Multiple Distinct Sets of Stereotyped Antigen Receptors Indicate a Role for Antigen in Promoting Chronic Lymphocytic Leukemia


1 North Shore–LIJ Research Institute and 2 Department of Medicine, North Shore University Hospital and 3 Department of Medicine, NYU School of Medicine, Manhasset, NY 11030
4 ICGEB Outstation-Monterotondo, CNR Campus “Adriano Buzzati-Traverso,” 00016 Rome, Italy
5 Dipartimento di Medicina Sperimentale, Sezione di Anatomia Umana, 6 Division of Medical Oncology C, Istituto Nazionale per la Ricerca sul Cancro, and 7 Dipartimento di Oncologia Clinica e Sperimentale, Università di Genova, 16132 Genova, Italy
8 Ematologie Dipartimento di Biotecnologie Cellulari ed Ematologia, University “La Sapienza,” 00161 Rome, Italy
9 Department of Medicine, Long Island Jewish Medical Center and 10 Department of Medicine, Albert Einstein School of Medicine, New Hyde Park, NY 11040

Abstract

Previous studies suggest that the diversity of the expressed variable (V) region repertoire of the immunoglobulin (Ig)H chain of B-CLL cells is restricted. Although limited examples of marked constraint in the primary structure of the H and L chain V regions exist, the possibility that this level of restriction is a general principle in this disease has not been accepted. This report describes five sets of patients, mostly with unmutated or minimally mutated IgV genes, with strikingly similar B cell antigen receptors (BCRs) arising from the use of common H and L chain V region gene segments that share CDR3 structural features such as length, amino acid composition, and unique amino acid residues at recombination junctions. Thus, a much more striking degree of structural restriction of the entire BCR and a much higher frequency of receptor sharing exists among patients than appreciated previously. The data imply that either a significant fraction of B-CLL cells was selected by a limited set of antigenic epitopes at some point in their development and/or that they derive from a distinct B cell subpopulation with limited Ig V region diversity. These shared, stereotyped Ig molecules may be valuable probes for antigen identification and important targets for cross-reactive idiotypic therapy.

Key words: Ig variable region genes • human • B lymphocytes • antibodies • autoantibodies

Introduction

The B lymphocyte clone expanded in chronic lymphocytic leukemia (B-CLL) expresses low levels of surface membrane Ig, the B cell antigen receptor (BCR). The genetics of this Ig have clinical relevance, as patients with a clone whose Ig variable (V) region has no or few mutations have a significantly worse outcome than those with significant numbers of Ig V mutations (1, 2). The biology underlying this association is unclear.

Several lines of evidence support a role for the BCR in the evolution of B-CLL (for review see reference 3). The distribution of individual IgVH in B-CLL clones differs from that found in normal cells (4), with an increased frequency of VH1-69, VH4-34, and VH3-07 (4–6). In addition, the distribution of mutations among B-CLL cases using these specific VH genes is selectively biased (4, 5, 7).

Recently, two subgroups of B-CLL cases with remarkable similarity of the entire BCR (V regions of the H and L chains) were identified (8, 9). Although these findings are
provocative, they have been considered rare and potentially anomalous because, in one instance, the clones expressed IgG (9) and in the other geography and ethnicity may be relevant (10). This report describes another five groups of B-CLL patients that express BCRs of strikingly similar primary structure defined by highly similar IgV regions in the H and L chains, and, in particular, distinct H and L CDR3 configurations. Thus, a significant fraction of B-CLL clones derive from B lymphocytes with constrained antigen-binding sites that could recognize individual, discrete antigens or classes of structurally similar epitopes.

Materials and Methods

IgV Gene Sequencing. V<sub>H</sub>D<sub>H</sub>J<sub>H</sub> and V<sub>L</sub>JK sequences were determined by previously described methods (4, 9).

