Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Immunology 42 (2005) 1085–1097
Molecular Immunology
www.elsevier.com/locate/molimm

Structural analysis of substitution patterns in alleles of human immunoglobulin VH genes

Tania Romo-González, Enrique Vargas-Madrazo*

Departamento Biología Sistémica, Instituto de Investigaciones Biológicas, Universidad Veracruzana, Morelos 9 La Pitaya, Zoncuantla, 91500 Coatepec, Xalapa, Veracruz, Mexico

Received 16 August 2004
Available online 21 December 2004

Abstract

The diversity in repertoires of antibodies (Abs) needed in response to the antigen challenge is produced by evolutionary and somatic processes. The mechanisms operating at a somatic level have been studied in great detail. In contrast, neither the mechanisms nor the strategies of diversification at an evolutionary level have yet been understood in similar detail. Particularly, the substitution patterns in alleles of immunoglobulin genes (Igs) have not been systematically studied. Furthermore, there is a scarcity of studies which link the analysis at a genetic level of the diversification of repertoires with the structural consequences at the protein level of the changes in DNA information. For the purpose of systematically characterizing the strategies of evolutionary diversification through sequence variation at alleles, in this work, we built a database for all the alleles of the IGHV locus in humans reported until now. Based on these data, we performed diverse analyses of substitution patterns and linked these results with studies at the protein level. We found that the sequence diversification in different alleles does not operate with equal intensity for all V genes. Our studies, both of the number of substitutions and of the type of amino acid change per sub-segment of the V-REGION evidenced differences in the selective pressure to which these regions are exposed. The implications of these results for understanding the evolutionary diversification strategies, as well as for the somatic generation of antibody repertoires are discussed.

Keywords: R/S ratio; Amino acid substitution pattern; Molecular recognition; Antibody repertoires

1. Introduction

The main function of the immune system in vertebrates is to protect them from foreign organisms (Mazumdar, 1995; Cohn and Langman, 1996). Since the variety of the antigenic world is enormous, efficient immunological mechanisms of defense must be provided with repertoires of highly diversified receptors able to contribute efficiently to the recognition and processing of foreign agents. Diversity in the repertoire of antibodies (Abs) is produced by genetic and somatic processes (Max, 1998). Generally, organisms have extensive multigenetic families with diverse members which encode numerous V-REGIONs. Such diversity increases by somatic mechanisms like recombination, inexact binding of genetic segments, hypermutation (Tonegawa, 1983) and germline conversion (Becker and Knight, 1990; McCormack et al., 1991).

The mechanisms operating to diversify the Abs repertoire at somatic level have been studied in great detail, but little is known about the genetic contribution to the diversification of the Ab repertoire. This lack of understanding is partly due to the very complex configuration, composition and evolution of the immunoglobulin (Ig) genes (Li et al., 2002). However, it is fundamental to understand the factors that have shaped the germline repertoire and its evolutionary diversification processes. This knowledge should lead to an understanding of the differential expression of V genes and their association with some pathologies (Milner et al., 1995).

Abbreviations: Ab, antibody; CDR, complementarity determining region; FR, framework region; Ig, immunoglobulin; R/S ratio, replacement/silent ratio

* Corresponding author. Tel.: +52 28 125757; fax: +52 28 125757.
E-mail address: envargas@yahoo.com (E. Vargas-Madrazo).

0161-5890/$ – see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.molimm.2004.11.004

In the human Ig loci, two types of polymorphism have been observed for the V genes: (i) variation in nucleotide substitution and (ii) insertion and/or deletion of genes in the locus (Cook and Tomlinson, 1995; Lefranc et al., 1999; Hammarstrom et al., 1990; Matsuda et al., 1993, 1998). Although a correlation between insertion/deletion polymorphism variations and pathologies has not been found (Pargent et al., 1991; Schaible et al., 1993), such events may affect the function that genetic segments play in the Repertoire (Li et al., 2002). In regions I and II of gene IGHV, this kind of polymorphism has been observed (Cui and Li, 1998; Pramanik and Li, 2002), whereas in region III this type of variations has not been found (Cui and Li, 1997, 2000).

Similar processes take place in evolution through nucleotide substitution where the genes situated in the most J_H-distal regions present a high degree of substitution polymorphism (genes 1-12L, 1-6/1-69, 1-3/l-68, 1-12R and 4-11/DP-66), in contrast to genes located in the most J_H-proximal portion that are highly conserved (Cook et al., 1994a, 1994b). Differences in substitution polymorphism depend not only on the physical location of the IGHV genes; it has been observed that within the V-REGIONs, there also exist fragments in which diversification is either favored or restricted (Cook and Tomlinson, 1995; Milner et al., 1995). Within the mutations observed in germline genes, processes of diversification and selection have been seen to stimulate the variability of amino acids at the antigen-binding site and to conserve the amino acids in the framework regions (FRs). This highly marked conservation of the FR is related to its function of maintaining the folding of the domain, and consequently conserve the general structure of the antigen-binding site (Tanaka and Nei, 1989; Kirkham and Schroeder, 1994; Vargas-Madrazo et al., 1994; Tomlinson et al., 1996). Furthermore, variability in the VH domain is not only favored in the CDRs, as some FRs are more susceptible to processes of diversification while others are highly conserved (Kirkham et al., 1992). Differences in conservation may be attributed to the distinct functions performed by these sub-segments in the antibody (Kirkham et al., 1992; Kirkham and Schroeder, 1994).

Within the V-REGIONs, the substitution of nucleotides permits the creation of alleles with amino acid changes which imply structural variants with respect to its parent allele. This process permits the exploration of variants of antigen-binding sites, allowing a better adaptation to the changing antigenic universe in an evolutionary period (Ota et al., 2000).

At present a detailed characterization of the implications of the substitution patterns in alleles for the mechanism of antigen-antibody recognition does not exist. There are several levels at which the evolutionary strategies of the Ig genes can be characterized (Sasso et al., 1990; van Dijk et al., 1991; Sasso et al., 1993; Cui and Li, 1998). For example, the evolutionary diversification of Ig genes can be studied by the number of mutations per allele (Li, 1997), as well as by the distribution and type of substitutions along the gene (Eigen et al., 1989). There is a marked difference in evolutionary patterns between FRs and CDRs in the antibodies (Kabat, 1978; Lara-Ochoa et al., 1995), but even within each region there are very diverse patterns of variation among distinct positions (Vargas-Madrazo et al., 1994). For instance, hypervariable positions that have a high frequency of contact with the antigen are present in the CDRs (Padlan et al., 1995; MacCallum et al., 1996; Ramirez-Benites and Almagro, 2001), whereas other highly conserved positions preserve the structure of the antigen-binding site (Chothia and Lesk, 1987; Padlan, 1990; Vargas-Madrazo et al., 1994). Therefore, an integral and detailed characterization of the substitution patterns in alleles in the V genes of Igs that considers its functional peculiarities, will allow a deeper comprehension of the strategies of molecular evolution in the repertoire of antibodies (Vargas-Madrazo et al., 1997).

In this paper we analyzed some aspects of molecular evolution by allelic variation in the human IGHV genes and the consequences of these alterations on the structural properties of the antigen-binding site. This will permit a more detailed understanding of the strategies that the immune system develops to create diverse repertoires of Abs with high affinity and specificity. Allele mutations of the human functional IGHV genes were analyzed, based on the alignments of alleles available in IMGT Repertoire from IMGT, the international ImMunoGenetics information system, http://imgt.cines.fr (Lefranc, 2003, 2004), and published in The Immunoglobulin FactsBook (Lefranc and Lefranc, 2001). Through a detailed study of the original reports on the sequences, we evaluated the quality of the available information. Based on this database, the following aspects are analyzed: (i) distribution of alleles per gene; (ii) number of substitutions per allele; (iii) distribution and type of amino acid substitution in the V-REGION; (iv) analysis of the replacement/silent substitutions ratio (R/S ratio) for the complete V-REGION and by sub-regions (FRs and CDRs).

