

Geographic patterns and pathogenetic implications of IGHV gene usage in chronic lymphocytic leukemia: the lesson of the IGHV3-21 gene

Paolo Ghia, Kostas Stamatopoulos, Chrysoula Belessi, Carol Moreno, Stefania Stella, Giuseppe Guida, Ariane Michel, Marta Crespo, Nikolaos Laoutaris, Emili Montserrat, Achilles Anagnostopoulos, Guillaume Dighiero, Athanasios Fassas, Federico Caligaris-Cappio, and Frédéric Davi

We studied immunoglobulin variable heavy-chain (IGHV) repertoire and mutational status in 553 patients with chronic lymphocytic leukemia (CLL) from the Mediterranean area to gain insight into the potential pathogenetic role of antigenic stimulation. The most commonly represented IGHV genes mirrored the usage of normal B cells, with the exception of IGHV1-18, IGHV3-30.3, and IGHV4-59 that were underrepresented. The IGHV3-21 gene, frequently expressed in Northern European CLL, was present only in 16 cases (2.9%). Based on HCDR3 cluster analysis, cases using IGHV3-21 could be

grouped in 2 subsets of similar frequency. The first one (7 of 16 cases) carried a similar HCDR3 amino acid sequence (common-HCDR3 subset), virtually identical to the Scandinavian IGHV3-21 CLL. These cases used the *IGHJ6* gene; 4 of 7 were unmutated; 6 of 7 carried the $V_{\lambda}2-14$ (*IGLV3-21*) light-chain gene with a similar LCDR3. All expressed CD38 and had a progressive disease. The second subset (9 of 16) was characterized by heterogeneous HCDR3 rearrangements (nonhomogeneous-HCDR3 subset), diverse IGHJ and IGV light-chain gene usage, variable IGHV mutational sta-

tus (5 of 9 unmutated), variable CD38 expression, and variable clinical course (4 of 9 progressed). The first subset suggests a potential antigenic element rarely encountered in the Mediterranean area, possibly responsible for a negative outcome. The second subset may reflect the physiologic heterogeneity of expression of IGHV3-21 rearrangements in the normal repertoire and is characterized by a variable clinical outcome. (*Blood*. 2005;105:1678-1685)

© 2005 by The American Society of Hematology

Introduction

The study of immunoglobulin variable heavy-chain (IGHV) genes in B-lymphoproliferative disorders has helped in tracing the developmental stage of neoplastic transformation by assigning a normal B-cell counterpart to individual B-cell malignancies and has provided a new prognostic tool. This progress is well illustrated by chronic lymphocytic leukemia (CLL)¹ where the analysis of the IGHV repertoire and mutation pattern has helped to define at least 2 disease subsets, one with unmutated IGHV genes and one with somatically mutated genes, that is, less than 98% homology with the closest germline gene.^{2,3} The mutational status translates into a different clinical outcome, with patients with unmutated IGHV CLL having a considerably poorer survival and an overall worse prognosis.

IGHV repertoire analysis in CLL has demonstrated biases both at the IGH subgroup level, with IGHV1, IGHV3, and IGHV4 being the most frequently used subgroup, and at the specific gene level, with IGHV1-69, IGHV3-07, IGHV3-23, and IGHV4-34 being the most frequently used IGHV genes.³⁻⁶ Importantly, somatic mutations do not appear to occur uniformly among IGHV gene subgroups; rather they display a hierarchy of mutations (IGHV3>IGHV4>IGHV1).⁴ Differences appear even more striking when considering individual IGHV genes, with IGHV1-69

carrying very few mutations as opposed to IGHV3-07, IGHV3-23, and IGHV4-34, which show a high load of mutations.

Even if IGH subgroup frequencies are similar in different cohorts, relative frequencies of individual genes vary among series. Possible reasons for these differences include size of cohorts, heterogeneity of the patient's ancestry, and the type of medical facility (referral versus primary centers) where the patients are evaluated, which may bring in unwanted selection biases.

Recently an overrepresentation of the IGHV3-21 gene, accounting for more than 11% of cases, has been described in a Scandinavian cohort of patients with CLL.^{7,8} Most cases (68%) were mutated, albeit with a low mutation "load" (between 2% and 5%). Interestingly, these mutated IGHV3-21 cases had a worse overall survival than the remaining patients with mutated CLL. Similar results were shown in a study of 83 British patients where IGHV3-21 was used in 9.6% of cases (most having 2% to 5% mutation) and was associated with poor survival⁹ although this series might carry a particular bias due to its focus on p53 dysfunction. A similar frequency was also reported in a Belgian series of 91 patients with 8 cases (8.8%) expressing the IGHV3-21 gene, mainly mutated.¹⁰

From the Department of Oncological Sciences, University of Torino and Institute for Cancer Research and Treatment, Torino, Italy; Hematology Department and HCT Unit, G. Papanicolaou Hospital, Thessaloniki, Greece; Hematology Department, Nikea General Hospital, Athens, Greece; Institute of Hematology and Oncology and Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Hospital Clinic, Barcelona, Spain; Unité d'Immunohématologie et d'Hématopathologie, Institut Pasteur, Paris, France; Department of Oncology, Università Vita Salute, San Raffaele, Milano, Italy; Laboratory of Hematology and University Paris 6, Hôpital Pitié-Salpêtrière, Paris, France.