Database Searches. B-CLL Ig H chain V sequences from our collection (n = 255) and the public databases (n = 197) were subjected to BLAST searches of both nucleotide and protein databases to identify similar sequences. The criteria used to define “sets” of similar rearranged V<sub>H</sub>D<sub>H</sub>J<sub>H</sub> were as follows: (a) use of the same V<sub>H</sub>, D, and J<sub>H</sub> germline genes; (b) use of the same D segment reading frame and position relative to the V<sub>H</sub>, plus or minus one codon; and (c) an amino acid similarity within the HCDR3 of ≥60% identity. In addition, all B-CLL Ig H protein sequences were aligned and clustered using the ClustalW alignment algorithm. Sequences clustering tightly were visually inspected for similarity. All of these searches used the complete V<sub>H</sub>D<sub>H</sub>J<sub>H</sub> and as such were weighted toward sequences that used the same V<sub>H</sub> gene. To identify sequences with similar HCDR3 but different V<sub>H</sub> genes, CDR3 motifs from the various sets were used to search the public databases with the ProteinInfo search engine (http://prowl.rockefeller.edu). The criteria for the members of set V were altered to permit the use of different IgV<sub>H</sub> genes that were members of the same IgH<sub>H</sub> clan, while retaining the criteria for the rearranged V<sub>L</sub>. Use of the same specific IgV<sub>L</sub> gene and ≥85% LCDR3 identity was required for the inclusion of a companion rearranged V<sub>L</sub>JK in a set.

538 V<sub>H</sub> sequences from CD5<sup>+</sup> and CD5<sup>-</sup> peripheral B lymphocytes (10, 11) were downloaded from the public database. These 538 sequences were compared independently with the translated databases using tblastn on the BlastMachine at the AMDeC Bioinformatics Core Facility at the Columbia Genome Center at Columbia University.

Online Supplemental Material. Figs. S1–S7 depict detailed nucleotide and amino acid sequence alignments of the junctional regions and complete protein sequence alignments of the sequences described here. Online supplemental material is available at http://www.jem.org/cgi/content/full/jem.20040544/DC1.

Table I. Sets of B-CLL Cases That Share Ig V Region Genes

<table>
<thead>
<tr>
<th>Set</th>
<th>Internal</th>
<th>Public</th>
<th>Non–B-CLL</th>
<th>Isotype</th>
<th>V&lt;sub&gt;H&lt;/sub&gt;</th>
<th>D; reading frame</th>
<th>J&lt;sub&gt;H&lt;/sub&gt;</th>
<th>V&lt;sub&gt;L&lt;/sub&gt;</th>
<th>J&lt;sub&gt;L&lt;/sub&gt;</th>
<th>Median % mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>3</td>
<td>2</td>
<td>1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>IgM&lt;sub&gt;k&lt;/sub&gt;</td>
<td>1-69</td>
<td>3-16; philic</td>
<td>3</td>
<td>kA27</td>
<td>k1/k4/k5</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>II</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>IgM&lt;sub&gt;k&lt;/sub&gt;</td>
<td>1-69</td>
<td>2-2; philic</td>
<td>6</td>
<td>kL6</td>
<td>k3</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>III</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>IgM&lt;sub&gt;A&lt;/sub&gt;</td>
<td>1-69</td>
<td>3-10; philic</td>
<td>6</td>
<td>k1-16</td>
<td>k1/k3</td>
<td>0.0 0.2</td>
</tr>
<tr>
<td>IV</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>IgG&lt;sub&gt;k&lt;/sub&gt;</td>
<td>4-34</td>
<td>5-5; philic</td>
<td>6</td>
<td>kA17</td>
<td>k1/k2</td>
<td>3.0 2.3</td>
</tr>
<tr>
<td>Va</td>
<td>4</td>
<td>2</td>
<td>1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>IgM&lt;sub&gt;k&lt;/sub&gt;</td>
<td>1-02</td>
<td>6-19; stop</td>
<td>4</td>
<td>kO12/2</td>
<td>k1/k2</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Vb</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>IgM&lt;sub&gt;k&lt;/sub&gt;</td>
<td>1-03</td>
<td>6-19; stop</td>
<td>4</td>
<td>kO12/2</td>
<td>k1/k2</td>
<td>0.8 0.0</td>
</tr>
<tr>
<td>Vc</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>IgM&lt;sub&gt;k&lt;/sub&gt;</td>
<td>1-18</td>
<td>6-19; stop</td>
<td>4</td>
<td>kO12/2</td>
<td>k1</td>
<td>1.4 0.0</td>
</tr>
<tr>
<td>Vd</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>IgM&lt;sub&gt;+&lt;/sub&gt;=</td>
<td>1-46</td>
<td>6-19; stop</td>
<td>4</td>
<td>–</td>
<td>–</td>
<td>0.0 –</td>
</tr>
<tr>
<td>Ve</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>IgM&lt;sub&gt;k&lt;/sub&gt;</td>
<td>5-51</td>
<td>6-19; stop</td>
<td>4</td>
<td>kO12/2</td>
<td>k2</td>
<td>0.0 0.0</td>
</tr>
</tbody>
</table>

