0
Total		37 (12)	13	2.0	2.0	0.8	0.3	1.8
Mouse								
Clan I	IGKV8	8 (1)	1	1/0	1.0	1.0	–	0/2
	IGKV7	1 (0)	0	–	–	–	–	–
Clan II	IGKV1	8 (2)	2	1.0	1.0	0	2.0	1/0
	IGKV2	4 (2)	4	0/2	2/0	0.8	3/0	0.5
Clan III	IGKV3	9 (0)	0	–	–	–	–	–
Clan IV/VI	IGKV4	24 (0)	0	–	–	–	–	–
Clan V	IGKV5	5 (0)	0	–	–	–	–	–
	IGKV6	12 (1)	1	1/0	–	–	–	–
	IGKV9	4 (0)	0	–	–	–	–	–
	IGKV10	3 (2)	16	0/4	3.2	0.3	9/0	0/22
	IGKV11	1 (0)	0	–	–	–	–	–
	IGKV12	7 (1)	1	0/1	–	–	–	–
	IGKV13	2 (1)	2	2/0	4/0	0/1	–	0.5
	IGKV14	3 (0)	0	–	–	–	–	–
	IGKV16	1 (0)	0	–	–	–	–	–
	IGVK18	1 (0)	0	–	–	–	–	–
	IGKV19	1 (0)	0	–	–	–	–	–
Clan VII	IGKV17	2 (0)	0	–	–	–	–	0
Total		96 (10)	27	1.8	2.3	0.4	14/0	0.1

IGKV genes grouped according to family and clan.
[a] The number of genes integrating each family is specified. The number of genes presenting alleles is reported in parentheses.
[b] No silent substitutions were found for some regions and families, while for others no replacement substitutions were found. In such cases, the number of replacement and silent substitutions is explicitly reported (numerator and denominator, respectively).
[c] Those positions for which no substitution of any kind was encountered are indicated with a hyphen.

Table 4
R/S ratio for sub-segments of VL domain in humans and mice

Clans and families		Number of genes	Number of alleles	FR1	CDR1	FR2	CDR2	FR3
Human								
Clan I	IGLV1	5 (3)[a]	4	1.0	–[d]	0/1[c]	2/0[b]	2/0
	IGLV2	5 (6)	13	1.0	0.4	0/2	5/0	4.0
	IGLV3	10 (5)	8	1.7	5.0	2/0	5.0	1.0
	IGLV6	1 (0)	0	–	–	–	–	–
	IGLV10	1 (1)	1	0/1	1/0	2.0	–	1/0
Clan II	IGLV4	3 (2)	3	–	–	–	0/1	2.0
	IGLV5	5 (2)	3	2/0	0/2	0/1	–	–
	IGLV9	1 (1)	2	–	–	0/1	–	0/1
	IGLV11	1 (0)	0	–	–	–	–	–
Clan III	IGLV7	2 (1)	1	–	–	–	–	1/0
	IGLV8	1 (1)	2	–	–	–	1.0	–
Total		33 (22)	38	1.7	1.3	2.2	6.0	2.0
Mouse								
IGLV1		2 (2)	2	–	1.0	0/1	–	–
IGLV2		1 (0)	0	–	–	–	–	–
IGLV3		5 (4)	5	2/0	2.5	0.3	3.0	–
Total		8 (5)	7	2/0	2.0	0.3	2.0	–

IgλV genes grouped according to family and clan.
[a] The number of genes integrating each family is specified. The number of genes presenting alleles is reported in parentheses.
[b] No silent substitutions were found for some regions and families, while for others no replacement substitutions were found. In such cases, the number of replacement and silent substitutions is explicitly reported (numerator and denominator, respectively).
[c] Those positions for which no substitution of any kind was encountered are indicated with a hyphen.

lective pressure for diversification in the CDRs and in some codons in particular (Romo-González and Vargas-Madrazo, 2005), the latter being associated with residues frequently in contact with the Ag (MacCallum et al., 1996).

At this point, it is important to note that although up to now there are more complex methods of estimating the *R*/*S* ratio (Li, 1993; Yang and Nielsen, 2000; Massingham and Goldman, 2005), there is also evidence in which the use of different methods, including the "classical"