Submitted July 9, 2004; accepted September 30, 2004. Prepublished online as *Blood* First Edition Paper, October 5, 2004; DOI 10.1182/blood-2004-07-2606.

Supported in part by Associazione Italiana per la Ricerca sul Cancro (AIRC), Progetto Oncologia CNR-MIUR, MURST 40%, Association Claude Bernard, the José Carreras International Leukemia Foundation (EM-03; CR-03), and Generalitat de Catalunya 2001/SGR/00381.

Reprints: Paolo Ghia, Laboratory of Cancer Immunology, IRCC, Strada Provinciale 142, 10060 Candiolo (TO), Italy; e-mail: paolo.ghia@unito.it.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.

© 2005 by The American Society of Hematology

It remains to be elucidated whether the IGHV3-21 gene by itself escapes the general rule of mutated versus unmutated IGHV in predicting clinical prognosis in CLL¹¹ and thus represents a so far unique example of a mutated gene that is associated with reduced survival.

In this present study, we retrospectively analyzed and compared the IGHV repertoire and mutation status in a series of 553 patients with CLL followed in different institutions from the Mediterranean area (France, Greece, Italy, and Spain). Specifically, we defined the commonly used IGHV repertoire in CLL patients of Mediterranean descent and focused our attention on the frequency, molecular features, and clinical significance of IGHV3-21 gene usage in this large series of patients. Our analysis provides evidence that IGHV3-21–expressing CLL cases in Southern Europe can be divided into 2 groups: one with similar molecular and clinical features to the Northern European cases, though much less frequent, and another one, more heterogeneous at both the biologic and clinical levels.

Patients, materials, and methods

Patient group

A total of 553 unselected patients with CLL from different institutions, located in Paris (France, 150 cases), Thessaloniki and Athens (Greece, 191 cases), Torino (Italy, 97 cases), and Barcelona (Spain, 115 cases) were studied for IGHV gene usage and mutational status (Table 1). All patients met the diagnostic criteria of the National Cancer Institute-Working Group (NCI-WG).¹² The mean age of the 337 men and 216 women was 64.1 years (range, 24-91 years; Table 1). Written informed consent was obtained at study entry. The study was approved by the local Ethics Review Committee of each institution.

Within this cohort, 16 patients carrying the IGHV3-21 gene in the expressed immunoglobulin (8 men, 8 women; mean age, 70.7 years) were studied in detail by analyzing the following parameters measured at diagnosis or during the follow-up: CD38 expression, immunoglobulin light-chain (LC) rearrangement, disease stage according to Binet¹³ or Rai modified criteria,¹⁴ history of treatment, and progressive or stable disease as defined by the NCI-WG.¹²

Determination of IGHV, IGKV, and IGLV gene mutational status

In all cases, the analysis of IGHV-D-J genes was done on leukemic cells obtained from peripheral blood samples after isolation by Ficoll gradient. Total RNA was extracted using the guanidinium thiocyanate method and reverse-transcribed into cDNA using Moloney murine leukemia virus (MMLV) reverse transcriptase and random hexamers as primers. gDNA was isolated by standard phenol chloroform extraction followed by ethanol precipitation.

The IGHV-D-J gene sequences were determined as previously described,¹⁵⁻¹⁸ by amplifying with polymerase chain reaction (PCR) either cDNA, using the appropriate sense IGHV leader primer and the antisense IGHC primer,⁴ or DNA with consensus primers for the framework 1 and IGHJ genes, as previously described¹⁹ or following the BIOMED-2 protocol.²⁰ Amplification of the immunoglobulin κ and λ LC gene rearrangements (IGKV-J and IGLV-J) was performed according to BIOMED-2 protocols²⁰ and as previously described.²¹

Purified PCR products were sequenced either directly or after a cloning procedure using an automated DNA sequencer. Sequence data were

analyzed using the international Immunogenetics database (IMGT, <http://imgt.cines.fr>; initiator and coordinator: Marie-Paule Lefranc, Montpellier, France) and more particularly, the IMGT/V-QUEST²² and IMGT/Junction Analysis²³ to determine the respective homology of each gene with the closest germline counterpart. The sequences with a germline homology 98% or higher were considered unmutated, and those with a homology less than 98% as mutated.^{2,3} The mutational status of 242 of 553 sequences was published previously.^{17,18,24}

To facilitate direct reference and comparison with previous studies (mainly those by Tobin et al^{3,4}), in the only case of the IGLV3-21 gene we have followed the Kawasaki nomenclature, after which this gene is named V λ 2-14. Assessment of the distribution of mutations and possible impact of antigen selection on the hypermutation process (for cases with > 2% deviation from the closest germline V gene) was performed by the multinomial model proposed by Lossos et al.²⁵ Each codon was assessed separately for the presence of somatic mutations. When more than one nucleotide substitution was encountered within a 3-bp codon sequence, it was postulated that mutations were introduced sequentially from the first through the third nucleotide in order to be assigned as being of the replacement (R mutation; leading to amino acid substitution in the encoded polypeptide) or silent (S mutation; leaving the encoded polypeptide unaltered) type.