<sup>a</sup>Anticardiolipin antibody.

<sup>b</sup>Splenic marginal zone lymphoma.
motif common to these three subsets to search public databases for additional sequences with the same HCDR3 configuration potentially associated with a different IgVH segment. This search was not restricted to B-CLL sequences. The approach confirmed the previously identified subsets and identified two additional subsets of set V (Vc and Vd).

The public database searches identified 21 VHDJH sequences, belonging to one of the five individual sets, bringing the total number of sequences among these sets to 43. Interestingly, only 2 of the 21 sequences culled from the public databases were not derived from B-CLL cells. These two were from an anticardiolipin antibody-producing B cell (set I) and from a splenic marginal zone lymphoma (set Va).

This distribution of similar sequences is particularly striking because, at the time of this search, the public databases contained only 197 Ig H chain V region sequences from B-CLL patients (excluding those from our laboratories) out of a total of >8,500 H chain V region sequences (search of Entrez with terms "human immunoglobulin heavy chain variable" produced 8,874 hits in the nucleotide database and >6,183 hits in the protein database on 12/16/03).

Pairing Restricted VLJL Rearrangements with VHDJH Segments in Sets. VLJL sequences corresponding to the shared VHDJH of the five sets were available for most of our B-CLL cases and for a few of those identified in the public databases. Remarkably, the available IgVL were highly conserved within the sets, and the corresponding JL were very restricted (Table I and Table S1, available at http://www.jem.org/cgi/content/full/jem.20040544/DC1). Four out of the five sets with available L chains expressed the κ isotype.

IgV Gene Mutation Status and Isotype Restrictions of Individual Sets. Most of the IgVH sequences in each set differed by <2.0% from the most similar germline gene, with the exception of set IV in which the median level of mutation was 3.0%. Notably, the deduced protein structures in those sequences that were considered "mutated" using the typical 2% threshold differed from the germline by relatively low levels. Only one sequence, from set IV (Table S1, CLL ID47), differed by >5% from its germline counterpart. The corresponding IgVL in each set exhibited low levels of mutation; in some cases, VH displayed <2.0% difference, whereas VL had ≥2% difference from the germline sequence (Table I and Table S1).

The H chain isotype was the same among members of a set. All sets expressed IgM, except for set IV that consisted of IgG+ cases, similar to a patient group reported previously (9).

H and L CDR3 Characteristics of the Individual Sets. We identified trends in the chemical, structural, or functional nature of the residues that comprise the H and L CDR3s, and in particular their VH-D and D-JH junctions (Fig. 1). For example, the D segments in the HCDRs of these sets were read in the hydrophobic and stop reading frames more often than in normal (12) and B-CLL (13) cells. For all cases in set V, the D6-19 segment is read in a nonproductive reading frame (Table I). However, the stop codon, located in the region of overlap with the terminal IgVH sequence, was trimmed, allowing productive rearrangements with the JF segment (Fig. S7, available at http://www.jem.org/cgi/content/full/jem.20040544/DC1).

