Selection of Ig \( \mu \) Heavy Chains by Complementarity-Determining Region 3 Length and Amino Acid Composition

Denise A. Martin, Harald Bradl, Tara J. Collins, Edith Roth, Hans-Martin Jäck and Gillian E. Wu

*J Immunol* 2003; 171:4663-4671; doi: 10.4049/jimmunol.171.9.4663

http://www.jimmunol.org/content/171/9/4663

Why *The JI*?

- **Rapid Reviews! 30 days** from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Speedy Publication!** 4 weeks from acceptance to publication

*average

References This article cites 57 articles, 24 of which you can access for free at: http://www.jimmunol.org/content/171/9/4663.full#ref-list-1

Subscription Information about subscribing to *The Journal of Immunology* is online at: http://jimmunol.org/subscription

Permissions Submit copyright permission requests at: http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts Receive free email-alerts when new articles cite this article. Sign up at: http://jimmunol.org/alerts
Selection of Ig μ Heavy Chains by Complementarity-Determining Region 3 Length and Amino Acid Composition

Denise A. Martin,* Harald Bradl,§ Tara J. Collins,* Edith Roth,§ Hans-Martin Jäck,‡ and Gillian E. Wu*‡

Although it is generally accepted that Ig heavy chains (HC) are selected at the pre-B cell receptor (pre-BCR) checkpoint, the characteristics of a functional HC and the role of pre-BCR assembly in their selection have remained elusive. We determined the characteristics of HCs that successfully passed the pre-BCR checkpoint by examining transcripts harboring Vμ81X and Jμ4 gene segments from Jμ4+/− and Δ5−/− mice. Vμ81X-Jμ4-HC transcripts isolated from cells before or in the absence of pre-BCR assembly had no distinguishing complementarity-determining region 3 traits. In contrast, transcripts isolated subsequent to passage through the pre-BCR checkpoint had distinctive complementarity-determining regions 3 of nine amino acids in length (49%) and a histidine at position 1 (73%). Hence, our data define specific structural requirements for a functional HC, which is instrumental in shaping the diverse B cell repertoire. The Journal of Immunology, 2003, 171: 4663–4671.

B lymphocyte development follows a complex series of events and decisions in which genetics, environmental, and stochastic processes all play a role in the outcome. Integral to the survival of a developing B cell is the productive rearrangement of the VH, DI, and JH, and VL and Jλ gene segments of Ig heavy chain (HC)1 and light chain (LC) loci, respectively (1, 2). The joining of VH, DI, and Jλ gene segments generates the variable domain of a HC with its complementarity-determining region 3 (CDR3) encompassing the sites of the joins. Formation of a μHC demarcates a crucial checkpoint in B cell development. At this checkpoint, called the pre-B cell receptor (pre-BCR) stage, the μHC and the surrogate LC (SLC) components, VpreB and Δ5, assemble to form the pre-BCR. The μHC, if capable of pairing with the SLC, appears on the surface of the B cell, where much evidence implicates the successful formation of a pre-BCR in signaling events and the proliferative expansion associated with B cell progression (3–5). The ability or inability of a particular μHC to pair with SLC as well as the competence of that interaction determines whether a μHC protein and ultimately a B cell are selected to proliferate and survive/persist into the mature pool (4, 6–8).

The importance of μHC deposition and the SLC in B cell progression has been demonstrated in various murine models with targeted deletions. Mice with a deletion in the transmembrane exon of the μHC are unable to produce cells that progress beyond the pro-B cell stage of development (9). Mice that cannot make a complete SLC due to the absence of Δ5 protein also have a block in generating mature B cells, with most cells arrested at the pro-B stage (10).

Evidence is accumulating that the μHC is assessed for its ability to functionally associate with a LC before LC rearrangement, with the SLC functioning as the assessor (8, 11, 12). The presence of the SLC before LC rearrangement as well as its structural similarity to a conventional Ig LC make it an ideal candidate for this function. If a cell has passed the fitness test overseen by the pre-BCR, placement of a successful HC-LC pair in the plasma membrane provides the B cell with a receptor necessary for survival and persistence in the periphery (13). The features in the μHC protein important for successful LC pairing, however, remain to be determined.

Some evidence suggests that the V-D-J joint that encodes CDR3 can impact HC folding and the ability to form HC-LC heterodimers (6). In particular, rearrangements using the most D-proximal V gene segment, Vμ81X, have been found to generate HC proteins largely incapable of pairing with SLC (8). Evidence that impaired pairing has in vivo consequences comes from the finding that Vμ81X is under-represented in productive rearrangements (7, 12, 14, 15). In addition, few mature cells have been found with Vμ81X-containing BCRs, despite its presence in >50% of the initial HC rearrangements in the bone marrow (11, 12, 14, 16, 17).

A diverse repertoire is paramount for immunocompetency. Thus, it is crucial to understand the mechanisms restricting selection while ensuring an effective, functional repertoire. We have used a system that allows us to track Vμ81X rearrangements through stages of B cell development to identify the structure and type of rearrangements that persist in the periphery. By breeding, we have generated a mouse that carries one wild-type Ig HC allele

---

1 Abbreviations used in this paper: HC, heavy chain of IgM; A-MuLV, Abelson murine leukemia virus; BCR, B cell receptor; BM, bone marrow; pl, isoelectric point; pre-BCR, pre-B cell receptor; SLC, surrogate light chain; Vμ81X-μHC, μHC using a Vμ81X-D-J rearrangement.

