ELSEVIER

# Novel Substitution Polymorphisms of Human Immunoglobulin VH Genes in Mexicans

Tania Romo-González, Jorge Morales-Montor, Mauricio Rodríguez-Dorantes, and Enrique Vargas-Madrazo

ABSTRACT: It has been proposed that the defense and recognition functions of the immune system, especially those mediated by antibodies, require a great diversity of receptors. Nonetheless, functional and structural evidence has demonstrated the presence of restrictions, both in the use of the repertoire and in the recognition of antigens. Fifty-one functional genes have been described in the IghV locus; however, there is a variety of evidences indicating that only a small fraction of the immunoglobulin genes plays a central role in determining the fundamental properties of the antibody repertoire of the immune system. On the basis of this functional and structural information, we selected four IghV genes and characterized their polymorphism in a sample of Mexican individuals. We also analyzed the implications for the recognition mechanism of the substitutions found in the sequenced alleles. We found that diversification through allelism varies from segment to segment, both in the amount of alleles encountered and in the nature and distribution of mutations in the codifying zone, which might depend on its importance for the repertoire. Such functional characteristics may be useful in the interpretation of differential gene usage in certain physiological, ontological, and/or pathological conditions. *Human Immunology 66, 732–740 (2005).* © American Society for Histocompatibility and Immunogenetics, 2005. Published by Elsevier Inc.

KEYWORDS: polymorphism; nucleotide substitutions; immunoglobulins; VH genes; immune recognition

## INTRODUCTION

It has been proposed that the genetic processes creating diversity in the repertoire of antibodies (Abs) can potentially generate $10^{11}$ specificities, including those virtually harmful [1]. This great diversity of specificities enables us to understand, at least partially, that vertebrates are able to cope with the great antigenic universe. Moreover, functional and structural evidence has demonstrated that not all variants generated by the diversification processes have been found naturally in the circulating Abs repertoires [2–7]. This fact indicates restrictions in using the Ab repertoire and therefore in the recognition of antigens (Ags).

For example, the analysis of the structural repertoire of immunoglobulins (Igs) has indicated the presence of restrictions and topological preferences. Of the 300 possible general forms of antigen-binding sites generated by the combination of diverse types of canonical structures (SC) for five of the six hypervariable loops [8], only 29 combinations (SC classes, or CSC) have been found in Igs [5]. Furthermore, not all hypervariable loops participate equally to generate diversity in the structural repertoire. Because H2, H3, and Ll vary in their length and conformation, these are the ones mainly responsible for the generation of structural variants that allow the recognition of different Ags [3, 5, 9, 10].

Similar to that observed in the structural repertoire, limitations and directions in the functional analyses of its variability have also been found. Studies evaluating gene use have revealed that of the 51 functional segments that codify the VH domain of Igs, very few are expressed in circulating Abs repertoires, although some of them are preponderant in certain ontological, physiological, and pathological states [2, 4, 6]. For instance, in humans, the gene V3-23 (VH26) of the IghV3 family (one of the most numerous and expressed of IgV genes) is the most pre-

*From the Departamento de Biología Sistémica, Instituto de Investigaciones Biológicas, Universidad Veracruzana, Xalapa, Veracruz, México (T.R.-G., E.V.-M.); and Departamento de Inmunología, Instituto de Investigaciones Biomédicas, UNAM, México, México (J.M.-M., M.R.-D.).*

*Address reprint requests to: Dr. Tania Romo-González, 2a Schubert No. 4 Indeco Animas, C.P. 91190, Xalapa, Veracruzana, México; Tel: +52 (228) 8418900 ext. 13401; Fax: +52 (228) 8418911 ext. 15911; E-mail: romisnaider@yahoo.com.mx.*

*Received January 10, 2005; revised February 26, 2005; accepted March 1, 2005.*

Human Immunology 66, 732–740 (2005)
© American Society for Histocompatibility and Immunogenetics, 2005
Published by Elsevier Inc.

0198-8859/05/$–see front matter
doi:10.1016/j.humimm.2005.03.002

Material may be protected by copyright law (Title 17, U.S. Code)

ponderant one throughout life, not only of its family but of all the IghV locus, with an expression of 20% to 30% [11–13]. This expression constancy of gene V3-23 is not observed in any other gene or gene family because in other cases, preeminence varies according to the age and physiopathology of the organism in question. Although gene V6-1 is preferentially used in earlier developmental stages and in autoimmune processes in the adult [6, 14, 15], the IghV4 family genes (mainly 4-59, 4-34, and 4-39) are more frequent in the adult's repertoire [1, 2, 16, 17]. These and other results imply that only a small fraction of the repertoire of Ig genes plays an important role in determining the fundamental properties of the immune system's repertoires of antibodies [7, 13, 18–24], and they outline the existence of an important selective force in terms of structural, and hence functional, characteristics acting on Ig genes [5, 25–29]. In this light, we think that the study of some aspects of the molecular evolution of V genes and the consequences of these alterations on the structural properties of the antigen-binding site will help us understand the strategies of the immune system for creating repertoires of diverse antibodies.

Although the mechanisms operating for the diversification of the Abs repertoire at a somatic level have been studied in great detail, very little is known about the genetic contribution for diversifying this repertoire [29–32]. Furthermore, it has been established that knowing and understanding the germline repertoire and its processes of evolutive diversification is fundamental because such knowledge can contribute to the understanding of the differential expression of V genes and their relationship to certain pathologies [31]. We have recently analyzed the substitution patterns in alleles of IgV genes in humans and mice [33, 34] and found that in general polymorphism is considered as one of the phenomena introducing diversity in the germ line repertoire. Nevertheless, this fact only indicates how little is known about the contribution of polymorphism to the formation of repertoires of circulating antibodies, because most of the studies evaluating polymorphism concentrate on the calculation of allelic frequencies in a population. Additionally, those studies come mostly from Caucasian populations [35–38]. By a detailed analysis of this phenomenon by means of parameters describing information of the genetic and structural properties of the substitutions, it is possible to observe patterns and evolutionary strategies for diversification acting on those genes [33, 34]. In view of the above, it is important to extend the characterization of polymorphism, at least of certain genes of interest [39].

