Immunoglobulin Heavy Chain Gene Expression in Peripheral Blood B Lymphocytes

Chichi Huang,* A. Keith Stewart,1 Robert S. Schwartz, and B. David Stollar*

*Department of Biochemistry, Tufts University Schools of Medicine, Dental Medicine, and Veterinary Medicine and Sackler School of Graduate Biomedical Sciences, and Division of Hematology/Oncology, Department of Medicine, New England Medical Center Hospitals, Boston, Massachusetts 02111

Abstract

cDNA libraries for IgM heavy chain variable regions were prepared from unmanipulated peripheral blood lymphocytes of two healthy people. Partial sequencing of 103 clones revealed VH gene family use and complete CDR3 and JH sequences. The libraries differed in the two subjects. In one person's cDNA the VH,5 family was overexpressed and the VH,3 family underexpressed relative to genomic complexity. In the second person's cDNA, VH,3 was most frequently expressed. In both libraries, JH4 was most frequent. VH segments of several clones were closely related to those in fetal repertoires. However, there was also evidence of mutation in many cDNAs. Three clones differed from the single nonpolymorphic VH,6 germline gene by 7–13 bases. Clones with several differences from VH,5 germline gene VH,251 were identified. CDR3 segments were highly diverse. JH portions of several CDR3's differed from germline JH sequences. 44% of the clones had D,1,3 genes related to the D,1,2 and D,3,2 families, most with differences from germline sequences. In 11 D,1,2-related sequences, several base substitutions could not be accounted for by polymorphism. Thus, circulating IgM-producing B cell populations include selected clones, some of which are encoded by variable region gene segments that have mutated from the germline form. (J. Clin. Investig. 1992. 89:1331–1343.) Key words: antibody • B cell • cDNA library • diversity • repertoire

Introduction

The extensive diversity of antibody variable regions is due in large measure to the division of germline coding regions into segments, e.g., the VH, D, and JH segments which together encode the heavy chain variable region (1, 2). Random combinations of the V gene segments give the immune system a vast potential repertoire. In the mouse, for example, the potential repertoire exceeds 1010, and perhaps 1015, different antigen binding sites (3). But because there are only 108 B cells in a mouse, only certain elements of the potential repertoire are represented at any given time in the actual repertoire of the animal. Our understanding of how B cells use the tremendous capacity of the potential repertoire to generate the actual repertoire is limited.

Results of previous studies suggest that the actual, or expressed, immunoglobulin repertoire is not simply a random representation of the germline V gene potential. Nonrandom VH gene utilization is especially marked in the early stages of fetal development in both mice and humans (4–10), in malignant B cells (11, 12), in CD5+ B cells (13, 14), and in autoantibody-forming B cells (15, 16). However, the lack of information about the B cell repertoire in normal adults makes it difficult to assess the significance of the restricted use of immunoglobulin V genes during development and in disease. It is not known, for example, whether the preferential expression of VH,5 and VH,6 gene families early in ontogeny (4) is a peculiarity of fetal B cells, or whether the B cell repertoire of normal adults can also manifest such a bias.

Until recently, investigations of the human B cell repertoire were, for technical reasons, confined to EBV-transformed B cell clones, neoplastic B cells, and a relatively small number of hybridoma-produced monoclonal antibodies (17, 18). In situ hybridization with VH gene probes can greatly increase the number of B cells that can be surveyed (19, 20), and polymerase chain reaction (PCR)-based analyses have increased the number even more (21–26). Nevertheless, all these methods introduce their own bias. For example, neither the mitogen-stimulated B cells used for most in situ hybridization studies nor EBV-transformed B cells are representative of the entire population (27). cDNA amplification by PCR has allowed analysis of CDR3 sequences of human immunoglobulin cDNA populations (21, 24, 26), and it has been used in mice to estimate the frequency of rearrangement or expression of members of a given gene family (23, 28). But the lack of universal primers that would enable unbiased amplification of all V gene families has limited the scope of the PCR technique for studies of the expressed immunoglobulin repertoire.