Also of note was the repeated occurrence of certain non-germline encoded amino acids within D segments in some of the sets. For example, in all members of set II, a change to M is found at the 3′ end of the D segment (Fig. S4, available at http://www.jem.org/cgi/content/full/jem.20040544/DC1), a position that is not known to be polymorphic.
Three out of seven sequences in set III had an R to Q change within the D3-10 segment that is also not listed as polymorphic (Fig. S5, available at http://www.jem.org/cgi/content/full/jem.20040544/DC1). In four out of five cases in set IV, P replaced A in the portion of HCDR3 encoded by the canonical D5-5 segment. Although this is most likely a polymorphism of the D segment rather than a common mutation, the last of the five sequences in this set (CLL ID47) also deviates from the canonical D5-5 sequence at this codon, substituting a D (Fig. S6, available at http://www.jem.org/cgi/content/full/jem.20040544/DC1). Thus, even if these amino acid changes represent polymorphisms, their relative consistency within each set suggests a selection for these residues.

Members of several sets have common junctional residues that were not templated by any known germline gene segments and, therefore, presumably arose from trimming and/or addition during recombinational assembly. The sequences in set I all contain a pair of Gs at the V_{H}-D junction and an N at the D-J_{H} junction (Fig. 1 and Fig. S3, available at http://www.jem.org/cgi/content/full/jem.20040544/DC1). A very similar V_{H}-D junctional finding exists in set II (Fig. 1 and Fig. S4). All sequences in set IV contain an aromatic residue at the V_{H}-D and a pair of basic residues (R or K) at the D-J_{H} junction (Fig. 1 and Fig. S6).

Other trends in the composition of the H and L LCDR3s are found in the other sets. These and the fine details of the nucleotide and amino acid sequences of the V_{H}D_{H} and V_{L}J_{L} junctions for each set are shown and discussed in the online supplemental data (Figs. S3–S7).

**Structural Similarities of the BCR among Members of the Sets.** The deduced V_{H}D_{H} and V_{L}J_{L} protein sequences for each member of the stereotyped sets are presented in Figs. S1 and S2 (available at http://www.jem.org/cgi/content/full/jem.20040544/DC1). Because most members of the sets use the same Ig_{V}V_{H} primarily in an unmutated form, associated with the same D and J_{H} segments and because these rearrangements are virtually always paired with an identical VLJL junctions for each set are shown and discussed in the online supplemental data (Figs. S3–S7).

These data indicate a much more marked constraint on the primary structure of the BCR in B-CLL than appreciated previously. They also indicate that this principle occurs in a sizeable number of patients. Collectively, ~12% (31 out of 255: 22 from this work, 5 from our previous paper [9], and 4 that match another described set [8, 10]) of all of sequences in our internal laboratory B-CLL database and ~20% (27 out of 131) of those with unmutated Ig_{V} belong to one out of the five stereotyped sets described here or one out of the two aforementioned patient groups (8–10). Approximately the same overall frequency (~12%) was encountered among the sequences from the public databases (21 out of 197), although the proportion of the public B-CLL sequences that are unmutated was not determined. Most of the rearrangements in these sets lack or have few somatic mutations, and even those whose V_{H} surpass the 2% threshold commonly used as the criterion to define significant Ig_{V} gene mutations (4, 5) are only slightly above that level. This suggests that restricted BCR structure is primarily a feature of those patients with the worse clinical course and outcome (1, 2). It appears that one out of five B-CLL cases with unmutated BCRs fit into one of these defined sets. Additional sets will likely be uncovered as more Ig_{V} region sequences are defined in B-CLL, and all unmutated cases may be similar to one of a discrete number of archetypal sets. Although sets I–III use unmutated 1–69, they differ from previously described 1–69–expressing B-CLL cases that have restrictions in specific D and J_{H} segments associations (4, 6). These differences include J_{H} (set I, J_{H}3 vs. J_{H}6), D (set II, D2 vs. D3 family and V_{K}L6 with an extremely short LCDR3), and L chain (set III, λ vs. κ) gene use.