The analysis of the R/S ratio with the model of genetic change through aleatory punctual mutation provides a powerful tool for determining the type of selective pressure that is operating on positions or specific sub-regions of genes (Jukes and King, 1979; Shlomchik et al., 1987). This characteristic is very appropriate for the objectives of the present study.

2. Methods

2.1. Construction of the database

Based on the IMGT (http://imgt.cines.fr) database, all alleles of the human IGHV locus were compiled and compared in detail with the germline genes. We aligned each of the alleles with the allele representing each IGHV 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 re-

viewed 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) and Go and Miyazawa (1980). 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 of one amino acid by another or preservation of the same residue (a silent 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 the database for human IGHV alleles

Table 1 reports the genes for which alleles have been found. This database includes 158 sequences, which contribute alleles for 42 of the 51 functional genes that form the IGHV locus (Cook and Tomlinson, 1995). It is remarkable that most of the studies reporting alleles were not realized for the purpose of characterizing allelic polymorphism. It should also be mentioned that most of the samples of genetic material for studies of IGHV genes have been obtained from Caucasian individuals.

The main features that we found from the detailed analysis of the sequence database are:

(i) Only part of the data comes from studies specifically designed to characterize polymorphism. Although the remaining studies report sequences of alleles, this was not the purpose of their research. For instance, the alleles found for genes V7-4-1, V5-51, V5-a and V3-23 resulted from systematic studies to characterize their polymorphism (Sanz et al., 1989; Willems van Dijk et al., 1992; Rubinstein et al., 1994; Sasso et al., 1995).

(ii) The polymorphism found for genes V3-11, V3-15, V3-30, V3-30-3, V3-33, V3-49, V3-64, V4-4, V4-30-2, V4-30-4, V4-31, V4-34, V4-39, V4-59, and V4-61 was reported in studies whose objective was to characterize the polymorphism of their respective families, but not specifically that of these genes (Sasso et al., 1990, 1992; Olee et al., 1991; Weng et al., 1992; Winkler et al., 1992; Adderson et al., 1993).

(iii) The alleles found for the remaining genes (21) may be considered circumstantial, since they were found in studies of other kinds, such as the mapping of the IGHV locus, or gene usage and its association with pathologies.

(iv) Genes V1-18, V3-7, V3-13, and V3-30-3 have been studied independently by different research groups that always found the same allele for each gene (Berman et al., 1988; Olee et al., 1991; Kuppers et al., 1992; Tomlinson et al., 1992; Winkler et al., 1992; Matsuda et al., 1993; Sasso et al., 1992, 1995).

(v) For gene V3-30, 18 different alleles have been reported, and they resulted from four independent studies (Chen, 1990; Olee et al., 1991; Sasso et al., 1992; Harmer et al., 1995); for gene V3-23, in contrast, only two alleles have been found also as a result of four independent studies (Chen et al., 1988; Tomlinson et al., 1992; Matsuda et al., 1993; Sasso et al., 1995).

(vi) In the case of the genes with more reported alleles (2-70, 4-34, 4-59 and 3-30), the different alleles resulted from independent studies (Kodaira et al., 1986; Lee et al., 1987; Baer et al., 1988; Chen, 1990; Olee et al., 1991; Pascual et al., 1990; Campbell et al., 1992; Sasso et al., 1992; Tomlinson et al., 1992; van Es et al., 1992; Weng et al., 1992; Andris et al., 1993; Matsuda et al., 1993; van der Maarel et al., 1993; Cook et al., 1994a, 1994b; Brezinschek et al., 1995; Harmer et al., 1995; Voswinkel et al., 1997).

The fifth and sixth columns of Table 1 show the number of alleles reported for each gene. A great variability in the number of alleles per gene exists, for example, gene 3-30 presents 18 alleles, whereas gene 3-13 has only one. This range implies great differences in the degree of variability for the genes, which allows us to classify the genes variability as follows: (i) highly variable (seven or more alleles per gene); (ii) variable (from three to six alleles) and (iii) conserved (one to two alleles). Of the 42 functional genes of the IGHV gene for which alleles have been reported 7 (17%) are highly variable, 12 (28%) are variable and 23 (55%) are conserved.

At a family level, marked differences are also observed in the number of reported alleles (Table 2, columns three

Table 1
Number of alleles reported per gene for human IGHV

Clans and families		Number of germline genes	Number of allelic segments	Gene name	Number of alleles
Clan I	IGHV1	11	7	1-2	3
				1-3	1
				1-18	1
				1-45	2
				1-46	2
				1-69	6
				1-f	1
	IGHV5	2	2	5-5-51	4
				5-a	3
	IGHV7	1	1	7-4-1	2
Clan II	IGHV2	3	2	2-5	8
				2-70	11
	IGHV4	11	10	4-4	5
				4-28	4
				4-30-2	3
				4-30-4	5
				4-31	9
				4-34	12
				4-39	5
				4-59	9
				4-61	5
				4-b	1
	IGHV6	1	1	6-1	1
Clan III	IGHV3	22	19	3-7	1
				3-11	2
				3-13	1
				3-15	7
				3-21	1
				3-23	2
				3-30	18
				3-30-3	1
				3-33	4
				3-38	1
				3-43	1
				3-47	2
				3-48	2
				3-49	2
				3-53	1
				3-64	4
				3-66	2
				3-72	1
				3-74	2
Totals		51	42	158	

and four). Evidently, this amount is strongly determined by the number of genes in each family. Within the complex families, it is noteworthy that, although the IGHV4 family contains fewer genes than IGHV3 (11 and 22, respectively), the first proved to be more variable in the number of alleles reported (58 and 55 genes, respectively). In the small families, IGHV2 and IGHV5 are highly variable, for they possess only three and two genes, respectively, yet 19 and 7 alleles have been reported, respectively. This contrasts with the IGHV6 and IGHV7 families, each consisting of only one gene, for which only one allele has been found.

3.2. Number of substitutions per allele

Fig. 1 shows the number of substitutions (total and replacements) per allele. The distribution is highly heterogeneous, with extreme values ranging from 1 to 12 replacement substitutions. We propose a classification of three groups of alleles according to the number of replacement

Table 2
R/S ratio by sub-segments of the VH domain grouped according to family and clan

Clans and families		Number of genes	Number of alleles	FR1	CDR1	FR2	CDR2	FR3
Clan I	IGHV1	11(7)[a]	16	0.1	2/0[b]	1/0	11.0	2.6
	IGHV5	2(2)	7	0/3[c]	1/0	0.5	–	2/0
	IGHV7	1(1)	2	–	–	0/1	–	4/0
Clan II	IGHV2	3(2)	19	4.0	9/0	1.0	20/0	2.4
	IGHV4	11(10)	58	3.0	5/0	1.3	3.4	0.7
	IGHV6	1(1)	1	–	–	–	–	–
Clan III	IGHV3	22(19)	55	1.7	12.5	0.2	3.0	2.0
Totals		51(42)	158	1.8	21.0	0.6	4.3	1.5

[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 replacements substitutions were found. In such cases, the number of replacements 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.

substitutions: (i) highly mutated (seven or more substitutions); (ii) moderately mutated (from three to six substitutions); (iii) slightly mutated (one or two substitutions). Of the 158 alleles studied, 107 (68%) are slightly mutated; 38 (24%) moderately mutated and 13 (8%) highly mutated.