We have recently described a sensitive method for amplifying the variable regions of immunoglobulin cDNAs of all VH families in a diverse mixture of B cells (29). The cDNA is amplified without using primers from variable region sequences, thus avoiding technical bias in the selection of amplified cDNA populations. The representative sampling allowed by the method permits analysis of immunoglobulin genes expressed by unmanipulated B cells, and gives a "snapshot" of the actual immunoglobulin repertoire at a given time. We report here an analysis of 103 unique clones from IgM libraries obtained by this method from two normal healthy adults. The clones were from cDNA libraries prepared from peripheral blood lymphocytes that were not stimulated in vitro, and whose only manipulation was centrifugation through Ficoll-Hypaque.

1. Abbreviation used in paper: PCR, polymerase chain reaction.
Methods

Preparation of cDNA Libraries from human peripheral blood lymphocytes. cDNA libraries were prepared from peripheral blood lymphocytes as described previously (29). Lymphocytes were centrifuged through a Ficoll-Hypaque medium and washed with PBS; they were not further manipulated before preparation of RNA. Double-stranded cDNA was synthesized from total cellular RNA according to the method of Gubler and Hoffman (30) and blunt-ended with T4 DNA polymerase. The primer for cDNA synthesis was complementary to a sequence within the Cδ1 region (29). Two steps of PCR amplification were performed, as described previously (29). The first step was primed by oligonucleotide linkers attached to the ends of the double-stranded (ds) cDNA. The products were ligated into M13mp19 RF DNA. A second amplification used a downstream-nested Cδ primer and an upstream primer within the M13 vector DNA. The second PCR products were again ligated to M13 RF DNA. This ligation mixture was transformed into DH5α bacteria to form the cDNA library for screening.

Analysis of the libraries. Libraries were screened for hybridization with a degenerate human Jh gene oligonucleotide probe and Vh family-specific oligonucleotide probes (17). M13 plaques were lifted onto GeneScreen membranes, which were then prehybridized, hybridized, and washed as described (31). Inserts in M13 phage were sequenced by chain termination with dideoxynucleotide triphosphates and Sequenase (US Biochemical Co., Cleveland, OH). For full V region analysis, sequencing was performed with two or three different primers, giving large overlaps that verified sequencing accuracy. Sequences were compared to those in the human Genbank database with the FASTA program of the GCG software package. The BESTFIT, LINEUP, and TRANSLATE programs were used for further sequence analysis.

Results

Amplified IgM cDNA libraries were prepared from RNA of unstimulated peripheral blood lymphocytes from two healthy adult donors (35 and 36 yr old). In both libraries, > 85% of the plaques hybridized with a degenerate Jh oligonucleotide probe. Sequencing of randomly picked Jh-positive clones began from the 5'-end of the Cδ region and continued through the Jh, CDR3 and at least the FR3. Complete V region sequences were obtained in selected cases. The sequence data allowed assignment of Vh families and full analysis of Dh and Jh gene segments. All clones discussed below had unique CDR3 sequences. All but four of the sequences corresponded to functional rearrangements with open reading frames through the Vh, CDR3, and Jh segments. A total of 103 clones from the two libraries were examined.

Use of Vh gene families in the normal adult repertoire. The 54 randomly picked clones of the first normal subject (Aα) included more Vh1 than Vh3 family genes—28% vs. 24% (Table 1). This result was surprising because the Vh3 family gene has the greatest genomic complexity and was the most frequent family detected in studies of expressed Vh genes from 104- and 130-d fetal liver cells (5, 6), in adult peripheral B cells examined by in situ hybridization (19, 20), and in EBV-transformed B cells (17, 33). The higher frequency of Vh1 than Vh3 family genes in the Aα cDNA library was confirmed by hybridization to plaque lifts with FR3 specific oligonucleotide probes; with this assay, 35% and 25% of 400 Jh-positive clones were members of the Vh1 and Vh3 families, respectively.