*Department of Immunology, University of Toronto, and Ontario Cancer Institute, Toronto, Canada; 1Division of Molecular Immunology, Department of Internal Medicine III, Nikolaus Fiebiger Center, University of Erlangen-Nürnberg, Erlangen, Germany; and 2Facility of Pure and Applied Science, York University, Toronto, Canada; †Division of Molecular Immunology, Department of Internal Medicine, University of Toronto, and Ontario Cancer Institute, Toronto, Canada; ‡Division of Molecular Immunology, Department of Internal Medicine, University of Toronto, and Ontario Cancer Institute, Toronto, Canada; §Division of Molecular Immunology, Department of Internal Medicine, University of Toronto, and Ontario Cancer Institute, Toronto, Canada.
and one allele with a targeted deletion at the J₅₆ locus. This arrangement allows each B cell to make only one productive VDJ₄₅ rearrangement, ensuring that we are monitoring the gene structure of only the used HC in various B cell compartments. By isolating productive rearrangements using V₄₈X1 and J₄₅ gene segments and determining their potential to associate with SLC and be expressed on the cell surface, we were able to determine which specific variable region structures, differing only at the CDR3, render an HC successful.

These studies have revealed two characteristics common to successful HCs using V₄₈X1 and J₄₅ gene segments: a CDR3 length of nine amino acids and the presence of histidine in position 1. These traits have significant roles in determining which V₄₈X1 µHCS are selected into the peripheral pool. Other parameters, including isoelectric point (pl) and motifs, were not found to contribute substantially to HC selection. These data suggest a level of selection before µHC deposition in the plasma membrane, followed by further selection subsequent to cell surface expression. Our model implicates SLC in this paradigm, functioning to select protein by shape and fit before surface expression and signaling efficiency.

Materials and Methods

**Mice**

₅/₆⁻⁻ and A⁻/⁻ mice have been previously described (10, 18). J₅/₆⁻⁻ mice were generated by breeding C57BL/6 and J₅/₆ mice. J₅/₆⁻⁻ mice were identified by Southern blot analysis as previously described (18). Briefly, genomic DNA from mouse tails was digested with StuI, electrophoresed, and transferred to Hybond nylon membranes (Amersham Pharmacia Biotech, Little Chalfont, U.K.). Membranes were hybridized with a 475-bp EcoRI-StuI fragment of J₅, which reveals a band of 4.7 kb for the wild-type allele and of 3.9 kb for the mutant allele. Mice identified by Southern analysis as having one mutant allele and one wild-type allele were used for further analysis.

**Flow cytometry and cell sorting**

Single-cell suspensions were prepared from bone marrow (BM) and spleen of the appropriate mice and were stained using standard procedures. BM cells were stained with FITC-conjugated anti-B220 (mAb RA3-6B2; BD Pharmingen, San Diego, CA), PE-conjugated anti-CD43 (mAb 57; BD Pharmingen), and biotinylated anti-µ IgH (mAb 33-60). Spolenocytes were stained with FITC-conjugated anti-IgD (mAb 11-26c2a; BD Pharmingen), PE-conjugated anti-B220 (mAb RA3-6B2; BD Pharmingen), and biotinylated anti-µ IgH (mAb 33-60). Staining with biotinylated Abs was revealed by the secondary reagent streptavidin Quantum Red (Sigma-Aldrich, St. Louis, MO). All collected cells were B220⁻. To isolate B cell precursors, BM populations were defined by the cell surface markers CD43 and µ; collected cells were CD43⁻ µ⁻ or CD43⁻ µ⁺. To isolate newly emerging B cells, spleen populations were defined by cell surface expression of IgM and IgD; collected cells were IgM⁺ IgD⁻⁻. Cells were sorted using FACStar Plus (BD Biosciences, Mountain View, CA) and MoFlo (Cytomation, Fort Collins, CO) instrumentation. Abelson murine leukemia virus (A-MuLV)-transformed pre-B cell lines, before (membrane staining) or after (cytoplasmic staining) fixation and permeabilization with formaldehyde/Tween 20, were stained using standard staining protocols (6). Unconjugated Abs were detected with the appropriate fluorochrome-conjugated secondary Abs. Cells were washed, and fluorescence was analyzed with a FACS Calibur (BD Biosciences). Flow diagrams were obtained by analyzing the primary data with the CellQuest software program (BD Biosciences). Rat IgG1 mAbs directed against the pre-BCR (clone SI-156) (19) and the hamster anti-mouse A5 mAb, FS1 (20), have been previously described (21). FITC-conjugated as well as unconjugated affinity-purified goat Abs against mouse µHC were purchased from Southern Biotechnologies Associates (Birmingham, AL).

**RNA isolation, amplification, and sequencing**

Total RNA was extracted from sorted cells using Tri-Reagent (Molecular Research Center, Cincinnati, OH). RNA was reverse transcribed using Superscript II (Life Technologies, Gaithersburg, MD), and the cDNA was used for RT-PCR. The primary PCR globally amplified >80% of VH gene rearrangements; the secondary PCR specifically amplified transcripts harboring the V₄₈X1 gene segment. cDNA was amplified for the primary PCR using V₄₈ALL sense 5'-AGGTG(C/G)AA(C/A)(A/G)CT(G/A/C/G)GA(G/T/A)(G/G)GG-3' and J₄₅mH T antisense 5'-GAGGAGACGCTGACTGAGTTCTCTGTTG-3', followed by a secondary PCR amplification using either V81X145 sense 5'-TGTTGGACGGCACTTCAATGAT-3' or V81X148 reverse 5'-GAATTTCCCTTTACCAGTCAAGTCTGG-3' and the J₄₅mH T antisense primer. RT-PCR products were identified by resolution on a 1.0% agarose gel and fragments of expected length were ligated into TA cloning vectors (Invitrogen, Carlsbad, CA). Plasmid clones were sequenced with the T7 sequencing kit (Amersham Pharmacia Biotech) and the T7 CDR3 regions were identified as the amino acids between codon 94 and codon 102 (22). Codon 94 is the second amino acid residue following the conserved YYC motif coded by the VH gene segment (usually an arginine residue), and codon 102 is the conserved tryptophan encoded by the J₅₆ gene segment.