To date, 51 IghV functional genes have been described [40]; however, not all of them participate in the repertoire of circulating antibodies. In previous work, we have proposed bidimensional maps that permit the grouping and identification of V genes codifying for antibodies with distinctive properties of recognition [39]. On the basis of this information, and on the basis of analyses of the use of genes mentioned above, we selected four IghV genes and characterized their polymorphism in a sample of Mexican individuals. Additionally, we analyzed the implications for the recognition mechanism of the substitutions found in the sequenced alleles.

## MATERIALS AND METHODS

### Donor Selection

All subjects were invited to participate voluntarily. Participants and parents (in the case of minors) received an explanation of the project objectives and the specific procedures included in the study. They then gave written informed consent.

*Natives.* Ten unrelated individuals belonging to the Nahua population from Zongolica, Veracruz, were recruited for this study. We chose this population on the basis of its genetic homogeneity, a criterion they fulfill because it has been geographically isolated from genetic recombination with other populations, a characteristic that guarantees (at least in part) the pure ethnicity required to fulfill our study objectives. An expert anthropologist of the CIESAS Golfo in Xalapa, Veracruz, designed a methodology in order to characterize the indigenous population. The methodology included determining the ethnicity of all subjects by direct questioning. Each subject was asked to state the ethnicity, country of birth, and language of each grandparent. For each subject, this information was used to construct a family tree of ethnicity.

*Mestizos.* A group of 10 unrelated individuals was included. The group consisted of student volunteers, all of whom were born and living in Xalapa and surroundings. Although they are an admixed sample (Native American with white) of individuals, they have an important native background, as revealed in the anthropological study. In Mexico, determination of ethnicity is relatively easy because most of the genetic mixes have been with the Spanish, and in few cases, with Africans, both of which have characteristic phenotypes. However, in order to assure mestizo origin, during the interview, we asked patients about their ancestors. All subjects whose grandparents were not Mexicans by birth were rejected for inclusion in the study.

### Selection of Genes for Study

The following genes were selected in accordance with the structural criteria for the use of genes: gene V3-23 (VH26), which is expressed with great frequency in fetal and adult B lymphocytes and in autoantibodies [4, 30]; gene V6-1, the only member of the family IghV6, which

Material may be protected by copyright law (Title 17, U.S. Code)

appears preferentially in fetal lymphocytes and autoantibodies [6, 14, 15] and which is also the only gene that codifies for CSC 3-5, a class with unique structural characteristics [5, 39, 41]; and genes V4-59 and V4-61, which presented 100% identity in those relevant positions to the interaction with the Ab but codify for different CSC (1-1 and 3-1, respectively) [39].

### Primer Design

We designed family-specific primers for IghV segments based on the methodology of Tomlinson et al. [42], in which the priming regions are located on the heptamer and part of the recombination spacer at the 3′ end of the VH exon, and regions of the leadering exon or intron at the 5′ end. Priming on the heptamer has the advantage that since the heptamer is lost during recombination, rearranged IghV genes are not amplifed.

On the other hand, "internal" primers for IghV3 and IghV4 were designed on the basis of those regions of CDR1 and CDR2 that display the greatest diversity within the family and are the target of the V3-23, V4-59, and V4-59 sequences. This strategy allowed us to identify the gene of interest. The sequences of the primers are listed in Table 1.

### Preparation of Genomic DNA

Genomic DNA was isolated from peripheral white blood cells obtained from 20 healthy volunteers (10 Mexican native and 10 mestizo subjects) by means of a previously described method [43]. Briefly, 5 ml of whole blood was collected in a tube with 0.057 EDTA at 15%. The whole blood was transferred to a 20-ml polypropylene tube containing 10 ml of ice-cold phosphate-buffered saline, 10 mM phosphate, and 150 mM NaCl, pH 7.2. The nuclear pellet was isolated by centrifugation at 800 *g* for 7 minutes at 4°C and then resuspended in 10 ml of extraction buffer (300 mM sodium acetate, 50 mM EDTA pH 7.5, RNaseA to a final concentration of 25 μg/ml), and 50 μl proteinase K. The samples were incubated at 50°C for 5 hours. The pellet was extracted twice with phenol/chloroform and once with chloroform, and precipitated with ice-cold ethanol. Then the samples were resuspended in 500 μl of water and quantified by measuring their absorbance at 260 nm.

**TABLE 1** Family and gene-specific primers

| Primer | Sequence |
|---|---|
| VH3 5′ | 5′-CTGAATTCCATGGAGTTTGGGCTGAG-3′ |
| VH3 3′ | 5′-GACTCTAGACAATGAACTTCCCCTCACT-3′ |
| VH4 5′ | 5′-CTGAATTCCATGAAACACCTGTGGTTCTT-3′ |
| VH4 3′ | 5′-GACTCTAGAGGGCTCACACTCACCTCCCCT-3′ |
| VH6 5′ | 5′-GACTCTAGAATGTCTGTCTCCTTCCTCAT-3′ |
| VH6 3′ | 5′-GGAATTCCTGACTTCCCCTCACTGTG-3′ |
| V3-23[a] | 5′-AGCAGCTATGCCATGAGCTGG-3′ |
| V4-59 | 5′-TCCATCAGTAGTTACTACTGG-3′ |
| V4-61 | 5′-TCCGTCAGCAGTGGTAGTTAC-3′ |

[a] M8, [11].