An important question at this point is whether some correlation exists between the presence of a great number of alleles in a gene and the fact that these alleles are highly substituted. It was found that there is a strong correlation between these two events. Thus, of the 13 highly mutated alleles, eight belong to highly variable genes, which represents a frequency far greater than expected (estimated X^2 is 46.6 and the expected $X^2_{0.95}$ is 3.84); that is, out of a total of 158 alleles, 74 alleles belong to highly variable genes, which implies a probability of 0.46 for the occurrence of alleles in highly variable genes. Therefore, if there are 13 highly mutated alleles, the mathematical expectation that these alleles should appear in highly variable genes is 5.98, which is less than the eight observed.

3.3. Type of substitutions per allele

Not only the number of substitutions is important to create diversity at the protein level; the type of amino acid substitution (alteration of physico-chemical properties)

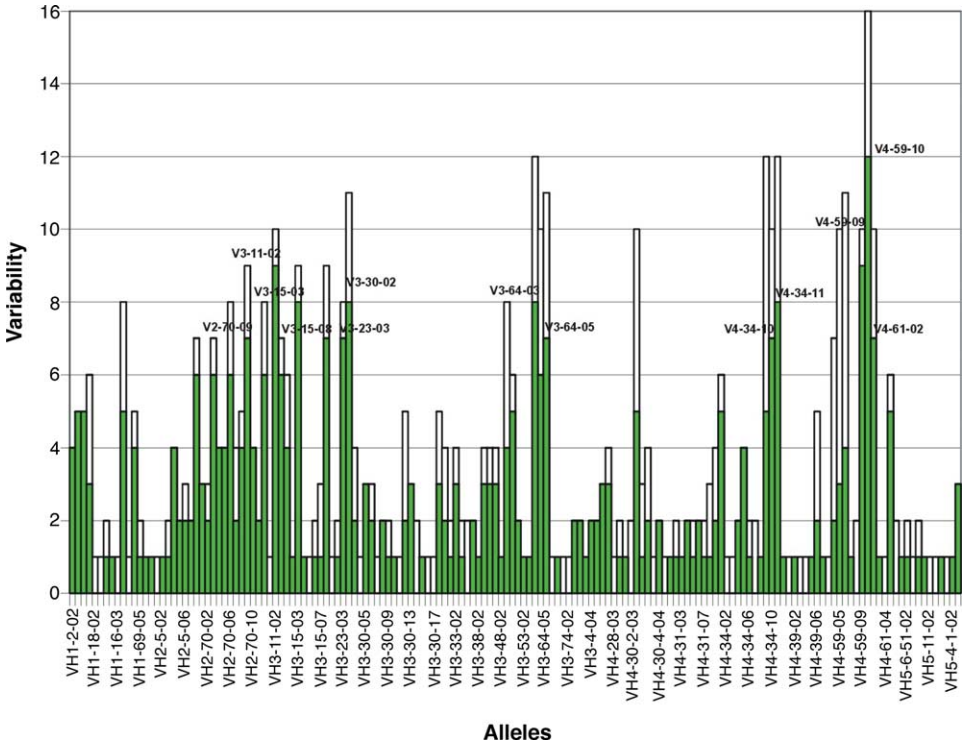

Fig. 1. This figure shows the number of substitutions occurring in each allele. The total of substitutions is shown in white, while those changes that produced a replacement at the amino acid level are shown in green.

presented in each residue also plays a fundamental role (see Section 2).

In the 158 reported alleles, 484 substitutions were found, of which 300 implied amino acid changes. Of these 300, 183 (61%) imply a conservative change, 107 (36%) non-conservative, and 10 (3%) radical. Almost two thirds of amino acid substitutions (61%) do not imply drastic alterations in physico-chemical properties, which seems to be partly associated with the preponderance of conservative substitutions in the FRs (see section on analysis of R/S ratio). Nevertheless, the fact that 39% of the changes (non-conservative and radical) imply partial or radical alterations in the properties of the residues suggests intense pressure toward diversification, at least in some sub-regions of the domain.

The types of amino acid replacement were estimated by grouping the genes according to their degree of variability (number of alleles per gene). The very variable genes present the following proportions: 62% of conservative substitutions, 35% non-conservative and 3% radical. For the variable genes, the percentages are: 65%, 31% and 4%, respectively, and for the conserved genes, the percentages are: 60%, 38% and 2%. This indicates that the proportions of the three types of amino acid replacements remain the same, whether the gene is highly variable or very conserved. Very similar proportions were found when grouping the alleles according to the number of substitutions they present. Thus, the highly mutated alleles have the following percentages per type of substitution: 72% with conservative substitutions, 23% non-conservative and five radical. The percentages for the moderately mutated alleles are: 60%, 37% and 3%, respectively, and for the slightly mutated alleles: 60%, 36% and 4%. These results suggest that the pressure to conserve the appropriate proportion of residues operates similarly on the different alleles, although, as we mentioned before, the intensity of the variation is different among the distinct genes. In the following sections, we will see that the types of amino acid substitutions are not equally distributed among the CDRs and the FRs.

3.4. Substitutions by position in the V-REGION

The previous results show that, through a detailed characterization of the substitution patterns at alleles, peculiar forms of variations associated with the specific functions of the protein under study become evident. A distinctive characteristic of antibodies is the great diversity of variability at different positions dependent on the function performed by the residues (Kabat, 1978; Vargas-Madrazo et al., 1994). Therefore, the number of substitutions per position in the V-REGION was analyzed for the 158 alleles, and the results are shown in Fig. 2. Both the total number of substitutions and those that imply amino acid replacement are reported.