Another notable feature of the Aα library was that the two-member Vh5 family was highly represented (Table 1), occurring in 10 (19%) of the 54 sequenced clones. The high representation of this small family was confirmed in two different IgM libraries prepared from the same RNA sample. Results with the two Aα preparations are combined in Table 1. 3 of the 54 Aα clones were related to the single germline Vh6 family. The distribution of Vh family gene usage in the Aα library was at the borderline of being significantly different from that expected from the genomic complexity of the families (32) (P < 0.05).

In contrast with the Aα library, the Tα library was a closer reflection of the genomic complexity of Vh gene families; Vh3 members were most frequent, and the Vh5 family was not prominent (Table 1). Statistically, the distribution of Vh gene usage in the Tα library was not different than expected from the genomic complexity of the gene families (P > 0.05). Only 4 of the 103 IgM sequences in both cDNA libraries could be assigned to the Vh2 family. The frequency of expression of genes of the Vh2 family (~4%), which is estimated to contain five genes (34), was confirmed by plaque hybridization with a Vh2-specific FR3 oligonucleotide probe (17) (not shown). Our result is consistent with previous observations (19).

A distinct subgroup or a possible new Vh gene family. The Vh segments of three clones (Aα4.1, Aα92.1, and Aα2.2) differed substantially from any known member of the Vh1 to Vh6 families. These three clones were similar to each other in the Vh segment but each used a different D gene, so they were distinct clones. Their Vh region sequences had 78% overall identity with a known Vh1 gene, 20P3 (5), which was the most closely related gene among reported members of the 6 Vh gene families. Their FR1 and FR2 sequences were, in fact, 93% identical to highly conserved Vh1 gene sequences. However, they had only 67% identity with any known Vh1 sequence in CDR2 and FR3 (Fig. 1). These three Vh sequences had closest overall identity (96%) with that of a previously described autoantibody with dual rheumatoid factor and anti-DNA activity, Ab47, which was considered to be a subgroup of the Vh1 family (18) (Fig. 1). PCR amplification of nonlymphoid genomic DNA was used to test whether related genes were present in the germ-line or whether these novel sequences may have arisen from a somatic process such as gene conversion (35). A sequence related to the three new clones was indeed found in the nonlymphoid genomic DNA of the donor for the Aα library (data not shown). One primer for this PCR was in the unique region of the FR3 and the other was in the Vh1-like FR1. This combination of primers amplified a product of appropriate size from genomic DNA. By contrast, a control combination of the

<table>
<thead>
<tr>
<th>Vh gene family</th>
<th>Library</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aα</td>
<td>n = 54</td>
<td>Tα</td>
<td>n = 49</td>
</tr>
<tr>
<td>1</td>
<td>28</td>
<td>25</td>
<td>33</td>
<td>3, 4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>4</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>49</td>
<td>40</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>17</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>6</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>t</td>
<td>5</td>
<td>0</td>
<td>7</td>
<td>10</td>
</tr>
</tbody>
</table>

* From Berman et al. (32).

Vh genes with < 78% identity to members of Vh1 to Vh6 families.
unique FR3 primer with a V_H3 FR1 primer did not yield an amplification product.

Expressed V_H genes in circulating B cells have mutations in CDR1 and CDR2. For several clones in which partial sequences were very similar to those of known germline V_H genes, sequencing was extended at least through the CDR1. Three clones from the A(H)H libraries (Au34.2, Au46.2, and Au51.1) differed from the single, highly conserved germline V_H6 gene by 7, 13, and 7 bases (Fig. 2 a). To ensure that these variations were indeed mutations, the germline V_H6 gene of the A(H) donor was cloned and sequenced. It was identical to the published sequence of germline V_H6. In Au34.2, five of the seven V_H base substitutions were in CDR2 and 2 were in FR2. The J_H segment had one difference from germline J_H5; this change was in the 5'-end, which forms part of CDR3. The D_H segment could not be assigned to a known germline gene.