### PCR Amplification and Sequencing

Genomic DNA was amplified by pairs of PCR primers designed as previously described [42] in a thermal cycler with Promega *Thermus aquaticus* (Taq) DNA polymerase. Reaction mixtures (50 μl) were prepared containing 25 pmol of each primer, 5–10 μg of genomic DNA, 2.5 units of Taq polymerase, 20 μM each dNTPs and the recommended buffer (Promega: 50 mM KCl, 10 mM Tris-HCl pH 8.8, 1.5 mM $MgCl_2$, 0.1% Triton X-100). The reaction mixture was performed with 30 cycles of amplification. Each cycle consisted of denaturation (94°C for 1 minute), annealing (55°C for 1 minute), and extension (72°C for 2 minutes). At the end of 30 cycles, there was a final extension at 72°C for 5 minutes. The product was analyzed by running 5 μl on a 2% agarose gel. The remainder was extracted with phenol/chloroform, precipitated with ethanol, and digested with restriction enzymes *Eco*RI and *Xba*I. A band of the expected size was cut from a 1.5% low melting point agarose gel and then purified and precipitated with ethanol.

The product was ligated into Puc-19 that had been digested with *Eco*RI and *Xba*I. The ligation mix was used to transform *Escherichia coli* XL1-blue cells by electroporation [44]. A single-stranded template from selected plaques was prepared and sequenced by the dideoxy chain termination method [45].

Several precautions were taken to avoid cross-contamination. Negative controls (no genomic DNA added) were always included in all amplifications to check for DNA contamination. Independent amplification with identical sets of primers was simultaneously undertaken to avoid clones isolated from one amplification contaminating the next. In all cases, we imposed the requirement that each germline VH segment had to be observed in at least two independent amplifications.

### Compilation of Germline Database and Structural-Functional Analysis

DNA sequences were aligned and translated by a sequence analysis program (Lasergene). The sequences were compared in detail with the germline genes. Once the fixed mutations (substitutions, insertions, or deletions) had been identified, each allele sequence was analyzed on the basis of its implications for evolution and for the mechanism of molecular recognition. This scheme has been developed by our research group [5, 28, 29].

## RESULTS

### Allelism Analysis per IghV Gene

Genes V3-23, V4-59, V4-61, and V6-1 were characterized in 20 individuals (10 natives and 10 mestizos) through

Material may be protected by copyright law (Title 17, U.S. Code)

sequencing in order to determine the degree of polymorphism in those genes. The sequences obtained for each one of the four studied genes were aligned and compared in detail with the predominant allele of each one.

We did not identify new polymorphisms in the V6-1 and V4-61 genes in any of the examined samples. The predominant allele, previously designated as V6-1*01 and V4-61*01, respectively [46], was present in the studied individuals in all cases. Regarding this, it is important to point out that alleles have been reported for both genes in the Caucasian population. For gene V6-1, only one allele has been reported; five alleles have been found for gene V4-61 [33, 46].

For V3-23 and V4-59 (Figures 1 and 2), two new or unreported alleles were identified. In the case of gene V4-59, the variant found was designated as V4-59*11 because currently there are nine reported alleles in addition to the predominant allele. All the individuals (natives or mestizos) examined in this study presented the allele described herein. This result in each individual comes from homologous sequences of amplified DNA from at least two independent PCRs, which provides reliability of the results and discards possible contamination.

Also, upon characterizing gene V3-23, we identified a new allele in addition to the three previously reported and we designated it as V3-23*04. It is important to point out that unlike that observed in gene V4-59, in the samples examined here, both the predominant allele of this gene (V3-23*01) and the new variant (V3-23*04) were present. In the 20 individuals sampled (10 natives and 10 mestizos), allele V3-23*04 was present in 5 individuals (2 natives and 3 mestizos), whereas in the other 15, the sequenced allele was V3-23*01.

### Number, Type, and Localization of Substitutions in the Alleles Found

In the alleles reported here (V3-23*04 and V4-59*11), we identified six and eight substitutions, respectively (Figures 1 and 2). Such substitutions were located in the following positions: 52 AG/TA; 52a G-T/A-C; 55 G/A and 59 C/T in V3-23*04 and 41 C-A/G-C; 50 TA/CG; 53 TA/AC; and 69 A/G; 84 T/C in V4-59*11. In allele V3-23*04, all substitutions were located in CDR2, and five of the six were replacement substitutions.

It is important to note that the amino acid replacement in positions 52 and 52a was the product of a double substitution (Figure 1). Such positions have been observed in frequent contact with the Ag [47]. Regarding allele V4-59*11, seven of its eight substitutions generated an amino acid change, six of them occurring through a double substitution in the codifying codon. Although the number of substitutions is similar to that encountered in V3-23*04, and double substitutions occur repeatedly, the substitutions in this allele are more heterogeneously distributed because they were located not only in CDR2, but also in FR2 and FR3 (Figure 2). Despite the large number of replacement substitutions seen in both cases, upon analyzing the type of modification, only 2 of the 17 replacements implied a nonconservative change in physicochemical properties of the residue [33, 48, 49], whereas the other 15 were of the conservative type (Figures 1 and 2).

## DISCUSSION

It has been proposed that allelic polymorphism is one of the fundamental sources of evolutionary variation in Ig genes [30]. However, the functional and structural implications of this mechanism are little understood. Because understanding