Immunophenotypic analysis

Immunophenotypic analysis was performed on fresh blood samples or cells cryopreserved at diagnosis to define the isotype and the LC clonal restriction and analyze the CD38 expression of the leukemic population.

The antibodies used and, in particular, the procedure to determine the percentage of leukemic B cells that expressed CD38 on their surface have been previously described.^{17,18} Cases with less than 7% CD38⁺ cells were considered negative.^{17,18}

Statistical analysis

Differences between median cumulative survivals and median times to undergo progressive disease were analyzed by the Kaplan-Meier method. The Fisher exact test was performed to determine the significance of differences among frequencies of events between patient subsets.

Results

IGHV gene usage in CLL is similar in 4 Mediterranean countries

Productive IGHV-D-J rearrangements were sequenced in 553 patients with CLL followed in different clinical centers from 4 different Mediterranean countries (Table 1). The IGHV genes used by these leukemic cell populations were derived from all the 7 human IGHV gene families. As shown in Figure 1A-B, among the 47 functional IGHV genes used, the most frequently encountered were IGHV1-69 (59; 10.7%), IGHV4-34 (59; 10.7%), IGHV3-23 (51; 9.2%), IGHV3-30 (37; 6.7%), IGHV3-07 (30; 5.4%), IGHV4-39 (29; 5.2%), and IGHV1-02 (22; 4.2%); collectively, these genes accounted for more than 50% of cases. As a whole, this large and geographically diverse series confirmed previously published results obtained in smaller series.³⁻⁶ Considering the percentage of somatic mutations present in the sequences and using the 98% cut-off value,^{2,3} 320 (57.9%) cases were considered mutated and the remaining 233 (42.1%) had a germline configuration. The most represented IGHV gene within the mutated subset was IGHV4-34, which was used in 59 cases overall (50 mutated), whereas IGHV1-69 was the most frequently used in the unmutated group, being expressed in 59 patients overall (52 unmutated) as shown in Table 2.

Table 1. Characteristics of all CLL patients analyzed

	France	Greece	Italy	Spain	Total
No. of cases	150	191	97	115	553
Mean age, y	61.1	67.2	69.8	58.0	64.1
Sex, no. M/no. F	89/61	116/75	60/37	72/43	337/216
No. mutated/no. unmutated	91/59	113/78	59/38	57/58	320/233

Table 3. Characteristics of all CLL patients using IGHV3-21 gene

	Total	France	Greece	Italy	Spain
No. IGHV3-21 cases	16	4	7	4	1
Frequency, %	2.9	2.7	3.7	4.1	0.9
Mean age, y	70.7	63.5	67	79.5	90*
Sex, no. F/no. M	8/8	2/2	4/3	1/3	1/0
No. mutated/no. unmutated	7/9	2/2	2/5	2/2	1/0
IGHV homology, no.					
At least 98%	9	2	5	2	0
98%-96%	3	0	1	1	1
Less than 96%	4	2	1	1	0
LC, no. κ/no. λ	7/9	2/2	4/3	1/3	0/1
IGV LC homology, no.					
At least 98%	14	3	7	3	1
98%-96%	2	1	0	1	0
Less than 96%	0	0	0	0	0
CD38, %					
Neg	5	0	4	1	0
Pos	10	3	3	3	1
Stage at diagnosis					
A	8	1	5	2	0
B	5	3	1	1	0
C	3	0	1	1	1
Clinical course					
Stable	5	0	4	1	0
Progressive	11	4	3	3	1
Need for treatment	11	4	3	3	1
CLL-related death	4	2	0	1	1

*Actual age of the single patient of the Spanish series.

sequence is virtually identical to the one published earlier by Tobin et al^{3,4} (ARDANGMDV). Interestingly, despite carrying a short HCDR3 sequence, all cases used the *IGHJ6* gene that is the longest IGHJ gene in the human genome. The first 2 amino acids in the conserved HCDR3 sequence (AR) derive from the IGHV3-21 gene, whereas the last 4 (AMDV) from the *IGHJ6* gene. The sixth residue was either a glycine (G) in 4 of 7 of our cases as in most of Scandinavian sequences, and is germline encoded by the *IGHJ6* gene, or an alanine (A; 3 of 7 cases), both being amino acids with an aliphatic side chain.

Four of these sequences were unmutated ($\geq 98\%$ homology), whereas the remaining 3 were mutated: 2 cases were, respectively, 96.3% and 96.9% homologous with the closest germline gene, whereas one was 94.9% homologous (Table 4). The impact of antigen selection on the hypermutation process for these 3 mutated cases was assessed by determining replacement (R) as opposed to silent (S) mutations in the CDRs and S as opposed to R mutations

in framework regions (FRs). The probability that excess (for CDR) or scarcity (for FR) of R mutations occurred by chance was calculated by the multinomial distribution model described by Lossos et al²⁵ and following IMGT definitions for FR and CDR lengths. Statistically significant evidence for positive selection by antigen was obtained in all 3 mutated cases; there were 2 cases with higher R/S mutation ratios in CDRs and one case with higher and lower R/S mutation ratios in CDRs and FRs, respectively, than those theoretically expected given the inherent germline $R_{f_{CDRs}}$ and $R_{f_{FRs}}$ values.