In Au46.2, seven differences from V_H6 were clustered within an 11-base segment in FR3.2. The other differences were scattered, with two in FR and four in CDR sequences. The CDR3-encoding portion of its J_H gene segment differed by one base from the germline J_H4 sequence. The heavy chain variable region codons of clone Au46.2, therefore, contained 12 base substitutions from germline V_H and J_H gene segments; five of those differences were in CDRs. The D segment of this clone differed by three bases from a 17-base portion of a germline D_H gene. The third V_H6-related clone, Au51.1, differed by seven bases from the germline V_H6 sequence. One difference was in CDR1, one was in CDR2, and five were in framework regions. The CDR3-encoding region of the J_H segment differed by one base from a germline J_H5 sequence. The D gene segment of Au51.1 could not be assigned to a known germline D_H sequence.

Five clones were very closely related to either V_H251 or V_H32, the two functional germline members of the small and minimally polymorphic V_H5 family (Fig. 2 b and c). Two clones from the A(H) library (Au59.1 and Au99.1) differed by 15 (Au59.1) and 5 (Au99.1) positions from V_H251. The substitutions tended to occur in the hypervariable regions; 10 of the 20 substitutions in these two clones were in either the CDR1 or CDR2. The D segment of Au59.1 could not be assigned, but that of Au99.1 had a six-base sequence identical to part of D_H5. Clone Au2.1 in the A(H) library differed at one position (in CDR2) from V_H32, had an unassignable D segment, and had a J_H segment differing by one base in the CDR3-encoding portion from J_H1 (Fig. 2 c).

There were fewer differences from V_H5 germline genes among the T_H library clones. One clone in this library, T_H16, differed by a single base from V_H251. Its D segment differed by two bases from a 15-base portion of the D_H family sequence. It had only one "N" base, at the D_H junction, and it had an unmutated J_H2 gene. A second T_H clone, T_H0, differed from V_H251 at four positions (one in CDR1). Its CDR3 had an 11-base sequence identical to a portion of D_H9, and its J_H6 sequence had one base change, which was in the CDR3-encoding portion of the gene.

cDNA of gene V_H26 (also termed V_H18/2 and 30P1 [36]) was represented in two clones selected at random for sequenc-
Figure 2. The heavy chain variable region cDNA sequences of clones in the 
$\upalpha_2$ and $\tau_2$ libraries containing $\upgamma_6$ gene fragments related to (a) the single 
germline $\upgamma_6$ gene, (b, opposite page) the germline $\upgamma_5$ family gene 
$\upgamma_5$, and (c) the germline $\upgamma_5$ family gene $\upgamma_5$. 1334 Huang et al.
Biased use and diversity of the J_{µ4} gene. The J_{µ4} gene was overrepresented in the cDNA libraries of both subjects. It occurred in 48% of all sequenced clones (Fig. 3). Use of other J_{µ} genes differed in the two subjects. For example, J_{µ5} occurred in 24% of the A_{µ} sequences and 13% of the T_{µ} sequences. The J_{µ1} and J_{µ2} families were present infrequently, an observation also made by others (19, 21).

Considerable diversity was found in the 5' end of the 49 sequenced J_{µ4} gene segments (Fig. 4). This region of the gene contributes to the CDR3 of the heavy chain. The diversity arose both from variation in the point of J_{µ4} joining to the rest of the CDR3 and from nucleotide differences from the germ-line sequence. Apart from a previously recognized polymorphic G/A variation (37), 15/28 A_{µ} clones and 11/21 T_{µ} clones contained substitutions in J_{µ4}, ranging from one to three in number, that cannot be accounted for by polymorphism. Most (81%) of the substitutions were clustered in the CDR3-coding region of the gene and most led to amino acid substitutions (Fig. 4 b).

Heterogeneity of D_{µ} regions. The D_{µ} segment sequences in both libraries were highly diverse (Figs. 5–7). Parts of some CDR3 sequences, ranging in length from 6 to 26 bases, were identical to portions of known germline D_{µ} gene segments; such germline sequences were found in 10/54 A_{µ} clones (18%) and in 16/49 T_{µ} clones (30%). The D_{H} region sequences of 12 other A_{µ} clones (21%) and 9 other T_{µ} clones (19%), from 15 to 29 bases long, had only one or two differences from germline D_{H} genes. The remaining 56 clones had more substantial differ-

Figure 2 (Continued)
ences from known germline $D_H$ sequences (Fig. 5 a). 21 (Fig. 6) were not assigned to known $D_H$ genes because they had either no identifiable sequence identity or, in some cases, because they had $<$ 75\% identity with a known $D_H$ gene.