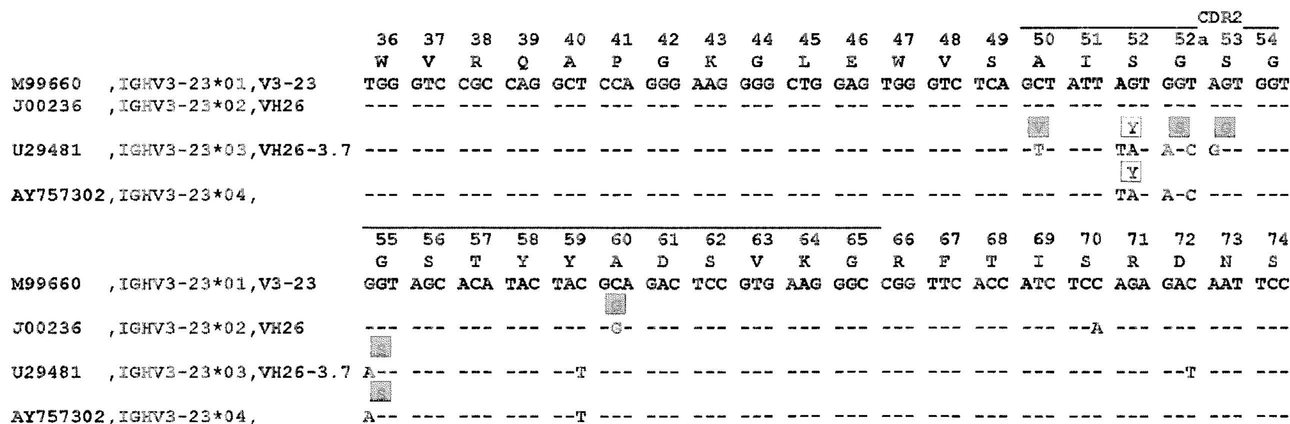


FIGURE 1 Alignment of the alleles of gene IGHV3-23 in humans. The alleles previously reported and their substitutions are indicated in red, the novel allele and its substitutions in blue. Among the substitutions, those generating amino acid replacements were classified in three groups, according to the type of change in the physicochemical properties of the residue [33, 47, 48]: red for radical, yellow for nonconservative, and green for conservative substitutions.

Material may be protected by copyright law (Title 17, U.S. Code)

the germline repertoire and its processes of evolutionary diversification is fundamental [31], in the present work, we studied germline diversification by substitution polymorphism in 4 of the 51 functional genes. This selection is possible because not all genes participate equally, nor do they have the same importance in the repertoire.

As a result of the study of substitution polymorphism in the four genes selected, two new alleles were identified in the codifying regions of genes V3-23 and V4-59. In gene V4-59, the same allele was identified in all the individuals analyzed. This preponderance might indicate an evolutionary advantage for this population. The same preponderance was not observed in gene V3-23, in which case the individuals studied presented both the predominant allele of the gene (V3-23*01) and a new variant (V3-23*04). Because the latter was present in the natives as well as in the mestizos, it may be assumed that this allele originated before the mixing of the races.

Although this study does not evaluate allelic frequencies and the number of individuals sample is not extensive, the following should be noted.

No novel alleles have been identified in gene V6-1. The conservation observed herein is consistent with previous reports because this gene has generally been described as demonstrating little polymorphism [50–52] and few mutations (the only mutation observed in the only allele reported is of the silent type) [53]. Furthermore, this conservation occurs not only at the germline level but also at the somatic level [2, 6, 54, 55]. The importance of this gene for the repertoire [14, 56–59] and its marked conservation make it possible to consider the presence of dynamic and evolutionary forces acting to maintain the structural and functional characteristics of this gene.

Of the 20 individuals sampled here, no new alleles were found for gene V4-61. This is a surprising result, not only because this gene, together with V4-59, is among the most polymorphic ones of the IghV locus and one of the most susceptible to mutation [33, 46], but also because for gene V4-59, we found an allele (V4-59*11) for this sample.

In contrast to the observed in gene V4-59, in which the same allele was identified in all individuals, the new allele reported herein for gene V3-23 is not present in all the samples. Instead, some subjects presented the preponderant allele of the gene (V3-23*01), whereas others presented the variant (V3-23*04). Several studies have demonstrated that this gene is one of the most expressed in humans throughout life, regardless of their ontological, physiological, or pathological condition [4, 11–13]. In addition, it has been demonstrated that the antibodies codified by this gene, though great plasticity, recognize both self-antigens as well as nonself ones (autoreactive and multireactive antibodies) [11, 13].

Previous studies of the IgV locus [33, 34] and of other loci [60, 61] have made it evident that polymorphism

**TABLE 2** R/S ratio for subregion of human VH domain of genes V3-23, V4-59, V4-61, and V6-1

| Families and genes | Number of alleles | FR1 | CDR1 | FR2 | CDR2 | FR3 |
|---|---|---|---|---|---|---|
| IghV3 | | | | | | |
| V3-23 | 3 | — | — | — | 6.0 | 0/2 |
| IghV4 | | | | | | |
| V4-59 | 10 | 9.0 | — | 1.6 | 2.8 | 0.2 |
| V4-61 | 5 | 4.0 | 2/0 | 2.0 | 4/0 | 0.5 |
| IghV6 | | | | | | |
| V6-1 | 1 | 0/1 | — | — | — | — |
| Totals | | 4.3 | 2/0 | 1.6 | 4.1 | 0.3 |

can be characterized in several ways, not only by the number of alleles and their frequency in a determined population, but also by the number of mutations per allele [60]—as well as by the distribution and type of substitutions occurring throughout the gene [61].

Upon analyzing the number, type, and location of substitutions in the alleles of the sampled individuals, we found that the number and the distribution of mutations is not homogeneous and varies among these genes (Figures 1 and 2). Moreover, there is not only a biased distribution of mutations, but also a repeated substitution of some residues that occurs through double mutations at the same codon. Given that the probability of having a double substitution in a codon is very small (0.0036 if exon V consists of 282 nucleotides), this phenomenon may be associated with the presence of mutation hotspots over these positions [62].

Furthermore, if the distribution of substitution per subregion is analyzed by calculating the replacement/silent (R/S) ratio [63, 64] (Table 2), it is possible to observe diversification tendencies proper to each genetic segment. Particularly in gene V3-23, substitutions are favored in CDR2 (R/S = 6.0); in gene V4-59, diversification is favored on FR1 (R/S = 9.0), whereas the diversification in gene V4-61 is more homogenous among the regions (Table 2). These results imply that diversification is not homogenous throughout the subregions between different gene segments and, more important, that it is not centered only on CDRs. This observation is interesting because it has been postulated that certain subsegments of the FR that are distant from the antigen-binding site could play an important recognition role because they affect the antigen binding [65–67] or the interaction with other nonspecific ligands [68, 69]. For instance, position 71 in FR3 participates in the conformation of CDR2 [70, 71] or affects VH:VL pairing [72, 73].