The second subset of IGHV3-21 cases (9 of 16; Table 6) carried HCDR3 rearrangements that were totally unrelated to each other and to the consensus sequence (nonhomogeneous-HCDR3 subset). Their mean amino acid length was almost twice longer (16.9 codons) than the consensus (Table 5). The use of IGHJ genes was also heterogeneous with 6 cases using *IGHJ4*, 1 using *IGHJ5*, and 2 using *IGHJ6*. Five cases were unmutated and the remaining 4 were mutated, with 3 of 4 exhibiting a high mutation "load" ($< 95\%$).

LC usage is different in the 2 IGHV3-21 subsets

As defined by cytofluorographic analysis, 9 of the 16 IGHV3-21 cases expressed monoclonal λ LCs on the cell surface and the remaining 7 cases expressed κ LCs. Considering only mutated IGHV3-21 cases (7 of 16), regardless of their HCDR3 composition, 3 expressed a clonal κ LC rearrangement and 4 cases expressed a λ LC rearrangement.

To facilitate direct reference and comparison with previous studies (mainly Tobin et al^{3,4}), we have followed the Kawasaki nomenclature, after which this gene is named as $V_{\lambda}2-14$, only in the case of the IGLV3-21 gene. The sequence of LC rearrangements of all IGHV3-21 showed that among the 9 λ -expressing cases, 8 cases expressed $V_{\lambda}2-14$ (IGLV3-21), and the remaining one expressed IGLV1-51 ($V_{\lambda}1-19$). All λ -expressing cases used the *IGLJ3* gene except one using *IGLJ1*. In contrast, the 7 κ -expressing cases used heterogeneous IGKV genes: IGKV3-15 (L2), IGKV3D-15 (L16), IGKV4-1 (B3), IGKV1-27 (A20), IGKV1-17 (A30), IGKV1-33/1D-33 (O8/O18), and IGKV1-12 (L5). In these cases, IGKJ usage was also quite heterogeneous: *IGKJ1* was used in 2 cases, *IGKJ2* in 3 cases, and *IGKJ3* in 2 cases.

Taken together, 14 immunoglobulin LC sequences showed more than 98% homology to their respective germline gene. In particular, all $V_{\lambda}2-14$ (IGLV3-21)-using cases were unmutated.

Table 4. Characteristics of the IGHV3-21-expressing CLL patients belonging to the common-HCDR3 subset

	Fra1	Fra3	Gre1	Ita1	Ita3	Ita4	Spa1
No. of cases	1	1	1	1	1	1	1
Age, y	69	46	56	72	69	91	90
Sex	F	M	M	M	M	F	F
IGHJ gene	<i>IGHJ6</i>	<i>IGHJ6</i>	<i>IGHJ6</i>	<i>IGHJ6</i>	<i>IGHJ6</i>	<i>IGHJ6</i>	<i>IGHJ6</i>
IGHV homology, %	94.9	98.3	98.0	96.3	98.6	98.3	96.9
LC	$V_{\lambda}2-14$	IGKV3-15	$V_{\lambda}2-14$	$V_{\lambda}2-14$	$V_{\lambda}2-14$	$V_{\lambda}2-14$	$V_{\lambda}2-14$
IGJ LC gene	<i>IGLJ3</i>	<i>IGKJ1</i>	<i>IGLJ3</i>	<i>IGLJ3</i>	<i>IGLJ3</i>	<i>IGLJ3</i>	<i>IGLJ3</i>
IGV LC homology, %	99.7	100	100	99	98.6	98.3	99.7
CD38, %	7	47	7.7	99	53	99	46
Stage at diagnosis	B	B	B	B	C	A	C
Progression	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Need for treatment	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CLL-related death	No	Yes	No	Yes	No	No	Yes

Fra indicates France; Gre, Greece; Ita, Italy; Spa, Spain.

Table 5. Heavy- and light-chain CDR3 amino acid sequences of the IGHV3-21-using CLL cases

Case	HCDR3 AA sequence	IG light chain gene	LCDR3 AA sequence
Cases with consensus sequence*			
Ita3	ARDANAMDV	V _λ 2-14 (IGLV3-21)	QVWDSGSDHPWV
Ita4	. . . G . G . . .	V _λ 2-14 (IGLV3-21)
Ita1	. . . M	V _λ 2-14 (IGLV3-21)
Gre1	. I . R . G . . .	V _λ 2-14 (IGLV3-21) S
Spa1	. T . R . G . . .	V _λ 2-14 (IGLV3-21)
Fra1 DG . . .	V _λ 2-14 (IGLV3-21) S
Fra3	IGKV3-15	QQYNNWPPET
Cases with completely unrelated sequence†			
Fra4	ARDRTYPDAYFDF	IGKV3D-15	QQYYNVWPPGR
Gre3	AKDYDSSGANFQH	IGLV1-51	GTWDTSLSAWL
Ita2	ARVGNVDRYFFDF	IGKV1-12	QQANSFPYS
Gre5	ARIWDIVVVVDPAMIWA	IGKV1-27	QKYNQAQFT
Gre4	ARDRRSIVVVVAATPVY	IGKV4-1	QQSYSTLVEA
Fra2	ATSGKDTTMVQGGHFDY	V _λ 2-14 (IGLV3-21)	QVWDSGSDHPFV
Gre7	ASQLWFDRVNNYYMDV	IGKV1-33/1D-33	QQYDNLPRYT
Gre2	AREPLSVLLWFGESFGGGFDY	V _λ 2-14 (IGLV3-21)	QVWDSGSDHPWV
Gre6	ARDRLGYCSSTSCWDYYYYYGMVDV	IGKV1-17	LQHNSYPWYT

Seven cases showed a consensus HCDR3 sequence or a high homology to it, differing by 1, 2, or 3 amino acids at different positions (common-HCDR3 subset), whereas 6 of 7 had a consensus LCDR3 sequence, differing by 1 amino acid in 2 cases. The remaining 9 patients expressed a completely unrelated sequence (nonhomogeneous-HCDR3 subset). Dots indicate homology.