Initial studies that considered only Ig_{V_{H}} or V_{L}D_{H} (4–6, 14) pointed toward limited structural diversity in the antigen-binding sites of B-CLL. However, our current results are much more striking because of the remarkable similarity of the sequences within a set and the virtual mathematical impossibility that this similarity arose by chance. If gene segment use in B-CLL was random, the probability of finding the same combination of V_{H}D_{H} and V_{L}J_{L} segments in independent leukemic (or normal) B cells would be >10^{-6}. Therefore, one would not expect to identify two B-CLL patients with BCRs comprised of the same V_{H}D_{H}/V_{L}J_{L} until 10^6 cases were analyzed. This calculation is conservative because it does not account for diversity at the V_{H}-D, D-J_{H}, and V_{L}-J_{L} junctions that can be quite extensive (potentially exceeding 10^{-5} and reaching 10^{-12}), although receptor editing and revision could limit these possibilities somewhat. Nevertheless, the level and frequency of BCR structural restriction in clusters of patients reported here is extraordinary and appears to be higher than any other B or T cell lymphoproliferative disorder reported to date.

Finding similar Ig_{H} chain V region sequences by homology searches of the public databases is not, in itself, completely surprising because some Ig_{V} are expressed in a biased fashion and ~6,600 different V_{H}-D-J_{H} combinations can occur. Because the databases contain more than that number of Ig_{H} chain V region sequences, identifying the same recombined gene segments is not improbable. When we analyzed 538 CD5+ and CD5– B cell–derived H chain V region sequences, we identified many pairs of similar sequences and some groups of similar sequences. However, these groups derived from B cells of diverse sources, as would be expected if the similarities were the product of random chance. In contrast, the similarity to a given B-CLL–derived sequence detected in our database comparisons arose almost exclusively from other B-CLL sequences (19/21) or other lymphoproliferative disorders (1/21), even though the entire database was searched. Only one identified sequence...
was from a non–B-CLL clone, one that coded an autoantibody (Table I and Table S1). Although the proper normal B cell repertoire against which B-CLL clones should be compared remains an open question (3), these results demonstrate that sequence sets of restricted cellular origin are not a generalized phenomenon in the public database.

Therefore, the development of B-CLL must involve B cell clones with restricted IgV and/or BCR structure. Although it seems unlikely that the expression of particular BCR gene combinations could be the sole promoting factor for leukemogenesis, a strong inherent bias in gene segment association and VHDJH/VLJL pairing in the B cell tor for leukemogenesis, a strong inherent bias in gene segregation has been suggested (3). Although evidence exists in mice for biases in the recombination of particular IgV gene segments before antigen experience (15), the extent of restriction imposed by recombination biases at both the H and L chain V gene loci in those instances, especially at the V-(D)-J junctions, are not as severe as in the sets described here. To our knowledge, there is no known subpopulation of human B cells in which the frequency of similar rearrangements, independent of antigen selection, is as great as among these B-CLL cases.

Therefore, antigen selection probably has a major restrictive influence on the transformation of a normal B lymphocyte to a B-CLL cell. A simple model would postulate that the transforming event is coupled with antigen specificity; i.e., an individual B lymphocyte from a highly diverse population could bind and internalize a transforming agent (e.g., virus) via its BCR. Although this seems unlikely, such a mechanism has been proposed for B-CLL (16).

Alternatively, antigen could be a promoting factor for transformation, selecting specific clones for expansion from an initially diverse population of B lymphocytes and fostering their development to and in the transformed state (3). This would be the case if the B-CLL–susceptible cell population were preselected for antigen reactivity and, therefore, BCR structure, by exposure to distinct antigens or classes of antigens during their development. These clones could differ among patients, especially if the selecting antigens were foreign or autologous and possibly polymorphic. From within these clonal expansions, one member could develop an initial transforming lesion that would promulgate the leukemogenic cascade independent of antigen.

Finally, the initial transforming events could occur at random within a diverse B cell population or a previously antigen–selected population, and the subsequent nurturing of the transformed clone to clinical B-CLL could require ongoing BCR engagement by antigen (3). Recently, clonal expansions of B cells with phenotypic characteristics of B-CLL were found in normal elderly individuals (17, 18). The clinical relevance of these clones is not established. However, they may represent clones that have some of the genetic lesions of B-CLL, but lack BCR specificities that would result in sufficient ongoing stimulus to mature them into clinical B-CLL.