Consequently,the distinctive substitution patterns of each gene segment could generate substantial changes in the affinity between alleles, which is possible with only a

Material may be protected by copyright law (Title 17, U.S. Code)

```
                                          CDR1 region and other annotations as printed

                                      1   2   3   4   5   6   7   8   9   10  11  12  13  14  15  16  17  18  19  20
                                      Q   V   Q   L   Q   E   S   G   P   G   L   V   K   P   S   E   T   L   S   L
L10088   ,IGHV4-59*01, 3d197d        CAG GTG CAG CTG CAG GAG TCG GGC CCA GGA CTG GTG AAG CCT TCG GAG ACC CTG TCC CTC
M29812   ,IGHV4-59*02, V71.4         --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
M95114   ,IGHV4-59*03, H4            --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
M95117   ,IGHV4-59*04, H7            --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
M95118   ,IGHV4-59*05, H8            --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
M95119   ,IGHV4-59*06, H9            --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

X56360   ,IGHV4-59*07, VH4.16        --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --C --- --- --- ---
X87091   ,IGHV4-59*08, DP-71RB                                                                               --- ---
Z75359   ,IGHV4-59*09, VH4-GL7       --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

Z14243   ,IGHV4-59*10                --- --- --- --A --- C-- -G- --- G-- --- --- T-- --- --- --- --- --- --- --- ---
AY757303,IGHV4-59*11                 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                                                                          ____________CDR1____________
                                      21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39  40
                                      T   C   T   V   S   G   G   S   I   S   S   Y   Y   W   S   W   I   R   Q   P
L10088   ,IGHV4-59*01, 3d197d        ACC TGC ACT GTC TCT GGT GGC TCC ATC AGT AGT TAC TAC TGG AGC TGG ATC CGG CAG CCC

M29812   ,IGHV4-59*02, V71.4         --- --- --- --- --- --- --- --- G-- --- --- --- --- --- --- --- --- --- --- ---
M95114   ,IGHV4-59*03, H4            --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
M95117   ,IGHV4-59*04, H7            --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
M95118   ,IGHV4-59*05, H8            --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --G

M95119   ,IGHV4-59*06, H9            --- --- --- --- A-- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
X56360   ,IGHV4-59*07, VH4.16        --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
X87091   ,IGHV4-59*08, DP-71RB       --- --- --- --- --- --- --- --- --- --X --- --- --- --- --- --- --- --- --- ---
Z75359   ,IGHV4-59*09, VH4-GL7       --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

Z14243   ,IGHV4-59*10                --- --- G-- --- -A- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
AY757303,IGHV4-59*11                 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                                                                          ________________CDR2________________
                                      41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60
                                      P   G   K   G   L   E   W   I   G   Y   I   Y   Y   S   G   S   T   N   Y   N
L10088   ,IGHV4-59*01, 3d197d        CCA GGG AAG GGA CTG GAG TGG ATT GGG TAT ATC TAT TAC AGT GGG AGC ACC AAC TAC AAC
M29812   ,IGHV4-59*02, V71.4         --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
M95114   ,IGHV4-59*03, H4            --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

M95117   ,IGHV4-59*04, H7            --- --- --- --- --- --- --- --- --- --- --- --- --T --- --- --- --- T-- --- ---
                                                                          R
M95118   ,IGHV4-59*05, H8            --G --- --- --- --- --- --- --- --- CG- --- --- --T --- --- --- --- T-- --- ---

M95119   ,IGHV4-59*06, H9            G-T --- --- --C --- --- --- --- --- --C --- --- --- --- --- --- --- T-- --- ---
X56360   ,IGHV4-59*07, VH4.16        --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
X87091   ,IGHV4-59*08, DP-71RB       --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
                                                                          R
Z75359   ,IGHV4-59*09, VH4-GL7       G-C --- --- --A --- --- --- --- --- CG- --- --- AC- --- --- --- --- --- --- ---
                                                                          R
Z14243   ,IGHV4-59*10                G-C --- --- --G --- --- --- --- --- CG- --- --- AC- --- --- --- --- --- --- ---
                                                                          R
AY757303,IGHV4-59*11                 G-C --- --- --- --- --- --- --- --- CG- --- --- AC- --- --- --- --- --- --- ---

                                     ___________________
                                      61  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76  77  78  79  80
                                      P   S   L   K   S   R   V   T   I   S   V   D   T   S   K   N   Q   F   S   L
L10088   ,IGHV4-59*01, 3d197d        CCC TCC CTC AAG AGT CGA GTC ACC ATA TCA GTA GAC ACG TCC AAG AAC CAG TTC TCC CTG
M29812   ,IGHV4-59*02, V71.4         --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
M95114   ,IGHV4-59*03, H4            --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

M95117   ,IGHV4-59*04, H7            --G --- --- --- --- --- --- --- --G --- --- --- --- --- --- --- --A --- --- ---
M95118   ,IGHV4-59*05, H8            --G --- --- --- --- --- --- --- --- --C --- --- --- --- --- ---
M95119   ,IGHV4-59*06, H9            --G --- --- --- --- --- --T --- --- --- --- --- --- --T --- ---
X56360   ,IGHV4-59*07, VH4.16        --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
X87091   ,IGHV4-59*08, DP-71RB       --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

Z75359   ,IGHV4-59*09, VH4-GL7       --- --- --- --- --- --- --- --- --G --- --- --- --- --- --- ---

Z14243   ,IGHV4-59*10                --- --- --- --- --- --- --- --- --G --- --- --- --- --- --- ---

AY757303,IGHV4-59*11                 --- --- --- --- --- --- --- --- --G --- --- --- --- --- --- ---

                                      81  82  82a 82b 82c 83  84  85  86  87  88  89
                                      K   L   S   S   V   T   A   A   D   T   A   V
L10088   ,IGHV4-59*01, 3d197d        AAG CTG AGC TCT GTG ACC GCT GCG GAC ACG GCC GTG
M29812   ,IGHV4-59*02, V71.4         --- --- --- --- --- --- --- --- --- --- --- ---
M95114   ,IGHV4-59*03, H4            --- --- --- --- --- --- --- --- --- --- --- ---
M95117   ,IGHV4-59*04, H7            --- --- --- --- --- --- --C --A --- --- --T ---
M95118   ,IGHV4-59*05, H8            --- --- --- --- --- --- --C --A --- --- --T ---
M95119   ,IGHV4-59*06, H9            --- --- --- --- --- --T --C --- --- --- --- ---
X56360   ,IGHV4-59*07, VH4.16        --- --- --- --- --- --- --- --- --- --- --- ---
X87091   ,IGHV4-59*08, DP-71RB       --- --- --- --- --- --- --C --A --- X-- --- ---
Z75359   ,IGHV4-59*09, VH4-GL7       --- --- --- --- --- --- --C --- --- --- --- ---
Z14243   ,IGHV4-59*10                --- --- --- --- --- --- --C --- --- --- --- ---
AY757303,IGHV4-59*11                 --- --- --- --- --- --- --C --- --- --- --- ---
```