*Common-HCDR3 subset.

†Nonhomogeneous HCDR3 subset.

Considering only the cases belonging to the common-HCDR3 subset, 6 of 7 used a V_λ2-14 (IGLV3-21) rearrangement (Table 4), and the remaining case expressed the IGKV3-15 gene. In these 6 V_λ2-14 (IGLV3-21)-expressing cases cluster analysis of the CDR3 of the LC (LCDR3) rearrangements showed a conserved sequence (QVWDSGSDHP) that was present in 4 cases, whereas 2 cases differed by only one amino acid (QVWDSGSDHP; Table 5). This sequence is virtually identical to the consensus published earlier by Tobin et al^{3,4} (QVWDGSSDHP), differing only by 2 amino acids at the fifth and sixth positions.

Correlation between clinical outcome and IGHV3-21 gene expression

When considering all IGHV3-21 patients together (Table 3), 4 of 16 died of CLL-related causes, leading to a median survival time of 98 months, which overlaps the median survival of unmutated CLL patients.³ Given the low number of IGHV3-21 cases in our series, the comparative analysis of median survival curves of mutated versus unmutated cases did not reach statistical significance.

Five patients had stable disease, not requiring any treatment; the remaining 11 had clinical progression during follow-up and needed

therapy. Among patients who progressed, 6 had a germline configuration in both heavy-chain and LC IGV genes; 3 showed a low level of mutations on their IGHV genes (homologies ranging between 97.6% and 96.3%) but expressed unmutated IGV LC genes. Finally, the remaining 2 patients had highly mutated IGHV genes (94.9% and 93.2% homology), which were associated in one with unmutated IGV LC and in the other with low levels of somatic mutations (97.0%) of the LC variable gene. Interestingly, 9 of 10 tested patients with progressive disease were CD38⁺.

If patients are grouped according to HCDR3 amino acids sequence, all those belonging to the common-HCDR3 subgroup (Table 4) had a progressive disease (100%) as compared to 4 of the 9 cases of the nonhomogeneous subset (44.4%; $P = .03$; Table 6). In addition, the patients with a common-HCDR3 IGHV3-21 seemed to experience also a shorter median time to progression (3.5 months versus 60.4 months) though statistical significance was not completely reached ($P = .06$), due to the low number of progressions in the nonhomogeneous subset.

All 7 cases with a common-HCDR3 expressed CD38 on the cell surface, though only 4 cases had unmutated IGHV genes.

Table 6. Characteristics of the IGHV3-21-expressing CLL patients belonging to the nonhomogeneous-HCDR3 subset

	Fra2	Fra4	Gre2	Gre3	Gre4	Gre5	Gre6	Gre7	Ita2
No. of cases	1	1	1	1	1	1	1	1	1
Age, y	54	85	65	67	70	60	81	70	86
Sex	M	F	F	F	F	M	F	M	M
IGHJ gene	IGHJ4	IGHJ4	IGHJ4	IGHJ5	IGHJ4	IGHJ4	IGHJ6	IGHJ6	IGHJ4
IGHV homology, %	98.3	93.2	99.0	94.6	97.6	98.0	99.0	98.6	92.9
LC	V _λ 2-14	IGKV3D-15	V _λ 2-14	IGLV1-51	IGKV4-1	IGKV1-27	IGKV1-1	IGKV1-33	IGKV1-12
IgJ LC gene	IGLJ1	IGKJ1	IGLJ3	IGLJ3	IGKJ3	IGKJ3	IGKJ2	IGKJ2	IGKJ2
IGV LC homology, %	100	97.0	100	98.3	100	99.0	100	100	96.5
CD38, %	35	ND	37	2.5	1.4	12	27	3.2	0.5
Stage at diagnosis	B	A	A	A	A	A	C	A	A
Progression	Yes	Yes	No	No	Yes	No	Yes	No	No
Need for treatment	Yes	Yes	No	No	Yes	No	Yes	No	No
CLL-related death	No	Yes	No	No	No	No	No	No	No