**Retroviral vector construction for expression of V₄₈X1 gene rearrangements**

The retroviral vector pELV81Cp (Fig. 1A) was generated from pELVC (a gift from F. Melchers, Basel, Switzerland) (4). pELVC contains the leader (L) exon/intron sequence of the S6 µ gene (23) and a functional VDJ₄₅ sequence fused to the cDNA sequence encoding the membrane form of µHC. The rearranged V₄₈X1 segments were isolated from a number of sources and inserted into pELV81. One set of V₄₈X1 segments originated from genomic V₄₈XD1 clones reported previously (24). TdT refers to clones isolated from the BM of an adult transgenic TdT mouse (25) and MTLF and MTBM refer to clones isolated from fetal liver and adult BM, respectively, of a µMT mouse (9). Genomic V₄₈XD1X1 segments were converted into cDNA-type sequences (the J₅ abutting the µH) by the following PCR cloning strategy. The forward primer was GAATTCCTTCCATGACATGTC (containing an EcoRI site that is present in all VH mRNAs; Fig. 1A), and the backward primer is GAACCTTTGACTCTCTGAGGAGACTGAGGTTCCTTG (containing HindIII, C, and J₅, and four sequences; Fig. 1A). The amplified products were TA cloned, and EcoRI/HindIII V₄₈X1 fragments and inserted in the corresponding restriction sites of pELV81 Cp. TdT₄-pELV81Cp, TdT₄-pELV81Cp, MTFL-pELV81Cp, MTTFL-pELV81Cp, MTTF-pELV81Cp, MTBM-pELV81Cp, and MTBM-pELV81Cp were constructed in this manner. Plasmid expression vectors encoding BC2-8V1X-Cp and 31AV-8X-Cp were gifts from J. Kearney (Birmingham, AL). F-V81X was cloned from the 18-81 subclone F, an A-MuLV pre-B cell line, and inserted into a conventional Ig expression vector (6). All VDJ₄₅ regions were verified by complete sequencing.

**Retroviral infection of the A-MuLV-transformed pre-B cell lines 38B9 and BINE4.8**

The ectopic packaging line GP+E (26), grown to 80–90% confluence in 25-cm² flasks, was first transfected with 5 µg of the respective retroviral plasmid vector with the calcium phosphate method using the Cal-Phos-Maximizer kit (Clontech, Palo Alto, CA) for 2 h at 37°C. After infection, the medium was removed, and the cells were washed with PBS and incubated in RPMI/10% heat-inactivated FCS at 37°C and 5% CO₂. After 24 h, the medium was replaced with 5 × 10⁶ pre-B cells (BINE 4.8 or 32B9) in 5 ml of RPMI/10% heat-inactivated FCS supplemented with 40 µg/ml of polybrene (Sigma-Aldrich). The culture was incubated for 24 h and then supplemented with 5 µg/ml puromycin. Single clones were generated from puromycin-resistant bulk cultures by the limiting dilution method and were analyzed for µHC production by flow cytometry.

**Cell lines and culture conditions**

The ectopic retroviral packaging line GP+E (26) and the A-MuLV-transformed mouse pre-B cell lines BINE 4.8, TK, TK₄ (27), and 38B9 (2, 4) were maintained at 37°C and 5% CO₂ in RPMI medium supplemented with 50 U/ml penicillin, 50 µg/ml streptomycin, 5% FCS, 1 mM sodium pyruvate, and 2 mM l-glutamine. BINE 4.8, TK, and 38B9 produce A5 and Vp6B proteins, but lack a µHC. TK₄ produces a fully assembled pre-BCR and serves as a positive control for µHC/SLC pairing.

**Immunoprecipitation and gel electrophoresis**

Metallo labeling was performed as previously described (8). Briefly, 5 × 10⁶ of retrovirus-infected TK, TK₄, or 38B9 cells in 1 ml of methionine-free labeling medium were incubated overnight with 50 µCi/ml of Tran³5 S label (ICN Biomedicals, Eschwege, Germany). Cells were lysed for 30 min on ice in 1 ml of lysis buffer. Cell lysates were incubated with either 5 µg of affinity-purified goat anti-mouse µHC Abs (Jackson ImmunoResearch Laboratories, West Grove, PA) or 5 µg of anti-human F(ab')₂ fragments (Jackson ImmunoResearch Laboratories) for 1 h at 4°C. Lysates were then precipitated with protein A-Sepharose (Pharmacia) and washed three times with 1 M NaCl-PBS (pH 8.0). The bound proteins were resolved on SDS-PAGE gels and analyzed by Coomassie blue staining. Gel slices were excised and digested with trypsin. The resulting peptides were analyzed by matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF-MS) using a Voyager DE-STR mass spectrometer (PerSeptive Biosystems, Framingham, MA).
The theoretical models were then assembled identically and displayed together. The pI of the modeled \( \text{V}_{H}81X\text{-HC} \) was calculated using Compute pI/Mw, a tool, which allows computation of the theoretical pI and Mw (http://www.expasy.ch/tools/pi_tool.html and (30, 31)).

Statistics

Statistics were performed by the Statistics Department at University of Toronto. Values were calculated using sampling distributions for counts and proportions. Significance was calculated under the null hypothesis of no difference in the proportion of occurrence of CDR3 lengths equal to 27 nt between the two populations or the null hypothesis of no difference in the proportion of occurrence of amino acid histidine at position 1 between the two populations.