FIGURE 2 Alignment of the alleles of gene IGHV4-59 in humans. The alleles previously reported and their substitutions are indicated in red, the novel allele and its substitutions in blue. Among the substitutions, those generating amino acid replacements were classified in three groups, according to the type of change in the physicochemical properties of the residue [33, 47, 48]: red for radical, yellow for nonconservative, and green for conservative substitutions.



Material may be protected by copyright law (Title 17, U.S. Code)

few, but significant, changes. For example, in a previous study that compared the affinity of two alleles of gene V3-23 (V3-23*01 and 03), it was observed that the variation (although small) in the positions substituted between alleles produced deviations that made the Ag binding as much as 20 times more effective [74]. This means that the differences in CDR2 presented by the alleles V3-23*03 confer a marked increase on its affinity [74]. With respect to gene V4-59 (the other gene in which we found an allele), there is no affinity test that compares the effects of mutations on its alleles that postulates a correlation between the mutations found on it and their functionality. However, it important to note that this gene has been found encoding Abs molecules that bind a wide array of three-dimensionally similar ligands such as streptococcal M protein, myosin, other α-helical proteins, and carbohydrate epitopes [73]; it has also typically been found rearranged in mutated mantle-cell lymphoma [76, 77].

To summarize, diversification by allelism in the genes analyzed varies from segment to segment, both in the number of alleles found and in the distribution and nature of mutations in the codified area, depending on the importance of each one for the repertoire. In previous studies by our research group, similar strategies have been observed in the evolution of the families of Ig genes [28] and their relationship with the topology of the antigen-binding site [39].

ACKNOWLEDGMENTS

We thank Dr. Edda Sciutto and Dr. Edmundo Lamoyi for their guidance and comments and Galileo Escobedo and Nelly Téllez for their methodological and experimental supervision. Irene Marquina and Warren Haid translated and revised the article in manuscript. This work was made possible through financial support from the Institute of Biological Research of the University of Veracruz, and through grants 40072 from Consejo Nacional de Ciencia y Tecnología (CONACYT) from México and IN-208103 from Programa de Apoyo a Proyectos de Investigación e Inovación Tecnológica (PAPIIT) from Dirección General de Asuntos del Personal Académico (DGAPA), UNAM, both to J M-M. Tania Romo-González is supported by Ph.D. scholarships from CONACYT and DGEP-UNAM.

## REFERENCES

1. Pinchuk GV, Alexander CM, Glas AM, Armitage RJ, Milner EC: VH repertoire in human B lymphocytes stimulated by CD40 ligand and IL-4: evidence for positive and negative selection mechanisms coupled to CD40 activation. Mol Immunol 33:1369, 1996.
2. Cuisinier AM, Gauthier L, Boubli L, Fougereau M, Tonnelle C: Mechanisms that generate human immunoglobulin diversity operate from the 8th week of gestation in fetal liver. Eur J Immunol 23:110, 1993.
3. Wilson IA, Stanfield RL: Antibody-antigen interactions: new structures and new conformational changes. Curr Opin Struct Biol 4:857, 1994.
4. Schroeder HW Jr, Mortari F, Shiokawa S, Kirkham PM, Elgavish RA, Bertrand FE 3rd: Developmental regulation of the human antibody repertoire. Ann N Y Acad Sci 764:242, 1995.
5. Vargas-Madrazo E, Lara-Ochoa F, Almagro JC: 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: 497, 1995.
6. van Dijk-Hard I, Soderstrom I, Feld S, Holmberg D, Lundkvist I: Age-related impaired affinity maturation and differential D-JH gene usage in human VH6-expressing B lymphocytes from healthy individuals. Eur J Immunol 27:1381, 1997.
7. Coutinho A: Germ-line selection ensures embryonic autoreactivity and a positive discrimination of self mediated by supraclonal mechanisms. Semin Immunol 12:205, 2000.
8. Chothia C, Lesk AM: Canonical structures for the hypervariable regions of immunoglobulins. J Mol Biol 196:901, 1987.
9. James LC, Roversi P, Tawfik DS: Antibody multispecificity mediated by conformational diversity. Science 299:1362, 2003.
10. Zemlin M, Klinger M, Link J, Zemlin C, Bauer K, Engler JA, Schroeder HW Jr, Kirkham PM: Expressed murine and human CDR-H3 intervals of equal length exhibit distinct repertoires that differ in their amino acid composition and predicted range of structures. J Mol Biol 334: 733, 2003.
11. Sasso EH, Buckner JH, Suzuki LA: Ethnic differences of polymorphism of an immunoglobulin VH3 gene. J Clin Invest 96:1591, 1995.
12. Mageed RA, Harmer IJ, Wynn SL, Moyes SP, Maziak BB, Bruggemann M, MacKworth-Young CG: Rearrangement of the human heavy chain variable region gene V3-23 in transgenic mice generates antibodies reactive with a range of antigens on the basis of VHCDR3 and residues intrinsic to the heavy chain variable region. Clin Exp Immunol 123:1, 2001.
13. Mackworth-Young CG, Harmer IJ, Mageed RA: The role of antigen in the selection of the human V3-23 immunoglobulin heavy chain variable region gene. Clin Exp Immunol 134:420, 2003.
14. Berman JE, Nickerson KG, Pollock RR, Barth JE, Schuurman RK, Knowles DM, Chess L, Alt FW: VH gene usage in humans: biased usage of the VH6 gene in immature B lymphoid cells. Eur J Immunol 21:1311, 1991.
15. Raaphorst FM, Langlois van den Bergh R, Waaijer JL, Vossen JM, van Tol MJ: Expression of the human immunoglobulin heavy chain VH6 gene element by fetal B lymphocytes. Scand J Immunol 46:292, 1997.
16. Fumoux F, Guigou V, Blaise D, Maraninchi D, Fougereau M, Schiff C: Reconstitution of human immunoglobulin

Material may be protected by copyright law (Title 17, U.S. Code)

VH repertoire after bone marrow transplantation mimics B-cell ontogeny. Blood 81:3153, 1993.