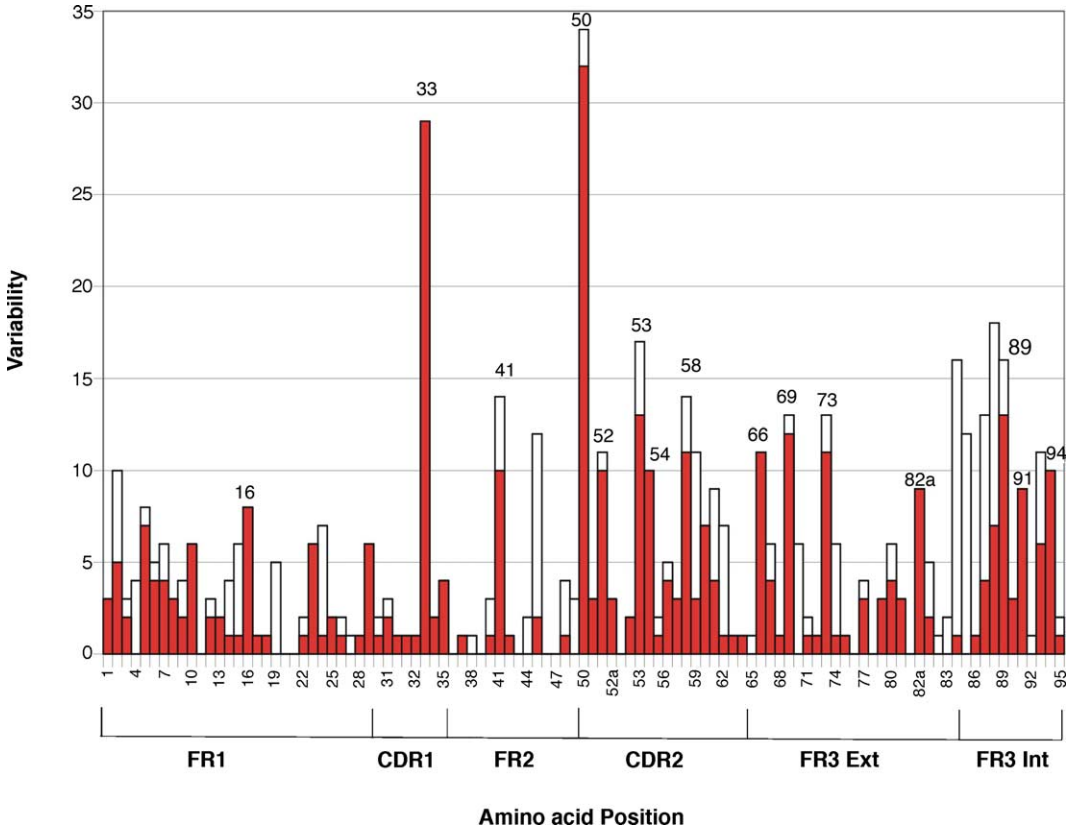

Fig. 2. This figure denotes the substitutions occurring in the VH exon by positions from 1 to 94. Those substitutions originating from replacement are shown in red, and white is for all the substitutions accounted for.

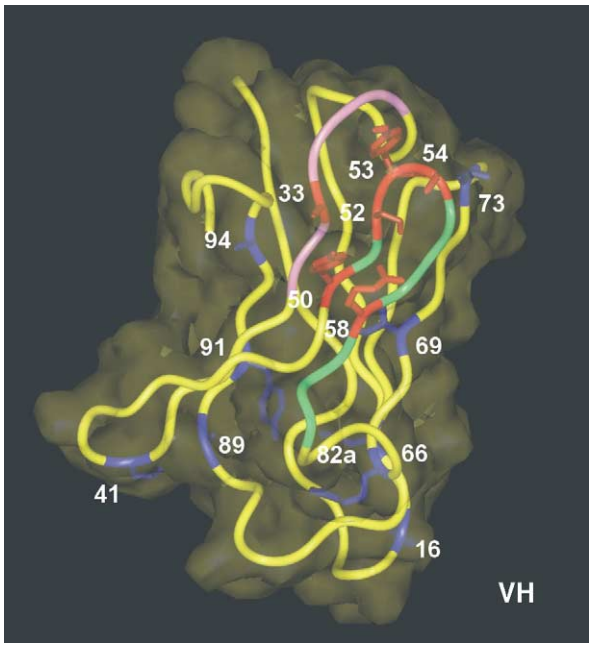

Fig. 3. Within the VH domain of the Ig, the positions produced by allelic polymorphism that were replaced most frequently are marked with blue; within these same positions those that were reported as frequently in contact with the Ag are marked with red (35). Pink represents CDR1 and green CDR2.

It can be seen that some positions (12 out of a total of 94) admit neither silent nor replacement substitutions, in any of the alleles (positions 11, 20, 21, 36, 39, 43, 46, 47, 52b, 76, 78, 82). Taking into account only the replacement substitutions, we find another group of positions, 37 of them, that mutate occasionally, that is, present one or two replacements. In contrast with this group, 27 positions presented three or more replacement substitutions; among these, 15 residues stand out for their high frequency of replacement (eight or more substitutions for positions 16, 33, 41, 50, 52, 53, 54, 58, 66, 69, 73, 82a, 89, 91 and 94). Within this group, positions 33 and 50 deserve special attention for their extraordinary hypervariability, having 29 and 32 replacements, respectively. It should be noted that this group of positions with an elevated number of replacements includes various residues that have been identified as having frequent contact with the antigen (positions 33, 50, 52, 53, 54, 58) (MacCallum et al., 1996). In Fig. 3 the location of the positions identified herein as being frequently replaced is presented in the three-dimensional structure of the VH domain, with emphasis on those positions showing frequent contact with the antigen. It can be seen that most of the frequently replaced positions (10 out of 15) are located in the CDRs or in nearby areas (Figs. 2 and 3).

The presence of a region in the FR3 with a high replacement frequency (positions 89, 91, 94) is also noteworthy; it corresponds with the interior region (Int) of the FR3, according to a classification by Kirkham and Schroeder (1994).

During our analysis to determine mutations by positions, we found that some positions of the V-REGION showed double and triple mutations within one and the same codon (Fig. 4). The expected probability of a double or triple mutation occurring within a codon is extremely small (0.0036 for a double mutation and 0.000013 for a triple one, if we assume an average length of 282 nucleotides per V-REGION). We also found that these double and triple mutations occur repeatedly in the same position in different alleles (Fig. 4), and that the great majority of such multiple mutations per codon take place in highly mutated positions. The previous observations suggest that these events of multiple mutations are associated with mutation hotspots.

3.5. R/S ratios per region in the V-REGION

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) and comparing them with the values obtained through a strict consideration of the redundance of the genetic code (Jukes and King, 1979; Shlomchik et al., 1987).

The R/S ratio for the different sub-regions (FR1, FR2, FR3, CDR1, CDR2) of the V-REGION in IGHV genes was calculated, and the results per family and region are summarized in Table 2. The analysis of the total R/S ratio per region (last row in Table 2) shows that the R/S ratio values for the FRs are far below the point of equilibrium (R/S = 2.9), whereas the values for the CDRs are above said value. This same tendency has been reported for the variation between germline genes in humans and mice, where the R/S ratio for the CDRs was above the point of equilibrium (Kirkham and Schroeder, 1994; Ota et al., 2000). This shows that in allelic variation there is also a selection for diversification in the CDRs and for conservation in the FRs, similar to that shown for orthologous and paralogous diversification for IGHV genes. FR2 is notable for its R/S ratio value of 0.6, which implies that almost all of the substitutions occurring in this region are silent. Although both CDRs favor replacement substitutions, the R/S ratio value is much higher for CDR1 (21.0) than for CDR2 (4.3). The reason for such a high value for CDR1 is that nearly all of the substitutions in this region take place in position 33, and all of those occurring in this position are replacements (Fig. 2). Even though CDR2 contains position 50, the most mutated position, several other positions in this region present substitutions, many of which are silent. Consequently, the R/S ratio for CDR2 shows a value closer to the point of equilibrium (4.3), though it also indicates a selection toward diversification.