Discussion

In the present multicenter study, we analyzed the immunoglobulin gene repertoire of 553 patients who had CLL and were of Mediterranean origin (Table 5). Our analysis shows that several of the most common IGHV genes in CLL are used at a frequency that is not disease-related, but mirrors the repertoire of normal peripheral blood IgM⁺ B cells (both CD5⁺ and CD5⁻)²⁶ with the exception of the well-known overrepresentation of the IGHV1-69 gene,³⁻⁶ especially among unmutated cases. Similar to normal peripheral blood IgM⁺ B cells, the CLL cases studied were found to use frequently the IGHV3-23, IGHV3-07, and IGHV4-34 genes (as previously reported in smaller series of patients^{3,4}) as well as the IGHV4-39 and IGHV1-02 genes. The frequency of the IGHV4-39 gene resembled that of normal CD5⁻ B cells, whereas the frequency of the IGHV1-02 gene was comparable to that of normal IgM⁺ CD5⁺ peripheral blood B cells. In addition, in our series the frequency of expression of other genes previously indicated as suppressed in the CLL repertoire in 2 distinct but smaller series of patients^{3,4} was found similar to that of normal B lymphocytes (both CD5⁺ and CD5⁻).²⁶ In particular, IGHV3-30 gene frequency (6.7%) was similar to that of both CD5⁺ and CD5⁻ B cells, whereas the IGHV3-33 gene was present in 2.5% of the cases showing a usage similar to normal CD5⁺ B cells.

In contrast, the IGHV3-49 gene, which is rarely expressed in normal B cells, was rather frequently expressed in our cohort (2.4% of cases), especially among Italian patients where it was present in 5.1% of the cases. Conversely, the IGHV1-18, IGHV3-30.3, and IGHV4-59 genes that are common in normal B cells were rarely represented in our series. This lack of representation has not been previously reported in CLL.

The differences in IGHV gene usage between our study and previously reported series may be explained, at least in part, by the small size of the patient populations previously analyzed. However, considering that the expressed immunoglobulin repertoire of both normal and leukemic B cells is shaped by molecular processes as well as by antigenic selection and affinity maturation mechanisms. Such discrepancies may be real and reflect a geographic difference that could allow peculiar genetic or environmental elements to differently shape both the normal and the “leukemic” repertoire. This possibility is highlighted by the different usage of the IGHV3-21 gene among the CLL patients in our multicenter, Mediterranean-based study, as compared to the previously published study originating from Scandinavia.⁸ In the latter series, this gene was second in frequency (> 11%) only surpassed by IGHV1-69. Furthermore, it was the most frequently used gene in mutated rearrangements. Similar results were reported in a second larger study from the same group (265 cases),⁷ thereby excluding a bias attributed to small sampling. In addition, a high frequency was also reported in British (9.6%) and Belgian studies (8.8%), though on smaller series of patients.^{9,10} In contrast, the IGHV3-21 gene barely ranked among the top 10 used IGHV genes in our cohort, being the 10th overall (2.9% of the cases) and the 5th among the unmutated cases. This frequency is at least 3 times lower than the frequency observed in patients from Northern Europe. To further underscore this finding in the Spanish series there was only one IGHV3-21 case among the 115 cases analyzed. Similar to our series, a low frequency for IGHV3-21 usage was suggested in a previous study that combined CLL IGHV sequences from unrelated and geographically diverse cohorts.^{3,4} The report included CLL cases from

heterogeneous sources and referred to IGHV sequences present in databases, not allowing to draw any conclusion regarding the actual gene frequency in terms of population or geographic distribution. Therefore, more studies are needed to finally elucidate whether the reported differences in terms of frequencies may be more widespread than so far reported.

The apparent geographic bias shown by our own data may theoretically be due to specific and peculiar genetic background, depending on variations in germline composition of the IGHV locus²⁸ or to the effect of a potential environmental variable less frequently encountered in different regions. Because normal IgM⁺ peripheral blood B cells and in particular CD5⁺ B cells have been described to express the IGHV3-21 gene at a low frequency,^{4,26} it would be interesting to analyze the normal usage of this gene among the Scandinavian population to evaluate whether a bias also exists in the normal repertoire or if it is restricted to CLL cells.

A detailed analysis of the Scandinavian series^{7,8} revealed that, besides a higher frequency of usage, several peculiarities characterize Scandinavian IGHV3-21-expressing CLL.²⁸ The authors show that an individual IGHV gene can have by itself a prognostic impact in CLL. In their cohort the patients using IGHV3-21 gene, although usually mutated, showed a poor overall prognosis and were therefore an exception to the rule of mutated CLL having good prognosis. In addition IGHV3-21 genes are mainly expressed in combination with a λ LC (28 of 31 cases), which is almost exclusively the V λ 2-14 (IGLV3-21) gene (24 of 27) and demonstrates remarkable similarities in their CDR3s. In our Mediterranean cohort, collectively IGHV3-21 cases (Table 3) express κ and λ LC chain in similar frequencies and usually have IGHV sequences in germline configuration, which explains the occurrence of a therapy-requiring progressive disease in several cases. Interestingly, the few patients with mutations who had progressive disease were mainly CD38⁺.^{2,17} These discrepancies may be reconciled by considering that, based on the HCDR3 sequences, we here describe the existence of 2 distinct IGHV3-21 subsets of similar frequency. Less than 50% of the cases do carry an almost identical amino acid consensus sequence (common-HCDR3 subset), together with a similar LCDR3 on the IGLV gene. Both are virtually identical to those commonly expressed in Scandinavian CLL.⁸ In this respect it is interesting to note a repetitive glycine (G) to alanine (A) variation at HCDR3 position 6 in both Scandinavian and Mediterranean sequences. Glycine is the germline configuration of the *IGHJ6* gene. However, the G>A substitution does not seem to bring significant charge or conformation differences, as both G and A amino acids have nonpolar side chains and similar size.