Results

The elements that dictate the configuration of biological systems remain important issues in understanding the development of the mature immune system. In any developing system, identifiable characteristics of a population may reflect the outcome of stochastic or selective mechanisms. In this study we sought to identify specific characteristics of \( \mu \text{HC} \) from splenic B cells and determine whether these characteristics would yield insights into the structural requirements of Ig HC. These HCs exhibit representative traits that allowed the cell to passage through selection and persist in the spleen. We addressed these issues using a mouse with a single rearrangeable IgH allele that would allow us to compare HCs that passed the pre-BCR checkpoint and persisted in the periphery to those present before selection in the BM. These experiments were followed up with analysis of functional Ig rearrangements isolated from peripheral B cells of \( \text{A}^{\text{w}-/-} \) mice, to examine the properties of HCs that persist in the absence of SLC. We then used an A-MuLV-transformed pre-B cell line that lacked its own HC (but had SLC protein) to assess by flow cytometry and immunoprecipitation the ability of \( \mu \text{HCs} \) to associate with the SLC and appear on the cell surface. The availability of modeling programs allowed us then to determine the characteristics of a functional, i.e., pairing, \( \mu \text{HC} \) that passes the pre-BCR checkpoint and persists in the periphery.

\( \text{V}_{H}81X \) HC rearrangements in BM and spleen

Mice heterozygous for a deletion of the \( \text{I}_{\text{H}} \) gene segments are only able to generate one productively rearranged allele per cell. In all other respects examined, the B cell populations isolated from spleen and BM were comparable to B cell compartments found in wild-type animals (18). We examined the expressed \( \mu \)HC allele first in cells from whole BM and spleen as described in Materials and Methods (15). For each reaction, 2–4% of the product was cloned, sequenced, and evaluated. Using the guidelines of Kabat and Wu (22), translated sequences were anchored at the conserved YYC motif and ending at (but not including) the conserved W (tryptophan) encoded by the JH gene segment. From the YYC motif and ending at (but not including) the conserved W (tryptophan) encoded by the JH gene segment. The theoretical models were then assembled identically and displayed together. The pI of the modeled \( \text{V}_{H}81X\text{-HC} \) was calculated using Compute pI/Mw, a tool, which allows computation of the theoretical pI and Mw (http://www.expasy.ch/tools/pi_tool.html and (30, 31)).

Statistics

Statistics were performed by the Statistics Department at University of Toronto. Values were calculated using sampling distributions for counts and proportions. Significance was calculated under the null hypothesis of no difference in the proportion of occurrence of CDR3 lengths equal to 27 nt between the two populations or the null hypothesis of no difference in the proportion of occurrence of amino acid histidine at position 1 between the two populations.

Results

The elements that dictate the configuration of biological systems remain important issues in understanding the development of the mature immune system. In any developing system, identifiable characteristics of a population may reflect the outcome of stochastic or selective mechanisms. In this study we sought to identify specific characteristics of \( \mu \text{HC} \) from splenic B cells and determine whether these characteristics would yield insights into the structural requirements of Ig HC. These HCs exhibit representative traits that allowed the cell to passage through selection and persist in the spleen. We addressed these issues using a mouse with a single rearrangeable IgH allele that would allow us to compare HCs that passed the pre-BCR checkpoint and persisted in the periphery to those present before selection in the BM. These experiments were followed up with analysis of functional Ig rearrangements isolated from peripheral B cells of \( \text{A}^{\text{w}-/-} \) mice, to examine the properties of HCs that persist in the absence of SLC. We then used an A-MuLV-transformed pre-B cell line that lacked its own HC (but had SLC protein) to assess by flow cytometry and immunoprecipitation the ability of \( \mu \text{HCs} \) to associate with the SLC and appear on the cell surface. The availability of modeling programs allowed us then to determine the characteristics of a functional, i.e., pairing, \( \mu \text{HC} \) that passes the pre-BCR checkpoint and persists in the periphery.

\( \text{V}_{H}81X \) HC rearrangements in BM and spleen

Mice heterozygous for a deletion of the \( \text{I}_{\text{H}} \) gene segments are only able to generate one productively rearranged allele per cell. In all other respects examined, the B cell populations isolated from spleen and BM were comparable to B cell compartments found in wild-type animals (18). We examined the expressed \( \mu \text{HC} \) allele first in cells from whole BM and spleen as described in Materials and Methods (15). For each reaction, 2–4% of the product was cloned, sequenced, and evaluated. Using the guidelines of Kabat and Wu (22), translated sequences were anchored at the conserved YYC (tyrosine-tyrosine-cysteine) motif and ending at (but not including) the conserved W (tryptophan) encoded by the JH gene segment. The theoretical models were then assembled identically and displayed together. The pI of the modeled \( \text{V}_{H}81X\text{-HC} \) was calculated using Compute pI/Mw, a tool, which allows computation of the theoretical pI and Mw (http://www.expasy.ch/tools/pi_tool.html and (30, 31)).

Statistics

Statistics were performed by the Statistics Department at University of Toronto. Values were calculated using sampling distributions for counts and proportions. Significance was calculated under the null hypothesis of no difference in the proportion of occurrence of CDR3 lengths equal to 27 nt between the two populations or the null hypothesis of no difference in the proportion of occurrence of amino acid histidine at position 1 between the two populations.
necessary to isolate populations at specific stages. For the BM populations, we focused on the B220+/CD43− surface μ⁺ cells (pro-B cells), and the large B220−CD43− surface μ− cells, (transition, early pre-B cells) as cells before and during pre-BCR selection, respectively. For the spleen populations, we isolated the B220+IgMbright(IgDlow−/− cells (newly emerging, immature B cells) as cells after pre-BCR selection. Fig. 2A shows typical profiles of the presorted populations. Four independent experiments, with one or two mice per group, generated the results presented. For each group at least 2×10⁶ cells were sampled. Sorted populations were consistently >98% pure (Fig. 2B).

