The pattern and distribution of immunoglobulin V_H gene mutations in chronic lymphocytic leukemia B cells are consistent with the canonical somatic hypermutation process

Bradley T. Messmer, Emilia Albesiano, Davorka Messmer, and Nicholas Chiorazzi

The overexpanded clone in most B-cell–type chronic lymphocytic leukemia (B-CLL) patients expresses an immunoglobulin (Ig) heavy chain variable (V_H) region gene with some level of mutation. While it is presumed that these mutations were introduced in the progenitor cell of the leukemic clone by the canonical somatic hypermutation (SHM) process, direct evidence of such is lacking. Nucleotide sequences of the Ig V_H genes from 172 B-CLL patients were analyzed. Previously described V_H gene usage biases were noted. As with canonical SHM, mutations found in B-CLL were more frequent in RGYW hot spots (mutations in an RGYW motif = 44.1%; germ line frequency of RGYW motifs = 25.6%) and favored transitions over transversions (transition-transversion ratio = 1.29). Significantly, transition preference was also noted when only mutations in the wobble position of degenerate codons were considered. Wobble positions are inherently unselected since regardless of change an identical amino acid is encoded; therefore, they represent a window into the nucleotide bias of the mutational mechanism.

B-CLL V_H mutations concentrated in complementarity-determining region 1 (CDR1) and CDR2, which exhibited higher replacement-to-silent ratios (CDR R/S, 4.60; framework region [FR] R/S, 1.72). These results are consistent with the notion that V_H mutations in B-CLL cells result from canonical SHM and select for altered, structurally sound antigen receptors. (Blood. 2004;103:3490-3495)

© 2004 by The American Society of Hematology

Introduction

B-cell–type chronic lymphocytic leukemia (B-CLL) is a disease of clonally expanded CD5+ lymphocytes that express intact antigen receptors. Roughly half of B-CLL patients have a clone that expresses an immunoglobulin (Ig) variable (V) region gene that differs from the germ line sequence by at least 2% in the V segment of the heavy chain (V_H). These patients have a significantly better clinical outcome than patients whose B-CLL clone expresses a relatively unmutated V_H gene. The biologic basis for the difference in clinical outcomes is not understood, although gene expression and surface membrane phenotypic differences have been described.

Normal B cells accumulate mutations in their Ig V genes during their affinity maturation in the germinal center by a T-cell–dependent process termed somatic hypermutation (SHM). These mutations alter the structure of the antibody molecule and allow for the selection of variants with greater affinity for the driving antigen. As such, the presence of Ig V gene mutations produced by SHM in B-CLL cells would indicate that the progenitor cell was antigen experienced. While the precise mechanistic basis for the SHM process is currently being elucidated, the characteristics of the resulting mutations have been delineated in a number of systems. The final spectrum of mutations observed in V_H sequences deviates from randomness because of the intrinsic mutability of the germ line V_H genes and the effects of selection for the antibody molecule’s structure.

The intrinsic mutability of V_H genes has been the subject of much theoretical and empirical study but in essence resolves to 2 aspects. The first is the relative likelihood that a given nucleotide will be mutated and the second is the relative chance that it will mutate into the 3 other nucleotides. The latter is firmly established as a transition (purine-to-purine or pyrimidine-to-pyrimidine mutation) bias. The former involves a role for sequence context and hot spots such as RGYW (R = A or G, Y = C or T, W = A or T) and is probably also influenced by position and local DNA or RNA structure. Since precise mechanistic models for SHM are still evolving, empirical descriptions of mutational spectra remain the best available method for establishing the likelihood that a given set of sequences were affected by the same or similar processes and selections.

It is often assumed or inferred that the mutations present in V_H genes of B-CLL cells developed by the standard SHM process that is initiated during a T-cell–dependent germinal center (GC) reaction. However, this has not been formally proven and may be suspect. Ig V gene mutations can develop in the absence of T-cell help and outside of classical germinal centers. The finding of V_H gene mutations in B-CLL cells was initially surprising because...
most CD5+ B lymphocytes in the blood or follicular mantles of humans lack V gene changes. For this and other reasons, it has been suggested that some or all B-CLL cells are derived from a subset of B cells that can accumulate Ig V gene mutations in a T-cell–independent fashion (eg, marginal zone [MZ] B cells). The characteristics of the mutations acquired in this manner and their similarities or differences with those occurring during a T-cell–dependent process are not yet defined.