Upon itemizing the values by gene family, it can be seen that both in the FRs and in the CDRs the R/S ratio values vary considerably among regions for the different families (Table 2). It should be noted that the small families, IGHV6 and IGHV7, for which only one or two alleles have been reported, contribute a very small sample for this type of analysis. There is a great variation among the different families for FR1 and FR3; in FR1, for example, the R/S ratio value for IGHV2 is 0.1, whereas for IGHV2 and IGHV4 the values are 4.0 and 3.0, respectively. Similarly, in FR3 a high conser-

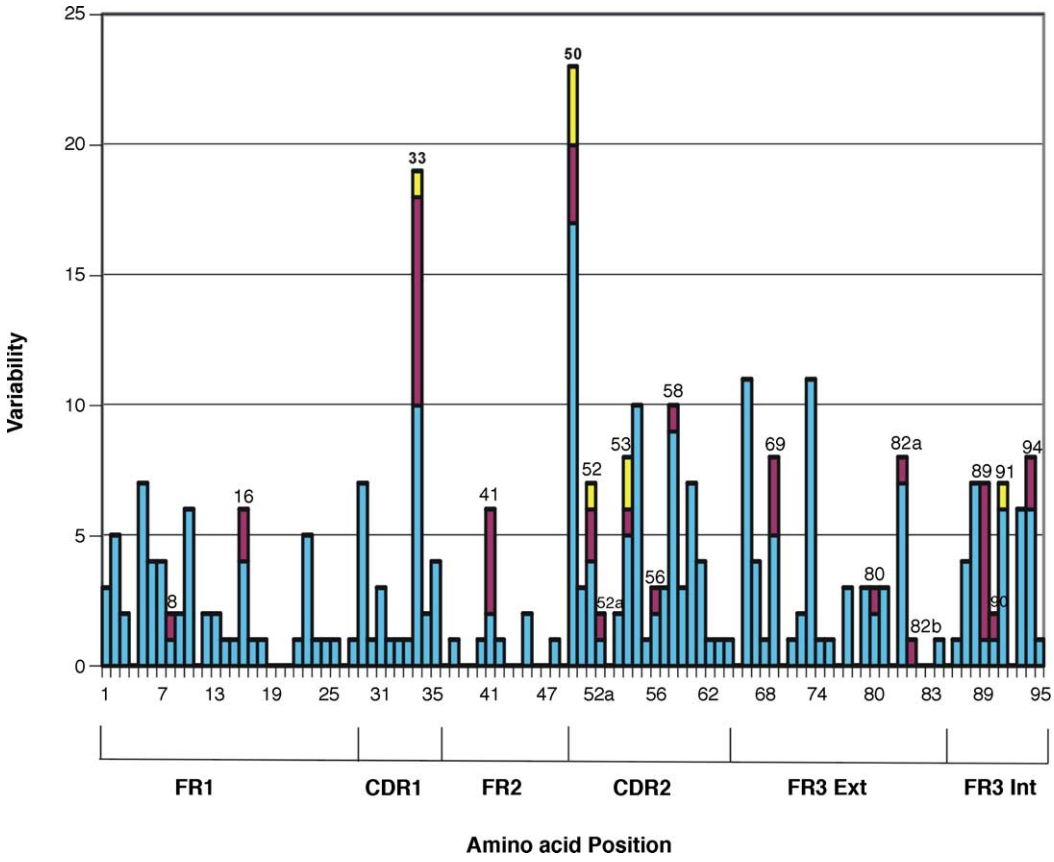

Fig. 4. This figure shows the replacement substitutions occurring within a single codon in the VH exon by positions from 1 to 94. They are grouped by color according to the type of substitutions: yellow for triple substitutions, purple for double, and blue for single ones.

vation is seen for the IGHV4 family (R/S = 0.7); in contrast, families IGHV5 and IGHV7 present only replacement substitutions (2.0/0 and 4.0/0, respectively).

The CDRs also present considerable variations in the R/S ratio values. In CDR1, families IGHV1 and IGHV5, for which 16 and 7 alleles, respectively, have been reported, show only two and one respective replacement substitutions. In contrast, families IGHV2, IGHV3 and IGHV4 in this same CDR show high ratios of replacement. Considerable differences are also observed in CDR2, in which there are families with very high R/S ratio values (IGHV1 and IGHV2), while others have values quite close to the point of mutational equilibrium (Table 2).

The simultaneous occurrence of opposing R/S ratio values between CDR1 and CDR2 in the same family presents an even more striking contrast; in family IGHV3, for example, the R/S ratio for CDR1 is 12.5, whereas the value for CDR2 is 3.0. Another noteworthy case is IGHV1, which as previously mentioned, has only two substitutions in CDR1 (both replacements); whereas 12 substitutions occur in CDR2, 11 of them replacements, due to which a R/S ratio value of 11.0 is obtained (Table 2).

Considering the marked differences observed in the R/S ratio among the different regions, the types of amino acid substitutions among these segments may also be expected to be significantly different. Therefore, the distribution of the three types of amino acid substitution previously described for the different CDR and FR regions was also calculated, the results per region being as shown in Table 3.

FR1 and FR2 show a clear tendency to restrict their substitutions toward the conservation of physico-chemical properties. Nonetheless, 17% of radical changes occurring in FR2 deserve attention. This value is contributed almost entirely by two replacements in positions 41 and 42 of alleles V1-45-03 and V5-51-05, respectively. Since FR2 forms part of the core of the VH:VL interphase and is therefore fundamental for maintaining the stability of the variable superdomain (Vargas-Madrazo and Paz-Garcia, 2003), these alleles

Table 3
Percentage of replacement type for sub-regions of the Exon-V

Region	Radical (%)	Non-conservative (%)	Conservative (%)
FR1	6	12	82
FR2	17	8	75
FR3	2	40	58
CDR1	0	38	63
CDR2	2	51	46
Exon-V	3	36	61

may not be functional. On the other hand, FR3 presents four times as many non-conservative substitutions as do FR1 and FR2.

Although the percentage of radical changes in the CDRs is also quite low, the value of their non-conservative substitutions rises considerably, the latter percentage being greater than that for conservative changes in the case of CDR2.

In general, the types of amino acid substitutions are not observed to occur in the same proportions in the different regions. This is congruent with observations in previous sections indicating the existence of diverse evolutionary restrictions for the different sub-segments.

4. Discussion

In the previous sections we analyzed the substitution patterns at alleles in the IGHV locus. The evolutionary strategy followed by the processes of diversification and/or conservation acting upon IGHV genes seems to be subject to mechanisms of selection and DNA dynamics, since the greatest incidence of substitutions tends to occur in those regions and residues that are important for the recognition of antigens, while 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). 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.

This work shows a detailed study of variation patterns, highlighting the fact that the diversification process does not occur homogeneously among the genes, nor among different regions of V genes (Kirkham and Schroeder, 1994; Blanden et al., 1998; Ota et al., 2000). From the analysis of allelism per gene performed here, we notice a clear tendency to conserve certain genes and diversify others. Nevertheless, evolution does not occur through the direct selection of individual genes having unique characteristics associated with susceptibility or resistance to diseases, but rather at the level of the complete Abs repertoire (Blanden et al., 1998). It is also necessary to consider that, in the evolution of Igs, not only selective pressure molds the repertoire; diverse mechanisms associated with DNA dynamics are also capable of selectively altering certain genes or their sub-regions. This implies the existence of evolutionary dynamics of optimization operating between different forces, and this will be expressed at different levels of organization in locus IGHV. Therefore, it is necessary to be careful when assuming that the same selective pressures will be acting for all the members of a family, be it small or large. For example, in the early studies of the evolution of Ig families it was postulated that, with the objective of increasing the diversity of repertoires, the genetic strategy was to generate large populations of redundant genes (complex families, e.g., family IGHV3), along with the persistent conservation of those small families with unique characteristics (e.g., family IGHV6) (Pascual and Capra, 1991; Ota et al., 2000).

Nevertheless, more detailed studies (like the one presented here) show the existence of much finer strategies. For example, for small families it is observed that families IGHV5, but especially IGHV2, are highly redundant for the different analyses made here, whereas IGHV6 and IGHV7 are highly conserved. Correspondingly, certain genes in large families are highly variable in the number of alleles reported and in other parameters analyzed, while others are conserved (Table 1). For instance, we found that the evolutionary diversification strategies differ greatly between genes V3-23 and V4-34, both of which are expressed very frequently in humans (Milner et al., 1995). Gene V3-23 has only three alleles, and its amino acid substitutions are restricted to CDR2. In contrast, V4-34 has 12 alleles and its substitutions occur all along the V-REGION (data not shown).