The common-HCDR3 subset shows a strong bias in the expression of λ LCs and in particular of the V λ 2-14 (IGLV3-21) gene. In addition, a worse prognosis due to progressive disease is also more frequent in this subset as compared to the remaining cases ($P = .03$). Nevertheless, most cases are actually unmutated and those that are not express some levels of CD38 on the cell surface. This makes it challenging to define the actual role of the IGHV3-21 rearrangement by itself in predicting the outcome as compared to the other biologic prognostic factors.

More than 50% of our IGHV3-21-expressing cases show HCDR3 rearrangements that are totally unrelated to each other and to the consensus sequence (nonhomogeneous-HCDR3 subset). They use diverse IGHJ and IGV LC genes and 5 of 9 cases are unmutated. They also show variable CD38 expression and have a

heterogeneous clinical course with only 4 of 9 patients experiencing progressive disease. This subset may mirror the heterogeneity and low frequency of expression of IGHV3-21 rearrangements in the normal repertoire of peripheral blood B cells, as shaped by the germline composition, molecular mechanisms, and distinct selective influences.

The first subset, with the evidence of a similar antigen-binding site in so many different CLL cohorts, indicates that a potential environmental variable may be the selecting element for the clonal founder of this leukemic progeny in all European regions. We obtained statistically significant evidence for positive selection by antigen for all 3 mutated cases with a short HCDR3. This subset may be differently selected and expanded in different regions, according to the prevalence of the selective element in the specific geographic environment. Given the usually bad prognosis of this subset of patients it becomes interesting to hypothesize that a potential antigenic stimulation may be eventually responsible for an aggressive clinical course.

Interestingly, no consensus HCDR3 or LCDR3 sequences were identified among 5 multiple myeloma cases, one acute lymphoblastic leukemia case, and one follicular lymphoma case using the IGHV3-21 gene (K.S. and C.B., unpublished observations, 2004) supporting the idea of a selective pressure acting preferentially in CLL. According to this hypothesis, it is notable that among 89 IGHV3-21 (+) sequences from the IMGT database, mainly obtained from libraries of normal B cells as well as from antibodies against known pathogens, there were only 3 IGHV3-21 sequences with a 9 codons-long HCDR3 similar to the common-HCDR3 subset sequence, all using the *IGHJ6* gene and all derived from CLL cases.

In contrast, the second subset includes cases heterogeneous at both the biologic and clinical levels, carrying diverse HCDR3 rearrangements and showing variable clinical outcomes. This subset may mirror the low frequency of expression of IGHV3-21

rearrangements in the normal repertoire of peripheral blood B cells as well as their biologic heterogeneity, shaped by the germline composition, molecular mechanisms, and distinct selective influences, thereby likely favoring a variable clinical course.

These findings bring in important implications. It was recently reported that unrelated CLL patients from different cohorts using other IGHV genes may share common CDR3s.²⁹⁻³¹ Consequently, it would not be unreasonable to speculate that important biologic and (perhaps) prognostic information might be gained by defining not only the usage of specific genes (eg, IGHV3-21) and their mutational load but also the configuration of HCDR3 and the associated light chain of each patient among a subset of cases using similar IGHV genes.

In conclusion, the striking difference in the overall frequency of IGHV3-21 usage between North and South European CLL may be explained by the overweight in terms of percentage of the common-HCDR3 subset as compared to the nonhomogeneous IGHV3-21-expressing subset in the Scandinavian series. The lower frequency of the common-HCDR3 CLL subset in the Mediterranean region suggests that the hypothetical antigenic element may be less frequently encountered around the Mediterranean Sea. Still, it is unequivocally present, as demonstrated by those patients from France, Greece, Italy, and Spain who share strikingly similar HCDR3 sequences with Scandinavian patients.

Acknowledgments

We wish to thank to Prof Marie-Paule Lefranc and Dr Veronique Giudicelli (Laboratoire d'Immunogenetique Moleculaire, LIGM, Universite Montpellier II, UPR CNRS) for sharing with us a wealth of insight on IG genes, offering valuable help in data analysis, and assisting us in navigating through the IMGT database and resources.