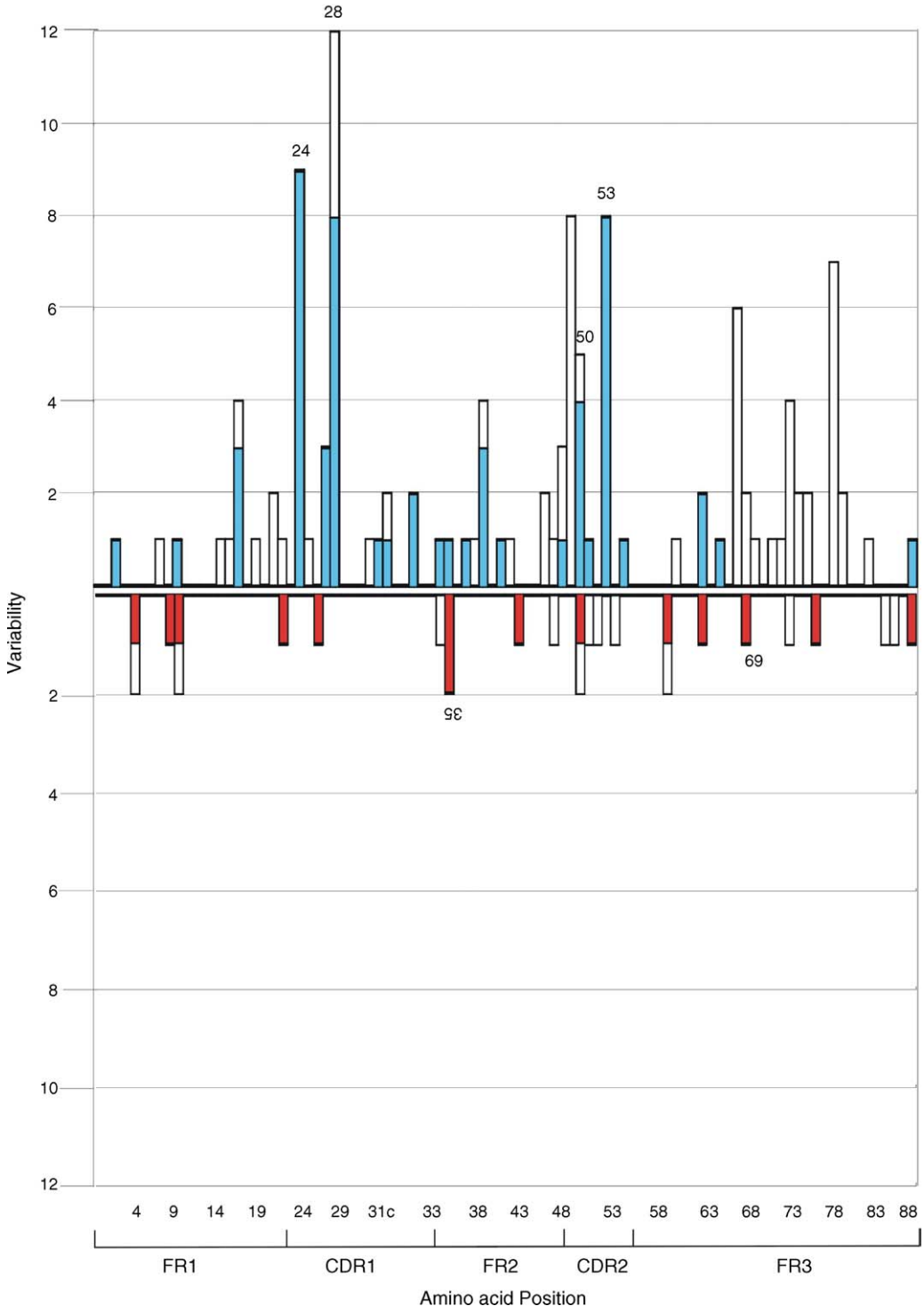

Fig. 3. Substitutions occurring by position in the Vκ exon of positions 1–88. For both species, all the substitutions are shown in white, while only those changes leading to a replacement at the amino acid level are represented in blue for mice and in red for humans. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

used here, yielded essentially identical results (Hughes, 2004).

The results of calculating the *R*/*S* ratio per family and sub-region are summarized in Tables 3 and 4. The analysis of the total *R*/*S* ratio per sub-region (last row of Tables 3 and 4) showed differences with respect to the human IGHV locus (Romo-González and Vargas-Madrazo, 2005), that is, the *R*/*S* ratio for the VL domain was generally below the point of