**CDR3 length in B cell developmental compartments**

The sorted population of cells from the BM before pre-BCR expression (B220+, CD43+/−, surface μ−) had μHC rearrangements exhibiting random Dμ gene segment usage and variable junctional sequences (n = 39; Fig. 3A). CDR3 lengths ranged from 5−15 aa in length: 5% of clones had regions 5 aa long, 2.5% of clones had regions 6 aa long, 9% were 7 aa long, 10% were 8 aa long, 23% were 9 and 10 aa long, 28% were 11 aa long, and ~8% were 12 and 13 aa long (Fig. 3A). Statistically, there was no significant bias toward any one particular CDR3 length. These data are consistent with our observations in total BM samples.

The splenic B220+IgMbright(IgDlow−/− B cell population represents cells that have survived selection in the BM and persisted in the periphery. In 37 clones analyzed from the IgH heterozygous mouse, Dμ gene segment usage was random and comparable to Vp81X rearrangements in the BM compartments (Fig. 3B), since Dμ gene segments from all families were represented in both populations, and over-representation of the absence of specific Dμ gene segments was not evident. In addition, there were no significant differences in hydrophobicity values between the populations before (average hydrophobicity, 0.82 ± 0.8) and after (average hydrophobicity, 1.08 ± 1.19) pre-BCR selection. The CDR3 length profiles were, however, distinct (Fig. 3B). There was a predominance of CDR3 regions 9 aa in length. Approximately 49% of rearrangements in this immature B cell population encode a CDR3 of 9 aa. Other CDR3 lengths isolated from this population were never >14% of the total. The occurrence of CDR3 lengths of 9 aa in the newly emerging splenic B cell compartment is significantly different from those in B220−, CD43+/−, surface μ− pro-B, and early pre-B cells in the BM (p < 0.02).

**Histidine in the first position in CDR3**

The CDR3 of a VDJH gene rearrangement is the only region encoded by all three gene segments. This circumstance allows for immense diversity due to combinatorial joining and end processing. Despite this extensive variability, examination of the amino acid sequences of CDR3 allowed us to identify constraints on the CDR3 structure imposed by the developmental selection process.

The 3’ end of the VH18X gene segment has nucleotides CA in its germline configuration. In the absence of processing of the VH gene during V to DJH joining, these nucleotides will encode histidine or glutamine at the first position of the CDR3. Processing, theoretically, could introduce any amino acid. The earliest compartment analyzed, corresponding to the pro-B and pre-B cell compartments, appears to have no bias for any particular junctional sequence or encoded residue (Fig. 4A). Position 1 of these CDR3s encoded a range of amino acids varying in charge and polarity; 36% were histidine. The more mature population, however, exhibited an increasing propensity to encode histidine in CDR3 position 1 (Fig. 4B). Seventy-three percent of the HCs from cells of the selected, newly emerging splenic B cell population, had histidine at position 1. The predominance of histidine at this position is a clear bias that was not evident in the pre- and pre-B cell population (p < 0.00012). Glutamine, encoded by CAG and CAA, is not found at equal frequency, as it would be if there were no preference for CAC or CAT (both encode histidine). The presence of histidine with a pI of 6.0 is interesting because it is unusual.

### Table 1. Vp81X-μHcs and their potential to pair with SLC and to be transported to the cell surface.

<table>
<thead>
<tr>
<th>Vp81X Sequence</th>
<th>Reference No.</th>
<th>FR3</th>
<th>CDR3</th>
<th>FR4</th>
<th>Intra μHC</th>
<th>Surface μHC</th>
<th>CDR3 Length (aa)</th>
<th>pl</th>
<th>GRAYV*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tdt4</td>
<td>24</td>
<td>YCAR</td>
<td>HANWDAMDY</td>
<td>WGOGT</td>
<td>+</td>
<td>+</td>
<td>9</td>
<td>4.20</td>
<td>−1.156</td>
</tr>
<tr>
<td>BC2</td>
<td>6, 56</td>
<td>YCAR</td>
<td>HGWSYWFVDV</td>
<td>WAGGT</td>
<td>+</td>
<td>+</td>
<td>8</td>
<td>5.08</td>
<td>−0.433</td>
</tr>
<tr>
<td>31A</td>
<td>6, 58</td>
<td>YCAR</td>
<td>HYYGYAMDY</td>
<td>WGOGT</td>
<td>+</td>
<td>+</td>
<td>9</td>
<td>6.08</td>
<td>−0.956</td>
</tr>
<tr>
<td>MTBM4</td>
<td>24</td>
<td>YCAR</td>
<td>HPRMVTTPYAMDY</td>
<td>WGOGT</td>
<td>+</td>
<td>+</td>
<td>12</td>
<td>6.64</td>
<td>−0.658</td>
</tr>
<tr>
<td>MTFIL8</td>
<td>24</td>
<td>YCAR</td>
<td>HFTTATDY</td>
<td>WGOGT</td>
<td>+</td>
<td>+</td>
<td>8</td>
<td>5.08</td>
<td>−0.686</td>
</tr>
<tr>
<td>MTFIL4</td>
<td>24</td>
<td>YCAR</td>
<td>HAMVTYYAMDY</td>
<td>WGOGT</td>
<td>+</td>
<td>+</td>
<td>12</td>
<td>5.06</td>
<td>−0.033</td>
</tr>
<tr>
<td>Tdt1</td>
<td>24</td>
<td>YCAR</td>
<td>HGNYAMDY</td>
<td>WGOGT</td>
<td>+</td>
<td>+</td>
<td>8</td>
<td>5.08</td>
<td>−1.188</td>
</tr>
<tr>
<td>F</td>
<td>8</td>
<td>YCAR</td>
<td>HDYYYNF</td>
<td>WGOGT</td>
<td>+</td>
<td>+</td>
<td>7</td>
<td>5.08</td>
<td>−0.829</td>
</tr>
<tr>
<td>BFL23</td>
<td>6</td>
<td>YCAR</td>
<td>EGNYAMDY</td>
<td>WGOGT</td>
<td>+</td>
<td>+</td>
<td>8</td>
<td>3.67</td>
<td>−1.225</td>
</tr>
</tbody>
</table>

* GRAYV = Grand average of hydrophylicity. 