Direct comparisons between observed and germline sequences were possible with the $D_{\text{LR}}$ and $D_{\text{XP}}$ families, for which the expected germline members are known (38, 39), and with the single $D_{\text{Q2}}$ gene (40). The majority of the expressed members of these families in both $C_{\mu}$ libraries contained differences, ranging from one to five bases, from germline sequences (Fig. 5). All assignable clones used only part of the germline $D_H$ gene segments, and $N$ insertions were observed in all of them (Figs. 5 and 7). The average length of $N$ at the $V_H$-$J_H$ junctions was 5.7 and at the $D_{\text{H}}$-$J_H$ junctions was 4.7 bases. Among clones with long $N$ regions, eight CDR3 sequences might be accounted for by $D$-$D$ or $D$-$D$ fusion (Fig. 7).

$D_H$ gene usage was not random. In the combined libraries, the $D_{\text{LR}}$ and $D_{\text{XP}}$ gene families were used with high frequency; these two families accounted for 54\% of the assignable clones (44\% of all clones). Sequences related to the $D_{\text{LR}}$2 gene alone were present in 11 clones; sequences related to $D_{\text{K4}}$ and $D_{\text{K1}}$ also occurred at high frequency (Fig. 5). The $D_{\text{Q2}}$ gene segment, which is overrepresented in human fetal liver (5), was present only once in the $A_{\mu}$ library and twice in the $T_{\mu}$ library.

Discussion

The sampling procedure. We have used a sensitive cDNA/PCR cloning method to examine usage of Ig heavy chain variable region genes in peripheral B cells of two normal adult donors. The procedure uses no variable region primers and, therefore, does not itself bias the V gene sampling (29). Moreover, since the lymphocytes were not stimulated in vitro, the results provide an insight into the V gene repertoire of circulating B cells in their native state at the time blood is drawn.

We do not know whether all B cells synthesize enough mRNA for a cell to be scored in this analysis. In humans, many of the circulating human B cells appear to be resting cells. Only 0.1–1\% of peripheral blood mononuclear cells synthesize mRNA at levels that can be detected by in situ hybridization (19, 41). PCR has a high sensitivity and probably samples a larger population than is detected by in situ hybridization.

The inherent error in PCR-based sequencing. The total number of nucleotides in the fully sequenced $J_H$, $V_H$, and $V_P$ genes was 4,065, among which there were 96 differences from germline sequences, for a rate of $\sim$ 24 bases per 1,000 ($4 \times 10^{-4}$ per nucleotide incorporated in the two PCR steps of 30 cycles). That frequency of base substitutions is much higher than the error rate of the PCR technique, which is $\sim$ 5 $\times 10^{-5}$ per nucleotide incorporated, both in the reported experience of others (25, 42, 43) and in our own experience. For example, several clones in different libraries from one individual that we studied had identical CDR3 sequences and were probably multiple copies of a single cDNA. The substitution frequency among those sequences was $\sim$ 1/300 bases ($2 \times 10^{-5}$ per nucleotide incorporated), a level at which PCR error could not be distinguished from clonal divergence.

$V_H$ gene family usage. Previous studies of $V_H$ gene family usage by circulating B cells from human adults, carried out by in situ hybridization with $V_H$ family-specific oligonucleotide probes, have drawn different conclusions, perhaps because of variations in technique and in sampling procedures. In the experiments of Guigou et al. (19), there were differences among individuals, but an average pattern of $V_H$ gene family expression by unstimulated cells could be defined. $V_H$3 family genes were the most frequently expressed, and $V_H$5 family usage correlated roughly with their genomic complexity. Zouali and Theze (20) averaged results of protein A–stimulated B cells from eight adults. They observed that the $V_H$ gene families were not represented in a random way; the $V_{H1}$ family was under-represented, whereas the $V_{H3}$ family was overrepresented relative to genomic complexity.