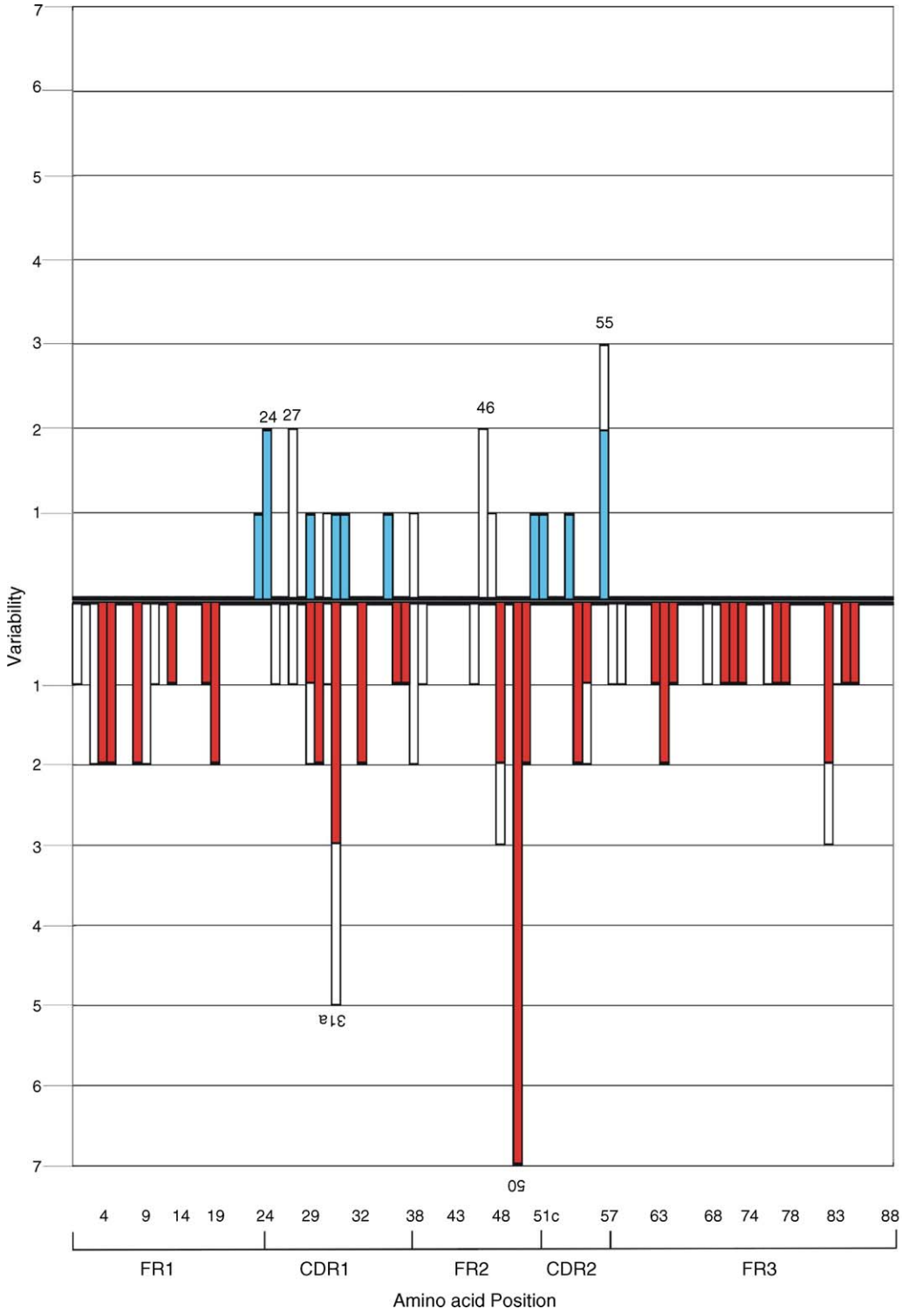

Fig. 4. Substitutions occurring by position in the Vλ exon of positions 1–88. For both species, all the substitutions are shown in white, while only those changes leading to a replacement at the amino acid level are represented in blue for mice and in red for humans. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

equilibrium ($R/S = 2.9$) in almost all the sub-regions of the four loci analyzed. Thus, there is not only a low ratio of alleles and mutations with respect to the human IGHV locus, but in general the sub-regions of the variable exon in IGKV and IGLV loci tend to be conserved. A notable example in this respect is that of CDR2 in the IGKV loci in mice and IGLV in humans, which showed values of 14/0 and 6.0, respectively. These high R/S ratios in CDR2 were observed to be the result of substitution in only a few families (see Tables 3 and 4); they contrast not only with that occurring in the other sub-regions of the variable exon, but particularly with the observations on CDR1 in the same loci (values of 2.3 and 1.3, respectively). There would seem to be a tendency to diversify CDR2 more than CDR1.

In view of the differences observed in the R/S ratios among the different sub-regions, the type of amino acid substitutions occurring among these segments may also be expected to be significantly different. Therefore, we calculated the distribution of the three types of amino acid substitution previously described for the distinct CDR and FR regions. We found that the preponderance of non-conservative substitutions in the four loci analyzed tended to be centered on the CDRs, just as we expected, considering the functions of those regions (data not shown). A similar tendency had previously been observed in the human IGHV locus, although it was not as evident as in the loci studied herein (Romo-González and Vargas-Madrazo, 2005).

Finally, the analysis of substitutions per residue showed a bias in the replacement tendency not as marked as that observed in the IGHV; however, for the loci analyzed in this study, not all the positions were equally replaced, since we encountered some positions with a high number of substitutions. This is noteworthy in view of the limited number of substitutions (2.76 on the average) presented by each allele in the four loci studied here (Figs. 3 and 4). In the IGKV locus of mice, residues 24, 28, 50 and 53 (with 9, 8, 8 and 4 replacements, respectively) were outstanding. Also, in the IGLV locus of humans, only positions 50 and 31a presented more than three replacements. Among these positions, L50 has been described as one having frequent contact with the Ag. Another noteworthy case is that of IGLV locus in mice, for in spite of its low ratio of substitutions, the majority of those occurring were localized in the CDRs or in nearby areas (Fig. 4).