References

- Kipps TJ. Immunobiology of chronic lymphocytic leukemia. *Curr Opin Hematol*. 2003;10:312-318.
- Damle RN, Wasil T, Fais F, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood*. 1999;94:1840-1847.
- Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood*. 1999;94:1848-1854.
- Fais F, Ghiotto F, Hashimoto S, et al. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *J Clin Invest*. 1998;102:1515-1525.
- Kipps TJ, Tomhave E, Pratt LF, Duffy S, Chen PP, Carson DA. Developmentally restricted immunoglobulin heavy chain variable region gene expressed at high frequency in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A*. 1989;86:5913-5917.
- Schroeder HW Jr, Dighiero G. The pathogenesis of chronic lymphocytic leukemia: analysis of the antibody repertoire. *Immunol Today*. 1994;15:288-294.
- Tobin G, Thunberg U, Johnson A, et al. Chronic lymphocytic leukemias utilizing the VH3-21 gene display highly restricted Vlambda2-14 gene use and homologous CDR3s: implicating recognition of a common antigen epitope. *Blood*. 2003;101:4952-4957.
- Tobin G, Thunberg U, Johnson A, et al. Somatic mutated Ig V(H)3-21 genes characterize a new subset of chronic lymphocytic leukemia. *Blood*. 2002;99:2262-2264.
- Lin K, Manocha S, Harris RJ, Matrai Z, Sherrington PD, Pettitt AR. High frequency of p53 dysfunction and low level of VH mutation in chronic lymphocytic leukemia patients using the VH3-21 gene segment. *Blood*. 2003;102:1145-1146.
- Philippe J, Janssens A, Smits K, et al. Prognostic value of specific VH-genes in CLL [abstract]. *Leuk Lymph*. 2003;44:S40.
- Nollet F, Cauwelier B, Billiet J, et al. Do B-cell chronic lymphocytic leukemia patients with Ig VH3-21 genes constitute a new subset of chronic lymphocytic leukemia? *Blood*. 2002;100:1097-1098.
- Cheson BD, Bennett JM, Grever M, et al. National Cancer Institute-sponsored Working Group guidelines for chronic lymphocytic leukemia: revised guidelines for diagnosis and treatment. *Blood*. 1996;87:4990-4997.
- Binet JL, Leproprier M, Dighiero G, et al. A clinical staging system for chronic lymphocytic leukemia: prognostic significance. *Cancer*. 1977;40:855-864.
- Rai KR, Han T. Prognostic factors and clinical staging in chronic lymphocytic leukemia. *Hematol Oncol Clin North Am*. 1990;4:447-456.
- Davi F, Maloum K, Michel A, et al. High frequency of somatic mutations in the VH genes expressed in prolymphocytic leukemia. *Blood*. 1996;88:3953-3961.
- Stamatopoulos K, Kosmas C, Papadaki T, et al. Follicular lymphoma immunoglobulin kappa light chains are affected by the antigen selection process, but to a lesser degree than their partner heavy chains. *Br J Haematol*. 1997;96:132-146.
- Ghia P, Guida G, Stella S, et al. The pattern of CD38 expression defines a distinct subset of chronic lymphocytic leukemia (CLL) patients at risk of disease progression. *Blood*. 2003;101:1262-1269.
- Crespo M, Bosch F, Villamor N, et al. ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. *N Engl J Med*. 2003;348:1764-1775.
- Aubin J, Davi F, Nguyen-Salomon F, et al. Description of a novel FR1 IgH PCR strategy and its comparison with three other strategies for the detection of clonality in B cell malignancies. *Leukemia*. 1995;9:471-479.
- van Dongen JJ, Langerak AW, Bruggemann M, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia*. 2003;17:2257-2317.
- Songsivilai S, Bye JM, Marks JD, Hughes-Jones NC. Cloning and sequencing of human lambda immunoglobulin genes by the polymerase chain reaction. *Eur J Immunol*. 1990;20:2661-2666.

22. Lefranc M, Lefranc G. *The Immunoglobulin FactsBook*. London: Academic Press; 2001.
23. Lefranc M. IMGT databases, web resources and tools for immunoglobulin and T cell receptor sequence analysis. <http://imgt.cines.fr>. *Leukemia*. 2003;17:260-266.
24. Vasconcelos Y, Davi F, Levy V, et al. Binet's staging system and VH genes are independent but complementary prognostic indicators in chronic lymphocytic leukemia. *J Clin Oncol*. 2003;21:3928-3932.
25. Lossos I, Tibshirani R, Narasimhan B, Levy R. The inference of antigen selection on Ig genes. *J Immunol*. 2000;165:5122-5126.
26. Brezinschek HP, Foster SJ, Brezinschek RI, Dorner T, Domiati-Saad R, Lipsky PE. Analysis of the human VH gene repertoire. Differential effects of selection and somatic hypermutation on human peripheral CD5(+)/IgM+ and CD5(-)/IgM+ B cells. *J Clin Invest*. 1997;99:2488-2501.
27. Lin K, Sherrington PD, Dennis M, Matrai Z, Cawley JC, Pettitt AR. Relationship between p53 dysfunction, CD38 expression, and IgV(H) mutation in chronic lymphocytic leukemia. *Blood*. 2002;100:1404-1409.
28. Milner EC, Hufnagle WO, Glas AM, Suzuki I, Alexander C. Polymorphism and utilization of human VH Genes. *Ann N Y Acad Sci*. 1995;764:50-61.
29. Ghiotto F, Fais F, Valetto A, et al. Remarkably similar antigen receptors among a subset of patients with chronic lymphocytic leukemia. *J Clin Invest*. 2004;113:1008-1016.
30. Widhopf G II, Rassenti L, Toy T, Gribben J, Wierda W, Kipps T. Chronic lymphocytic leukemia B cells of over one percent of patients express virtually identical immunoglobulins. *Blood*. Prepublished on June 24, 2004, as DOI 10.1182/blood-2004-03-0818.
31. Tobin G, Thunberg U, Karlsson K, et al. Subsets with restricted immunoglobulin gene rearrangement features indicate a role for antigen selection in the development of chronic lymphocytic leukemia. *Blood*. Prepublished on June 24, 2004, as DOI 10.1182/blood-2004-01-0132.