Therefore, we undertook an empirical study on a large database of B-CLL V\(\text{H}\) sequences to determine whether the mutations in the Ig V gene exhibited features that were similar or different from canonical SHM. The targeting preference for RGYW motifs, base change bias for transitions, and focusing of replacement mutations in the complementarity-determining regions (CDRs) and away from the framework regions (FRs) in B-CLL sequences were consistent with SHM. Thus, regardless of the T-cell or GC dependence of these mutations, our data do not support a necessity to invoke a mutational process other than the normal SHM as the mediator of Ig V gene mutations in B-CLL cells.

**Patients, materials, and methods**

**Patients**

The Institutional Review Boards of North Shore University Hospital (Manhasset, NY) and Long Island Jewish Medical Center (New Hyde Park, NY) approved these studies. Informed consent was provided by all participants in accordance with the Declaration of Helsinki. The patients in this study were diagnosed with typical B-CLL based on clinical criteria and laboratory features. Peripheral blood mononuclear cells (PBMCs) were separated from heparinized venous blood of B-CLL patients by density gradient centrifugation using Ficoll-Paque (Amersham Biosciences, Piscataway, NJ). PBMCs were cryopreserved with a programmable cell-freezing machine (Cryomed, Mt Clemens, MI).

**Ig V\(\text{H}\) gene sequencing**

V\(\text{H}\) cDNA sequences were determined by the previously described method. Briefly, mRNA was purified from cryopreserved PBMCs of B-CLL patients. The cDNA was synthesized by reverse transcription and used as template in polymerase chain reaction (PCR) reactions with V\(\text{H}\) family FR-specific forward primers and \(\mu, \gamma, \text{and} \\alpha\) chain reverse primers. Once the V\(\text{H}\) family was determined, the PCR was repeated using a leader-specific forward primer, and this product was either directly sequenced or cloned and sequenced. Only complete (beginning of FR1 through FR4) B-CLL VH sequences were considered here.

**Other V\(\text{H}\) sequences**

A collection of mutated V\(\text{H}\) sequences from the public GenBank database was obtained. All of these sequences were entered by Dr P. E. Lipsky and used the V\(\text{H}\) genes 3-07, 3-23, 4-34, or 4-59.

**Sequence analysis**

The closest germ line V\(\text{H}\) for each B-CLL V\(\text{H}\) sequence was assigned using DNAPLOT (http://www.mrc-cpe.cam.ac.uk/DNAPLOT.php?menu=901) or IgBlast (http://www.ncbi.nlm.nih.gov/igblast/). Other sequence manipulations were done using DNASTAR v5.06 (DNASTAR, Inc, Madison, WI) and Microsoft Excel 2002. Nucleotide differences that were assignable to known alleles or that occurred independently among multiple B-CLL sequences, indicating allelic origin, were not scored as mutations. Deletions or insertions were not considered. CDRs and FRs were defined by the rules of Kabat et al.

**Average mutability calculations**

The average mutability of the “wobble” nucleotides was calculated for each germ line gene as follows. For a given gene, the redundant codon positions were identified and, for each one, the mutability scores of the 3 triplets in which the redundant nucleotide resides were averaged. The nucleotide mutability scores were then averaged for all positions containing the same nucleotide to create an average mutability score for every A, G, C, and T in each germ line gene. Finally the overall average mutability was determined by weighting the score for each nucleotide in each germ line gene by summing the products of each gene’s mutability values and the number of mutations in V\(\text{H}\) sequences of that gene and dividing the total by the overall number of mutations considered.

**Statistical analysis**

The percentage of mutations in RGYW motifs, the transition-to-transversion ratios, and the replacement-to-silent ratios were evaluated by the \(\chi^2\) test using the non-CLL dataset as the expected values. When the number of observed events was too low for the \(\chi^2\) test to be considered accurate, the Fisher exact test was employed. All calculations were performed using GraphPad InStat version 3.05 (GraphPad Software, San Diego, CA).

**Results**

**V\(\text{H}\) gene usage and mutation frequencies**

Complete V\(\text{H}\) sequences were obtained from 172 leukemic clones of B-CLL patients. Each sequence was assigned a germ line counterpart using the public tools (IgBlast and DNAplot). V\(\text{H}\) family genes were the most commonly used, with V\(\text{H}1\) and V\(\text{H}4\) used at somewhat lower frequencies. V\(\text{H}4\)-34, 1-69, 3-07, and 3-23 and their alleles were the most commonly used V\(\text{H}\) genes (Figure 1A), as expected. The frequency of cases using V\(\text{H}4\)-34 and 1-69 was greater than the frequency of those genes reported in a normal CD5+ B-cell repertoire, and the frequency of V\(\text{H}4\)-30– and 3-09 gene–using cases was lower than the frequency of those genes in the normal repertoire (\(P < .5\); data not shown). These specific V\(\text{H}\) gene frequencies are consistent with those described previously. However, these frequency comparisons may be spurious if the cell of origin of B-CLL cells is not the circulating CD5+ B cell found in most healthy subjects.