4. Discussion

As it is previously mentioned, diverse and complementary forces converge on the evolution of Igs. Since few studies deal integrally with this complex problem, a profound exploration of the nature of the evolutionary diversification processes in these proteins is important. Allelic polymorphism is one of the fundamental sources of the evolutionary variation in Ig genes (Cook and Tomlinson, 1995); however, this mechanism is poorly understood regarding its functional and structural implications. In this work, therefore, we analyzed the genetic and structural information contained in the alleles of IGKV and IGLV loci in humans and mice and compared the results with those observed previously in the human IGHV locus (Romo-González and Vargas-Madrazo, 2005).

In previous sections of our study, we observed that the evolutionary variation processes in Igs generally tend more toward conservation than toward diversification, just as it has usually been postulated in relevant literature, since most alleles show a tendency to substitute only one or two nucleotides. This rate of variation is very similar to that observed in the alleles of humans (Kruglyak, 1997; Halushka et al., 1999). In spite of this low substitution rate, it is to be noted that when these variations are examined taking into account the genetic and structural information, dissimilar tendencies are observed at distinct levels in the diversification of Ig genes. This means that their diversification is not homogeneous but reflects the diverse selective pressures and processes of DNA dynamics operating on the loci, genes, sub-regions and positions; since the greatest incidence of substitutions tends to occur in those regions and residues that are important for the recognition of antigens, it is restricted in the regions responsible for the general folding of the domain (Milner et al., 1995; Tomlinson et al., 1996; Ota et al., 2000). Thus, these preponderant substitutions of only certain attributes in IGV genes seem to be in accordance with the function that they perform. At the same time, it must be considered that in concertation with selective pressure there are several mechanisms of DNA dynamics, like “concerted” evolution, that co-determine the final properties of genes.

The first thing that becomes evident is that not all of the loci and alleles of Igs are equally substituted, for instance, the diversification processes by selective pressure operating on loci κ and λ in humans and mice are more restricted than those operating on the human IGHV locus. This agrees not only with previous reports on the germinal diversity in VH and VL (Chang and Casali, 1995), but also with structural (Vargas-Madrazo et al., 1995; Kabat and Wu, 1991; Wilson and Stanfield, 1993) and functional (Zouali, 1995; Ward et al., 1989) evidence that the VH domain plays a more preponderant role in the recognition mechanism and particularly in the diversification strategy. Indeed, in many specific immune responses, the VL domain remains constant, whereas the VH favors diversification (Kabat and Wu, 1991; Wilson and Stanfield, 1993). In this respect, it should be noted that the few replacement mutations occurring in loci κ and λ in humans and mice are mostly of the non-conservative type, whereas those in the human IGHV locus tend to be conservative (Romo-González and Vargas-Madrazo, 2005). Since the number of substitutions in loci κ and λ in humans and mice is very limited, the diversification of the VL domain appears to be subject to few but significant changes. In contrast, the strategy for diversification in the VH domain takes place through small and constant variations that could be modulating the plasticity of the antigen-binding site.

In VL loci, even though diversification regarding the number of alleles is very limited (mostly one or two alleles per gene), it is possible to note certain preferences for creating greater diversity in some of the genes (10–94, 10–96, 2–109, 2–14, 2–18). This tendency was also observed in the IGHV locus in humans (Romo-González and Vargas-Madrazo, 2005), which means that diversification via the number of alleles does not occur homogeneously among the genes. This tendency might be at least partly associated with strategies for the preferential gene usage observed in certain ontogenic or pathological states, or with the regulation of the immune network (Schroeder et al., 1995; Viale et al., 1994; Coutinho, 2000). Thus it is possible to postulate the existence of a certain correlation between the degree of polymorphism and the genetic functionality of IgV genes, as it has been observed in other multigenetic families (Noel et al., 1999; Obata et al., 2000). Nevertheless, it is important to consider that the evolution of Igs does not occur through the direct selection of individual genes, but rather of a complete repertoire of genes (Rothenfluh et al., 1995; Blanden et al., 1998). It is possible to suppose that mutual codetermination exist between the properties of the Ab repertoire and the extent of allelism in V genes, especially if we consider the diverse evidence indicating that only a small fraction of the repertoire of Ig genes plays a central role in determining the fundamental properties of the Ab repertoire in the immune system (Kearney et al., 1989; Stewart and Varela, 1989; Coutinho, 2000). As an example, studies on allogenic cells show that the self-recognition is mediated only by a small group of genes (Zinkernagel and Doherty, 1977; Schurmans et al., 1991; Nasman and Lundkvist, 1996). It is also known that in the repertoire of both humans and mice, there exists a preferential use of certain genes that varies with age, while others continue to be expressed throughout life (Kearney et al., 1989; Coutinho, 2000). It has been postulated that the latter are fundamental to the organization, maintenance and regulation of the immune network (Kearney et al., 1989; Viale et al., 1992; Nobrega et al., 1998; Lacroix-Desmazes et al., 1998; Coutinho, 2000).