The remarkable protein similarity of the entire BCR among members of each set (Figs. S1 and S2) suggests that they could recognize the same or similar antigens. Although the nature of the antigens cannot be directly deduced from the Ig sequences presented here, there are several reasons to suspect that they are autoantigens or carbohydrates possibly derived from bacterial or viral coats, or a combination of the two.

VH1-69 (sets I–III) and VH3-21 (previously described set [8, 10]) are enriched among rheumatoid factors (19, 20). VH4-34 (set IV) is used in every case of monoclonal cold agglutinin disease (21) and in autoimmune conditions. Indeed, the inherent autoreactivity of this VH segment elicits a major inhibitory process by the immune system that keeps 4-34+ B cells from diversifying into high affinity, isotype-switched B cells (22). The anticardiolipin antibody identified as a member of set I implies that the other members of that set may be specific for cardiolipin or DNA because some antibodies to the former react with the latter (23). In addition, restricted VH/DJH and/or VLJL gene segments are features of B cells that produce anticarbohydrate mAb in human (24) and mouse (25).

Characteristic junctional residues are also a feature of anticarbohydrate mAb and autoantibodies and basic junctional residues, as seen in sets I, IV, and Ve (Fig. 1 and Fig. S1), often indicate reactivity with acidic targets such as DNA (26). The synthesis of autoreactive Ig/BCR molecules by many B-CLL clones (27, 28) supports a link between the unique BCR structural features of these sets and autoantibodies.

The non–B-CLL Ig sequences that matched these B-CLL stereotypes may give insight into the identity of the B-CLL progenitor cells. One of those two was derived from a splenic marginal zone lymphoma (Table S1, set Va), and the other was derived from an autoantibody–producing B cell (Table S1, set I). Interestingly, normal MZ B cells produce mAb that can recognize thymus–dependent type II antigens and autoantigens (29). In addition, the IgV region repertoire of murine MZ B cells is very restricted in gene segment use and structure that requires intact BCR signal transduction to develop (30). MZ B cells appear to be progenitors for gastric MALT lymphoma (31) and have been proposed as precursors of B-CLL cells (3). If one infers common antigenic reactivity based on the similar sequences within a set, a significant fraction of B-CLL cases, and in particular those with unmutated IgV genes, could produce mAb that recognize one of a limited, discrete array of antigens or epitopes. With such an interpretation, some B-CLL cases may resemble gastric MALT lymphoma regarding the role of antigenic drive (in that instance. Helicobacter pylori) in the promotion of malignancy. The stereotyped Ig molecules reported here might be valuable probes to identify antigens that drive the leukemogenic process in B-CLL.

Finally, these sets of stereotyped Ig molecules may serve as therapeutic targets on B-CLL cells. A conceptual drawback to targeting the BCR as a tumor–specific antigen has been the apparent need to create an individualized reagent for each patient. However, because our data indicate that
there is potentially extensive overlap in BCR structure and specificities among groups of B-CLL cases, this approach may be far less daunting. Indeed, because ~20% of the cases with unmutated IgVH genes fall into one of these sets, such targeting might be most effective in those cases that have the worst prognosis, are least responsive to therapy, and have the most aggressive clinical courses (1, 2).

This work was supported in part by RO1 grants (nos. CA 81554 and CA 87956) from the National Cancer Institute, an M01 General Clinical Research Center Grant (no. RR018535) from the National Center for Research Resources, the Associazione Italiana Ricerca sul Cancro, and MIUR. The Peter J. Sharp Foundation, the Laurie Strauss Leukemia Research Fund, the Joseph Eletto Leukemia Research Fund, the Jean Walton Fund for Lymphoma and Myeloma Research, the Joseph Eletto Leukemia Research Fund, the Laurie Strauss Leukemia Foundation, and the S.L.E. Foundation, Inc. also provided support for this work.

The authors have no conflicting financial interests.

Submitted: 22 March 2004
Accepted: 1 July 2004

References


23. Kumar, S., S. Nagl, J.K. Kalsi, C.T. Ravirajan, D. Athwal,