The results of our study emphasize that there are indeed differences among single samples from different normal individuals, as Guigou et al. (19) found. In the cDNA library of one subject ($A_{\mu}$), a nonrandom representation was seen, with disproportionate representation of $V_{H5}$, and fewer than expected $V_{H3}$ gene family members. The $A_{\mu}$ library, which also contained 3 $V_H$6 members, resembles, in its overall composition, a fetal $C_{\mu}$ library (5, 6). Further study will be required to determine whether that pattern is stable for the donor of $A_{\mu}$ lymphocytes. It is possible that the nonrandom $V_H$ distribution in this library reflects an unknown, recent immunizing stimulus. $V_H$ gene usage in the cDNA library from the $T_{\mu}$ donor, by contrast, more closely paralleled the genomic complexity of the families; however, that single library does not exclude a continuously changing pattern of $V_H$ gene usage.

The three novel sequences, with FR1 and FR2 sequences characteristic of $V_{H1}$ genes and unique CDR2 and FR3 sequences, along with Ab47 (18), may represent a distinct subgroup of the $V_{H1}$ family, as suggested by Sanz et al. (18), or a new $V_{H}$ gene family. A closely related combination of sequences exists in the germline DNA, as shown by PCR amplification. This subset of genes may have arisen from a gene conversion or recombination in evolutionary time rather than as a somatic event. The clones in this group have 78\% overall sequence identity with the mouse immunoglobulin gene $V_{H9}$.

The normal $V_H$ gene repertoire contains genes used by fetuses and for autoantibodies. Table II summarizes the findings in seven clones (sequenced at least from CDR1 to the end of FR3) with 97\% or more identity to members of the set of $V_H$.
genes, such as 58P2, that has been a feature of the immunoglobulin V gene repertoire of fetal B cells. Five of the seven \( V_H \) genes represented in these clones are known to be used to form autoantibodies such as rheumatoid factor, cold agglutinins, and anti-DNA and anti-cardiolipin antibodies (\( V_H \) 251, 21/28, FL2-2, \( V_H \) 6, and \( V_H \) 32 (37). To those we can add the three genes closely related to Ab47, a rheumatoid factor/anti-DNA antibody. VH genes with one and three base differences from the germline \( V_H \) 26, used in anti-DNA autoantibodies, were also identified in the \( A_M \) library.

B cells capable of forming such autoantibodies are highly represented among human–human hybridomas (36), EBV-transformed cells (13, 14, 44), and B-cell malignancies (11, 12). Many of them, like those listed in Table II, use \( V_H \) genes that are also expressed by fetal B cells, with few or no mutations from the germline \( V_H \), \( D_H \), or \( J_H \) components. These results are compatible with the conclusion, drawn from studies of EBV-transformed B cells, that cDNAs associated with IgM autoantibodies are highly represented in the normal B cell repertoire (44). Some such immunoglobulins, encoded by \( V_H \) genes with few mutations, may bind to both autoantigens and foreign antigens such as bacterial polysaccharides (45, 46).

Evidence that circulating B cells have undergone selection. Several lines of evidence, when taken together, strongly suggest that many IgM* B cells in the circulating blood are not naive, but instead have undergone selection and clonal expansion. Four aspects of our results support that conclusion: overrepresentation of V region gene families, or of individual V region gene segments; somatic mutation of V genes; the high frequency of replacement substitutions compared to silent muta-
Figure 5. (a and b) Relationship of CDR3 base sequences of clones in the Aμ and Tα cDNA libraries to known germline DH genes. Numerically annotated clones: 1, Tα56 may be assigned equally well to DK4, DK1, or DK5; 2, clone Aα29.2 may be assigned equally well to DM1 or DM5; 3, clone Aα103.1 may be assigned equally well to DM1, DM2, or DM5; 4, clone Aμ47.2/r indicates that the sequence is reversed. Position 20 in DXP, I in b has been reported as C (39) or A (38).