Similarity to what we recently observed in the human IGHV locus, a detailed study of variation through allelism enabled us to observe some peculiarities of the structural diversification in these loci. From the analysis of substitution patterns (number and type of substitutions and *R*/*S* ratio) per region and position in distinct alleles, it was possible to distinguish a tendency to diversify only a certain region (CDR2) within the locus, though not as preponderantly as in the IGHV of humans. This tendency observed in the variation through allelism in loci κ and λ is opposed to that reported previously for the orthologous diversification of these genes, since this last diversification occurs mainly in CDR1 (Chang and Casali, 1995); that is, whereas the repertoire of VL genes multigenetic families developed through variation in CDR1, allelism has diversified this repertoire by means of substitutions in CDR2.

If we consider that most of the different alleles are the result of independent studies on unrelated individuals, the results listed below seem to indicate the presence of a concerted evolutionary and functional strategy acting on Ig genes (Zuckerkandl, 1976):

(i) each allele is unique evolutionary document;
(ii) only a limited number of substitutions take place in the different alleles, genes and loci;
(iii) they are localized only in CDR2 but not in CDR1.

In regard to the substitution patterns by position, the few mutations observed might possibly be distributed at random; however, it should be noted that, in the regions composing the V exon, only some positions were highly replaced (24, 28, 31a, 50 y 53), and these are in the CDRs or adjacent to them. Position L50 stands out prominently as one of the most frequently replaced positions in our study, and also as one observed to be highly mutated within orthologous diversity (Tomlinson et al., 1996); it has also been identified as being in frequent contact with the Ag (MacCallum et al., 1996). The possibility of mutation occurring by chance in one single codon for a large number of alleles in unrelated individuals is very low. This being the case, the presence of replacement mutations far above the expected values in position L50 of loci IGKV in mice and IGLV in humans may indicate that these events are associated with mutational "hotspots."In short, the detailed study of substitution patterns in alleles that we have presented here makes it evident that the nature of the selective pressures acting on a given protein become obscure upon examining only the complete protein, that is without considering the functions of each residue (Eigen et al., 1988; Booker and Haughton, 1994). This is important because not all amino acid substitutions have the same impact on protein evolution (Zuckerkandl, 1976).

Acknowledgments

We wish to thank Edda Sciutto and Edmundo Lamoyi for their guidance, dedication and comments during this creative research process, and Walter Locht for his counseling in the development of tools and algorithms for handling data. We especially thank Carlos Larralde, whose comments have revealed the complex stratagem for determining the subject and object of study during the process of knowing. Irene Marquina and Warren Haid translated and revised the manuscript. This work was made possible through the economic support from the Institute of Biological Research of the University of Veracruz. Tania Romo-González receives the support of scholarships from CONACYT and DGEP-UNAM for postgraduate studies.

References

Almagro, J.C., Hernandez, I., Ramirez, M.C., Vargas-Madrazo, E., 1998. Structural differences between the repertoires of mouse and human germline genes and their evolutionary implications. Immunogenetics 47 (5), 355–363.

Almagro, J.C., Hernandez, I., del Carmen Ramirez, M., Vargas-Madrazo, E., 1997. The differences between the structural repertoires of VH germ-line gene segments of mice and humans: implication for the molecular mechanism of the immune response. Mol. Immunol. 34 (16–17), 1199–1214.

Andersson, E., Matsunaga, T., 1995. Evolution of immunoglobulin heavy chain variable region genes: a VH family can last for 150-200 million years or longer. Immunogenetics 41 (1), 18–28.

Barbie, V., Lefranc, M.P., 1998. The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. Exp. Clin. Immunogenet. 15 (3), 171–183.

Blanden, R.V., Rothenfluh, H.S., Zylstra, P., Weiller, G.F., Steele, E.J., 1998. The signature of somatic hypermutation appears to be written into the germline IgV segment repertoire. Immunol. Rev. 162, 117–132.

Booker, J.K., Haughton, G., 1994. Mechanisms that limit the diversity of antibodies. II. Evolutionary conservation of Ig variable region genes which encode naturally occurring autoantibodies. Int. Immunol. 6 (9), 1427–1436.

Chang, B., Casali, P., 1995. A sequence analysis of human germline Ig VH and VL genes. The CDR1s of a major proportion of VH, but not VL, genes display a high inherent susceptibility to amino acid replacement. Ann. N.Y. Acad. Sci. 764, 170–179.

Chothia, C., Lesk, A.M., 1987. Canonical structures for the hypervariable regions of immunoglobulins. J. Mol. Biol. 196 (4), 901–917.