In the human genome the alleles of a gene generally differ by an average of one nucleotide, this level of diversity being associated with evolution through genetic drift (Kruglyak, 1997; Halushka et al., 1999). In locus IGHV, 73% of the alleles have one or two substitutions, which also suggests genetic drift; however, the remaining 27% (particularly the 8% of genes having more than eight substitutions) show evidence of intense selective pressure to diversify the repertoire of antibodies.

In previous studies, a great amount of deleterious mutations have been observed in said genes which come from the random genetic drift (Perlmutter et al., 1985; Ota et al., 2000). Additionally, the structural and functional evidences imply a significant role of diversifying selection preserving the germline V gene segment repertoire (Rothenfluh et al., 1995; Ota et al., 2000). Therefore, our results and the foregoing entail complex and contradictory forces in the evolution of human V gene segments.

From the analyses of substitution patterns (number and type of substitutions and R/S ratio) per region and position of the different alleles, a second evolutionary strategy can be distinguished; more explicitly, not only a selective force and "concerted" evolution acting to conserve and/or diversify some of the genes, but also a tendency to diversify only certain regions (mainly CDR2) within the gene may be observed. Furthermore, within those regions only some positions were frequently replaced (positions 33, 50, 53, 54 and 58), being positions in frequent contact, these have been seen to play a central role in the interaction with the antigen (MacCallum et al., 1996). This pattern is very similar to the one observed when the differences among the IGHV genes in humans are analyzed (Tomlinson et al., 1996) and with some patterns in the sequence data that are consistent with "signatures" of somatic events (Tomlinson et al., 1996; Blanden et al., 1998).

It was also found that the FR3 region contains several positions that are highly mutated (Fig. 2), and that the percentage of non-conservative substitutions is as much as four-fold higher than in FR1 and FR2. Certain sub-segments of the FR

that are distant from the antigen-binding site have been postulated as playing an important recognition role, since they affect the antigen binding (Jones et al., 1986; Riechmann et al., 1988; Gorman et al., 1991) or the interaction with other non-specific ligands (Kirkham et al., 1992; Zouali, 1995) For instance, position 71 in FR3 participates in the conformation of CDR2 (Chothia et al., 1989; Foote and Winter, 1992) or affects VH:VL pairing (Saul and Poljak, 1993; Vargas-Madrazo and Paz-Garcia, 2003).

On the other hand, we found that the forces and mechanisms to diversify the CDRs do not operate equally for the different families. That is, in some families CDR1 is more diversified and in others it is CDR2. The different CDRs may perform distinct functions in the recognition properties of antibodies (Rothenfluh and Steele, 1993; Vargas-Madrazo et al., 1995, 1997; Hande and Manser, 1997; Hemminki et al., 1998; Yazici et al., 1998; Decanniere et al., 1999; Roe et al., 1999; Xu and Davis, 2000). Consequently, differential strategies can operate for the CDRs in germinal diversification in accordance with the distinct functions fulfilled by each gene or family (Pascual et al., 1990; Berman et al., 1991; Wang and Stollar, 1999; Van Dijk-Hard and Lundkvist, 2002). Unfortunately, at the present time analyses similar to the ones described above for germline genes or somatic hypermutation are not available.

4.1. Global strategies for repertoire diversification

Various evidence indicates that only a small fraction of the repertoire of immunoglobulin genes is important for the immune response (Zinkernagel and Doherty, 1977; Viale et al., 1992; Rothenfluh et al., 1995; Nobrega et al., 1998; Lacroix-Desmazes et al., 1998; Coutinho, 2000; Zinkernagel and Hengartner, 2001; Zinkernagel, 2002). In previous works we have proposed bi-dimensional maps that make it possible to group and identify V genes that codify for antibodies having very peculiar properties of recognition (Vargas-Madrazo and Paz-Garcia, 2001). The detailed study of substitution patterns in alleles that we have presented here shows that diversification strategies at distinct levels of variation differ considerably among the genes and families, thereby suggesting differential evolutionary strategies for distinct genes in terms of their contribution to the general properties of the Abs repertoire.

This work points up the scarcity of systematic population studies for characterizing both the substitution patterns in alleles and in general the polymorphism of immunoglobulin genes, as well as the importance of more detailed research in the case of certain important genes (Vargas-Madrazo and Paz-Garcia, 2001). This will yield a wealth of information that will help to understand the evolutionary strategies and development of antibody repertoires. A similar analysis of the substitution patterns in alleles for the Vκ and Vλ genes in humans and mice, under the same parameters explored in this study, is presently underway and will be reported in a forthcoming article.

Acknowledgements

We wish to thank Edda Sciutto and Edmundo Lamoyi for their helpful guidance and comments throughout this creative research process. And we especially thank Carlos Larralde, who has been the pivot point for a substantial change in our mental landscapes. Irene Marquina and Warren Haid translated and revised the manuscript. This work was made possible through economic support from the Biological Research Institute of the University of Veracruz. Tania Romo González is supported by scholarships from CONACYT and DGEP-UNAM for postgraduate studies.

References

Adderson, E.E., Azmi, F.H., Wilson, P.M., Shackelford, P.G., Carroll, W.L., 1993. The human VH3b gene subfamily is highly polymorphic. J. Immunol. 151 (2), 800–809.

Andris, J.S., Brodeur, B.R., Capra, J.D., 1993. Molecular characterization of human antibodies to bacterial antigens: utilization of the less frequently expressed VH2 and VH6 heavy chain variable region gene families. Mol. Immunol. 30 (17), 1601–1616.

Baer, R., Forster, A., Lavenir, I., Rabbitts, T.H., 1988. Immunoglobulin VH genes are transcribed by T cells in association with a new 5′ exon. J. Exp. Med. 167 (6), 2011–2016.

Becker, R.S., Knight, K.L., 1990. Somatic diversification of immunoglobulin heavy chain VDJ genes: evidence for somatic gene conversion in rabbits. Cell 63 (5), 987–997.

Berman, J.E., Mellis, S.J., Pollock, R., Smith, C.L., Suh, H., Heinke, B., Kowal, C., Surti, U., Chess, L., Cantor, C.R., Alt, F.W., 1988. Content and organization of the human Ig VH locus: definition of three new VH families and linkage to the Ig CH locus. EMBO J. 7 (3), 727–738.

Berman, J.E., Nickerson, K.G., Pollock, R.R., Barth, J.E., Schuurman, R.K., Knowles, D.M., Chess, L., Alt, F.W., 1991. VH gene usage in humans: biased usage of the VH6 gene in immature B lymphoid cells. Eur. J. Immunol. 21 (5), 1311–1314.

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.

Brezinschek, H.P., Brezinschek, R.I., Lipsky, P.E., 1995. Analysis of the heavy chain repertoire of human peripheral B cells using single-cell polymerase chain reaction. J. Immunol. 155 (1), 190–202.

Campbell, M.J., Zelenetz, A.D., Levy, S., Levy, R., 1992. Use of family specific leader region primers for PCR amplification of the human heavy chain variable region gene repertoire. Mol. Immunol. 29 (2), 193–203.

Chen, P.P., 1990. Structural analyses of human developmentally regulated Vh3 genes. Scand. J. Immunol. 31 (3), 257–267.

Chen, P.P., Liu, M.F., Sinha, S., Carson, D.A., 1988. A 16/6 idiotype-positive anti-DNA antibody is encoded by a conserved VH gene with no somatic mutation. Arthritis Rheum. 31 (11), 1429–1431.