A database of the differences from the germ line sequences was compiled, with attention paid to eliminating possible and probable allelic variants from subsequent inclusion in mutation analysis. The total number of mutations in each B-CLL sequence was determined. Approximately 41% of the Ig V\(\text{H}\) genes analyzed had less than 2% difference from the most similar germ line gene, the commonly used threshold for considering a case “unmutated,” and 24% had no changes whatsoever (Figure 1B). Approximately 22% of the leukemic clones had between 2% and 5% mutations and 37% had more than 5% mutations. These mutations occurred in all V\(\text{H}\) families, although within each family the distribution of mutations was largely significantly.
of the 3-07 genes exhibit at least 2% mutations, the other \( V_H^3 \) genes have almost the same percentage of unmutated genes as the overall (38%). Finally, while \( V_H^4 \)-34 genes are deficient in unmutated sequences, the remaining \( V_H^4 \) family genes are slightly enriched in unmutated sequences (59%; Figure 1C).

### Nucleotide characteristics of B-CLL \( V_H \) mutations

**Targeting of mutations to specific sequences and areas within \( V_H \) genes.** The targeting and base change preferences associated with canonical SHM are nonrandom and have been empirically determined.\(^5\) A sampling of non-B-CLL mutated \( V_H \) sequences from the literature was analyzed and found to be consistent with expected biases toward mutations in RGYW hot spots, increased replacement-to-silent (R/S) ratios in the CDR and depressed ratios in the framework, and for transitions (\( A\rightarrow G, G\rightarrow A, C\rightarrow T, \) and \( T\rightarrow C) \) over transversions (Table 1). Overall, the mutations in the B-CLL sequences exhibited quantitatively similar biases that were not statistically different from the representative non-CLL sequences. There was an increased frequency of mutation in RGYW motifs, with a preference for transitions over transversions regardless of the nucleotide that was mutated (Figure 2A). Interestingly, the characteristics of the mutations occurring in “unmutated” cases (those with \(<2\%\) were also not significantly different from non-CLL, although the lower level of replacement mutations in the CDR was almost significant (\( P = .0501\)). A larger number of B-CLL–derived \( V_H \) gene sequences will be needed to answer this question definitively.

### Nucleotides biases in \( V_H \) gene mutations.

In order to assure that the transition/transversion biases were not the result of amino acid selection for antigen binding or undefined structural requirements, we analyzed the transition-to-transversion ratios at redundant wobble positions in the \( V_H \) sequences (Figure 2B). Wobble positions are the third nucleotide of redundant codons. Since any nucleotide at these positions encodes the same amino acid, these nucleotides are useful for deconvoluting the potential role of amino acid selection because they cannot be selected at the protein level. As such, these positions offer a window into the targeting and base change preferences of the mutational mechanism with no interference from protein structure–based selective pressures (although selection at the level of transcription or translation cannot be formally excluded). These analyses also indicated that nucleotide changes that are found in B-CLL \( V_H \) favored transitions.

**Bias against changes in C nucleotides at wobble positions of redundant codons.** One surprising result of the analysis of these redundant codons was an unexpectedly strong bias against mutations at \( C \) nucleotides that is not present when the mutations at all sites are considered. This bias was also found in redundant codons of the non-B-CLL \( V_H \) sequences analyzed here (Figure 3).

One possible explanation for such a bias could invoke sequence context effects. Other studies have demonstrated that the conventional SHM process has a hierarchy of preference for various dinucleotide and trinucleotide motifs.\(^4\) The difference between the

---

### Table 1. Targeting and base change characteristics of mutations in B-CLL, unmutated B-CLL, and non-B-CLL \( V_H \) genes

<table>
<thead>
<tr>
<th>Type</th>
<th>No. of sequences</th>
<th>No. of mutations</th>
<th>R/S ratio</th>
<th>Germ line RGYW, %</th>
<th>Mutation in RGYW, %</th>
<th>Transitions-to-transversions ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-CLL</td>
<td>172</td>
<td>1789</td>
<td>2.35*</td>
<td>4.60*</td>
<td>1.72*</td>
<td>25.6</td>
</tr>
<tr>
<td>B-CLL with less than 2% mutation</td>
<td>71</td>
<td>71</td>
<td>1.86*</td>
<td>3.50†</td>
<td>1.40*</td>
<td>24.7</td>
</tr>
<tr>
<td>Non-B-CLL</td>
<td>67</td>
<td>452</td>
<td>2.23</td>
<td>4.27</td>
<td>1.50</td>
<td>24.9</td>
</tr>
</tbody>
</table>

*Not a significant difference compared with non-CLL (\( P > .4\)).