Chothia, C., Lesk, A.M., Gherardi, E., Tomlinson, I.M., Walter, G., Marks, J.D., Llewelyn, M.B., Winter, G., 1992. Structural repertoire of the human VH segments. J. Mol. Biol. 227 (3), 799–817.

Cocho, G., Lara-Ochoa, F., Vargas, E., Jiménez-Montaño, M.A., Rius, J.L., 1993. Structural patterns in macromolecules in thinking about biology. In: Stein, W., Varela, F.J. (Eds.), Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley Pu. Co., Reading, Mass, pp. 105–119.

Cook, G.P., Tomlinson, I.M., 1995. The human immunoglobulin VH repertoire. Immunol. Today 16 (5), 237–242.

Coutinho, A., 2000. Germ-line selection ensures embryonic autoreactivity and a positive discrimination of self mediated by supraclonal mechanisms. Semin. Immunol. 12 (3) 205–13; discussion 257–344.

Dayhoff, M.O., 1972. Atlas of Protein Sequence and Structure, 5. Natl. Biomed. Res. Found, Washington, DC.

Eigen, M., Winkler-Oswatitsch, R., Dress, A., 1988. Statistical geometry in sequence space: a method of quantitative comparative sequence analysis. Proc. Natl. Acad. Sci. U.S.A. 85 (16), 5913–5917.

Ferguson, A., 1980. Biochemia Systematcis and Evolution. Blackie, Glasgow.

Go, M., Miyazawa, S., 1980. Relationship between mutability, polarity and exteriority of amino acid residues in protein evolution. Int. J. Pept. Protein Res. 15 (3), 211–224.

Grantham, R., 1974. Amino acid difference formula to help explain protein evolution. Science 185 (4154), 862–864.

Halushka, M.K., Fan, J.B., Bentley, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R., Chakravarti, A., 1999. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat. Genet. 22 (3), 239–247.

Harvey, P.H., Holmes, E.C., Nee, S., 1994. Model phylogenies to explain the real world. Bioessays 16 (10), 767–770.

Hughes, A.L., 2004. Evidence for abundant slightly deleterious polymorphisms in bacterial populations. Genetics 169 (2), 533–538.

Jukes, T.H., King, J.L., 1979. Evolutionary nucleotide replacements in DNA. Nature 281 (5732), 605–606.

Kabat, E.A., Wu, T.T., 1991. Identical V region amino acid sequences and segments of sequences in antibodies of different specificities. Relative contributions of VH and VL genes, minigenes, and complementarity-determining regions to binding of antibody-combining sites. J. Immunol. 147 (5), 1709–1719.

Kataoka, T., Nikaido, T., Miyata, T., Moriwaki, K., Honjo, T., 1982. The nucleotide sequences of rearranged and germline immunoglobulin VH genes of a mouse myeloma MC101 and evolution of VH genes in mouse. J. Biol. Chem. 257 (1), 277–285.

Kearney, J.F., Vakil, M., Solvason, N., 1989. The role of idiotypic interactions and B-cell subsets in development of the B-cell repertoire. Cold Spring Harb. Symp. Quant. Biol. 54 (1), 203–207.

Kirkham, P.M., Schroeder Jr., H.W., 1994. Antibody structure and the evolution of immunoglobulin V gene segments. Semin. Immunol. 6 (6), 347–360.

Klein, J., Satta, Y., O'hUigin, C., Takahata, N., 1993. The molecular descent of the major histocompatibility complex. Annu. Rev. Immunol. 11, 269–295.

Klein, J., 1986. Natural History of the Mayor Histocompatibility Complex. Wiley, New York.

Krawinkel, U., Christoph, T., Blankenstein, T., 1989. Organization of the Ig VH locus in mice and humans. Immunol. Today 10 (10), 339–344.

Kruglyak, L., 1997. The use of a genetic map of biallelic markers in linkage studies. Nat. Genet. 17 (1), 21–24.

Lacroix-Desmazes, S., Kaveri, S.V., Mouthon, L., Ayouba, A., Malanchere, E., Coutinho, A., Kazatchkine, M.D., 1998. Self-reactive antibodies (natural autoantibodies) in healthy individuals. J. Immunol. Meth. 216 (1–2), 117–137.

Lara-Ochoa, F., Almagro, J.C., Vargas-Madrazo, E., Conrad, M., 1996. Antibody-antigen recognition: a canonical structure paradigm. J. Mol. E 43 (6), 678–684.

Lefranc, M.P., Lefranc, G., 2001. The Immunoglobulin FactsBook. Academic Press, p. 458.

Li, W.-H., 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. E 36 (1), 96–99.

Li, W.-H., 1997. Molecular Evolution. Sinauer Associates, Sunderland, MA.

Litman, G.W., Anderson, M.K., Rast, J.P., 1999. Evolution of antigen binding receptors. Annu. Rev. Immunol. 17, 109–147.