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., Tramontano, A., Levitt, M., Smith-Gill, S.J., Air, G., Sheriff, S., Padlan, E.A., Davies, D., Tulip, W.R., Colman, P.M., Spinelli, S., Alzari, P.M., Poljak, R.J., 1989. Conformations of immunoglobulin hypervariable regions. Nature 342 (6252), 877–883.

Cohn, M., Langman, R.E., 1996. The immune system: a look from a distance. Front Biosci. 1, d318–d323.

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

Cook, G.P., Tomlinson, I.M., Walter, G., Riethman, H., Carter, N.P., Buluwela, L., Winter, G., Rabbitts, T.H., 1994a. A map of the human immunoglobulin VH locus completed by analysis of the telomeric region of chromosome 14q. Nat. Genet. 7 (2), 162–168.

Cook, G.P., Tomlinson, I.M., Walter, G., Riethman, H., Carter, N.P., Buluwela, L., Winter, G., Rabbitts, T.H., 1994b. A map of the human immunoglobulin VH locus completed by analysis of the telomeric region of chromosome 14q. Nat. Genet. 7 (2), 162–168.

Coutinho, A., 2000. Germ-line selection ensures embryonic autoreactivity and a positive discrimination of self mediated by supraclonal mechanisms. Semin. Immunol. 12 (3), 205–213 (discussion 257–344).

Cui, X., Li, H., 1997. Discriminating between allelic and interlocus differences among human immunoglobulin VH4 sequences by analyzing single spermatozoa. Hum. Genet. 100 (1), 96–100.

Cui, X., Li, H., 1998. Determination of gene organization in individual haplotypes by analyzing single DNA fragments from single spermatozoa. Proc. Natl. Acad. Sci. U.S.A. 95 (18), 10791–10796.

Cui, X., Li, H., 2000. Human immunoglobulin VH4 sequences resolved by population-based analysis after enzymatic amplification and denaturing gradient gel electrophoresis. Eur. J. Immunogenet. 27 (1), 37–46.

Decanniere, K., Desmyter, A., Lauwereys, M., Ghahroudi, M.A., Muyldermans, S., Wyns, L., 1999. A single-domain antibody fragment in complex with RNase A: non-canonical loop structures and nanomolar affinity using two CDR loops. Struct. Fold Des. 7 (4), 361–370.

Eigen, M., Lindemann, B.F., Tietze, M., Winkler-Oswatitsch, R., Dress, A., von Haeseler, A., 1989. How old is the genetic code? Statistical geometry of tRNA provides an answer. Science 244 (4905), 673–679.

Foote, J., Winter, G., 1992. Antibody framework residues affecting the conformation of the hypervariable loops. J. Mol. Biol. 224 (2), 487–499.

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.

Gorman, S.D., Clark, M.R., Routledge, E.G., Cobbold, S.P., Waldmann, H., 1991. Reshaping a therapeutic CD4 antibody. Proc. Natl. Acad. Sci. U.S.A. 88 (10), 4181–4185.

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.

Hammarstrom, L., Ghanem, N., Smith, C.I., Lefranc, G., Lefranc, M.P., 1990. RFLP of human immunoglobulin genes. Exp. Clin. Immunogenet. 7 (1), 7–19.

Hande, S., Manser, T., 1997. Single amino acid substitutions in V(H) CDR2 are sufficient to generate or enhance the specificity of two forms of an anti-arsonate antibody variable region for DNA. Mol. Immunol. 34 (18), 1281–1290.

Harmer, I.J., Loizou, S., Thompson, K.M., So, A.K., Walport, M.J., Mackworth-Young, C., 1995. A human monoclonal antiphospholipid antibody that is representative of serum antibodies and is germline encoded. Arthritis Rheum. 38 (8), 1068–1076.

Hemminki, A., Niemi, S., Hautoniemi, L., Soderlund, H., Takkinen, K., 1998. Fine tuning of an anti-testosterone antibody binding site by stepwise optimisation of the CDRs. Immunotechnology 4 (1), 59–69, Jun.

Jones, P.T., Dear, P.H., Foote, J., Neuberger, M.S., Winter, G., 1986. Replacing the complementarity-determining regions in a human antibody with those from a mouse. Nature 321 (6069), 522–525.

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

Kabat, E.A., 1978. The structural basis of antibody complementarity. Adv. Protein Chem. 32, 1–75.

Kirkham, P.M., Mortari, F., Newton, J.A., Schroeder Jr., H.W., 1992. Immunoglobulin VH clan and family identity predicts variable domain structure and may influence antigen binding. EMBO J. 11 (2), 603–609.

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

Kodaira, M., Kinashi, T., Umemura, I., Matsuda, F., Noma, T., Ono, Y., Honjo, T., 1986. Organization and evolution of variable region genes of the human immunoglobulin heavy chain. J. Mol. Biol. 190 (4), 529–541.

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

Kuppers, R., Fischer, U., Rajewsky, K., Gause, A., 1992. Immunoglobulin heavy and light chain gene sequences of a human CD5 positive immunocytoma and sequences of four novel VHIII germline genes. Immunol. Lett. 34 (1), 57–62.

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. Methods 216 (1–2), 117–137.

Lara-Ochoa, F., Vargas-Madrazo, E., Almagro, J.C., 1995. Distributions of the use frequencies of amino acids in the hypervariable regions of immunoglobulins. J. Mol. E 41 (1), 98–103.

Lee, K.H., Matsuda, F., Kinashi, T., Kodaira, M., Honjo, T., 1987. A novel family of variable region genes of the human immunoglobulin heavy chain. J. Mol. Biol. 195 (4), 761–768.

Lefranc, M.P., 2003. IMGT, the international ImMunoGeneTics database. Nucleic Acids Res. 31 (1), 307–310.

Lefranc, M.P., 2004. IMGT-ONTOLOGY and IMGT databases, tools and Web resources for immunogenetics and immunoinformatics. Mol. Immunol. 40 (10), 647–660.

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

Lefranc, M.P., Pallares, N., Frippiat, J.P., 1999. Allelic polymorphisms and RFLP in the human immunoglobulin lambda light chain locus. Hum. Genet. 104 (5), 361–369.

Li, W.-H., 1997. Molecular Evolution. Sinauer Associates Inc.

Li, H., Cui, X., Pramanik, S., Chimge, N.O., 2002. Genetic diversity of the human immunoglobulin heavy chain VH region. Immunol. Rev. 190, 53–68.

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.

Matsuda, F., Ishii, K., Bourvagnet, P., Kuma, K., Hayashida, H., Miyata, T., Honjo, T., 1998. The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. J. Exp. Med. 188 (11), 2151–2162.

Matsuda, F., Shin, E.K., Nagaoka, H., Matsumura, R., Haino, M., Fukita, Y., Taka-ishi, S., Imai, T., Riley, J.H., Anand, R., et al., 1993. Structure and physical map of 64 variable segments in the 3′0.8-megabase region of the human immunoglobulin heavy-chain locus. Nat. Genet. 3 (1), 88–94.

Max, E.E., 1998. Immunoglobulins: Molecular Genetics. In: Paul, W. (Ed.), Fundamental immunology, fourth ed. Lippincott-Raven, Philadelphia, pp. 111–182.

Mazumdar, P.M.H., 1995. Species and Specificity. An Interpretation of History of Immunology. Cambridge University Press, pp. 3–382.

McCormack, W.T., Tjoelker, L.W., Thompson, C.B., 1991. Avian B-cell development: generation of an immunoglobulin repertoire by gene conversion. Annu. Rev. Immunol. 9, 219–241.

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.