...
tion of selection in the libraries we tested is the over-representation of the D_{LR2} genes, present in 6 of the 54 A\mu clones (11%) and 5 of the 49 T\mu clones (10%). The assignment of D_{LR2} genes is possible because all five members of the D_{LR} family are known (38, 39).

Gu et al. (23) have analyzed members of a large V_{H} gene family (J558) expressed by B cells from three unimmunized CB.20 mice. In contrast to populations of pre-B cells, which expressed the ~100 J558 genes randomly, populations of mature surface IgM\(^+\) splenic B cells were found to express preferen-
Somatic mutation of V genes is a cardinal manifestation of clonal selection of B cells (49–56). In the case of the V₅ gene segments we analyzed, definitive evidence of somatic mutation was found in the case of the 3 V₅ clones (Fig. 2 a). These three clones differed from Αμ’s own germline sequence by 7 (Αμ34.2), 13 (Αμ46.2), and 7 bases (Αμ51.1). The Αμ34.2 and Αμ51.1 clones also had evidence of somatic mutation in their CDR3 sequences.

The sequences of the two functional V₅ genes, V₅251 and V₅32, are remarkably consistent in the human germline (37). The nucleotide sequences of all five examples of V₅ family genes in both libraries differed from V₅251 and V₅32 by 1–15 bases (Fig. 2 b). Given the conservation of V₅251 and V₅32 in the germline, it is highly likely that the variations we observed can be attributed to somatic mutation.

Adding to the evidence from the V₅ sequences for somatic mutation is the finding that the CDR3 portions of many J₄ sequences in the Αμ library differ from germline genes in a way that cannot be explained by polymorphism (Fig. 4). Whereas polymorphic sites, such as the G → A substitution in J₄, are identical in all clones from the same subject (Fig. 4), somatic mutations of V genes are typically clone-specific. Furthermore, the variations from the J₄ germline sequence were not random, but clustered at the 5’ end of the gene; of the 42 bases that differ from the germline (not counting the polymorphic G → A substitution), 81% occur at the 5’ end of the gene, in the region that contributes to the CDR3.

A mechanism other than polymorphism is also required to account for the several D₄ sequences that are closely related to D₄LR and D₄XP (Fig. 5). Even if subject Αμ had polymorphic

---

Figure 6. CDR3 base sequences that could not be assigned to germline D₄ gene segments. These cDNAs do not contain a portion with more than 75% identity with known germline D₄ genes.

---

Figure 7. CDR3 base sequences that may be accounted for by D-D or D-D₄ fusions. These include examples in which one of the fused segments is reversed (D₄LR3/r, D₄XP1/r, and D₄XP2/r).
differences from the published D_{H}2 germline sequence in both alleles—an unlikely proposition because a sequence identical to D_{H}3 was found in one clone—that would account for only two of the six A_{H} genes related to D_{H}.2.

Apart from the evidence compiled from the nucleotide sequences themselves, a cogent argument for the occurrence of somatic mutation in circulating IgM\(^+\) B cells is that the majority of the base substitutions we observed were not silent but resulted in a changed amino acid sequence. In studies of V gene sequences of antibodies produced during the secondary response of the mouse to several different classes of antigens, mutations causing amino acid substitutions ("replacement mutations") were found to exceed silent mutations by far, and were characteristically located in the CDRs (50, 52, 53, 57–64). That pattern is striking in the J_{H}4 gene of the A_{H} library. Of the 42 base differences from the germline sequence found among 49 J_{H} genes (not counting the polymorphic G \rightarrow A substitution), 5 were silent and 37 were replacement variants; of those 37 replacement variants, 31 (84%) occurred in the 5’ CDR3-coding region of the J_{H} genes (Fig. 4 b).

A similar, but less striking picture emerges from analysis of the seven V_{H}5 and V_{H}6 genes. The total number of base variations from germline sequences was 54; of those, 41 were replacement and 13 were silent (of the latter, 10 occurred in the three V_{H}6 genes). And of the 22 amino acid replacements in the V_{H}5 genes, 50% occurred in either CDR1 or CDR2. Although framework mutations can affect antibody binding properties (65), mutations in the CDRs, which are largely responsible for the ligand-binding surface of the immunoglobulin molecule, are the principal molecular signs of clonal selection. It is thus highly likely that the mutations we observed are a reflection of the selective effect of a ligand on the circulating B cell population.