MacCallum, R.M., Martin, A.C., Thornton, J.M., 1996. Antibody–antigen interactions: contact analysis and binding site topography. J. Mol. Biol. 262 (5), 732–745.

Martin, A.C., Cheetham, J.C., Rees, A.R., 1991. Molecular modeling of antibody combining sites. Meth. Enzymol. 203, 121–153.

Martinez-Jean, C., Folch, G., Lefranc, M.P., 2001. Nomenclature and overview of the mouse (Mus musculus and Mus sp.) immunoglobulin kappa (IGK) genes. Exp. Clin. Immunogenet. 18 (4), 255–279.

Massingham, T., Goldman, N., 2005. Detecting amino acid sites under positive selection and purifying selection. Genetics, [Epub ahead of print].

McCormack, W.T., Thompson, C.B., 1990. Somatic diversification of the chicken immunoglobulin light-chain gene. Adv. Immunol. 48, 41–67.

Milner, E.C., Hufnagle, W.O., Glas, A.M., Suzuki, I., Alexander, C., 1995. Polymorphism and utilization of human VH Genes. Ann. N.Y. Acad. Sci. 764, 50–61.

Nagylaki, T., Petes, T.D., 1982. Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. Genetics 100 (2), 315–337.

Nasman, I., Lundkvist, I., 1996. Evidence for oligoclonal diversification of the VH6-containing immunoglobulin repertoire during reconstitution after bone marrow transplantation. Blood 87 (7), 2795–2804.

Nobrega, A., Grandien, A., Haury, M., Hecker, L., Malanchere, E., Coutinho, A., 1998. Functional diversity and clonal frequencies of reactivity in the available antibody repertoire. Eur. J. Immunol. 28 (4), 1204–1215.

Noel, L., Moores, T.L., van Der Biezen, E.A., Parniske, M., Daniels, M.J., Parker, J.E., Jones, J.D., 1999. Pronounced intraspecific haplotype divergence at the RPP5 complex disease resistance locus of Arabidopsis. Plant Cell 11 (11), 2099–2112.

Obata, F., Shiiba, R., Iizuka, M., Kashiwagi, N., Kurosu, F., Shimada, N., Nishijima, M., Tozawa, H., 2000. Human T-cell receptor BV6 gene polymorphism in relation to expression level and CD4/CD8 skewness. Scand. J. Immunol. 51 (6), 543–547.

Ota, T., Sitnikova, T., Nei, M., 2000. Evolution of vertebrate immunoglobulin variable gene segments. Curr. Top. Microbiol. Immunol. 248, 221–245.

Pallares, N., Frippiat, J.P., Giudicelli, V., Lefranc, M.P., 1998. The human immunoglobulin lambda variable (IGLV) genes and joining (IGLJ) segments. Exp. Clin. Immunogenet. 15 (1), 8–18.

Pascual, V., Capra, J.D., 1991. Human immunoglobulin heavy-chain variable region genes: organization, polymorphism, and expression. Adv. Immunol. 49, 1–74.

Perlmutter, R.M., Berson, B., Griffin, J.A., Hood, L., 1985. Diversity in the germline antibody repertoire. Molecular evolution of the T15 VN gene family. J. Exp. Med. 162 (6), 1998–2016.

Poljak, R.J., Amzel, L.M., Avey, H.P., Chen, B.L., Phizackerley, R.P., Saul, F., 1973. Three-dimensional structure of the Fab' fragment of a human immunoglobulin at 2,8-A resolution. Proc. Natl. Acad. Sci. U.S.A. 70 (12), 3305–3310.

Reynaud, C.A., Dahan, A., Anquez, V., Weill, J.C., 1989. Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. Cell 59 (1), 171–183.

Romo-González, T., Vargas-Madrazo, E., 2005. Structural analysis of substitution patterns in alleles of human immunoglobulin VH genes. Mol. Immunol. 42 (9), 1085–1097.

Rothenfluh, H.S., Blanden, R.V., Steele, E.J., 1995. Evolution of V genes: DNA sequence structure of functional germline genes and pseudogenes. Immunogenetics 42 (3), 159–171.

Saul, F.A., Poljak, R.J., 1993. Structural patterns at residue positions 9, 18, 67 and 82 in the VH framework regions of human and murine immunoglobulins. J. Mol. Biol. 230 (1), 15–20.

Scaviner, D., Guiraudou, D., 1999. Gene table: Mouse (*Mus musculus*, *Mus spretus*) IGLV. "IMGT, the international ImMunoGeneTics information system® http://imgt.cines.fr (Initiator and coordinator: Marie-Paule Lefranc, Montpellier, France).

Schroeder Jr., H.W., Mortari, F., Shiokawa, S., Kirkham, P.M., Elgavish, R.A., Bertrand, F.E., 1995. Developmental regulation of the human antibody repertoire. Ann. N.Y. Acad. Sci. 29 (764), 242–260.