†Not a significant difference compared with non-CLL (\( P > .05\)).
most mutated and least mutated motifs can be significant. Therefore, the flanking sequences of each redundant codon were considered. The average mutability of the 3 trinucleotide windows (upstream, centered, and downstream) that contained the particular wobble nucleotide was calculated for each germ line redundant codon position using the mutability index of Shapiro et al., as described in “Patients, materials, and methods.” While C nucleotides had the lowest relative mutability, the difference was not as pronounced as the actual frequencies observed here (average mutability: A, 1.08; G, 0.84; C, 0.77; T, 0.98). Furthermore, the average mutability calculated for G nucleotides was only marginally greater than that for C, yet the normalized mutation frequency of G was slightly elevated in both the B-CLL and non-B-CLL sequences analyzed (Figure 3). Similar results were obtained when dinucleotide contexts were considered (data not shown).

The rate of mutation drops off exponentially toward the 3' end of the Ig V gene sequence, and therefore a clustering of redundant C nucleotides in one portion of the V<sub>H</sub> gene could lead to a lowered mutation rate as a function of their position. However, the various redundant nucleotides were evenly distributed throughout the germ line genes, excluding a gross positional effect (data not shown).

B-CLL V<sub>H</sub> mutations are selected

Selection for improved antigen binding sites or for the maintenance of antibody structure typically causes localized shifts in the R/S ratios of mutations during canonical SHM. The B-CLL mutations analyzed here follow a similar pattern (Table 1), with an elevated R/S in CDRs (mean, 4.60) and decreased R/S in FRs (mean, 1.72). This pattern mirrored that seen in the normal sequences (4.27 and 1.50, respectively).

When the mutations are mapped onto the gene, there is an apparent targeting of replacement mutations toward the CDR1 and CDR2 (Figure 4A). Although such mapping is helpful in identifying nucleotides of particular vulnerability, it is less valuable when considering targeting of larger stretches of DNA sequence. Therefore, to gain more resolution as to which areas were enriched or suppressed for replacement mutations, a “moving window” approach was applied to the positional R/S data. In this analysis, we determined the R/S ratio that surrounds each nucleotide by using the cumulative R/S of stretches of 5, 6, and 7 nucleotides upstream and downstream of the central position (moving windows of 11, 13, 15, and 17 nucleotides).

Figure 2. A transition bias is apparent in all mutated nucleotides. Graphs show the relative frequency with which each nucleotide mutates into each of the other 3. The number of mutations occurring in each nucleotide is shown in parentheses on the x-axis. (A) All mutations. (B) Mutations occurring at the third position of completely redundant codons. Data are from 172 B-CLL sequences from our internal laboratory database and 67 non-B-CLL sequences from the public databases.

Figure 3. Nucleotide biases of the mutational mechanism. Normalized mutation frequencies were calculated by comparing the fraction of mutations in a given nucleotide to the fraction of that nucleotide in the germ line sequences. If mutations occurred in a nucleotide at the same relative rate that it appeared in the germ line V<sub>H</sub> sequences, it would have a normalized mutation rate of 1. When all mutations are considered, there is a slight bias away from mutating T nucleotides. However, when only the silent redundant positions are considered, there is a pronounced bias against mutations in C nucleotides in both B-CLL (top) and normal (bottom) sequences.

Figure 4. CDRs of B-CLL sequences contain more replacement mutations. (A) Number of replacement mutations at each position in the V<sub>H</sub> genes of B-CLL. (B) Number of silent mutations observed at each position in the V<sub>H</sub> of B-CLL. (C) A “moving window” R/S ratio for each position was calculated by considering the cumulative R/S ratio of the centered position and the 5 (dotted line), 6 (dashed line), or 7 (solid line) positions on either side. CDR and FR are as defined by Kabat et al. 24

Ig V<sub>H</sub> MUTATIONS IN B-CLL RESEMBLE SHM 3493

BLOOD, 1 MAY 2004 • VOLUME 103, NUMBER 9
and 15 nucleotides). These windows make the differences from individual nucleotide to individual nucleotide (Figure 4A-B) less pronounced, and the larger the window the greater the “smoothing” becomes. This analysis clearly localized the CDR1 and CDR2 as having elevated R/S ratios and the FR1 and FR2 as having lowered R/S ratios. Interestingly an increased R/S ratio at the end of FR3 was also clearly identified with this approach (Figure 4C).