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.

Olee, T., Yang, P.M., Siminovitch, K.A., Olsen, N.J., Hillson, J., Wu, J., Kozin, F., Carson, D.A., Chen, P.P., 1991. Molecular basis of an autoantibody-associated restriction fragment length polymorphism that confers susceptibility to autoimmune diseases. J. Clin. Invest. 88 (1), 193–203.

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

Padlan, E.A., 1990. On the nature of antibody combining sites: unusual structural features that may confer on these sites an enhanced capacity for binding ligands. Proteins 7 (2), 112–124.

Padlan, E.A., Abergel, C., Tipper, J.P., 1995. Identification of specificity-determining residues in antibodies. FASEB J. 9 (1), 133–139.

Pargent, W., Schable, K.F., Zachau, H.G., 1991. Polymorphisms and haplotypes in the human immunoglobulin kappa locus. Eur. J. Immunol. 21 (8), 1829–1835.

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

Pascual, V., Randen, I., Thompson, K., Sioud, M., Forre, O., Natvig, J., Capra, J.D., 1990. The complete nucleotide sequences of the heavy chain variable regions of six monospecific rheumatoid factors derived from Epstein-Barr virus-transformed B cells isolated from the synovial tissue of patients with rheumatoid arthritis. Further evidence that some autoantibodies are unmutated copies of germ line genes. J. Clin. Invest. 86 (4), 1320–1328.

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.

Pramanik, S., Li, H., 2002. Direct detection of insertion/deletion polymorphisms in an autosomal region by analyzing high-density markers in individual spermatozoa. Am. J. Hum. Genet. 71 (6), 1342–1352.

Ramirez-Benites, M.C., Almagro, J.C., 2001. Analysis of antibodies of known structure suggests a lack of correspondence between the residues in contact with the antigen and those modified by somatic hypermutation. Proteins 45 (3), 199–206.

Riechmann, L., Foote, J., Winter, G., 1988. Expression of an antibody Fv fragment in myeloma cells. J. Mol. Biol. 203 (3), 825–828.

Roe, M., Norderhaug, I.N., Brandtzaeg, P., Johansen, F.E., 1999. Fine specificity of ligand-binding domain 1 in the polymeric Ig receptor: importance of the CDR2-containing region for IgM interaction. J. Immunol. 162 (10), 6046–6052.

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.

Rothenfluh, H.S., Steele, E.J., 1993. Origin and maintenance of germ-line V genes. Immunol. Cell Biol. 71 (Pt 3), 227–232.

Rubinstein, D.B., Schwartz, R.S., Guillaume, T., Leblanc, P., Stewart, A.K., 1994. Germline complexity, restriction fragment length polymorphism, and coding region sequences of the human VH7 gene family identified with family-specific FR3 segment oligonucleotides. Mol. Immunol. 31 (10), 713–721.

Sanz, I., Dang, H., Takei, M., Talal, N., Capra, J.D., 1989. VH sequence of a human anti-Sm autoantibody. Evidence that autoantibodies can be unmutated copies of germline genes. J. Immunol. 142 (3), 883–887.

Sasso, E.H., Buckner, J.H., Suzuki, L.A., 1995. Ethnic differences of polymorphism of an immunoglobulin VH3 gene. J. Clin. Invest. 96 (3), 1591–1600.

Sasso, E.H., van Dijk, K.W., Milner, E.C., 1990. Prevalence and polymorphism of human VH3 genes. J. Immunol. 145 (8), 2751–2757.

Sasso, E.H., Willems van Dijk, K., Bull, A.P., Milner, E.C., 1993. A fetally expressed immunoglobulin VH1 gene belongs to a complex set of alleles. J. Clin. Invest. 91 (6), 2358–2367.

Sasso, E.H., Willems van Dijk, K., Bull, A., van der Maarel, S.M., Milner, E.C., 1992. VH genes in tandem array comprise a repeated germline motif. J. Immunol. 149 (4), 1230–1236.

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.

Schaible, G., Rappold, G.A., Pargent, W., Zachau, H.G., 1993. The immunoglobulin kappa locus: polymorphism and haplotypes of Caucasoid and non-Caucasoid individuals. Hum. Genet. 91 (3), 261–267.

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.

Tanaka, T., Nei, M., 1989. Positive darwinian selection observed at the variable-region genes of immunoglobulins. Mol. Biol. E 6 (5), 447–459.

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.

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.

Tonegawa, S., 1983. Somatic generation of antibody diversity. Nature 302 (5909), 575–581.

van der Maarel, S., van Dijk, K.W., Alexander, C.M., Sasso, E.H., Bull, A., Milner, E.C., 1993. Chromosomal organization of the human VH4 gene family. Location of individual gene segments. J. Immunol. 150 (7), 2858–2868.

van Dijk, K.W., Sasso, E.H., Milner, E.C., 1991. Polymorphism of the human Ig VH4 gene family. J. Immunol. 146 (10), 3646–3651.

Van Dijk-Hard, I., Lundkvist, I., 2002. Long-term kinetics of adult human antibody repertoires. Immunology 107 (1), 136–144.

van Es, J.H., Heutink, M., Aanstoot, H., Logtenberg, T., 1992. Sequence analysis of members of the human Ig VH4 gene family derived from a single VH locus. Identification of novel germ-line members. J. Immunol. 149 (2), 492–497.

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.

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., Ramirez-Benites, M.C., Almagro, J.C., 1997. Evolution of the structural repertoire of the human V(H) and Vkappa germline genes. Int. Immunol. 9 (12), 1801–1815.

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., Paz-Garcia, E., 2001. Interpreting Gene Usage of immunoglobulins: Structural Classification of the Genes of Immunoglobulins According to Their Recognition Properties. EFIS 2000 International Proceedings Division. pp. 81–90.

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.

Voswinkel, J., Pfreundschuh, M., Gause, A., 1997. Evidence for a selected humoral immune response encoded by VH4 family genes in the synovial membrane of a patient with RA. Ann. N.Y. Acad. Sci. 815, 312–315.

Wang, X., Stollar, B.D., 1999. Immunoglobulin VH gene expression in human aging. Clin. Immunol. 93 (2), 132–142.

Weng, N.P., Snyder, J.G., Yu-Lee, L.Y., Marcus, D.M., 1992. Polymorphism of human immunoglobulin VH4 germ-line genes. Eur. J. Immunol. 22 (4), 1075–1082.

Willems van Dijk, K., Milner, L.A., Sasso, E.H., Milner, E.C., 1992. Chromosomal organization of the heavy chain variable region gene

segments comprising the human fetal antibody repertoire. Proc. Natl. Acad. Sci. U.S.A. 89 (21), 10430–10434.

Winkler, T.H., Fehr, H., Kalden, J.R., 1992. Analysis of immunoglobulin variable region genes from human IgG anti-DNA hybridomas. Eur. J. Immunol. 22 (7), 1719–1728.

Xu, J.L., Davis, M.M., 2000. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. Immunity 13 (1), 37–45.

Yazici, Z.A., Behrendt, M., Goodfield, M., Partridge, L.J., Lindsey, N.J., 1998. Does the CDR3 of the heavy chain determine the specificity of autoantibodies in systemic lupus erythematosus? J. Autoimmun. 11 (5), 477–483.

Zinkernagel, R.M., 2002. Uncertainties—discrepancies in immunology. Immunol. Rev. 185, 103–125.

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 (Pt. 2), 505–510.

Zinkernagel, R.M., Hengartner, H., 2001. Regulation of the immune response by antigen. Science 293 (5528), 251–253.

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