Schurmans, S., Merino, J., Qin, H.Y., Kramar, G., Duchosal, M., Skalli, O., Benzonana, G., Gabbiani, G., Lambert, P.H., 1991. Autoimmune syndrome after neonatal induction of tolerance to alloantigens: analysis of the specificity and of the cellular and genetic origin of autoantibodies. Autoimmunity 9 (4), 283–291.

Shlomchik, M.J., Marshak-Rothstein, A., Wolfowicz, C.B., Rothstein, T.L., Weigert, M.G., 1987. The role of clonal selection and somatic mutation in autoimmunity. Nature 328 (6133), 805–811.

Sitnikova, T., Su, C., 1998. Coevolution of immunoglobulin heavy- and light-chain variable-region gene families. Mol. Biol. Evol. 15 (6), 617–625.

Stewart, J., Varela, F.J., 1989. Exploring the meaning of connectivity in the immune network. Immunol. Rev. 110, 37–61.

Takano, T.S., 1998. Rate variation of DNA sequence evolution in the *Drosophila* lineages. Genetics 149 (2), 959–970.

Tomlinson, I.M., Cox, J.P., Gherardi, E., Lesk, A.M., Chothia, C., 1995. The structural repertoire of the human V kappa domain. EMBO J. 14 (18), 4628–4638.

Tomlinson, I.M., Walter, G., Marks, J.D., Llewelyn, M.B., Winter, G., 1992. The repertoire of human germline VH sequences reveals about fifty groups of VH segments with different hypervariable loops. J. Mol. Biol. 227 (3), 776–798.

Tomlinson, I.M., Walter, G., Jones, P.T., Dear, P.H., Sonnhammer, E.L., Winter, G., 1996. The imprint of somatic hypermutation on the repertoire of human germline V genes. J. Mol. Biol. 256 (5), 813–817.

Tourasse, N.J., Gouy, M., 1997. Evolutionary distances between nucleotide sequences based on the distribution of substitution rates among sites as estimated by parsimony. Mol. Biol. Evol. 14 (3), 287–298.

Tutter, A., Brodeur, P., Shlomchik, M., Riblet, R., 1991. Structure, map position, and evolution of two newly diverged mouse Ig VH gene families. J. Immunol. 147 (9), 3215–3223.

Vargas-Madrazo, E., Paz-Garcia, E., 2003. An improved model of association for VH-VL immunoglobulin domains: asymmetries between VH and VL in the packing of some interface residues. J. Mol. Recognit. 16 (3), 113–120.

Vargas-Madrazo, E., Lara-Ochoa, F., Jimenez-Montano, M., 1994. A skewed distribution of amino acids at recognition sites of the hypervariable region of immunoglobulins. J. Mol. E 38 (1), 100–104.

Vargas-Madrazo, E., Lara-Ochoa, F., Almagro, J.C., 1995. Canonical structure repertoire of the antigen-binding site of immunoglobulins suggests strong geometrical restrictions associated to the mechanism of immune recognition. J. Mol. Biol. 254 (3), 497–504.

Viale, A.C., Chies, J.A., Huetz, F., Malenchere, E., Weksler, M., Freitas, A.A., Coutinho, A., 1994. VH-gene family dominance in ageing mice. Scand. J. Immunol. 39 (2), 184–188.

Viale, A.C., Coutinho, A., Freitas, A.A., 1992. Differential expression of VH gene families in peripheral B cell repertoires of newborn or adult immunoglobulin H chain congenic mice. J. Exp. Med. 175 (6), 1449–1456.

Ward, E.S., Gussow, D., Griffiths, A.D., Jones, P.T., Winter, G., 1989. Binding activities of a repertoire of single immunoglobulin variable domains secreted from *Escherichia coli*. Nature 341 (6242), 544–546.

Weill, J.C., Reynaud, C.A., 1996. Rearrangement/hypermutation/gene conversion: when, where and why? Immunol. Today 17 (2), 92–97.

Wilson, A.C., Carlson, S.S., White, T.J., 1977. Biochemical evolution. Annu. Rev. Biochem. 46, 573–639.

Wilson, I.A., Stanfield, R.L., 1993. Antibody–antigen interactions. Curr. Opin. Struct. Biol. 3, 113–118.

Wu, T.T., Kabat, E.A., 1970. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. J. Exp. Med. 132 (2), 211–250.

Yang, Z., Nielsen, R., 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. 17 (1), 32–43.

Zinkernagel, R.M., Doherty, P.C., 1977. The concept that surveillance of self is mediated via the same set of genes that determines recognition of allogenic cells. Cold Spring Harb. Symp. Quant. Biol. 41 (2), 505–510.

Zouali, M., 1995. B-cell superantigens: implications for selection of the human antibody repertoire. Immunol. Today 16 (8), 399–405.

Zuckerkandl, E., 1976. Evolutionary processes and evolutionary noise at the molecular level. I. functional density in proteins. J. Mol. Evol. 7 (3), 167–183.