Discussion

The results presented here strongly support the hypothesis that the clonal V_{H} mutations seen in B-CLL cells are the product of the classical antigen-driven SHM process. In addition, previously observed mutation distribution and gene usage biases were confirmed and extended. While these observations are not unexpected, the large sample size analyzed here creates a more firm footing for models of B-CLL development that suggest that B-CLL cases with mutated V_{H} genes derived from cells whose V segments were somatically mutated (Table 1; Figure 1) via a process that strongly resembles or is (Table 1; Figures 2-4) the conventional SHM machinery in response to antigen stimulation.

The targeting preference and base change biases of the mutational mechanism responsible for the B-CLL V_{H} mutations were distinguished from the effects of selection by focusing on intrinsically unselectable wobble positions in redundant codons (Figure 2B). This approach confirmed that the overall pattern of transition bias was not excessively influenced by selecting for protein structure that provides advantages in antigen binding or eliminates the danger of destabilizing the B-cell–receptor scaffolding. In addition, this approach revealed an unexpected aspect of mutation targeting. Cytidine nucleotides at redundant wobble positions were mutated at approximately half the frequency of the other bases in both B-CLL and canonical SHM-generated sequences (Figure 3). Curiously, there was no overall bias against mutations in cytidine, suggesting that the wobble positions are somehow distinct in their mutation profile.

An inherent limitation of this analysis is the restriction to B-CLL cases with V_{H} mutations, the degree of which can differ from case to case. Sequences with higher numbers of mutations had a greater influence on the results, although this was unavoidable. The data reported here, however, in no way rule out a role for antigen selection in the unmutated cases, since many germ line or natural antibodies can have moderate affinity antigens and antibodies. The data reported here, however, in no way rule out a role for antigen selection in the unmutated cases, since many germ line or natural antibodies can have moderate affinity antigens and antibodies.

The few mutations observed in sequences with less than 2% differences from the most similar germ line gene (“unmutated”) also resembled the canonical SHM spectra, although a depressed R/S ratio in the CDR approached a statistically significant difference. This may indicate that the unmutated Ig V sequences occurred in cells that began, but quickly aborted, the SHM process, perhaps due to an already avid antigen receptor or impermissible autoreactivity. However, these are speculations in the absence of defined antigens and affinities.

It has been proposed that B-CLL is derived from marginal-zone (MZ) B cells.19 Although MZ B cells can display somatic mutations,40 these may occur via a T-independent mechanism outside of classical GCs.17,18 If MZ cells are the precursors of B-CLL cells, then the mechanism that generates mutations in MZ cells would presumably be mechanistically similar to the SHM that takes place in GCs. The results of the current study would be inconsistent with a model that attributes B-CLL mutations to a distinct mechanism with a different mutational spectrum.

Ongoing mutations in Ig V genes41 and BCL-642 have been observed in some B-CLL cases. At this time, there is an insufficient number of sequences of emerging variants available to perform the analyses similar to those in this study, although one study did suggest that targeting might differ in these ongoing.43 However, such an analysis would be valuable, as it would indicate whether the ongoing mutations are consistent with continued SHM, as has been postulated from demonstrations that activation-induced cytidine deaminase (AID) is expressed in some B-CLL cells.43,44 The alternative, that the ongoing mutations are not the product of SHM, would indicate that at least some B-CLL cells have a more elevated mutation rate. Elevated mutation rates are often conducive to increased aggressiveness46 and would be cause for clinical concern.

Acknowledgments

We thank David Thaler, Manlio Ferrarini, and Federico Caligaris-Cappio for useful discussions of the data.

References

42. Sahola SS, Davis Z, Hamblin TJ, Stevenson FK.\nSomatic mutation of bcl-6 genes can occur in the\nabsence of V(H) mutations in chronic lymphocytic\nleukemia. Blood. 2000;95:3534-3540.
The pattern and distribution of immunoglobulin V<sub>H</sub> gene mutations in chronic lymphocytic leukemia B cells are consistent with the canonical somatic hypermutation process

Bradley T. Messmer, Emilia Albesiano, Davorka Messmer and Nicholas Chiorazzi