These findings, when viewed as a whole, suggest that ligand-selected IgM\(^+\) B cells not only circulate in the blood of normal adults but they may comprise a substantial fraction of the B lymphocytes in human blood. They could correspond to long-lived memory cells that have been rescued from programmed cell death by contact with antigen (66). Indeed, it is likely that the B lymphocyte dies soon after it completes its differentiation, as a result of apoptosis, unless it undergoes selection by antigen (67). Our finding of somatic mutation in circulating B cells is of interest because B cells engaged in responses to specific antigens are generally thought to reside in the germinal centers of the spleen and lymph nodes (68). However, it was recently shown that during the secondary immune response of the mouse to horseradish peroxidase, B cells have been found to leave the germinal centers, enter the circulation, and seed the bone marrow where they mature into antibody-producing plasma cells (69). Presumably, those cells underwent at least the initial stages of antigen selection, although the molecular evidence to support that conclusion is presently lacking.

Another noteworthy aspect of our results is that V region genes in a C\(_\mu\) library showed evidence of somatic mutation. In the experiments of Gu et al. (23), no somatic mutations were found among 44 complete V region sequences in C\(_\mu\) libraries from young unimmunized mice; Manser and Gefter (70) also found no somatic mutations in naive mice. Even so, it is known that IgM antibodies can be encoded by mutated V region genes (71), and that somatic mutation can be detected very early in the immune response (63), independently of heavy chain class switching (72).

The molecular signs of clonal selection in circulating B cells suggest that the selective ligand was encountered after the pre-B cell stage of differentiation, when the maturing B cell has rearranged its V genes (73) and expressed at least a surface heavy chain. In the steady state, such B cells could represent long-lived circulating memory cells that provide an early defense against microbial reinvansion; or, in some instances they may represent selection of the repertoire by idiotypes or anti-idiotypes (74). In either case, any analysis of V gene repertoires in disease will have to take into account the variations in composition and structure of the normal repertoire.

Acknowledgments

This research was supported by grants AI-19794 and AI-26450 from the National Institutes of Health. A. K. Stewart is a fellow of the Medical Research Council of Canada.

References


Table II. cDNAs Closely Related to Germline V\(_H\) Genes

<table>
<thead>
<tr>
<th>Clone No.</th>
<th>Percent identity</th>
<th>Related gene</th>
<th>V(_H) family</th>
<th>No. of bases sequenced</th>
</tr>
</thead>
<tbody>
<tr>
<td>T(_{H}16)</td>
<td>99.6</td>
<td>VH251</td>
<td>5</td>
<td>292</td>
</tr>
<tr>
<td>T(_{H}59)</td>
<td>99.1</td>
<td>M60</td>
<td>2</td>
<td>222</td>
</tr>
<tr>
<td>T(_{H}73)</td>
<td>99.1</td>
<td>S8P2</td>
<td>4</td>
<td>230</td>
</tr>
<tr>
<td>T(_{H}74)*</td>
<td>99.1</td>
<td>21/28</td>
<td>1</td>
<td>229</td>
</tr>
<tr>
<td>T(_{H}49)</td>
<td>97.3</td>
<td>FL2-2</td>
<td>1</td>
<td>218</td>
</tr>
<tr>
<td>A(_{H}51.1)</td>
<td>97.8</td>
<td>VH6</td>
<td>6</td>
<td>315</td>
</tr>
<tr>
<td>A(_{H}21.1)</td>
<td>99.7</td>
<td>VH32</td>
<td>5</td>
<td>303</td>
</tr>
</tbody>
</table>

*Clone T\(_{H}74\) has the same N and D\(_{H}\) sequence as the anti-DNA autoantibody 21/28 (see reference 76).


63. Clarke, S. K., H. Huppi, D. Ruezinsky, L. Staudt, W. Gerhard, and M. 1342

Huang et al.