---

**FIGURE 2.** Flow cytometric sorting of B lymphoid populations from spleen and bone marrow. Flow cytometric profiles of presort (A) and post-sort (B) BM and spleen (Spl) populations are depicted. A, BM suspensions (upper panels) were triple-stained with fluorochrome-conjugated Abs against surface B220, surface CD43, and surface μHC. Cells were gated for surface μHC-negative cells (a), and B220+/CD43+ as well as CD43− cells were sorted from the surface μHC-negative population (see sorting gate in b). Spleen cell suspensions (Spleen; lower panels) were triple-stained with Abs against surface B220, IgM, and IgD. Cells were gated for B220-positive cells (c), and IgMbright(IgDlow−/− cells were sorted (see sorting gate in d). B, Staining profiles of representative populations isolated from BM of JH+/− mice (a and b) and from spleens of JH+/− (c) and JH−/− (d) mice. Collected populations were used to isolate Ig VDJ rearrangements for subsequent sequencing of CDR3 regions. The staining profiles are representative of four independent experiments. Sort purity was consistently >98%.
of V<sub>H</sub> regions rarely have charged amino acids and are generally neutral or slightly hydrophilic (35, 36). Potentially, histidine is selected for its association with SLC. This convergence of the types of junctions that are preferred for the peripheral pool inevitably limits the VH<sup>81X</sup>-expressing clones that will survive and may play a role in the reduced number of mature B cells that use the VH<sup>81X</sup> gene segment. While the majority of sequences isolated from cells that have traversed the pre-BCR checkpoint code for a histidine at position 1 of the CDR3, a small number of isolates do not conform to this strategy. However, hydrophobicity, pI, and CDR3 analyses of His-less VH<sup>81X</sup>-/H<sub>9262</sub> HC s that have successfully traversed the pre-BCR checkpoint did not reveal any distinguishing characteristics of these HCs.

Restrictions on pre-BCR assembly and cell surface expression

One requirement for a VH<sub>81X</sub> rearrangement to appear in the peripheral compartments includes the pairing-fitness of a VH<sub>81X</sub>-
μHC with SLC (8, 27). The repertoire changes that we observed as a cell passes through SLC-dependent and independent compartments suggest a role for both a μHC and the SLC in choosing the structure of the V_H regions that are maintained in the peripheral repertoire. To directly test the ability of various V_H81X-μHCs to pair with SLC, we first cloned V_H81X sequences (see Table I) isolated from primary T cells, A-MuLV pre-B lines, and hybridomas into either retroviral or conventional plasmid expression vectors and tested the pairing capability of V_H81X-μHC with the SLC after introducing the V_H81X-μHC genes either by retroviral transduction (clones TdT-4, MTBM-4, MTFL-8, MTFL-4, and TdT-1 in Fig. 1 and Table I) or electroporation (clones BC2, 31A, F, and BFL23 in Table I) in the μHC-negative A-MuLV pre-B lines 38B9 and Bine4.8 (see Materials and Methods). These V_H81X-μHC sequences differed from each other only in the CDR3 region, that is, they used various D_H gene segments with different joining sequences. μHC synthesis and association with SLC were evaluated in stably infected or transfected pre-B cell subclones by flow cytometry and immunoprecipitation of metabolically labeled cell lysates. An example of a typical analysis is shown in Fig. 1, B and C; Table I summarizes the findings of all analyzed sequences.

Nine V_H81X-μHCs expression constructs were tested (see Table I). Although intracellular μHCs could be detected in all stable clones (Fig. 1, B and C, and Table I), three had clear surface expression of μHC, as revealed by Abs against μHC (clone TdT-4 in Fig. 1B, and clones BC2 and 31A in Table I), one showed intermediate levels (clone MTBM-4 in Fig. 1B) and five were negative (clones TdT-1, MTFL-4, and MTFL-8 in Fig. 1B and F and BFL23 in Table I). Surface pre-BCR expression was verified with an mAb (SL156) that recognizes an assembled pre-BCR (data not shown). As expected, all four surface transport competent V_H81X-μHCs coprecipitated with the SLC components VpreB and A5 (results for TdT-4 and MTBM-4 are shown in Fig. 1C). Interestingly, three of the four V_H81X-μHCs (TdT-4, BC2, and 31A) that were transported to the cell surface (Fig. 1B and Table I) and coprecipitated with the SLC components VpreB and A5 (Fig. 1C and Table I) used a CDR3 region consisting of 9 aa (Table I; CDR3 defined as above, according to Kabat’s nomenclature (22)). Consistent with the trend observed in the splenic B cell compartment, all four pairing V_H81X-μHCs contained a histidine residue at position 1 of the CDR3 (Table I). One of the four pairing and surface-competent V_H81X-μHCs (clone MTBM-4 in Fig. 1, A and B) used a CDR3 with 12 aa (Table I). In contrast to the TdT-4-μHC, the MTBM-4-μHC showed a clearly reduced surface expression despite the fact that intracellular levels of both μHCs were very similar (compare fluorescence intensities for cytoplasmic and membrane staining in Fig. 1B). Therefore, these two V_H81X-μHC differ in their propensity to be transported to the surface, which might be due to CDR3-mediated differences in binding strength between a μHC and the SLC. However, compared with transport-incompetent V_H81X-μHCs, there was no bias in the pI or hydrophobicity values of those V_H81X-μHCs that were expressed on the cell surface (summarized in Table I).