17. Brezinschek HP, Brezinschek RI, Lipsky PE: Analysis of the heavy chain repertoire of human peripheral B cells using single-cell polymerase chain reaction. J Immunol 155:190, 1995.
18. Zinkernagel RM, Doherty PC: 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 41Pt 2505, 1977.
19. Viale AC, Coutinho A, Freitas AA: Differential expression of VH gene families in peripheral B cell repertoires of newborn or adult immunoglobulin H chain congenic mice. J Exp Med 175:1449, 1992.
20. Rothenfluh HS, Blanden RV, Steele EJ: Evolution of V genes: DNA sequence structure of functional germline genes and pseudogenes. Immunogenetics 42:159, 1995.
21. Nobrega A, Grandien A, Haury M, Hecker L, Malanchere E, Coutinho A: Functional diversity and clonal frequencies of reactivity in the available antibody repertoire. Eur J Immunol 28:1204, 1998.
22. Lacroix-Desmazes S, Kaveri SV, Mouthon L, Ayouba A, Malanchere E, Coutinho A, Kazatchkine MD: Self-reactive antibodies (natural autoantibodies) in healthy individuals. J Immunol Methods 216:117, 1998.
23. Zinkernagel RM, Hengartner H: Regulation of the immune response by antigen. Science 293:251, 2001.
24. Zinkernagel RM: Uncertainties—discrepancies in immunology. Immunol Rev 185:103, 2002.
25. Kirkham PM, Schroeder HW Jr: Antibody structure and the evolution of immunoglobulin V gene segments. Semin Immunol 6:347, 1994.
26. Tomlinson IM, Walter G, Jones PT, Dear PH, Sonnhammer EL, Winter G: The imprint of somatic hypermutation on the repertoire of human germline V genes. J Mol Biol 256:813, 1996.
27. Almagro JC, Hernandez I, del Carmen Ramirez M, Vargas-Madrazo E: 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 34:1199, 1997.
28. Vargas-Madrazo E, Lara-Ochoa F, Ramirez-Benites MC, Almagro JC: Evolution of the structural repertoire of the human V(H) and Vkappa germline genes. Int Immunol 9:1801, 1997.
29. Ota T, Sitnikova T, Nei M: Evolution of vertebrate immunoglobulin variable gene segments. Curr Top Microbiol Immunol 248:221, 2000.
30. Cook GP, Tomlinson IM: The human immunoglobulin VH repertoire. Immunol Today 16:237, 1995.
31. Milner EC, Hufnagle WO, Glas AM, Suzuki I, Alexander C: Polymorphism and utilization of human VH genes. Ann N Y Acad Sci 764:50, 1995.
32. Li H, Cui X, Pramanik S, Chimge NO: Genetic diversity of the human immunoglobulin heavy chain VH region. Immunol Rev 190:53, 2002.
33. Romo-González T, Vargas-Madrazo E: Structural analysis of substitution patterns in alleles of human immunoglobulin VH genes. Mol Immunol, 42:1085-1097, 2005.
34. Romo-González T, Vargas-Madrazo E: Substitution patterns in alleles of immunoglobulin V genes in humans and mice. J Mol Evol, in press.
35. Sasso EH, Van Dijk KW, Milner EC: Prevalence and polymorphism of human VH3 genes. J Immunol 145: 2751, 1990.
36. van Dijk KW, Sasso EH, Milner EC: Polymorphism of the human Ig VH4 gene family. J Immunol 146:3646, 1991.
37. Sasso EH, Willems van Dijk K, Bull AP, Milner EC: A fetally expressed immunoglobulin VH1 gene belongs to a complex set of alleles. J Clin Invest 91:2358, 1993.
38. Cui X, Li H: Determination of gene organization in individual haplotypes by analyzing single DNA fragments from single spermatozoa. Proc Natl Acad Sci U S A 95:10791, 1998.
39. Vargas-Madrazo E, Paz-Garcia E: Interpreting gene usage of immunoglobulins: structural classification of the genes of immunoglobulins according to their recognition properties. Mackiwicz A, Kurpisz M, Poznan JZ, eds. In International Proceedings Division, Poland, September 23-27, 2000.
40. Pallares N, Lefebvre S, Contet V, Matsuda F, Lefranc MP: The human immunoglobulin heavy variable genes. Exp Clin Immunogenet 16:36, 1999.
41. Pascual V, Capra JD: Human immunoglobulin heavy-chain variable region genes: organization, polymorphism, and expression. Adv Immunol 49:74, 1991.
42. Tomlinson IM, Walter G, Marks JD, Llewelyn MB, Winter G: The repertoire of human germline VH sequences reveals about fifty groups of VH segments with different hypervariable loops. J Mol Biol 227:776, 1992.
43. Blin N, Stafford DW: A general method for isolation of high molecular weight DNA from eukarocytes. Nucleic Acids Res 3:2303, 1976.
44. Dower WJ, Miller JF, Ragsdale CW: High efficiency transformation of E. coli by high voltage electroporation. Nucleic Acids Res 16:6127, 1988.
45. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463, 1977.
46. Lefranc MP, Lefranc G: The Immunoglobulin FactsBook, San Diego, CA, Academic Press, 2001.
47. MacCallum RM, Martin AC, Thornton JM: Antibody-antigen interactions: contact analysis and binding site topography. J Mol Biol 262:732, 1996.
48. Grantham R: Amino acid difference formula to help explain protein evolution. Science 185:862, 1974.
49. Go M, Miyazawa S: Relationship between mutability, polarity and exteriority of amino acid residues in protein evolution. Int J Pept Protein Res 15:211, 1980.