**Modeling of the CDR3 conformation**

To examine whether the structures of V_H81X-μHCs might reveal a clue as to why some were expressed on the surface of the A-MuLV and some were not, we used the SWISS-MODEL Protein Modeling Server. As described in Materials and Methods, we searched the protein databases for an Ig crystal structure with an amino acid sequence similar to that of V_H81X to use as a template for our modeling. A single HC Fv, named in the database Protein 31AR1,
was identified as having the most sequence similarity to the V_{H}81X variable domain (~72% identity). We then used protein 31AR1 as the template and modeled on it the 10 V_{H}81X-HCs. The data generated by the modeling program were converted, assembled, and viewed using the SWISS PDB viewer (28, 29). These theoretical models were positioned to align the identical regions (the β strands of the framework regions), while outlining their differences (the CDR3 regions). The results of this structure prediction analysis are presented in Fig. 5. The variable domains of the four molecules that were expressed on the cell surface were essentially superimposable in all regions, including the backbone tracing of the hypervariable loop of CDR3 (shown with R groups in blue). The CDR3 backbone is a simple loop in the single plane, although the R groups of the residues are distinct. On the other hand, the molecules that did not go on the surface have different CDR3 structures (backbone is shown in red). CDR1, CDR2, the framework region, and the C region, as expected by their near identities, are more similar to each other, although not completely superimposable; their positioning is probably influenced by the CDR3 residues. Thus, the four molecules that are expressed on the surface have a superimposable backbone structure despite differences in the sequences of their CDR3 regions.

Discussion

This study reveals specific characteristics in the V_{H}81X μHC that have been allowed to passage through selection and be expressed in splenic B cells. The majority of such HC have CDR3 lengths of 9 aa and a histidine in position 1. A computer-generated model pictures the CDR3 backbone structure as planar.

The paradigm of B cell development in mice and humans is characterized by checkpoints and selection events that ultimately shape the mature repertoire. The structural complexity of the BCR and its various functions necessitates a systematic and rigorous testing of the receptor and its component parts. Many studies have highlighted the importance of the pre-BCR checkpoint (9, 10, 18). This stage, immediately following assembly of the HC gene segments, serves to select cells that have made productive rearrangements, expand clones with suitable receptors on the surface, and regulate the structure of assembled HCs (6, 8, 10, 37). The SLC components, VpreB and λ5, are ideally positioned to play a role in testing and selecting the components, VpreB and λ5 protein appears to facilitate the proper folding and assembly of the SLC, while forming an Ig domain and a region that will be situated close to the CDR3 of the LC. These limitations imposed by the SLC may actually serve to maximize the efficiency of the developing immune system. The ultimate goal of the system is to generate a B cell population where cells produce HCs capable of pairing efficiently with LCs, forming receptors that broaden and expand Ag binding capabilities. Facilitating progression of only those cells capable of competent HCl/LC interactions later in development increases the probability of productive functional, viable cells. Testing through the SLC may also serve to recruit groups or families of preferred LCs that augment the diversity at the paired CDR3 domains. Studies have shown that particular LC and HC combinations are capable of assuming multiple conformations depending on the components involved (48). LCs that amplify this isomerism may have the potential to bind multiple Ags and thereby increase the efficiency of an immune response.

We found that the persisting CDR3 structures of maturing B cells are largely skewed to 9 aa. At this length, the hypervariable region of the molecule takes on a structure distinct from the configuration assumed by CDR3 domains that were unable to pair with SLC. Although other lengths can survive in vivo, the majority are this length. HC that have other CDR3 lengths might bind the SLC less efficiently, and as a result, fewer complexes might be deposited on the surface. With a lower pre-BCR density, a B cell clone would be signaled less efficiently, resulting in less expansion or even deletion of that clone (37, 41).

A reduction in the average CDR3 length has been observed, as human B cells progress through development (42). Some evidence suggests that longer CDR3 regions are associated with autoreactivity (43–45). Thus, testing of HCs by the SLC could function as a screen to limit the production of autoreactive cells. In the J_{H}5^{−/−} mouse splenic compartment, there was a notable reduction in the number of clones exhibiting longer CDR3s, a reduction not observed in the splenic compartment of A5^{−/−} animals. This observation provides support for the hypothesis that longer CDR3s are associated with autoreactivity and consequently may be more readily eliminated under normal circumstances (46). In this fashion, the SLC could function to minimize or regulate autoreactivity by modulating the HC repertoire.

CDR3 length has been directly correlated to the potential for interaction between Ig HC and LC (47). In these analyses there were notable differences in the shape assumed by the coupled CDR3 regions depending on the lengths involved. Both HC/LC pairing and HC/SLC pairing may play a role in shaping the repertoire in the newly emerging B cell population. The correlations and structural limitations of these studies can easily be extended to the pairing of an HC and SLC. The fitness of the interaction would be determined by the shape created by the CDR3.

Our experiments reveal that the repertoire shift observed in the splenic B cell population subsequent to selection across the pre-B cell transition does not occur in the absence of pre-BCR formation. This finding directly implicates the pre-BCR and the SLC in particular in determining the final configuration of the B cell repertoire. This central role for the pre-BCR in repertoire selection extends the function of the receptor in the developmental paradigm. The limitations imposed by the SLC may actually serve to maximize the efficiency of the developing immune system. The ultimate goal of the system is to generate a B cell population where cells produce HCs capable of pairing efficiently with LCs, forming receptors that broaden and expand Ag binding capabilities. Facilitating progression of only those cells capable of competent HCl/LC interactions later in development increases the probability of producing functional, viable cells. Testing through the SLC may also serve to recruit groups or families of preferred LCs that augment the diversity at the paired CDR3 domains. Studies have shown that particular HC and LC combinations are capable of assuming multiple conformations depending on the components involved (48). LCs that amplify this isomerism may have the potential to bind multiple Ags and thereby increase the efficiency of an immune response.