Material may be protected by copyright law (Title 17, U.S. Code)

50. Sanz I, Dang H, Takei M, Talal N, Capra JD: VH sequence of a human anti-Sm autoantibody: evidence that autoantibodies can be unmutated copies of germline genes. J Immunol 142:883, 1989.
51. Willems van Dijk K, Schroeder HW Jr, Perlmutter RM, Milner EC: Heterogeneity in the human Ig VH locus. J Immunol 142:2547, 1989.
52. Huang C, Stewart AK, Schwartz RS, Stollar BD: Immunoglobulin heavy chain gene expression in peripheral blood B lymphocytes. J Clin Invest 89:1331, 1992.
53. Campbell MJ, Zelenetz AD, Levy S, Levy R: Use of family specific leader region primers for PCR amplification of the human heavy chain variable region gene repertoire. Mol Immunol 29:193, 1992.
54. Andris JS, Brodeur BR, Capra JD: 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: 1601, 1993.
55. Insel RA, Varade WS: Bias in somatic hypermutation of human VH genes. Int Immunol 6:1437, 1994.
56. Cuisinier AM, Guigou V, Boubli L, Fougereau M, Tonnelle C: Preferential expression of VH5 and VH6 immunoglobulin genes in early human B-cell ontogeny. Scand J Immunol 30:493, 1989.
57. Logtenberg T, Young FM, van Es J, Gmelig-Meyling FH, Berman JE, Alt FW: Frequency of VH-gene utilization in human EBV-transformed B-cell lines: the most JH-proximal VH segment encodes autoantibodies. J Autoimmun 2(Suppl):203, 1989.
58. Hillson JL, Oppliger IR, Sasso EH, Milner EC, Wener MH: Emerging human B cell repertoire. Influence of developmental stage and interindividual variation. J Immunol 149:3741, 1992.
59. van Dijk-Hard I, Feld S, Holmberg D, Lundkvist I: Increased utilization of the VH6 gene family in patients with autoimmune idiopathic thrombocytopenic purpura. J Autoimmun 12:57, 1999.
60. Li W-H: Molecular Evolution. Sunderland, Massachusetts, Sinauer Associates, 1997.
61. Eigen M, Winkler-Oswatitsch R, Dress A: Statistical geometry in sequence space: a method of quantitative comparative sequence analysis. Proc Natl Acad Sci U S A 85:5913, 1988.
62. Milstein C, Neuberger MS, Staden R: Both DNA strands of antibody genes are hypermutation targets. Proc Natl Acad Sci U S A 95:8791, 1998.
63. Jukes TH, King JL: Evolutionary nucleotide replacements in DNA. Nature 281:605, 1979.
64. Shlomchik MJ, Marshak-Rothstein A, Wolfowicz CB, Rothstein TL, Weigert MG: The role of clonal selection and somatic mutation in autoimmunity. Nature 328:805, 1987.
65. Jones PT, Dear PH, Foote J, Neuberger MS, Winter G: Replacing the complementarity-determining regions in a human antibody with those from a mouse. Nature 321:522, 1986.
66. Riechmann L, Foote J, Winter G: Expression of an antibody Fv fragment in myeloma cells. J Mol Biol 203:825, 1988.
67. Gorman SD, Clark MR, Routledge EG, Cobbold SP, Waldmann H: Reshaping a therapeutic CD4 antibody. Proc Natl Acad Sci U S A 88:4181, 1991.
68. Kirkham PM, Mortari F, Newton JA, Schroeder HW Jr: Immunoglobulin VH clan and family identity predicts variable domain structure and may influence antigen binding. EMBO J 11:603, 1992.
69. Zouali M: B-cell superantigens: implications for selection of the human antibody repertoire. Immunol Today 16: 399, 1995.
70. Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR, Colman PM, Spinelli S, Alzari PM, Poljak RJ: Conformations of immunoglobulin hypervariable regions. Nature 342:877, 1989.
71. Foote J, Winter G: Antibody framework residues affecting the conformation of the hypervariable loops. J Mol Biol 224:487, 1992.
72. Saul FA, Poljak RJ: Structural patterns at residue positions 9, 18, 67 and 82 in the VH framework regions of human and murine immunoglobulins. J Mol Biol 230:15, 1993.
73. Vargas-Madrazo E, Paz-Garcia E: 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 16:113, 2003.
74. Liu L, Lucas AH: IGH V3-23*01 and its allele V3-23*03 differ in their capacity to form the canonical human antibody combining site specific for the capsular polysaccharide of *Haemophilus influenzae* type b. Immunogenetics 55336, 2003.
75. Adderson EE, Shikhman AR, Ward KE, Cunningham MW: Molecular analysis of polyreactive monoclonal antibodies from rheumatic carditis: human anti–N-acetylglucosamine/anti-myosin antibody V region genes. J Immunol 161:2020, 1998.
76. Camacho FI, Algara P, Rodriguez A, Ruiz-Ballesteros E, Mollejo M, Martinez N, Martinez-Climent JA, Gonzalez M, Mateo M, Caleo A, Sanchez-Beato M, Menarguez J, Garcia-Conde J, Sole F, Campo E, Piris MA: Molecular heterogeneity in MCL defined by the use of specific VH genes and the frequency of somatic mutations. Blood 101:4042, 2003.
77. Kienle D, Krober A, Katzenberger T, Ott G, Leupolt E, Barth TF, Moller P, Benner A, Habermann A, Muller-Hermelink HK, Bentz M, Lichter P, Dohner H, Stilgenbauer S: VH mutation status and VDJ rearrangement structure in mantle cell lymphoma: correlation with genomic aberrations, clinical characteristics, and outcome. Blood 102:3003, 2003.

Material may be protected by copyright law (Title 17, U.S. Code)