The biased representation of CDR3 lengths of 9 aa in the J_{H}5-de heterozygotes and the absence of this skewing in A5-deficient animals may represent structural restrictions that have evolved over time and are employed to augment the diversity of the HC in this context. Using 9 aa in the CDR3 loop may provide the greatest
diversity for the CDR3 without compromising the structural integrity of the receptor. Some studies have indicated that too many residues in the loop may affect the stability and rigidity of the Ig fold (49). The interaction between framework regions and CDRs may also impact upon molecular structure in the vicinity of HC/LC association, affecting LC compatibility as well as the ability to bind Ag. The size and composition of the CDR3 have the potential to alter the conformation of the molecule in ways that may or may not be optimal for these contacts (50, 51). It is possible that at the length of 9 aa, optimization of diversity and structural integrity converge.

Compositional analysis of selected CDR3 sequences reveals a propensity for the amino acid histidine at position 1 (position 95 of the variable domain). Although the germline sequence of V<sub>H</sub>81X encodes the CA necessary for the histidine codon, this could also code for glutamine, which is not found in this position as frequently. The only explanation is that this amino acid provides optimal selective advantage to the V<sub>H</sub>81X-HC at all stages of development. In addition to the effects on conformation, there may be a role for this residue in ligand binding. CDR3 residues have been implicated in binding specific Ags (44, 45, 52–54). In the context of V<sub>H</sub>81X, the histidine may play a role in contacts with the extracellular matrix of stroma cells or other putative ligands for the pre-BCR. It is notable that there is a strong selection for histidine at position 1 in the CDR3 of V<sub>H</sub>10 rearrangements, another HC variable region gene associated with restricted expression patterns (55).

Our conclusions differ slightly from those of Hayden et al. (56), perhaps due to the approach taken by the authors. While their studies looked at the characteristics of V<sub>H</sub>181X-HCs that used a number of different J<sub>H</sub> gene segments (i.e., J<sub>H</sub>1 through J<sub>H</sub>4), we have focused on V<sub>H</sub>81X rearrangements using J<sub>H</sub>4. In this manner, we were able to attribute the selection or enrichment observed as being due solely to the differences in the CDR3 regions used. Therefore, structural contributions from amino acids encoded by the J<sub>H</sub> region and their confounding effects can be negated. This system allows us to directly test and assess the contribution of the CDR3 to the HCs that are used/successful in the periphery. Based on this approach we can for the first time look at specific characteristics of the HC that influence its association with the SLC, and thereby assess a direct function for SLC in shaping the mature repertoire and the manner in which this specific mechanism is implemented.

In fact, if one looks closely at the data presented by Hayden et al. (56), selection biases become apparent. For example, in one group 70% of the sequences use histidine at position 1, whereas only 15% of the sequences use glutamine at this position. In another group, 45% of CDR3s used were 9 aa long, with very few CDR3 lengths >11 aa. In the group of sequences using the I<sub>H</sub>4<sup>4</sup> gene segment, there is skewing to the CDR3 length of 9 aa as well as the use of histidine at position 1 of the CDR3. However, one could only make these contentions if larger sample sizes for each group and appropriate statistics were applied to the data to make the findings significant.

The fact that these biases exist in this population does not imply that these same characteristics are important in all populations, but does imply that these traits can play a significant role in determining HC selection and that the SLC is capable of distinguishing Ig proteins based on these characteristics.

The under-representation of the V<sub>H</sub>81X gene segment in the peripheral B cell pool is due to both molecular and cellular factors (6–8, 12, 14–16, 27, 57, 58). Our studies indicate that there are limitations placed on the V<sub>H</sub>81X HCs that are permitted to persist based on the amino acids that are encoded by the V(D)J junctional sequence. Clearly, the shape assumed by the HC plays a role in selecting the cells that survive. In addition to the other factors that restrict the presence of V<sub>H</sub>81X in the periphery, germline-encoded features of the V<sub>H</sub>81X HC protein may disfavor a competent interaction with SLC. Only under certain conditions is the complex suitably organized to warrant maintenance of the cell.

Functional assembly of a pre-BCR is an integral step in promoting the efficient transition of cells through the pro-B cell stage to an immature stage of development. This facility with which a particular HC associates with SLC has a direct impact on the survival of that cell (4). The studies that we have undertaken here indicate that the length of the CDR3 region can have implications for the suitability of fit and the subsequent fate of the cell. In these studies the most striking evidence of CDR3 length selection was observed across the pre-B transition; this was further confirmed by the lack of length selection in the absence of A5 and the pre-BCR. Further studies will reveal whether the absence of an SLC selection mechanism affects the characteristics of LCs that are able to pair with HC at later stages in development. Comparison of the LC repertoire in wild-type animals, in which pre-BCR-selective programs are intact, and in A5<sup>−/−</sup> animals will reveal whether the CDR3 restriction directly imposes identifiable features relevant to compatibility and LC pairing.

Acknowledgments
We are grateful to Aaron Marshall, Anne Tran, Dirk Mielenz, and Ivan Fong for valuable comments. We thank Queenie Lam and Stacy Hirano for expert technical assistance. We thank Ana Cumano for critical reading of the manuscript and thoughtful comments, and Ton Rolink, Fritz Melchers and Jan Jongstra for providing mAbs against the pre-BCR and pre-BCR components.

References


